

# ГЕНОМИКА

---

Челомина Г. Н.





FEDERAL SCIENTIFIC CENTER OF  
THE FAR EAST ASIA TERRESTRIAL BIODIVERSITY  
FAR EASTERN BRANCH OF THE RUSSIAN ACADEMY OF SCIENCES

G. N. Chelomina

# GENOMICS

Vladivostok  
2025

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ  
«ФЕДЕРАЛЬНЫЙ НАУЧНЫЙ ЦЕНТР БИОРАЗНООБРАЗИЯ  
НАЗЕМНОЙ БИОТЫ ВОСТОЧНОЙ АЗИИ»  
ДАЛЬНЕВОСТОЧНОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК

Г. Н. Челомина

# ГЕНОМИКА

Владивосток  
2025

УДК: 577.2

**Челомина Г. Н.** Геномика. Владивосток: ФНЦ Биоразнообразия ДВО РАН, 2025. – 295 с.

Начиная с краткого исторического обзора, в работе обсуждаются основные положения геномики, включая методы секвенирования ДНК и разделы геномики, в том числе многообещающие направления, представленные палеогеномикой и синтетической геномикой. Приводятся примеры применения геномных данных в решении широкого круга научных и прикладных задач. Завершается монография кратким заключением и приложением с фактами из жизни известных ученых, открытия которых легли в основу геномных исследований. Книга адресована биологам различных специальностей и может представлять интерес для археологов, медицинских работников, а также студентов и аспирантов вузов.

Илл. 84 + 23 Приложение, табл. 8, библиограф. 762

**Chelomina G. N.** Genomics. Vladivostok: FSC of Biodiversity FEB RAS, 2025. – 295 pp.

Beginning with a brief historical overview, the book provides the review of the key concepts in genomics, such as DNA sequencing methods and subfields of genomics, including the promising directions represented by paleogenomics and synthetic genomics. Examples of the application of genomic data in solving a wide range of scientific and applied problems are given. The monograph is completed with a brief conclusion and an appendix with facts from the lives of famous scientists whose discoveries formed the basis of genomic research. The book is addressed to biologists of various specialties and may be of interest to archaeologists, medical staff, as well as students and postgraduates of universities.

Ill. 84 + 23 Suppl., tab. 8, bibl. 762

Рецензенты:

д.б.н. проф. О.И. Подгорная;

к.б.н. А.В. Недолужко

Утверждено к печати Ученым советом  
ФНЦ Биоразнообразия ДВО РАН

ISBN 978-5-6052787-3-3

<https://doi.org/10.25221/genomics>

<https://elibrary.ru/hjetbr>

© Г.Н. Челомина

© ФНЦ Биоразнообразия ДВО РАН

© Издательское оформление  
ИП Мироманова И.В.

## СОДЕРЖАНИЕ

Список сокращений . . . . .	6
Введение . . . . .	7
Глава 1. История геномики . . . . .	9
Глава 2. Технологии секвенирования ДНК. . . . .	14
Глава 3. Разделы геномики . . . . .	34
3.1. Структурная геномика . . . . .	35
3.2. Функциональная геномика . . . . .	43
3.3. Сравнительная геномика. . . . .	52
3.4. Популяционная геномика. . . . .	65
3.5. Ландшафтная геномика . . . . .	69
3.6. Метагеномика . . . . .	71
3.7. Палеогеномика . . . . .	76
3.8. Синтетическая геномика . . . . .	81
3.9. Веномика . . . . .	85
Глава 4. Применение геномики в решении научных задач . . . . .	90
4.1. Вариации размера геномов. . . . .	90
4.2. Нуклеотидный состав и предпочтение кодонов. . . . .	100
4.3. Малые интерферирующие РНК и РНК-интерференция . . . . .	106
4.4. Мобильные генетические элементы и горизонтальный перенос генов. . . . .	110
4.5. Эпистаз, плейотропия и гетерозис. . . . .	123
4.6. Гибридизация и гибридные зоны . . . . .	127
4.7. Эволюция, филогенез и биология видов . . . . .	147
Глава 5. Практическое применение геномики . . . . .	165
5.1. Медицина . . . . .	165
5.2. Антропология и археология . . . . .	189
5.3. Экология: сохранение биоразнообразия и управление биоресурсами. . . . .	203
5.4. Сельское хозяйство и биотехнология . . . . .	221
Заключение . . . . .	232
Литература . . . . .	235
Список терминов . . . . .	272
Приложение. . . . .	284

## СПИСОК СОКРАЩЕНИЙ

ДНК	Дезоксирибонуклеиновая кислота
РНК	Рибонуклеиновая кислота
л. н.	Лет назад
п. н.	Пары нуклеотидов
ПЦР	Полимеразная цепная реакция
BLAST	Basic Local Alignment Search Tool (базовый инструмент поиска локального выравнивания)
ChIA-PET	Chromatin Interaction Analysis by Paired-End Tag Sequencing (анализ взаимодействия хроматина с секвенированием парных концов)
ChIP	Chromatin ImmunoPrecipitation (иммунопреципитация хроматина)
CNV	Copy Number Variation (вариация числа копий)
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats (сгруппированные регулярно расположенные короткие палиндромные повторы)
CUB	Codon Usage Bias (отклонение в использовании кодонов)
EST	Expressed Sequence Tag (маркерная экспрессирующаяся последовательность)
ISS	In Situ Sequencing (секвенирование <i>in situ</i> )
GWAS	Genome-Wide Association Study (полногеномное исследование ассоциаций)
HGT	Horizontal Gene Transfer (горизонтальный перенос генов)
Hi-C	Chromatin Conformation Capture (захват конформации хроматина)
LINEs	Long Interspersed Nuclear Elements (длинные диспергированные повторы)
LTRs	Long Terminal Repeats (длинные концевые повторы)
NGS	Next Generation Sequencing (секвенирование следующего поколения)
ORF	Open Reading Frame (открытая рамка считывания)
PAV	Presence/Absence Variants (варианты присутствия/отсутствия)
QTLs	Quantitative Trait Loci (локусы количественных признаков)
RRS	Reduced Representation DNA Sequencing (секвенирование с уменьшенным представлением)
SBL	Sequencing By Ligation (секвенирование путём лигирования)
SBS	Sequencing By Synthesis (секвенирование путём синтеза)
SINEs	Short Interspersed Nuclear Elements (короткие диспергированные повторы)
SLR	Synthetic Long Read (синтетические длинные прочтения)
SMRT	Single Molecule RealTime Sequencing (секвенирование одной молекулы в реальном времени)
SMS	Single Molecule Sequencing (секвенирование одной молекулы)
SNP	Single Nucleotide Polymorphism (однонуклеотидный полиморфизм)
SNV	Single Nucleotide Variants (однонуклеотидные варианты)
SV	Structural Variants (структурные варианты)
TAD	Topologically Associating Domains (топологически ассоциированные домены)
TE	Transposable Elements (мобильные элементы)
WGD	Whole Genome Duplication (полногеномная дупликация)
WGS	Whole Genome Sequencing (полногеномное секвенирование)

## ВВЕДЕНИЕ

Существует множество причин для того, чтобы утруждать себя определением полных и точных геномных последовательностей ... полный набор инструкций обеспечивает потенциально мощную основу для разработки вакцин и других терапевтических средств ... даёт представление о ряде функций, которыми обладает организм, об относительной важности, которую естественный отбор придаёт каждой функции, и об истории эволюции организма ... доступность полных последовательностей генома породила огромное количество творческих подходов к глобальному функциональному анализу генов и генных сетей ... не только можно предсказать все или почти все белки, присутствующие в организме, но и то, что отсутствует, также становится значимым.

Eisen et al., 1997

Геномика – быстро развивающаяся и чрезвычайно многообещающая область биологии, и одна из самых ярких междисциплинарных наук, которая посвящена изучению генов и геномов, определяющих молекулярные структуры, функции и регуляцию биологических организмов. В полной мере используя инструменты и методы, первоначально разработанные в других областях науки, геномика фокусируется на структуре, функциях, эволюции, картировании и редактировании генома у человека и других организмов с конечной целью улучшения качества жизни человека. Хотя термин «геномика» впервые появился в 1986 году, его применение было тесно связано с полногеномным секвенированием живых организмов с использованием подхода «дробовика» в рамках Международного проекта генома человека, начавшегося в 1990 году. Технологии секвенирования следующих поколений предоставили экономичные и высокопроизводительные варианты анализа геномов. Новые методы секвенирования позволяют решать многие совершенно невыполнимые ранее задачи, в том числе полногеномное картирование метилирования и мутаций ДНК, выяснение регуляторных механизмов экспрессии генов, определение специфических ДНК–белковых взаимодействий, расшифровку структуры хроматина и т. д. Сочетание высокопроизводительной технологии секвенирования со всё более сложным программным обеспечением, включая использование искусственного интеллекта, и новыми подходами к подготовке образцов позволило изучать общность и специфичность отдельных организмов и предвидеть их реакции на внешние воздействия, тем самым значительно пополнив наши знания о генах и геномах (Korhonen et al., 2016). Достижения в области геномики расширили возможности передовых исследований для системной биологии и способствовали пониманию сложных биологических систем (таких как мозг человека) и таких явлений, как гетерозис, эпистаз и плейотропия (Kadakkuzha,

Puthanveettil, 2013; Pevsner, 2009). Тесно связанная с протеомикой и метаболомикой, геномика закладывает фундамент медицины будущего (преобразованная и высокоэффективная персонализированная медицина, новые методы диагностики и новые лекарства). Кроме того, усовершенствования в технологии секвенирования, позволяющие обнаруживать тонкие различия внутри и между видами (и даже на уровне отдельных клеток) животных и растений, помогут расширить применение геномики в сельском хозяйстве, сохранении редких и исчезающих видов и обеспечат более глубокое понимание эволюции жизни на Земле в целом (Touchman, 2010).

Книга содержит пять тематических глав и сопровождается иллюстративным материалом с поясняющими комментариями и списком литературы. В первой главе кратко изложена история становления геномики, включая развитие представлений о ДНК и технологий исследования ДНК и геномов. Во второй главе описаны методы секвенирования геномов, их принципы, возможности и ограничения. Третья глава посвящена основным направлениям геномики с выделением ряда подразделов, направленных на изучение более частных проблем. В четвёртой и пятой главах приведены примеры применения геномики в научных и прикладных целях соответственно. Завершается книга кратким заключением с приложением.

Автор выражает глубокую признательность за внимательный и доброжелательный анализ книги, ценные замечания и полезные советы академику РАН Юрию Николаевичу Журавлёву, д.б.н. профессору Ольге Игоревне Подгорной, к.б.н. Артёму Валерьевичу Недолужко и д.б.н. Юрию Николаевичу Шкрылю, а также к.х.н. Елене Васильевне Сундуковой за техническое редактирование и макетирование рукописи.

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 121031000154-4).

## ГЛАВА 1. ИСТОРИЯ ГЕНОМИКИ

Термин «геномика», предложенный генетиком Томасом Родериком на совещании по картированию генома, состоявшемся в Мэриленде в 1986 г., происходит от «гена» и «омики» (последнее предполагает изучение полных наборов биологических данных). По одной из существующих версий, *«слово «омикс» может быть связано с «ом» (произносится как «аум»), древней санскритской информацией, которая, как и музыка, преодолевает барьеры возраста, расы, культуры и даже видов. Считается, что он является основным звуком мира и содержит в себе все остальные звуки и сам по себе является мантрой»* (Yadav, 2007).

Развитие представлений о «молекуле жизни» имеет 150-летнюю историю и чёткое разделение на два этапа: до и после разработки технологий секвенирования ДНК. История изучения наследственного материала началась в XIX в., когда Грегор Иоганн Мендель опубликовал свою известную работу «Опыты на гибридах растений» о наследственных признаках гороха, в которой он описал основные законы наследственности, введя базовые в современной генетике человека термины «рецессивный» и «доминантный» (Mendel, 1866). Следующие десятилетия отмечены крупными научными открытиями, в том числе значительными достижениями в экспериментальной биологии и химии, а также зарождающимися представлениями о ДНК, которые подчёркивали теоретические исследования и предсказания Николая Константиновича Кольцова, Георгия Антоновича Гамова и др. (табл. 1.1). Через три года после публикации Менделя была открыта химическая природа ДНК, хотя только в 1944 г. удалось показать её роль в генетической наследственности (Avery et al., 1944). Систематическое исследование нуклеиновых кислот в нашей стране началось с 30-х годов работами академика Андрея Николаевича Белозерского по идентификации тимина в клетках и ДНК растений, которые позволили опровергнуть существовавшее в то время деление нуклеиновых кислот на «животные» и «растительные» (Белозерский, Чигирев, 1936). В 1935 г. произошёл значительный прогресс в понимании природы генных мутаций и структуры генов (Timofeeff-Ressovsky et al., 1935), а в 1942 г. был введён термин «эпигенетика» (Waddington, 1942). В 1953 г. Джеймс Уотсон и Фрэнсис Крик представили модель двухцепочечной структуры ДНК и подчеркнули важность определения точной последовательности оснований в цепи ДНК для полного понимания репликации ДНК и наследственного контроля клеточной активности (Watson, Crick, 1953). Их открытие привело к созданию нового научного направления – молекулярной биологии. Наконец, в 60-х годах Маршалл Ниренберг и Филипп Ледер расшифровали генетический код, который оказался универсальным для большинства организмов (Nirenberg, 2004). Это открытие обеспечило доступ к изучению взаимосвязи между генетической информацией и её биологическим смыслом (табл. 1.1).

Таблица 1.1. Развитие представлений о ДНК (по: Chelomina, 2024a)

Дата	Событие
1865	Грегор Иоганн Мендель указывает на «Наследственные признаки, содержащиеся в половых клетках» и описывает основные законы наследственности
1869	Фридрих Мишер идентифицировал «нуклеин», который позже получил название «нуклеиновая кислота»
1901	Альбрехт Коссель открыл пять азотистых оснований: аденин (А), цитозин (С), гуанин (G), тимин (Т) и урацил (U)
1906	Уильям Бэтсон предложил термин «генетика»
1909	Вильгельм Людвиг Йохансен ввёл термин «ген»
1910	Феб Левен выдвинул тетра-нуклеотидную гипотезу строения нуклеиновых кислот
1927	Николай Константинович Кольцов предсказывает, что врождённые признаки должны передаваться потомству через «гигантскую наследственную молекулу», состоящую из «двух зеркальных нитей, которые будут воспроизводиться полуконсервативным образом, используя каждую нить в качестве шаблона»
1935	Николай Васильевич Тимофеев-Ресовский, Карл Циммер и Макс Дельбрюк опубликовали работу «О природе генной мутации и структуре гена»
1936	Андрей Николаевич Белозерский опровергает деление нуклеиновых кислот на «животные» и «растительные»
1942	Конрад Уоддингтон ввёл термин «эпигенетика»
1944	Освальд Эйвери, Колин Маклауд и Маклин Маккарти определяют ДНК как материал, из которого состоят гены и хромосомы
1944	Эрвин Шрёдингер называет ДНК «транспортным принципом», а переносимую ДНК информацию – «шифровальным кодом»
1950	Эрвин Чаргафф разработал схему сопряжения оснований А, С, G и Т
1952	Альфред Херши и Марта Чейз показали, что генетическая информация передаётся ДНК, а не белком
1952	Розалинда Франклин фотографирует кристаллизованные нити ДНК
1953	Джеймс Уотсон и Фрэнсис Крик открывают двойную спираль ДНК
1953	Георгий Антонович Гамов предположил, что набор из 3-х оснований кодирует 20 стандартных аминокислот
1957	Конрад Уоддингтон сформулировал концепцию «эпигенетического ландшафта»
1959	Борис Фёдорович Ванюшин определил природу метилированных последовательностей ДНК растений и микроорганизмов
1961–1966	Маршалл Ниренберг и Филипп Ледер раскрывают триплетную природу генетического кода, определяют последовательности 54 из 64 кодонов

История секвенирования биополимеров (определение аминокислотной или нуклеотидной последовательности для белков и нуклеиновых кислот соответственно) началась в 1953 г., когда Фредерик Сенгер и его австралийский аспирант Тед Томпсон установили полную структуру белка инсулина (Sanger, Thompson, 1953). В 1965 г. Роберт Холли опубликовал первую рибонуклеотидную последовательность длиной 76 п. н., транспортной РНК аланина *Saccharomyces cerevisiae* Meyen ex E.C. Hansen 1883 (Holly et al., 1965), а в 1976 г. Уолтер Фирс и его коллеги секвенировали первый геном РНК из 3569 п. н. (Fiers et al., 1976). В 1968 г. Рэй Ву и Александр Кайзер описали первые сегменты ДНК, когезионные концы (12 п. н., известные как «cos» сайты) бактериофага лямбда (Wu, Kaiser, 1968), а в 1972 г. Уолтер Фирс с сотрудниками секвенировали ген длиной 510 п. н., кодирующий белок оболочки бактериофага MS215 (Min et al., 1972).

В 1977 г. Фредерик Сенгер с коллегами предложили ферментативный метод секвенирования ДНК, широко применяемый до настоящего времени, в котором используется ДНК-полимераза и модифицированные дезоксинуклеотиды, обрывающие цепь (т. е., терминирующие включение нуклеотида) (Sanger et al., 1977). Одновременно, Алан Максам и Уолтер Гилберт разработали альтернативный химический подход секвенирования путём химической дегградации, основанный на ограниченном расщеплении фрагмента ДНК с радиоактивной меткой на одном конце (Maxam, Gilbert, 1977). Однако техническая сложность и другие особенности этого метода сдерживали его применение. Наконец, идея секвенирования по принципу «дробовика» (Staden, 1979) (см. ниже), использование одноцепочечного фагового вектора M13 (Messing et al., 1981) и изобретение полимеразной цепной реакции (Mullis et al., 1986) ещё больше упростили процесс секвенирования ДНК. В результате стало возможным расшифровывать большие геномы за короткое время, особенно с появлением новых методов секвенирования. Эти разработки, поддерживаемые мощными инструментами биоинформатики, ознаменовали новую геномную эру фундаментальных и прикладных исследований с запуском Международного проекта генома человека в 1990 г. После публикации последовательностей геномов различных патогенных бактерий (Fleischmann et al., 1995) и модельных эукариотических видов, таких как пекарские дрожжи *S. cerevisiae* (Goffeau et al., 1996), почвенная нематода *Caenorhabditis elegans* (Maupas, 1900) (Seq. Cons., 1998), резуховидка Таля *Arabidopsis thaliana* (L.) Heynh., 1842 (Kaul et al., 2000), домовая мышь *Mus musculus* Linnaeus, 1758 (Waterston et al., 2002), в 2001 г. был опубликован черновой вариант генома человека (Lander et al., 2001; Venter et al., 2001). Окончательный геном был объявлен (с точностью менее одной ошибки на 20000 оснований) в 2007 г. (Clamp et al., 2007), хотя незавершённые гетерохроматические области и Y-хромосома были прочитаны совсем недавно (Nurk

et al., 2022; Rhie et al., 2023). В последние годы были расшифрованы геномы ряда вымерших видов (например, Miller et al., 2008; Slon et al., 2018; Gros-Balthazard et al., 2021), а также разработаны методы редактирования генома и моделирования его трёхмерной структуры (табл. 1.2).

Таблица 1.2. Развитие технологий исследования ДНК и геномов  
(по: Chelomina, 2024a)

Дата	Событие
1965	Роберт Холли с соавторами опубликовал первую последовательность РНК
1968	Рэй Ву и Александр Кайзер секвенируют первые сегменты ДНК (фаг лямбда)
1972	Уолтер Фирс с коллегами – секвенирование первого гена (белка оболочки бактериофага MS215)
1975	Фредерик Сенгер и Алан Коулсон – разработка метода «плюс–минус»; четыре ПЦР-реакции проводят в присутствии и в отсутствии каждого из четырёх типов дезоксирибонуклеозидтрифосфатов
1977	Фредерик Сенгер с коллегами создали ферментативный метод секвенирования ДНК
1977	Алан Максам и Уолтер Гилберт разрабатывают химический метод секвенирования ДНК
1979	Роджер Стаден предложил стратегию секвенирования «дробовик»; последовательность ДНК восстанавливают на основе случайной выборки клонированных фрагментов
1981	Иоахим Мессинг предложил использовать для секвенирования одноцепочечный фаговый вектор M13
1983	Кэри Мюллис изобретает метод полимеразной цепной реакции (ПЦР)
1985	Алек Джеффрис разработал технику ДНК-дактилоскопии (ДНК-фингерпринтинг)
1986	Томас Родерик вводит термин «геномика»
1987	На основе метода Сенгера разработан первый автоматический секвенатор ДНК
1990	Запущен проект «Геном человека» под руководством Джеймса Уотсона
1995	Получен первый геном бактерии ( <i>Haemophilus influenzae</i> )
1996	Компания Applied Biosystems автоматизировала секвенирование ДНК
1996	Пол Нирен с коллегами предложили технологию пиросеквенирования
1996	Секвенирован первый геном эукариотического организма ( <i>Saccharomyces cerevisiae</i> )
1997	Разработаны первые ДНК–микрочипы
1998	Иницирован первый проект генома многоклеточных эукариотических видов ( <i>Caenorhabditis elegans</i> )
1998	Крейг Меллоу и Эндрю Файер открыли РНК–интерференцию

окончание табл. 1.2

Дата	Событие
1999	Создан геномный браузер для позвоночных
2000	Определён геном цветкового растения ( <i>Arabidopsis thaliana</i> )
2001	Завершён черновик генома человека
2002	Расшифрован геном модельного вида млекопитающих ( <i>Mus musculus</i> )
2003	Применение подхода «Zero-mode waveguide» для секвенирования одиночных молекул ДНК
2003	Использование метода амплификации фрагментов ДНК в эмульсионной ПЦР
2005	Группой Джорджа Макдональда Черча разработана технология циклического лигазного секвенирования
2005	Коммерциализирована первая технология NGS (пиросеквенирование)
2007	Завершено секвенирование генома человека
2009	Предложены технологии NGS третьего поколения
2010	Получена черновая последовательность генома неандертальца
2010	Разработана первая технология неоптического (полупроводникового) секвенирования
2010	В лаборатории Крейга Вентера создана первая живая синтетическая клетка
2011	Дженифер Дудна и Эммануэль Шарпантье разработали технику редактирования генома с помощью системы CRISPR–Cas9
2012	Разработана и коммерциализирована технология нанопорового секвенирования
2015	Расшифрован первый геном вымершего вида (шерстистый мамонт <i>Mammuthus primigenius</i> )
2018	Проведено первое редактирование генома человека (с синдромом Хантера)
2018	Коммерческий выпуск нанопорового секвенатора PromethIon (ONT)
2021	Секвенированы геномы древних финиковых пальм ( <i>Phoenix dactylifera</i> ) из проросших 2000-летних семян
2023	Секвенированы незавершённые области генома человека

## ГЛАВА 2. ТЕХНОЛОГИИ СЕКВЕНИРОВАНИЯ ДНК

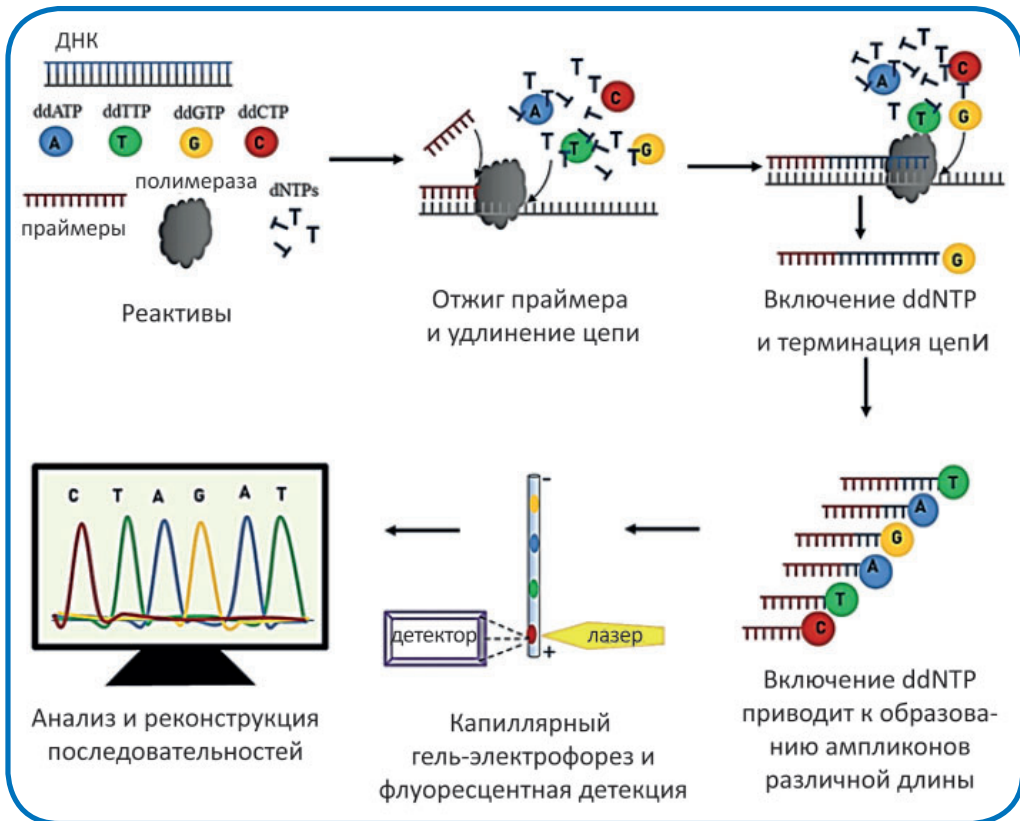
Геном – это совокупность наследственного материала, содержащегося в клетке организма; он включает все генные и внегенные участки ядерной и внеядерной ДНК и содержит весь объём информации, необходимой для развития и существования организма. Геномика – это междисциплинарная область исследований, которая использует рентгеновскую кристаллографию, ядерный магнитный резонанс, редактирование генома, молекулярную биологию, биохимию, биоинформатику и методы вычислительной биологии и искусственного интеллекта для изучения генов и геномов. Таким образом, геномика определяет все генетические характеристики клетки, чтобы раскрыть сущность организма. Сюда входит определение числа генов и порядка их расположения в геноме, размера и последовательности нуклеотидов каждого гена, функции каждого гена (связанной с регуляторными элементами метаболизма и видоспецифичными/ индивидуальными особенностями), потенциальных возможностей организма и предсказание его реакции на внешние воздействия. Анализ начинается с выбора целевого генома, за которым следует генерация его последовательностей (быстрое или высокопроизводительное секвенирование), сборка (*de novo* или



Рисунок 2.1. Основные этапы геномного анализа

сравнительная, т. е. на основе уже известных геномов, ресеквенирование) и аннотация (структурная или функциональная) (рис. 2.1). Поскольку современные методы секвенирования не способны считывать целые геномы как непрерывную последовательность, ДНК необходимо фрагментировать для прочтения и последующей сборки. Следовательно, чем больше геном, тем выше потенциальная частота ошибок.

Метод секвенирования первого поколения (Сенгера) основан на селективном включении дидезоксирибонуклеотидов, терминирующих цепь, с помощью ДНК-полимеразы (Sanger et al., 1977); значительно усовершенствован благодаря замене радиоактивной метки на флуоресцентную (Smith et al., 1986). Этот метод секвенирования



**Рисунок 2.2.** Метод секвенирования по Сенгеру: принцип и этапы (из: Singh et al., 2022).

(рис. 2.2) использовался в Международном проекте генома человека (хотя в таком масштабе он оказался очень дорогим). До настоящего времени секвенирование по Сенгеру (также известное как секвенирование праймеров с красителем или метод обрыва цепи) остаётся «золотым стандартом» и обеспечивает наиболее точные результаты. Поэтому его используют для валидации последовательностей ДНК, полученных на платформах секвенирования нового поколения, и целевого повторного секвенирования (Pervez et al., 2022). Высокопроизводительный капиллярный секвенатор может прочитывать последовательности тысячи фрагментов ДНК длиной до 900 п. н. за один день с точностью 95–98% (Liu et al., 2012). Секвенирование генома путём терминации цепи с использованием капиллярного электрофореза включает две стратегии метода дробовика: целногеномного и иерархического. Секвенирование с использованием первой стратегии (была использована компанией Celera для расшифровки генома человека) быстрее и дешевле. Геномную ДНК сначала фрагментируют до небольших размеров, клонируют и секвенируют, а затем длинные перекрывающиеся последовательности (контиги) сравнивают и объединяют с помо-

шью BLAST или других компьютерных программ. При секвенировании с использованием второй стратегии (применена в Международном проекте «Геном человека») ДНК сначала разрезают на большие фрагменты (около 150–200 т. п. н.) и клонируют с помощью бактериальных искусственных хромосом (BAC). Затем большие фрагменты (после выявления перекрывающихся клонов) разрезают на более мелкие и секвенируют. Преимущество этого метода заключается в высокой точности сборки отдельных фрагментов ДНК в контиги, соответствующие размеру полной хромосомы (Lander et al., 2001; Venter, 2003). Метод секвенирования по Сенгеру применяется для анализа мутаций и полиморфизмов небольших фрагментов генома, для микросателлитного анализа, видовой идентификации организмов. Высокая стоимость и низкая производительность ограничивают его использование для полногеномного секвенирования (Pervez et al., 2022).

Методы высокопроизводительного секвенирования второго или следующего поколения (NGS) основаны на одновременном секвенировании (параллельном прочтении) миллионов отдельных фрагментов ДНК для получения нуклеотидных последовательностей всего генома. Эти технологии, позволяющие генерировать огромное количество коротких последовательностей (прочтений/ридов) с высокой скоростью и хорошей точностью, основаны на достижениях в области микротехнологии, визуализации с высоким разрешением и усовершенствовании программного обеспечения (Pavlopoulos et al., 2013; Head et al., 2014; Giani et al., 2020). Методы NGS применяются как для ресеквенирования, если доступен референсный (эталонный) геном, так и для секвенирования *de novo*. Они легли в основу многих новых приложений, таких как секвенирование всего экзома, определение полногеномного эпигенетического ландшафта и ChIP-seq (иммунопреципитация хроматина с последующим секвенированием), а также анализа транскриптома (Slatko et al., 2018; Giani et al., 2020; Satam et al., 2023). Помимо высокой производительности, преимуществом NGS является более низкая стоимость по сравнению с методом секвенирования первого поколения. Недостатком являются короткая длина прочтений и сложная пробоподготовка, включающая ПЦР-амплификацию фрагментов ДНК (может быть причиной систематических ошибок) (Pervez et al., 2022).

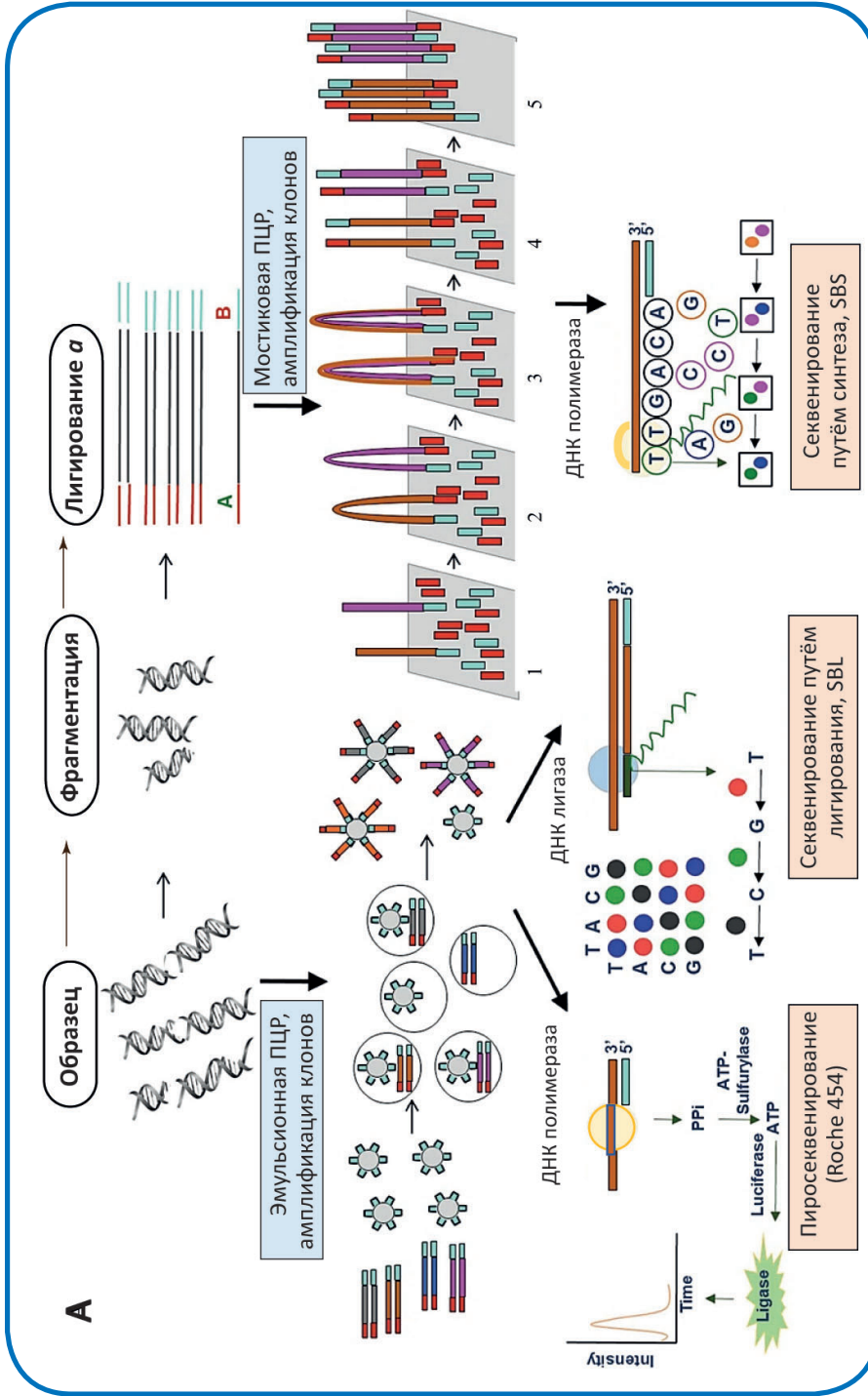
В пробоподготовке NGS опирается на приготовление ДНК-библиотек (библиотек ДНК-фрагментов), которое начинается с фрагментации (с помощью ультразвука или ферментов) длинных молекул ДНК с последующим лигированием адаптеров и амплификации, происходящей в основном на твёрдой поверхности. Подходы можно разделить на две группы: секвенирование путём синтеза (SBS) и секвенирование путём лигирования (SBL). В них клональные матричные популяции могут быть получены различными способами: на основе шариков (используется эмульсионная ПЦР, при этом один адаптер комплементарен иммобилизованному на шарике адаптеру), на твёрдой поверхности (прямой и обратный праймеры ковалентно привязаны к поверхности) и с помощью ДНК-наношариков (амплификация происходит в раство-

ре, по катящемуся кругу; синтезированная ДНК сворачивается в нанопарики, которые затем наносят на субстрат) (Goodwin et al., 2016) (рис. 2.3А, 2.3Б).

Методы SBS зависят от ДНК-полимеразы. При секвенировании молекулярных кластеров после нескольких циклов амплификации тысячи клональных копий группируются вокруг одной ДНК-матрицы, и на проточную ячейку могут быть получены миллионы кластеров, возникших из разных матриц. Секвенирование происходит в «дорожках» (микрожелобках) или лунках твёрдой поверхности; включение меченого нуклеотида прекращает удлинение ДНК. Сигнал (флуоресценция) обнаруживается после ферментативного расщепления метки и включения следующего нуклеотида. Помимо коротких прочтений, недостатком SBS является то, что, во-первых, разные фрагменты кластера секвенируются на разных фазах, и это приводит к получению фрагментов разной длины, а во-вторых, гомополимерные последовательности читаются плохо (Goodwin et al., 2016; Slatko et al., 2018).

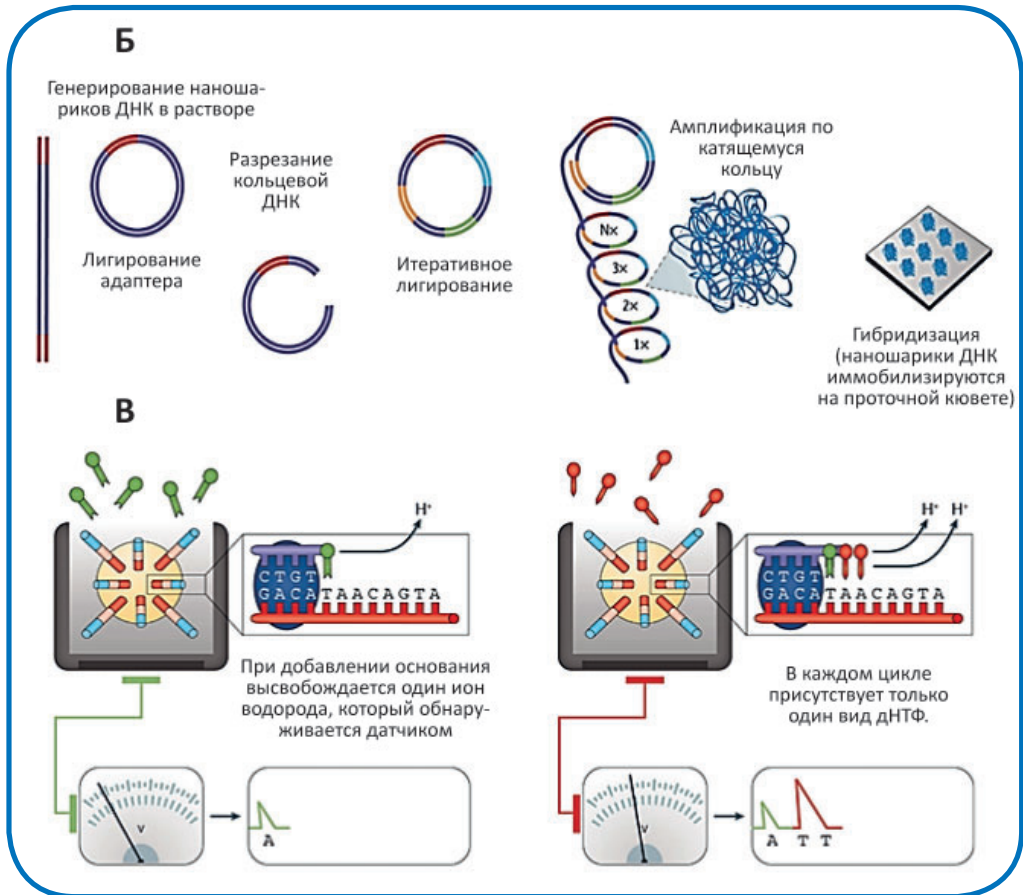
SBL – это ДНК-лигазозависимый подход, который основан на способности ДНК-лигазы обнаруживать и включать основания через 5'-фосфатную группу вместо 3'-гидроксильной группы, как в стандартной реакции, опосредованной полимеразой. В нём специфические праймеры гибридизуются с адаптерами ДНК-матриц для лигирования с флуоресцентными зондами (длиной 8 п. н., из которых первые два основания являются специфичными, а остальные – вырожденными), а лигирование сопровождается флуоресценцией. После серии циклов лигирования продукт удаляют, а матрицу перезагружают для второго раунда циклов лигирования со смещением праймера на один нуклеотид. Таким образом, каждый нуклеотид матрицы секвенируется дважды, что является одной из причин сравнительно более высокой точности этого метода. Однако по сравнению с секвенированием на кластерах SBL обеспечивает меньше выходных данных и даёт более короткие прочтения (Goodwin et al., 2016).

Пиросеквенирование – один из первых методов NGS и первая альтернатива методу Сенгера для секвенирования ДНК *de novo*, который основан на определении пирофосфата, образующегося после включения очередного нуклеотида при синтезе комплементарной цепи ДНК (рис. 2.3А). В пробоподготовке фрагменты ДНК лигируют с адаптерами и амплифицируют с помощью ПЦР в индивидуальных эмульсионных шариках. Обнаружение высвободившегося пирофосфата возможно благодаря его связыванию с люциферазой и получению светового сигнала. Метод позволяет эффективно секвенировать множество коротких последовательностей (до 200 п. н.), обнаруживать мутации одиночных нуклеотидов и проводить ресеквенирование (Ronaghi, 2001; Margulies et al., 2005; Montero et al., 2008). Технология пиросеквенирования нового поколения (анализатор ДНК 454 GS FLX) обеспечивает длину считывания 700 п. н. с точностью 99,9% (Liu et al., 2012; Slatko et al., 2018). Пиросеквенирование широко использовалось в разных приложениях (например, в молекулярной диагностике) и считалось «золотым стандартом» платформ для количественного locus-специфического анализа метилирования ДНК (Nyrén, 2015).



**Рисунок 2.3.** Схематическое изображение основных этапов секвенирования с короткими прочтениями (из: Goodwin et al., 2016; Gupta, Verma, 2019).

A – пиросеквенирование, секвенирование путём лигирования и путём синтеза: Лигирование a – репарация концов и лигирование адаптера; 1 – связывание с праймером, 2, 3 – ПЦР, 4 – диссоциация, 5 – формирование кластеров.



**Рисунок 2.3** (продолжение). Схематическое изображение основных этапов секвенирования с короткими прочтениями (из: Goodwin et al., 2016; Gupta, Verma, 2019). Б – секвенирование наночастиц ДНК; В – полупроводниковое секвенирование.

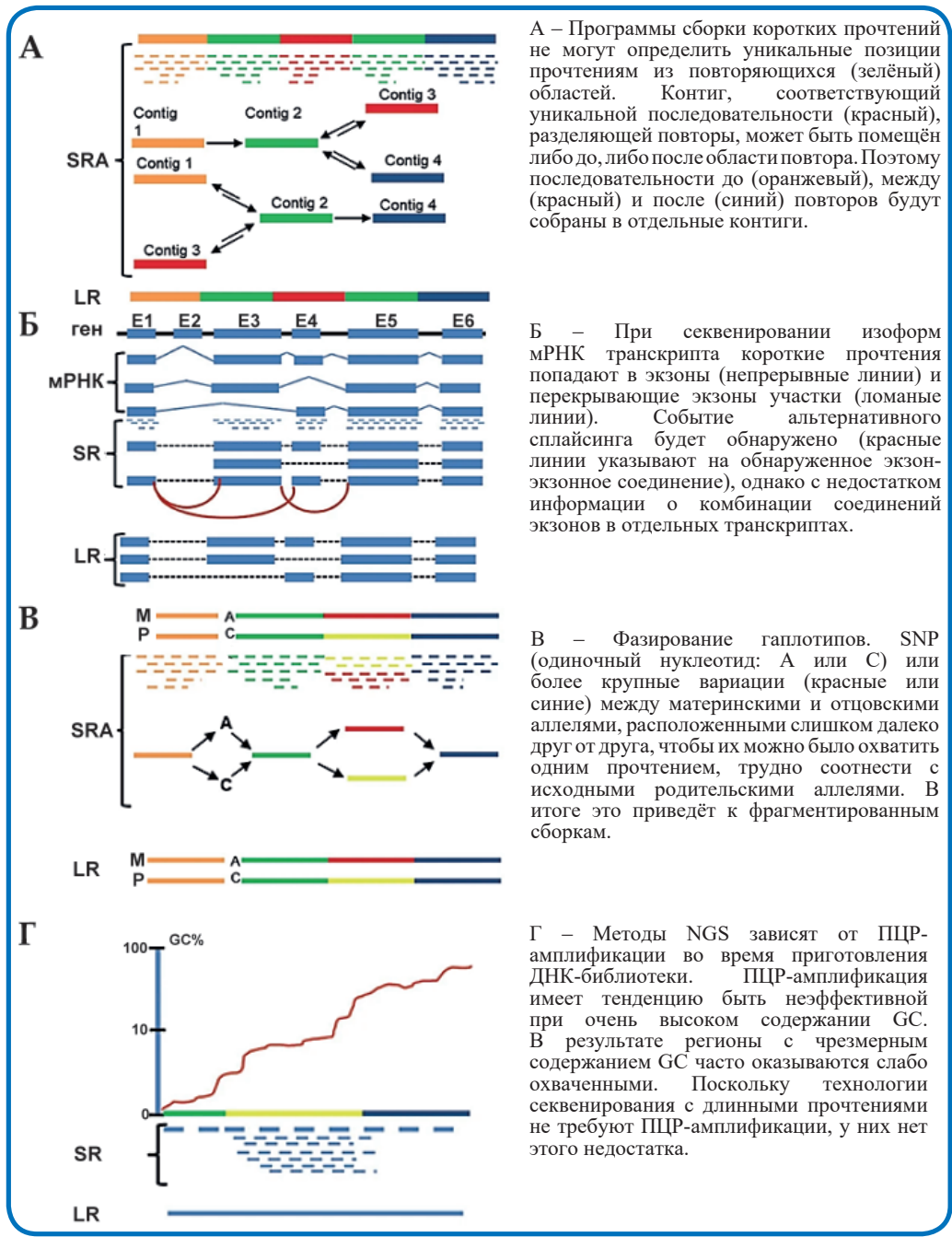
Полупроводниковое (или pH-индуцированное) секвенирование не требует модификации нуклеотидных оснований, хемилюминесцентных или флуоресцентных меток. Процесс основан на идентификации ионов водорода (протонов), образующихся во время синтеза дочерней нити ДНК в качестве побочного продукта, что приводит к изменениям pH раствора, которые легко регистрируются (рис. 2.3В). Этот метод может быть полезен для секвенирования ДНК/РНК *de novo*, анализа микросателлитов, выявления вариаций числа копий, проведения ChIP-seq и т. д. (Rusk, 2011; Slatko et al., 2018).

Хотя методы NGS значительно ускорили секвенирование генома и снизили стоимость его прочтения, они не могут обеспечить высокое качество генома, собранно-

го *de novo*, и повторное секвенирование также существенно ограничено. Причина в том, что в коротких прочтениях могут отсутствовать целые части геномов, содержащие некоторые важные гены, и это в ряде случаев приводит к меньшему размеру собранных последовательностей по сравнению с референсным геномом. Кроме того, методы NGS малоэффективны при поиске геномных перестроек и мобильных генетических элементов (Lee et al., 2016). При сборке геномов по данным NGS сложно определить точное положение для прочтений из повторяющихся областей; не обеспечивается получение достаточной информации об альтернативном сплайсинге; возникают трудности фазирования (отнесение прочтений к определённым родительским аллелям), приводящие к фрагментированным сборкам; а зависимость от ПЦР-амплификации ДНК (неэффективная при очень высоком содержании GC оснований) приводит к тому, что GC-богатые регионы при секвенировании оказываются слабо охваченными (Singh et al., 2022) (рис. 2.4).

Подходы секвенирования третьего поколения (NNGS/TGS) основаны на более простой пробоподготовке, не требуют амплификации для увеличения сигнала от каждого фрагмента и большого количества повторных секвенирований, и имеют сокращённое время цикла. В результате одиночные прочтения значительно длиннее, что обеспечивает более непрерывную реконструкцию генома, хотя частота ошибок при этом выше, чем при коротких прочтениях (Hook, Timp, 2023; Satam et al., 2023). Длинные прочтения могут улучшить анализ таких структурных изменений, как крупные вставки, делеции и транслокации. TGS вместе с технологиями картирования позволяет формировать суперконтиги (скаффолды), охватывающие почти все плечи хромосом, что значительно улучшает сборку *de novo* и структурный анализ геномов. Эти методы также дают возможность измерять эпигенетические модификации отдельных молекул, проводить полногеномное профилирование паттернов метилирования, анализ конформации хроматина и модификации гистонов; они широко используются для изучения альтернативного сплайсинга, метилирования РНК и анализа длинных некодирующих РНК (Lee et al., 2016; Pervez et al., 2022; Satam et al., 2023). Сочетание технологий секвенирования короткими и длинными прочтениями для секвенирования новых (ранее не изученных) геномов успешно применяется для точной сборки генома без референсной последовательности (Satam et al., 2023).

Секвенирование с длинными прочтениями способствует изучению повторяющихся последовательностей (тандемных и диспергированных), которые могут играть исключительно важную роль в организации, стабильности, регуляции и эволюции эукариотического генома. Они формируют домены хроматина и влияют на трёхмерную организацию ДНК, могут изменить экспрессию генов и эволюцию фенотипа, участвуют в формировании гибридной несовместимости между видами и т. д. (Sproul et al., 2023). Например, исследование геномов более 600 видов насекомых показало, что повторяющихся последовательностей в сборках с длинными прочтениями больше на 36%, чем в сборках с короткими прочтениями. При этом обнаружение длинных концевых повторов увеличилось на 162%, хотя для ДНК-транспозонов и длинных



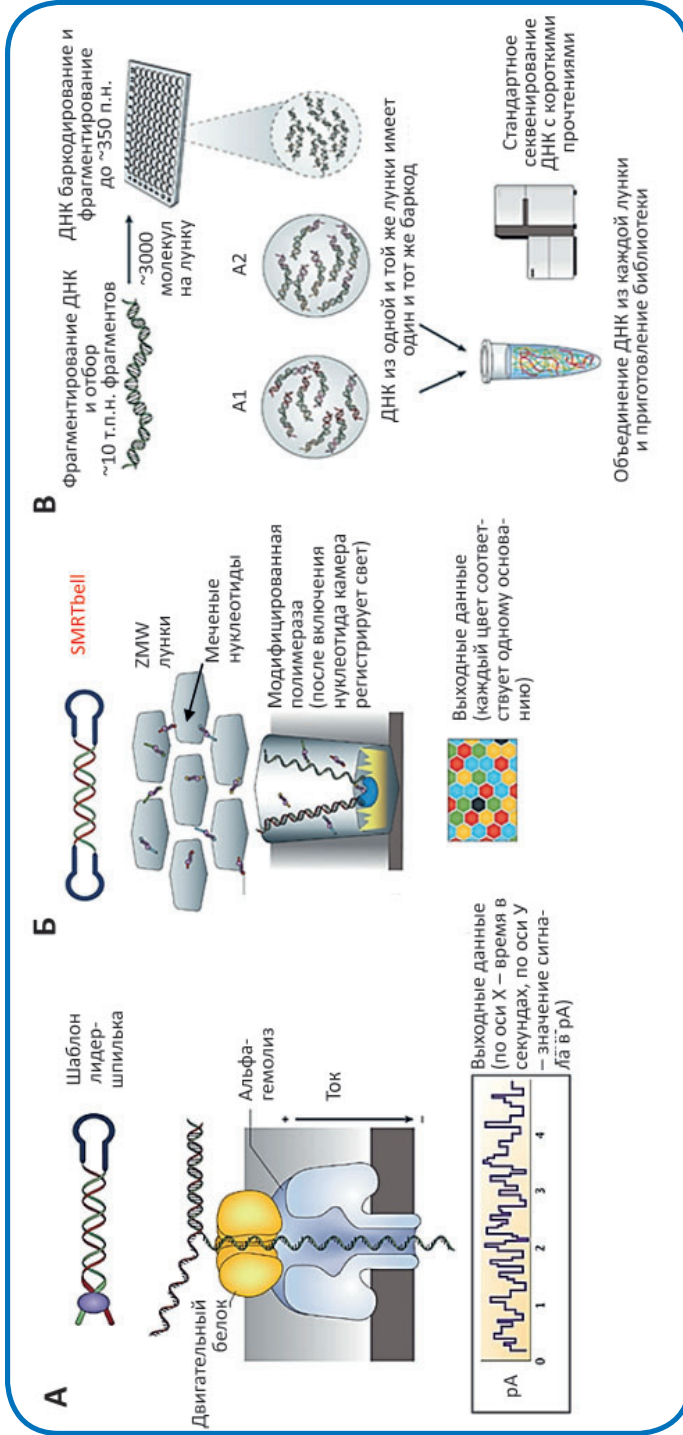
**Рисунок 2.4.** Недостатки технологий секвенирования с короткими прочтениями (из: Singh et al., 2022).

LR – длинные прочтения; SR – короткие прочтения; SRA – сборка коротких прочтений; E – экзон; М – материнские аллели, Р – отцовские аллели.

диспергированных повторов (LINEs) зависимость от метода секвенирования была менее выраженной (Sproul et al., 2023). Высокий потенциал длинных прочтений для понимания эволюции подчёркивают результаты анализа геномов некоторых видов рыб. В частности, в них были обнаружены доказательства исторических изменений активности мобильных элементов у полярных рыб, которые, вероятно, отражают их реакцию на изменения в окружающей среде. Интересно, что в одной из линий была выявлена специфическая утрата кластера генов антифризных белков (обычно располагаются в высокоповторяющихся tandemных массивах), связанных с появлением большого количества костистых рыб, места обитания которых обычно находятся ниже точки замерзания и покрыты льдом (Hotaling et al., 2023).

Одномолекулярное секвенирование с длинными прочтениями – подход, позволяющий считывать нуклеотидную последовательность непосредственно с отдельных молекул ДНК или РНК; включает наиболее часто используемый метод SMRT (секвенирование одной молекулы в реальном времени) и нанопоровое секвенирование (Eid et al., 2009; Eisenstein, 2012). В то время как в технологиях с короткими прочтениями ДНК-полимераза перемещается по связанной ДНК-матрице, в технологиях с длинными прочтениями, таких как SMRT, молекула полимеразы фиксируется на дне ячейки (ZMW – волноводы нулевой моды, сквозь дно которых подаётся свет), позволяя цепи ДНК проходить через ZMW. Включение каждого нуклеотида визуализируют с помощью лазерной и конфокальной микроскопии высокого разрешения; за включением нуклеотида следует высвобождение фрагмента флуорофора (рис. 2.5Б). Этот метод позволяет считывать фрагменты ДНК длиной до 100000 п. н. десятки раз с консенсусной точностью 99,9% (Lee et al., 2016). Кроме того, консенсусные последовательности генерируются из одномолекулярных прочтений с 15-кратным покрытием, демонстрируя медианную точность 99,3% без систематических ошибок, кроме частот ошибок, зависящих от флуорофора (Eid et al., 2009; Goodwin et al., 2016; Slatko et al., 2018).

Нанопоровые секвенаторы напрямую определяют нуклеотидный состав нативной молекулы одноцепочечной ДНК. Длина прочтений может достигать сотен тысяч и даже миллионов пар нуклеотидов, но их качество по сравнению с NGS ниже. ДНК проходит через пору (биологическую, состоящую из трансмембранных белков, встроенных в липидную мембрану, или из синтетического материала) под воздействием электрического поля, что приводит к изменению напряжения (рис. 2.5А). Оцениваются различные типы сигналов, включая блокаду дифференциального ионного тока через пору, туннельные токи через пору или изменения ёмкости поры (Goto et al., 2020). Скорость прохождения ДНК через поры может регулироваться ДНК-полимеразой  $\phi$ 29, ДНК-хеликазой и экзонуклеазой 1. Продолжительность и величина падения силы тока ионов зависят от того, какой нуклеотид в данный момент находится внутри поры (типичное время прохождения ДНК через нанопору составляет <1 мкс/нуклеотид) (Goto et al., 2020). Эти технологии (с использованием структуры библиотеки «лидер–шпилька») также позволяют считывать обе цепи ДНК



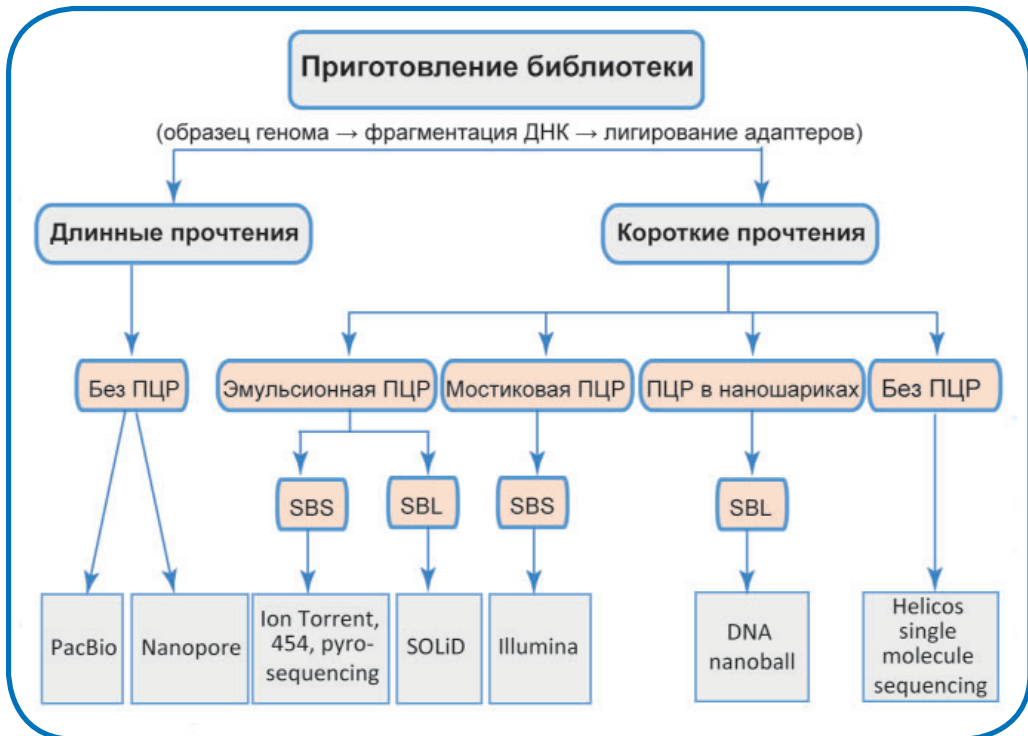
**Рисунок 2.5.** Схематическое изображение основных этапов секвенирования с длинными прочтениями (из: Goodwin et al., 2016).

A – нанопоровое (Oxford) секвенирование: лидерная последовательность взаимодействует с порой и «моторным» белком, направляя ДНК; шпилька позволяет проводить двунаправленное секвенирование;  
 Б – секвенирование в реальном времени (PacBio); использует SMRTbell – замкнутый круговой шаблон с двумя адаптерами-шпильками, позволяющими выполнять непрерывное циклическое секвенирование;  
 В – синтетическое секвенирование (Illumina).

(сначала прямую цепь ДНК пропускают через пору, а затем – обратную) для создания консенсусной последовательности «2D» прочтения и увеличения точности секвенирования (Eisenstein, 2012; Lee et al., 2016; Slatko et al., 2018). Благодаря портативности и простоте прибора, позволяющими работать даже в полевых условиях, а также низкой стоимости, метод имеет множество применений, включая генотипирование и мониторинг окружающей среды, а также надзор за инфекционными заболеваниями, такими как геморрагическая лихорадка Эбола и лихорадка Зика. Oxford Nanopore (MinION) – первая платформа для секвенирования ДНК с помощью нанопор, использованная в космосе (Slatko et al., 2018; Gupta, Verma, 2019).

Технологии синтетического длинного прочтения (SLR) основаны на системе штрихового кодирования. Сначала большие фрагменты ДНК распределяют в микротитровальные лунки или эмульсию (с очень небольшим количеством молекул). Затем каждый фрагмент матрицы разрезают на более мелкие части, к которым лигируют штрихкоды (в каждой лунке – свой штрихкод); SLR состоят из коротких прочтений с одним и тем же штрихкодом (рис. 2.5B). Такая стратегия секвенирования обеспечивает ассоциацию между небольшими фрагментами, полученными из одного большого, предотвращая появление неразрешённых точек ветвления в сборках (причина пробелов и укорочения смежных последовательностей) (McCoey et al., 2014; Liu et al., 2021). Новая технология Illumina с малой частотой ошибок, называемая синтетическими длинными считываниями TruSeq, позволяет размещать копии отдельных высокоповторяющихся последовательностей (например, мобильных элементов) в соответствующих геномных областях и точно реконструировать их последовательности. Это важно, потому что в настоящее время сборка сложных геномов остаётся трудной задачей, частично из-за присутствия диспергированных повторов (McCoey et al., 2014). Обобщенная схема технологий секвенирования второго и третьего поколений представлена на рисунке 2.6.

Благодаря методам NGS была создана новая область исследований – анализ генома отдельных клеток (single-cell DNA sequencing). Использование платформы TGS значительно повысила возможности метода (Kong et al., 2023). Протокол подготовки включает изоляцию и обогащение единичных клеток, приготовление их суспензий, а также полногеномную амплификацию для создания ДНК-библиотек. Используются разные подходы амплификации генома, например, с высокоточной и высокопроцессивной полимеразой phi29 в изотермической реакции, или с циклами множественного отжига на основе петель, когда полные ампликоны «запечатываются» в петли для предотвращения повторной амплификации. Несмотря на то, что полногеномная амплификация имеет ограничения, связанные с низким охватом, систематической ошибкой амплификации и «выпадением» аллелей, секвенирование генома отдельных клеток в целом является высокоэффективным и надёжным методом для выявления различных типов мутаций, эпигенетического профилирования и анализа транскрипции. Метод оказал большое влияние на различные области науки, и показательным примером является создание Атласа клеток человека (HCA), который представляет собой трёх-



**Рисунок 2.6.** Различные технологии секвенирования новых поколений с разными принципами и на разных платформах (из: Satam et al., 2023). SBS – секвенирование путём синтеза; SBL – секвенирование путём лигирования.

мерную карту каждого типа клеток. Секвенирование генома отдельной клетки, прежде всего, используется для анализа микробиомов (Pachiadaki et al., 2017), исследования нейронов и геномной гетерогенности злокачественных опухолей (Francis et al., 2014; Hu et al., 2016; Kong et al., 2023). Появилась возможность выявить редкие популяции клеток, раскрыть механизмы, лежащие в основе формирования клеточной идентичности, дать представление об эволюционных отношениях между популяциями клеток и т. д. Современные технологии позволяют одновременно секвенировать геномы сотен тысяч клеток в одном эксперименте. Полученные данные необходимо обрабатывать вычислительно эффективным и статистически обоснованным способом, и эта проблема активно решается (Lähnemann et al., 2021).

Однако при использовании метода секвенирования генома единичных клеток теряется гистологический контекст. Кроме того, некоторые клетки в процессе выделения могут разрушаться/повреждаться, т. е. становятся непригодными для анализа. Вместе с тем, органы и ткани состоят из множества клеток, выполняющих специализированные функции, окружающая среда которых может влиять на них, регулируя экспрессию генов (например, через внешние сигналы, возникающие в результате локальной

межклеточной коммуникации) (Magouloupolou et al., 2023). Поэтому для выяснения пространственно-временных изменений в экспрессии генов отдельных клеток были разработаны новые подходы. Методы секвенирования четвёртого поколения, такие как секвенирование *in situ* (ISS), сохраняют пространственные координаты последовательностей РНК и ДНК с субклеточным разрешением. Они сочетают в себе традиционные способы анализа изображений и современные технологии секвенирования, предоставляя новые возможности в изучении гетерогенности тканей. Методы интегрировали технологию нанопорового секвенирования (самая популярная платформа для секвенирования четвёртого поколения), которое выполняется в реальном времени без амплификации и повторных циклов (Pervez et al., 2022). Дальнейшие разработки, способные обеспечить пространственную и временную информацию в геноме отдельных клеток, откроют новые возможности в геномных исследованиях и приложениях. Например, в нейробиологии (для углублённого понимания развития мозга и неврологических расстройств), в онкологии (для определения сложности заболевания, составления графиков развития злокачественных опухолей и т.п.), при инфекционных заболеваниях (идентификация патогенов) и в персонализированной медицине (Ke et al., 2016; Moreno-Ayala, Junker, 2021; Magouloupolou et al., 2023).

Технологии NGS были разработаны коммерческими компаниями Illumina, Roche и Applied Biosystems, и они постоянно совершенствуются (таб 2.1). Например, новая система NovaSeq 6000 (Illumina) за один запуск прибора может секвенировать 48 геномов человека с более чем 30-кратным покрытием. Методы TGS, такие как Pacific Biosciences, Oxford Nanopore и Complete Genomics, позволяют (опираясь на референсный геном) проводить секвенирование генома человека по низкой цене и в течение всего нескольких часов (Pavlopoulos et al., 2013; Pervez et al., 2022). Платформа Oxford Nanopore ProMethION предлагает секвенирование по требованию, т.е. даёт возможность запускать и останавливать цикл по мере необходимости, или использовать более одной проточной кюветы для одного эксперимента, обеспечивая высокую производительность и скорость. Она позволяет использовать отдельные проточные ячейки с 3000 нанопорами, что гарантирует получение почти в шесть раз больше данных по сравнению с MinION и GridION. Бета-системный модуль ProMethION предусматривает использование 192 различных ДНК-библиотек внутри всего устройства, а длина прочтений при этом составляет 4Mb с выходными данными объёмом 14000 Gb (Gupta, Verma, 2019; Pervez et al., 2022).

**Сборка и аннотирование геномов.** NGS генерирует колоссальное количество последовательностей ДНК или РНК, которые необходимо обрабатывать, анализировать и интерпретировать для получения биологической информации. В этих целях используют подходы биоинформатики, охватывающие широкий спектр вычислительных методов, алгоритмов и инструментов. Они позволяют проводить сборку *de novo*, картирование на референсный геном и анализ транскриптома, а также облегчают идентификацию структурных вариантов, однонуклеотидных полиморфизмов, вариаций числа копий и т.д. Основные биоинформационные инструменты, используемые

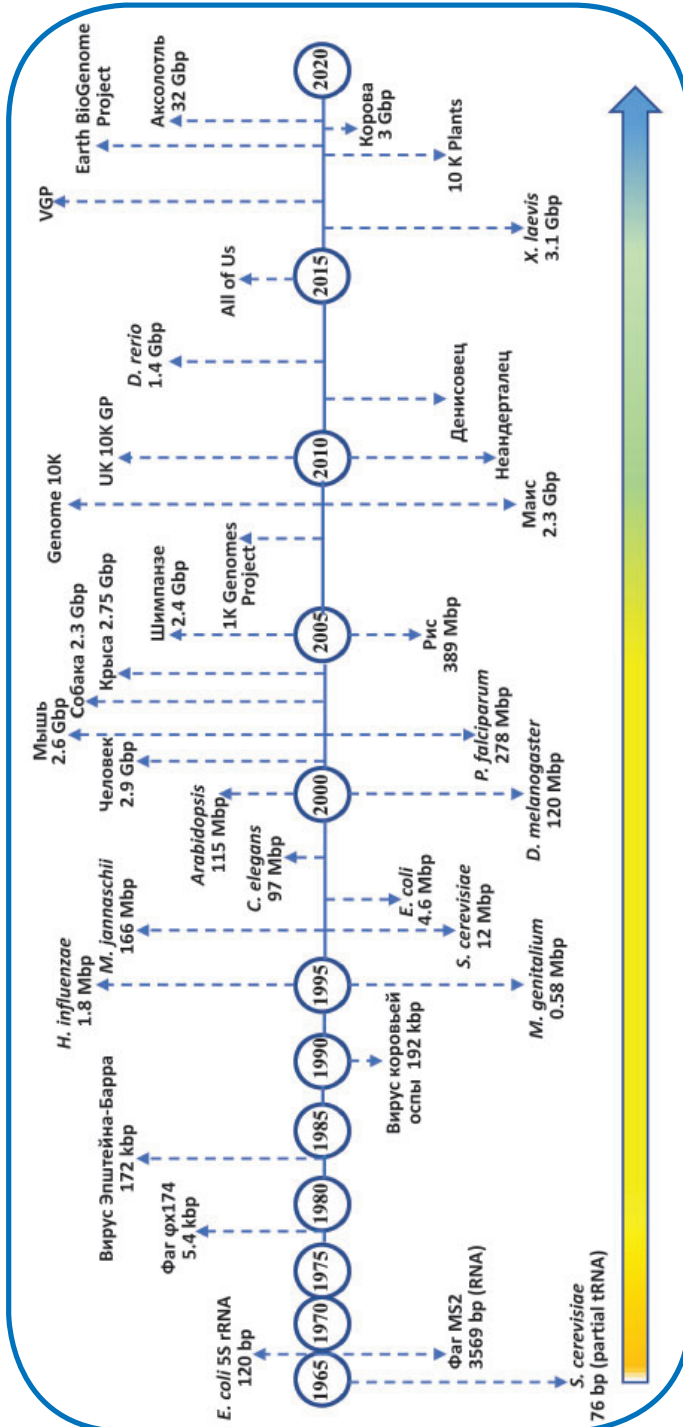
Таблица 2.1. Современные высокопроизводительные платформы для секвенирования геномов  
(из: Pervez et al., 2022)

№	Производитель	Длина прочтений (п.н.)	Выходные данные (Gb)	Максимальное время (часы)	Метод	Основное применение
1	Illumina (NovaSeq 6000)	300	6000	44	Секвенирование путём синтеза	Полногеномное секвенирование малых (бактерии, вирусы) и больших (животные, растения, человек) геномов, секвенирование и профилирование экспрессии целевых генов, полногеномное и метагеномное секвенирование, профилирование одиночных клеток, анализ профиля метилирования генома.
2	Ion Torrent (Ion GeneStudio S5 Prime)	600	50	12		Секвенирование полных геномов, экзоменов и целевых генов.
3	BGI/Complete Genomics	400	6000	40	Секвенирование с помощью ДНК наночастиц	Секвенирование больших и малых геномов, экзоменов и целевых генов.
4	RacBio (прочтения HiFi)	25 тыс.	66,5	30	Секвенирование в реальном времени	Секвенирование <i>de novo</i> , секвенирование транскриптома; идентификация изоформ альтернативного сплайсинга, мутаций и эпигенетических модификаций.
5	Nanopore (PromethION)	4 млн.	14000	72		Сборка генома; фазирование (определение принадлежности аллелей к одной из хромосом); идентификация структурных вариантов, изоформ, модифицированных оснований ДНК и РНК; секвенирование транскриптома.

для анализа данных NGS в геномных, эпигеномных, транскриптомных и метагеномных исследованиях приведены в работе (Satam et al., 2023). Интегративный анализ, объединяющий данные NGS/TGS с другими источниками геномных и функциональных данных, позволяет исследовать экспрессию генов и их регуляторные сети (Satam et al., 2023). Хронология достижений в сборке генома представлена на рисунке 2.7.

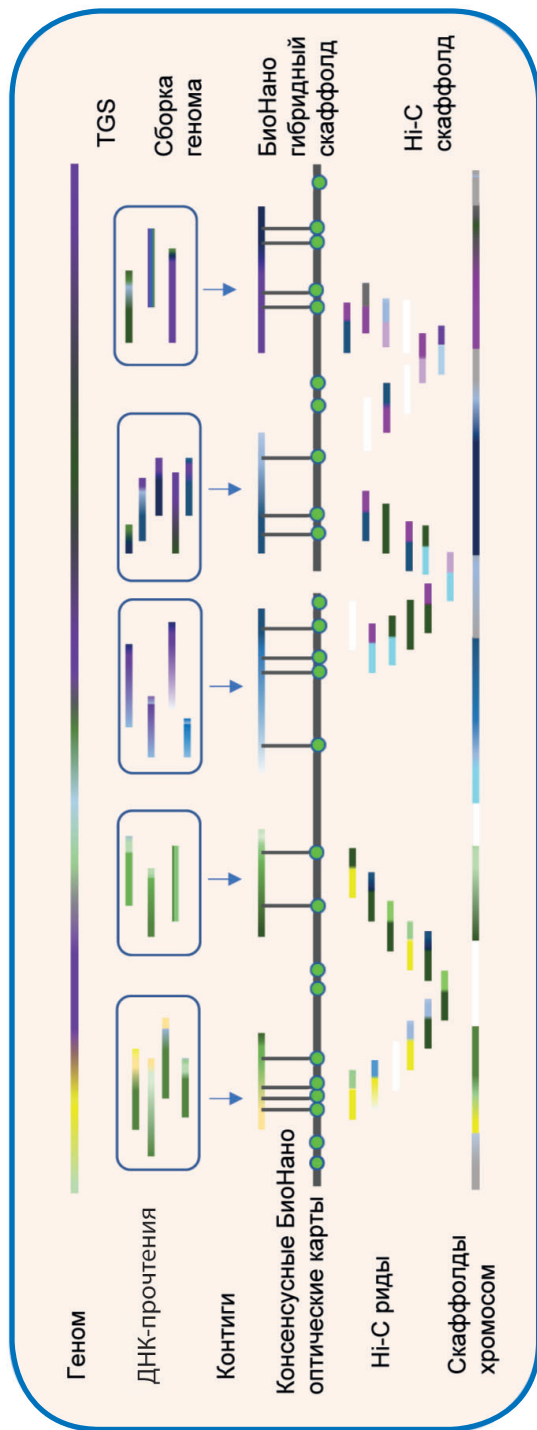
Сборка *de novo* сложна в вычислительном отношении (особенно при использовании NGS) и часто использует две основные стратегии: путь Эйлера и «перекрытие–макет–консенсус» (Pop, 2009). Для сборки большого количества прочтений было предложено множество алгоритмических и эвристических методов. Однако, многочисленные технические или вычислительные проблемы при сборке *de novo* всё ещё остаются. Ошибки при секвенировании (увеличивающиеся с длиной прочтений), а также гетерозиготность, повторяющиеся последовательности (повышают число пробелов в сборках и вызывают неравномерную глубину прочтений) и возможная контаминация образца могут существенно осложнить сборку генома (Sohn, Nam, 2018). Базовая стратегия сборки *de novo* для коротких прочтений NGS включает три этапа: сборка контигов, формирование скаффолда и заполнение пробелов (рис. 2.8). В зависимости от качества, сборки геномов подразделяют на стандартные черновые (первичная сборка), черновые высокого качества (охватывают до 90% целевого генома, не имеют загрязнений и подходят для составления аннотаций), улучшенные черновые геномы высокого качества (без ошибок сборки и без пробелов в скаффолдах; используется в сравнительной геномике), ориентированные на аннотации улучшенные сборки (ошибки в белок-кодирующих последовательностях сведены к минимуму), почти законченные (закрыты большинство пробелов, ошибок сборки, включая области с повторами) и законченные/финишные (менее 1 ошибки на 100000 нуклеотидов; целые хромосомы) сборки. Чтобы получить аннотированные геномы из контигов, используют такие инструменты (существенно улучшающие геномные сборки), как Abacas (для упорядочивания и ориентации контигов вдоль референсного генома), Image (для закрытия пробелов), Iscon (для коррекции мелких ошибок путём сравнения прочтений с референсным геном) и RATТ (для переноса аннотации из референсного или близких геномов в черновой вариант сборки) (Swain et al., 2012).

Следует отметить, что большая часть геномныхборок имеет отклонения от ожидаемого размера генома более чем на 10%, причём эти отклонения возрастают с увеличением размера генома (т. е., причиной могут быть структурные ДНК и небелковые кодирующие участки). В связи с этим была предложена новая пропорциональная метрика N50, при расчёте которой учитываются средние размеры хромосом каждого вида. Такой подход обладает более высокой степенью корреляции с полными сборками генома и позволяет проводить более точное сравнениеборок для геномов, различающихся по размеру и архитектуре (Hjelmen, 2024). Новые технологии значительно упрощают процесс сборки в масштабе хромосом, обеспечивая меньшие затраты и большую скорость по сравнению с традиционными методами (Rice et al., 2019). Недавние достижения в области алгоритмов и протоколов сборки стали основой для



**Рисунок 2.7.** Вехи сборки генома (по: Gianni et al., 2020).

Хронология, иллюстрирующая ряд крупных достижений в сборке генома, от зарождения эры секвенирования до нынешних крупномасштабных геномных проектов (оранжевый – ранние методы секвенирования, жёлтый – методом дробовика по Сенгеру, зелёный – NGS, и синий – TGS).  
*C. elegans* – *Caenorhabditis elegans*, *D. melanogaster* – *Drosophila melanogaster*, *E. coli* – *Escherichia coli*, *H. influenzae* – *Haemophilus influenzae*, *M. jannaschii* – *Methanococcus jannaschii*, *M. genitalium* – *Mycoplasma genitalium*, *P. faiciparum* – *Plasmodium falciparum*, *S. cerevisiae* – *Saccharomyces cerevisiae*, *X. laevis* – *Xenopus laevis*.



**Рисунок 2.8.** Сборка генома *de novo* на уровне хромосом одним из методов (по: Giani et al., 2020).

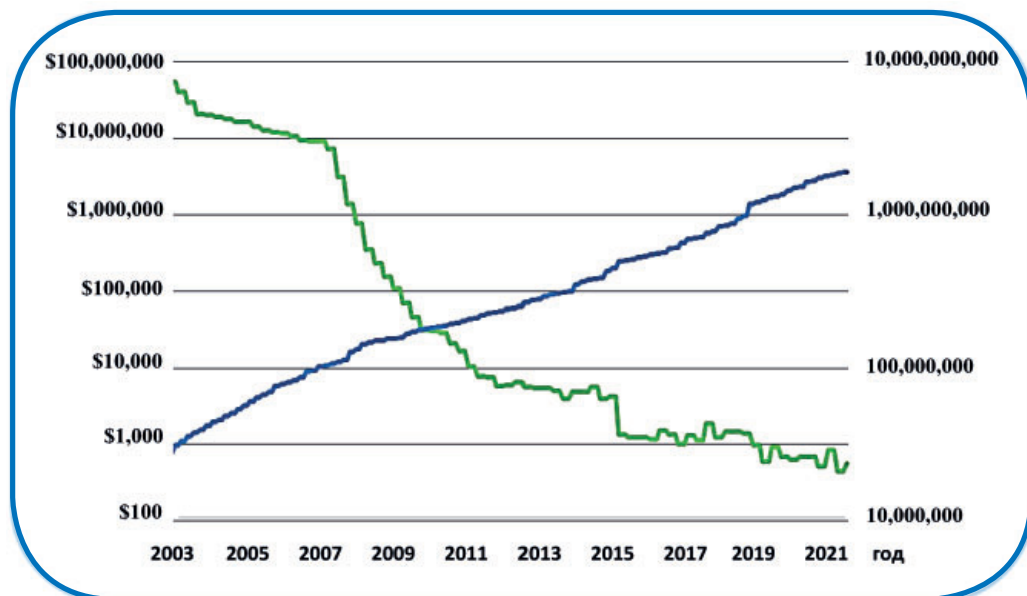
1 – ДНК-прочтения (риды) собираются вместе, чтобы сформировать контиги; 2 – контиги выравниваются по консенсусным оптическим картам BioNano (позволяет точно упорядочивать и ориентировать контиги относительно друг друга и назначать каждый контиг определенной хромосоме); оптические карты BioNano представлены черными линиями, а зелёные точки указывают на маркированные мотивы последовательности; 3 – дальнейшее упорядочение скаффолда возможно благодаря картам Hi-C, которые учитывают взаимодействия между смежными геномными локусами. В полученном скаффолде на уровне хромосом цветные сегменты представляют геномные области с известной последовательностью, а оставшиеся промежуточные каркасами неизвестной последовательности изображены серыми линиями.

разработки методов, позволяющих получать сборки с околотеლოмерными участками, а также устранять оставшиеся пробелы в процессе сборки и осуществлять сборку недиплоидных геномов (Li, Durbin, 2024).

Аннотация (*in silico* и/или вручную) зависит от качества сборки генома и включает идентификацию частей генома, которые не кодируют белки, предсказание генов (идентификацию кодирующих областей) и биологическую информацию (функции) к этим элементам (Brent, 2008). Структурная аннотация включает обнаружение экзонов, интронов, промоторов, транспозонов и т. д. Первый шаг структурной аннотации – проверка последовательности ДНК на наличие вкраплений повторов и последовательностей ДНК низкой сложности, например, с помощью программы RepeatMasker. Установление основных характеристик повторяющихся элементов до сих пор остаётся сложной задачей. Идентификация (предсказание) генов, кодирующих белки, занимает центральное место в аннотации геномов. Используются разные подходы к предсказанию генов: основанные на данных нуклеотидной последовательности (методы *ab initio*), меток экспрессируемых последовательностей (EST; короткие нуклеотидные последовательности, полученные из целого транскрипта), комплементарной ДНК (кДНК) или белка, а также сочетающие в себе подходы *ab initio* и подходы, основанные на фактических данных или гомологии (Ejigu, Jung, 2020). Базы данных для структурных аннотаций включают такие глобальные общедоступные ресурсы, как GenBank (поддерживается Национальным центром биотехнологической информации США, NCBI), ENA (Европейский архив нуклеотидов), DDBJ (Банк данных ДНК Японии), а также специализированные базы данных, например, NONCODE, Pseudogene.org, Dfam и miRbase, предоставляющие важную информацию для структурной аннотации некодирующих РНК, псевдогенов, мобильных элементов и микроРНК соответственно и др. (Ejigu, Jung, 2020).

Ассоциация биологической информации с последовательностями генов или белков, которые были идентифицированы с помощью структурной аннотации, называется функциональной аннотацией (также используется для оценки вариаций генов) (Ejigu, Jung, 2020). Базы данных для функциональных аннотаций включают ресурс геномной онтологии (GO; охватывает три аспекта функции гена: молекулярную функцию, клеточный компонент и биологический процесс), Киотскую энциклопедию генов и геномов (KEGG; связующее звено между геномными данными и функциональной информацией более высокого порядка), базу знаний Reactome Pathway (фокусируется на *Homo sapiens*, связывая белки человека с их молекулярными процессами), базу данных геномных вариантов (DGV; предоставляет исчерпывающую информацию о геномных вариациях размером более 50 п. н. из сегментов ДНК генома человека) и др. (Ejigu, Jung, 2020).

Сравнительная аннотация позволяет спроецировать аннотации хорошо изученного генома на эволюционно близкие виды; использует методы *ab initio* (прогнозирование делается на основе рассматриваемой геномной последовательности) и методы, основанные на гомологии (для прогнозирования и аннотирования генов путём выяв-



**Рисунок 2.9.** Линейная диаграмма, отображающая стоимость секвенирования последовательности генома (зелёная линия) и кумулятивное количество последовательностей генома (синяя линия) с течением времени в логарифмической шкале (из: National Human Genome Research Institute, Gen Bank, BlackRock, 2022).

Показано, как стоимость секвенирования генома резко снизилась за последние два десятилетия, что привело к экспоненциальному росту числа завершённых последовательностей генома.

ления значимых совпадений из хорошо аннотированной последовательности генома с помощью методов выравнивания) (Ejigu, Jung, 2020). Хотя ручное аннотирование по-прежнему считается золотым стандартом, инструменты локального выравнивания, такие как простой в использовании BLAST, во многих случаях дают хорошие результаты. Для анализа больших объёмов данных, генерируемых NGS, используются биоинформационные конвейеры аннотаций – наборы алгоритмов, которые обрабатывают данные секвенирования и выполняются в заранее определенном порядке (подробно описано в работе: Ejigu, Jung, 2020).

Развитие технологий значительно снизило затраты на секвенирование геномов. По данным Национального исследовательского института генома человека (США) стоимость секвенирования полного генома человека в 2001, 2005, 2010, 2015 и 2021 гг. составила в долларах примерно 95,6 млн, 13,8 млн, 29091, 1245 и 454 соответственно, что ожидаемо привело к увеличению темпов накопления полногеномных данных (рис. 2.9). В базе данных NCBI к началу 2025 г. были доступны более 49000 геномов эукариот (в том числе около 6600, 5500, 4000, 2800 и 2000 – для насекомых, высших растений, млекопитающих, костистых рыб и птиц), а также более 2,5 млн геномов бактерий. В рамках проекта Earth BioGenome предполагается секвенирова-

ние всех существующих эукариот (примерно 2 млн видов) с высоким качеством для создания цифровой библиотеки жизни на Земле. Эти ресурсы будут способствовать решению многих задач научного (например, понимание видообразования, эволюции и адаптации) и прикладного (таких как охрана природы, развитие сельского хозяйства и медицины) характера.

### ГЛАВА 3. РАЗДЕЛЫ ГЕНОМИКИ

Геномику можно разделить на «естественную» геномику – геномику видов природного происхождения, исследования в рамках которой проводятся на геномах современных организмов, а также древней ДНК вымерших видов/популяций (т. е. палео/музейную геномику), и «синтетическую» геномику – область исследований, направленных на создание и изучение искусственных геномов (рис. 3.1). Основные разделы геномики включают структурную, функциональную и сравнительную геномику, для которых необходимы полные последовательности генома. Другие разделы



Рисунок 3.1. Основные разделы геномики (по: Chelomina, 2024a).

представляют исследования всего генома без необходимости полных данных о последовательностях ДНК. Это метагеномика (сосредоточенная на сообществе видов), популяционная геномика (с концентрацией внимания на популяциях отдельных видов), ландшафтная геномика (интегрирующая данные популяционной геномики и информацию об окружающей среде) и филогеномика (основанная на эволюционном подходе). Кроме того, выделяют ряд подразделов геномики с целью изучения более частных проблем, например, когнитивная геномика, онкогеномика, веномика и т. д.

### 3.1. Структурная геномика

Структурная геномика исследует содержание и организацию геномной информации, при этом структурная аннотация классифицирует гены и некодирующие элементы и устанавливает их локализацию в геноме, создавая генетические и физические карты высокого разрешения (рис. 3.2). Основная цель структурной геномики состоит в том, чтобы идентифицировать все гены генома и определить трёхмерные формы всех важных биологических макромолекул с помощью как экспериментальных методов, так и методов компьютерного моделирования, сосредоточив внимание на белках, особенно на тех, укладки которых трудно распознать простым сравнением последовательностей с белками известной структуры (Terwilliger et al., 2009; Mat-Sharani et al., 2019). Интересно, что предсказания вторичной структуры белка начались ещё до того (в 1951 г. Лайнусом Полингом и Робертом Кори), как была определена первичная структура белка (в 1953 г. Фредериком Сенгером) (Yang et al., 2018).

Экспериментальный подход моделирования фолдинга *de novo* используется, когда каждая открытая рамка считывания (ORF) может быть клонирована и экспрессирована с последующим анализом очищенного и кристаллизованного белка с помощью рентгеновской кристаллографии или ядерного магнитного резонанса (ЯМР). Вычислительный подход прогнозирования структуры белка основан на сравнительном, потоковом и свободном (или *ab initio*) моделировании (Dukka, 2017; Bhattacharya et al., 2021). Сравнительное моделирование сопоставляет последовательность гена неизвестного белка с последовательностями белков известной структуры, т. е. использует гомологию последовательностей белков для создания структурной модели неизвестных белковых молекул. Такой подход к моделированию требует как минимум 50% идентичности аминокислотной последовательности между неизвестным белком и белком с решённой структурой для создания новой высокоточной структуры; 30–50% аминокислотной идентичности обеспечивает среднюю точность моделирования, а сходство менее 30% даёт низкую точность модели для структуры нового белка (Baker, Sali, 2001; Dukka, 2017).

Многопоточность (threading) или распознавание фолдинга – ведущий метод прогнозирования структуры белка, основанный на сходстве пространственной структуры, а не на идентичности последовательностей. Такой подход помогает идентифицировать отдалённо родственные белки и их молекулярные функции, а также выявлять древнее родство с другими белками (Baker, Sali, 2001; Bhattacharya et al., 2021). Метод

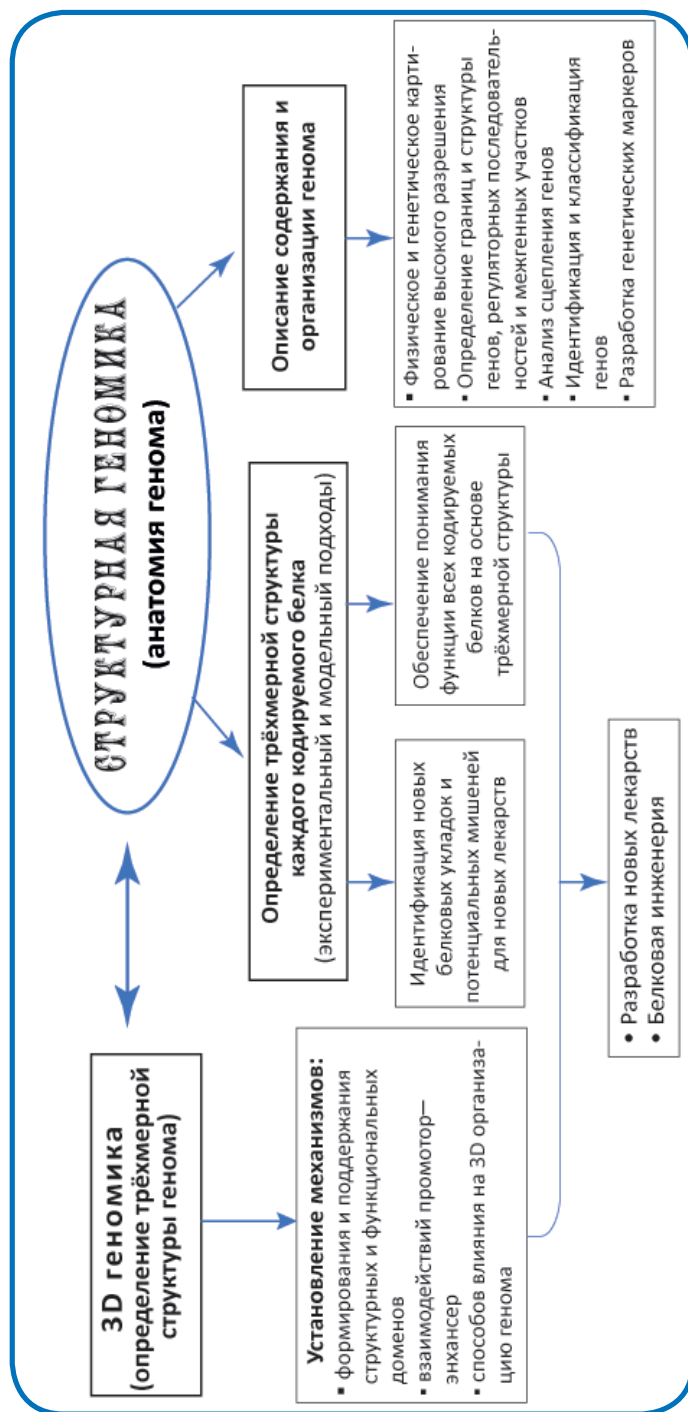


Рисунок 3.2. Структурная геномика

обычно включает программирование, сравнение на основе профилей и машинное обучение. Прогресс в прогнозировании карт взаимодействия между аминокислотными остатками с использованием коэволюции последовательностей и глубокого обучения открыл новые перспективы в отношении распознавания фолдинга для белков с отдалённой гомологией. Ожидается, что в дальнейшем разработки по предсказанию структуры белка с использованием интеграции глубокого обучения на разных этапах анализа должны значительно преобразовать и усовершенствовать процесс создания моделей (Bhattacharya et al., 2021).

При моделировании *ab initio* для предсказания трёхмерной структуры белка используются данные о его последовательности, а также о химических и физических взаимодействиях аминокислот, но без использования дополнительной информации о структурах схожих белков. Программа Rosetta (Rohl et al., 2004) является одним из достаточно успешных методов моделирования *ab initio*; белок делится на короткие сегменты и каждый сворачивается в низкоэнергетическую локальную конформацию, с последующей кластеризацией для получения наилучшей структуры (Xu, Zhang, 2012). В 2020 г. были достигнуты крупные успехи в моделировании *ab initio* на основе применения искусственного интеллекта. К ним относится AlphaFold – система, разработанная DeepMind, которая делает современные предсказания структуры белков на основе их аминокислотных последовательностей. База данных AlphaFold обеспечивает программный доступ и интерактивную визуализацию предсказанных атомных координат, включая оценки достоверности модели и предсказанных ошибок выравнивания (Varadi, Velankar, 2023). С самого начала выпуски AlphaFold были тематическими (в частности, связанными с глобальным здравоохранением и «забытыми» болезнями), а набор данных за 2021 год увеличился почти в 600 раз (Varadi, Velankar, 2023).

Таким образом, в отличие от традиционных экспериментальных методов, полногеномные последовательности облегчают быстрое предсказание модели структуры белка на основе либо ранее решённых структур, либо физических и химических принципов для белка без гомологичной структуры. Самые первые структуры могут способствовать пониманию функции белка и определению фолдинга для других членов семейства, созданию подробных моделей для наиболее похожих белков и выявлению ранее не охарактеризованных эволюционных взаимосвязей (Chandonia, Brenner, 2006). Подход к геномному моделированию может быть очень эффективным при выявлении новых укладок (фолдинга), поскольку ему не препятствуют экспериментальные ограничения, и он может выявить потенциальные мишени для новых лекарств (Chandonia, Brenner, 2006). В 2005 г. структурная геномика дала примерно половину новых структурно охарактеризованных семейств и примерно в пять раз больше новых фолдингов, чем доминирующая в то время структурная биология (Chandonia, Brenner, 2006; Marsden et al., 2007). Высокопроизводительное определение структуры белка выполняется преимущественно в специализированных центрах структурной геномики. Давно существующая открытая сеть аннотаций структуры

белка (TOPSAN) продемонстрировала перспективность аннотирования белковых структур, полученных в центрах структурной геномики (Weekes et al., 2010). По состоянию на 2022 г. универсальный ресурс белков (UniProt) насчитывает более 230 миллионов белковых последовательностей (из них 214 миллионов обеспечивает AlphaFold DB), а банк данных белков (PDB) содержит около 200000 структур примерно 60000 уникальных белков. Причём в настоящее время, благодаря наличию крупных баз данных последовательностей и структур белков, а также мощных методов глубокого обучения, наивысшая точность предсказания без опоры на шаблоны структур составляет 82–84%, что близко к теоретическому пределу (88–90%) – результат, немыслимый ещё совсем недавно (Yang et al., 2018).

Благодаря тесной связи между структурой и функцией, структурная геномика также раскрывает информацию о функции белка на основе различных уровней функционального предсказания и других ключей (Brenner, Levitt, 2000; Brenner, 2001; Chandonia, Breener, 2006). Например, сохранение последовательности может указывать на потенциально активный сайт; связанный кофактор, метаболит или молекула растворителя могут быть ассоциированы с функционально важным сайтом; а распознавание каталитического мотива может свидетельствовать об особой активности (Baker, 2007).












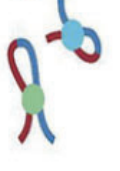


**3.1.1. Трёхмерная геномика.** Трёхмерную (3D) геномику можно рассматривать как часть или подраздел структурной геномики, появление которой совпало с открытием регуляторных эпигенетических механизмов на уровне укладки генома в начале XXI века. Помимо того, что трёхмерный фолдинг позволяет хранить хромосомы в ядре и обеспечивает точную сегрегацию хромосом в дочерние клетки, он тесно связан с тем, как генетическая информация преобразуется в программы транскрипции, являясь критически важным фактором, поддерживающим высокоспецифичные взаимодействия между цис-регуляторными элементами (CRE) и промоторами (Hehmeier et al., 2023). 3D геномика фокусируется на механизмах формирования и поддержания структурных и функциональных доменов генома, установлении взаимодействий промотор–энхансер и способах влияния на пространственную организацию генома (в том числе, посредством различных низкомолекулярных молекул) (Razin et al., 2019; Ulianov, Razin, 2021; Pei et al., 2021; Zhang, Li, 2020). Модификации в организации хроматина сопровождают или даже управляют созданием новых регуляторных стратегий, и вполне возможно, что 3D укладка *«катализовала разделение и специализацию генетических программ, которые обеспечили многоклеточность и сложность строения тела на уровне органов»* (Rowley, Corces, 2018; Hehmeier et al., 2023). Динамические изменения в архитектуре хроматина являются ключевыми в регуляции транскрипции кластеров гомеозисных *Нох*-генов – важнейших регуляторов эмбрионального развития (De Wit, De Laat, 2012). Организация 3D-хромосом может играть центральную роль в регуляции эпигенома. Например, геномные регионы, расположенные внутри компактных трёхмерных доменов, обладают более протяжёнными эпигенетическими профилями, чем локусы внутри слабокомпакт-

ных регионов. Следовательно, укладка генома может влиять на формирование эпигеномных доменов, сближая в пространстве два удалённых локуса и способствуя трёхмерному распространению эпигеномной метки на большие расстояния, а эпигеномная метка может воздействовать на фолдинг генома посредством рекрутирования архитектурных белков (Abdulla et al., 2023).

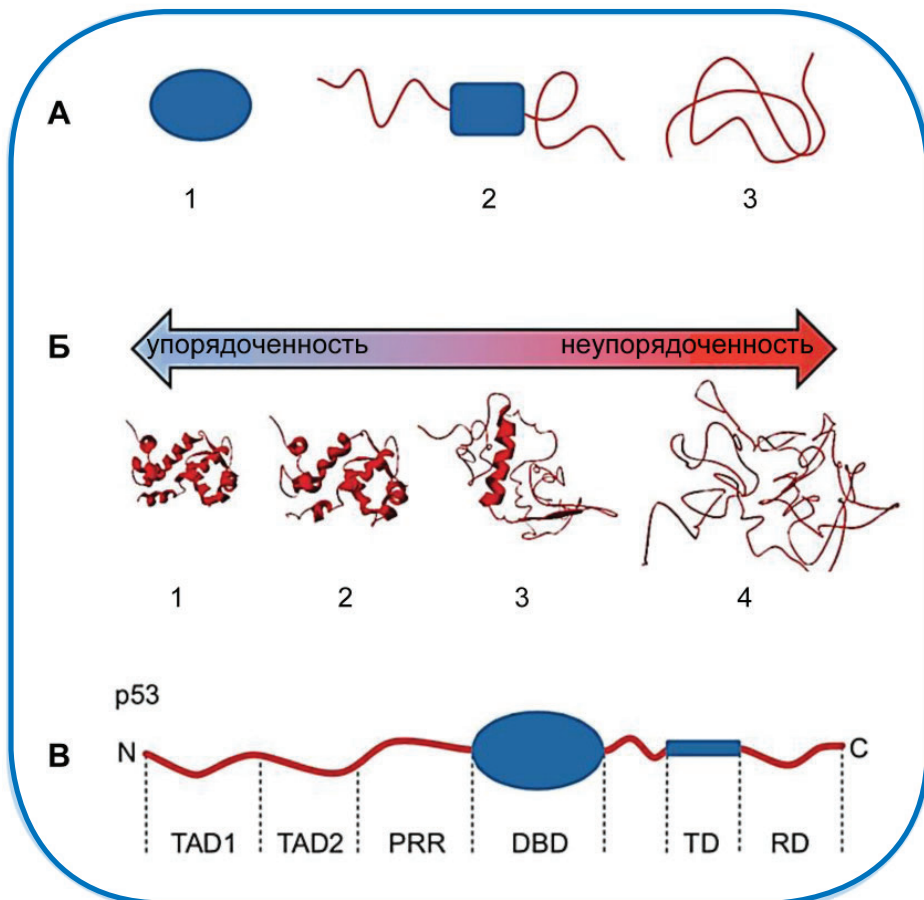
Выделяют четыре основные категории процессов, формирующих 3D-структуры: экструзия петель, образование хроматиновых нанодоменов (неструктурированных агломератов, включающие десятки нуклеосом), компарментализация и индуцированная транскрипцией суперспирализация (Hehmeyer et al., 2023). Для определения потенциальных структур и функций трёхмерного генома применяются новые улучшенные методы, включая Hi-C (высокопроизводительный захват конформации хромосом; обеспечивает полногеномное картирование конформации хромосом с высоким разрешением) и ChIA-PET (полногеномный анализ, включающий обогащение на основе иммунопреципитации хроматина, проксимальное лигирование хроматина, парные концевые метки и высокопроизводительное секвенирование для установления *de novo* взаимодействий между участками, расположенными на значительном удалении друг от друга и связанными с белковыми факторами), а также методы, основанные на визуализации, и новые подходы, базирующиеся на стратегиях неproxимального лигирования; методы на основе CRISPR (сгруппированные регулярно расположенные короткие палиндромные повторы) также полезны для функциональной проверки. В них используется подход «все против всех», который позволяет охарактеризовать сворачивание хроматина на уровне всего генома, тогда как методы 3C («один против одного») и 4C («один против всех») оценивают частоты взаимодействия двух интересующих областей, или одной области с остальной частью генома соответственно (Zhang, Li, 2020; Hauth et al., 2022; Davidson et al., 2023) (табл. 3.1).

Имеющиеся данные позволяют предполагать, что хромосомы теплокровных животных делятся на два типа структурных единиц: топологически ассоциированные домены (TAD) и компартментальные домены (возможно, более древнего типа), а конформация хромосом может отражать тип или состояние клетки. TAD играют важную роль в контроле активности генов. Удаление или повреждение участка ДНК, разделяющего соседние TAD, активирует гены, которые в физиологической норме не экспрессируются, что, в свою очередь, может вызвать серьёзные нарушения функции клеток (Разин и др., 2019; Hoencamp et al., 2021). В отличие от тканеспецифичных компартментов, TAD у млекопитающих, как правило, постоянны и их границы сохраняются для всех видов (Hauth et al., 2022). Изменения в трёхмерной структуре генома раковых клеток происходят в основном на уровне TAD и петель хроматина, поэтому новые эпигенетические противораковые препараты нацелены именно на трёхмерную организацию генома, меняющую топологию ДНК (Kantidze et al., 2020). Более того, если ранние исследования основывались на клеточных популяциях, то в настоящее время они фокусируются на отдельных клетках, что позволяет работать со сложной межклеточной гетерогенностью (Разин и др., 2019; Ulianov, Razin, 2021).

Таблица 3.1. Методы анализа конформации генома (по: Davidson et al., 2023; De Wit, De Laat, 2012).

Метод	3С	4С	Hi-C	Micro-C	Hi-ChiP	ChiA-PET	Tiled
Петли хроматина							
Подход	Один против одного	Один против всех					
Основные этапы	Расщепление рестриктазами			Двойное перекрёстное связывание	Все против всех		
	ПЦР	Микрочип	Обработка ультразвуком	Расщепление MНазой	Расщепление рестриктазами	Расщепление рестриктазами или MНазой	Расщепление рестриктазами или MНазой
Разрешение	Стандартные эксперименты дают разрешение 10–50 Кб (зависит от локуса)	От 200 до 400 п. н. в зависимости от частоты сайтов рестриктаз (для 4 или 6 п. н.)	От 400 п. н. до 4000 п. н. в зависимости от частоты сайтов рестриктаз (4 или 6 п. н.)	~200 п. н.	От 1000 до 50000 п. н.	От 100 п. н. до 1000 п. н.	От 200 п. н. до 2000 п. н. (зависит от локуса)
	сшивки хроматина	расщепление сшитого хроматина	расщепление сшитого хроматина	лигирование	лигирование	обратное шивание	обратное шивание
Этапы сшивания, расщепления и лигирования, общие для всех методов							

**Примечания.** 3С – захват конформации хромосомы; 4С – захват конформации хромосомы на чипе/захват конформации кольцевой хромосомы; Hi-C – сочетание захвата конформации хроматина и секвенирования следующего поколения для полного картирования конформации хромосом с высоким разрешением; Micro-C – процедура, при которой шпигый геном разрезается микрококовой нуклеазой (MНазой) вместо ферментов рестрикции, в результате чего образуются монохромосомы; Hi-ChiP – анализ взаимодействия хроматина, основанный на Hi-C *in situ* с этапом иммунопреципитации (ChiP); ChiA-PET – анализ взаимодействия хроматина с помощью секвенирования парных концевых меток; Tiled-C – метод малозатратного захвата конформации хромосом (3С); MНаза – фермент, расщепляющий ДНК в участках, которые не связаны стабильно белками; хроматин IP – иммунопреципитация хроматина (ChiP).



**Рисунок 3.3.** Отличия обычных белков с фиксированной структурой от внутренне неупорядоченных белков (из: Watson, Stott, 2019).

А: 1 – свернутый белок с фиксированной конформацией, 2 – свернутый белок с неупорядоченной частью, 3 – полностью неупорядоченный белок. Б: 1 – глобула, 2 – расплавленная глобула (белок имеет фиксированную часть), 3 – предрасплавленная глобула с большой неупорядоченной частью, 4 – случайный клубок, полностью неупорядоченная форма. В: частично упорядоченный белок на примере p53; внизу даны названия доменов.

Помимо традиционных белок–белковых, ДНК–белковых и РНК–белковых взаимодействий (стабильных и точных), движущей силой реализации функции клетки являются так называемые конденсаты, образующиеся в результате разделения фаз жидкость–жидкость (LLPS). Важной задачей предстоящих исследований в области 3D геномики считают понимание того, как эти LLPS можно модифицировать и как их контролировать (Razin, Ulianov, 2020).

В настоящее время широко обсуждаемой темой является ключевая роль белков, не имеющих (в отличие от большинства белковых молекул) строго заданную конформацию, в регуляции архитектуры хроматина и экспрессии генов (рис. 3.3).

Гибкость таких белков приводит к различным моделям взаимодействия с ДНК, от жёсткого связывания до образования высокодинамичных комплексов. Внутренне неупорядоченные белки и белки, содержащие внутренние неупорядоченные области, способствуют конденсации больших участков генома через разделение фаз жидкость–жидкость и таким образом создают новую «жидкообразную картину» сборки хроматина (Watson, Stott, 2019). Недавно было показано, что у тихоходки *Ramazzottius varieornatus* Bertolani and Kinchin, 1993 (тип Tardigrada), обладающей исключительной устойчивостью к различным формам стресса, в спонтанном разделении фаз жидкость–жидкость, вероятнее всего, участвует в высшей степени неструктурированный белок-супрессор повреждений ДНК/РНК (Dsup). Dsup имеет две области, которые способствуют образованию капель, что также указывает на его участие в организации хроматина (Zarubin et al., 2024). Кроме Dsup исследования *R. varieornatus* привели к открытию других уникальных неупорядоченных белков нескольких семейств, которые связаны с экстремальной толерантностью к стрессу. Эти белки имеют большие перспективы использования в медицине, биотехнологии и космических исследованиях (например, для создания стрессоустойчивых организмов, защитных средств при радиационной и химии терапии, компонентов фармацевтических препаратов, вакцин и т. д.) (Zarubin et al., 2024).

Укладка генома в разных клетках существенно отличается, что может указывать на постоянную сортировку различных геномных конфигураций в клетке, позволяющую ей быстро адаптироваться к изменяющимся условиям окружающей среды (Flyamer et al., 2017). На примере кластеров глобиновых генов позвоночных показано, что эволюция сопровождается потерей линейных сегментов хромосом при сохранении сегментов, организованных в глобулы. Это связано со значительным возрастанием роли удалённых энхансеров в регуляции активности генов у млекопитающих (Kovina et al., 2017). Исследования 3D генома показали, что (1) на уровне пространственной организации геном разделён на структурные и функциональные блоки (которые могут быть «разорваны» в линейном геноме), ограничивающие объём работы энхансеров; (2) пространственная реконфигурация протяжённого сегмента генома может служить механизмом подавления/активации генов, расположенных в этом сегменте; и (3) изменения трёхмерной организации генома часто связаны с возникновением различных заболеваний (Razin et al., 2019; Kantidze et al., 2020).

Сравнительная геномика в сочетании с моделями трёхмерной укладки генома выявила «перехват» энхансера, и позволила рассматривать его как объяснение быстрой эволюции «ускоренных» областей генома человека (HAR). HAR представляют собой эволюционно консервативные последовательности, которые неожиданно приобрели большое количество нуклеотидных замен в геноме человека. Многие из них являются энхансерами регуляции генов, которые функционируют во время эмбрионального развития, особенно при формировании нервной системы. Объединение глубокого обучения с экспериментами по захвату хроматина в нейронных клетках–предшественниках человека и шимпанзе, позволило обнаружить повышенное присутствие HAR

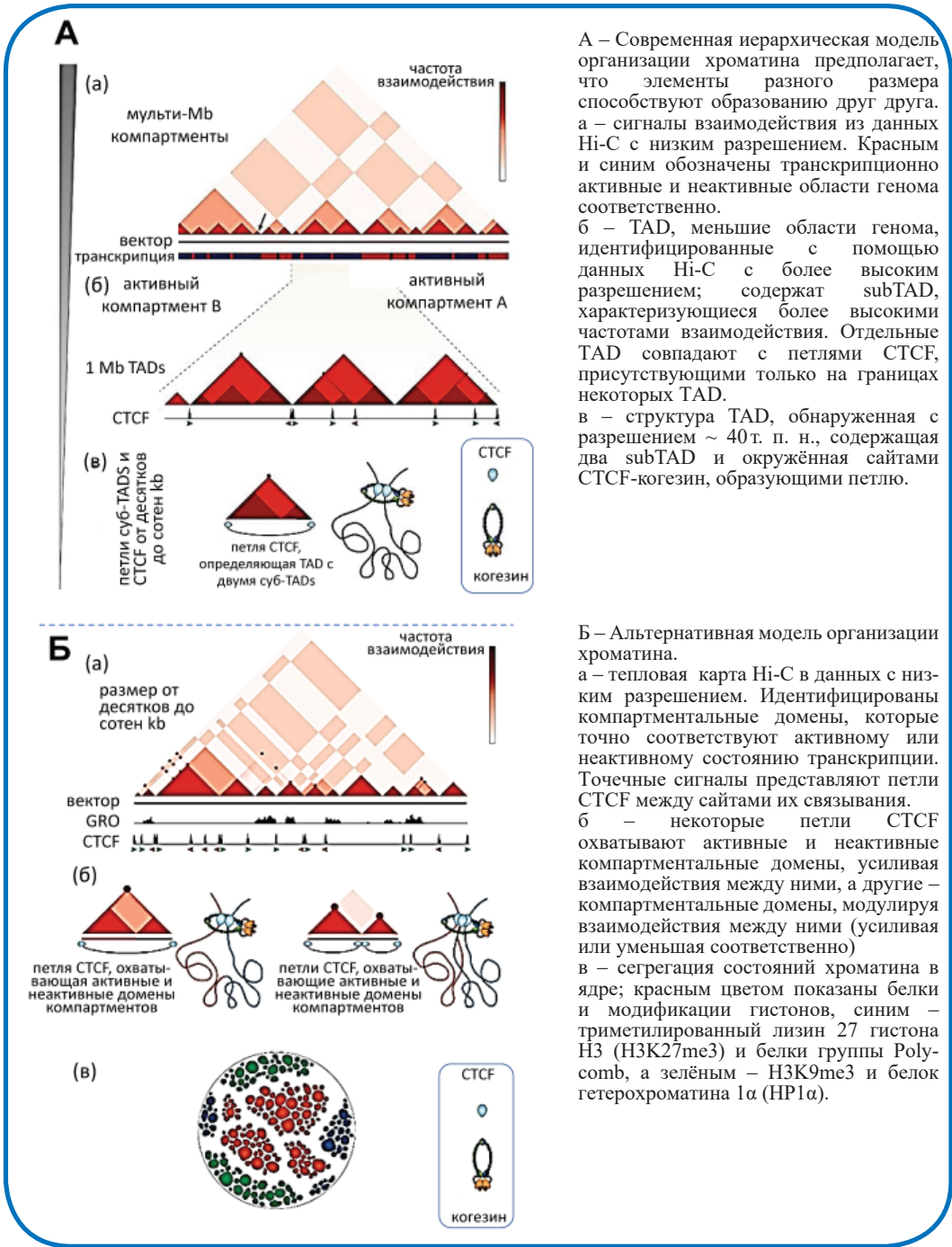
в TAD, содержащих специфичные для человека геномные варианты, способные изменять трёхмерную организацию генома. Эта способность подтверждается наблюдением того факта, что 30% ZooHAR встречаются в пределах 500 т. п. н. специфичных для человека структурных вариантов (hsSV), приводя к существенно изменяющимся локальным взаимодействиям хроматина. Некоторые HAR, вероятно, произошли в результате захвата энхансеров, вызванного соседними hsSV, вследствие чего эволюционно законсервированные энхансеры начинают адаптироваться к различным генам-мишеням и регуляторным доменам (Keough et al., 2023).

Трёхмерные структуры генома в настоящее время исследованы у многих многоклеточных животных, хотя из-за низкого разрешения или проблем сборки генома некоторые вопросы остаются открытыми. В эволюционных масштабах времени в результате параллельного действия различных механизмов возникают различные 3D топологические карты, детектируемые с помощью Hi-C (Helmeyer et al., 2023). Характерным признаком 3D генома являются ADTs (топологии, зависящие от активности), но их трудно распознать, если данные получены на смешанных клеточных популяциях. Типичные для позвоночных LEDs (домены экструзии петель), не были достоверно идентифицированы у ранних Eumetazoa, и есть данные о том, что LEDs могут не играть заметной роли в организации их трёхмерного генома. Недавняя идентификация TADs в геномах головохордовых с CTCF-подобным мотивом (CCCTC-связывающий фактор) на их границах предполагает участие экструзии петель в организации генома ещё до появления позвоночных (Helmeyer et al., 2023). На рис. 3.4 представлены две современные модели трёхмерной организации хроматина, иерархическая и альтернативная.

### 3.2. Функциональная геномика

Функциональная геномика направлена на описание биологической функции генов и их продуктов (белков, РНК) и, в конечном итоге, всех компонентов генома с долгосрочной целью понять взаимосвязь между геномом организма и его фенотипом (рис. 3.5). Таким образом, функциональная геномика стремится выяснить, как генетические элементы регулируются и экспрессируются, чтобы определить физиологию организма, фокусируясь на таких динамических аспектах, как транскрипция генов, трансляция, регуляция экспрессии генов и белок-белковые взаимодействия (Goh et al., 2018).

Функциональная геномика изучает функции генов на уровне всего генома с использованием высокопроизводительных полногеномных подходов и интеграции геномных, транскриптомных и протеомных данных, а также исследует естественные (т. е. в пределах физиологических норм) генетические изменения и функциональные нарушения. Кроме того, функциональная геномика изучает карты генетических взаимодействий на уровне ДНК, чтобы идентифицировать гены (физически взаимодействующие или нет) со связанными функциями и анализировать функционально связанные характеристики генома, такие как мутации и полиморфизм. Идентификация



А – Современная иерархическая модель организации хроматина предполагает, что элементы разного размера способствуют образованию друг друга. а – сигналы взаимодействия из данных Hi-C с низким разрешением. Красным и синим обозначены транскрипционно активные и неактивные области генома соответственно.

б – TAD, меньшие области генома, идентифицированные с помощью данных Hi-C с более высоким разрешением; содержат subTAD, характеризующиеся более высокими частотами взаимодействия. Отдельные TAD совпадают с петлями CTCF, присутствующими только на границах некоторых TAD.

в – структура TAD, обнаруженная с разрешением ~ 40 т. п. н., содержащая два subTAD и окружённая сайтами CTCF-когезин, образующими петлю.

Б – Альтернативная модель организации хроматина.

а – тепловая карта Hi-C в данных с низким разрешением. Идентифицированы компартментальные домены, которые точно соответствуют активному или неактивному состоянию транскрипции. Точечные сигналы представляют петли CTCF между сайтами их связывания.

б – некоторые петли CTCF охватывают активные и неактивные компартментальные домены, усиливая взаимодействия между ними, а другие – компартментальные домены, модулируя взаимодействия между ними (усиливая или уменьшая соответственно)

в – сегрегация состояний хроматина в ядре; красным цветом показаны белки и модификации гистонов, синим – триметилированный лизин 27 гистона H3 (H3K27me3) и белки группы Polycomb, а зелёным – H3K9me3 и белок гетерохроматина 1α (HP1α).

Рисунок 3.4. Модели организации хроматина (из: Rowley, Corces, 2018).

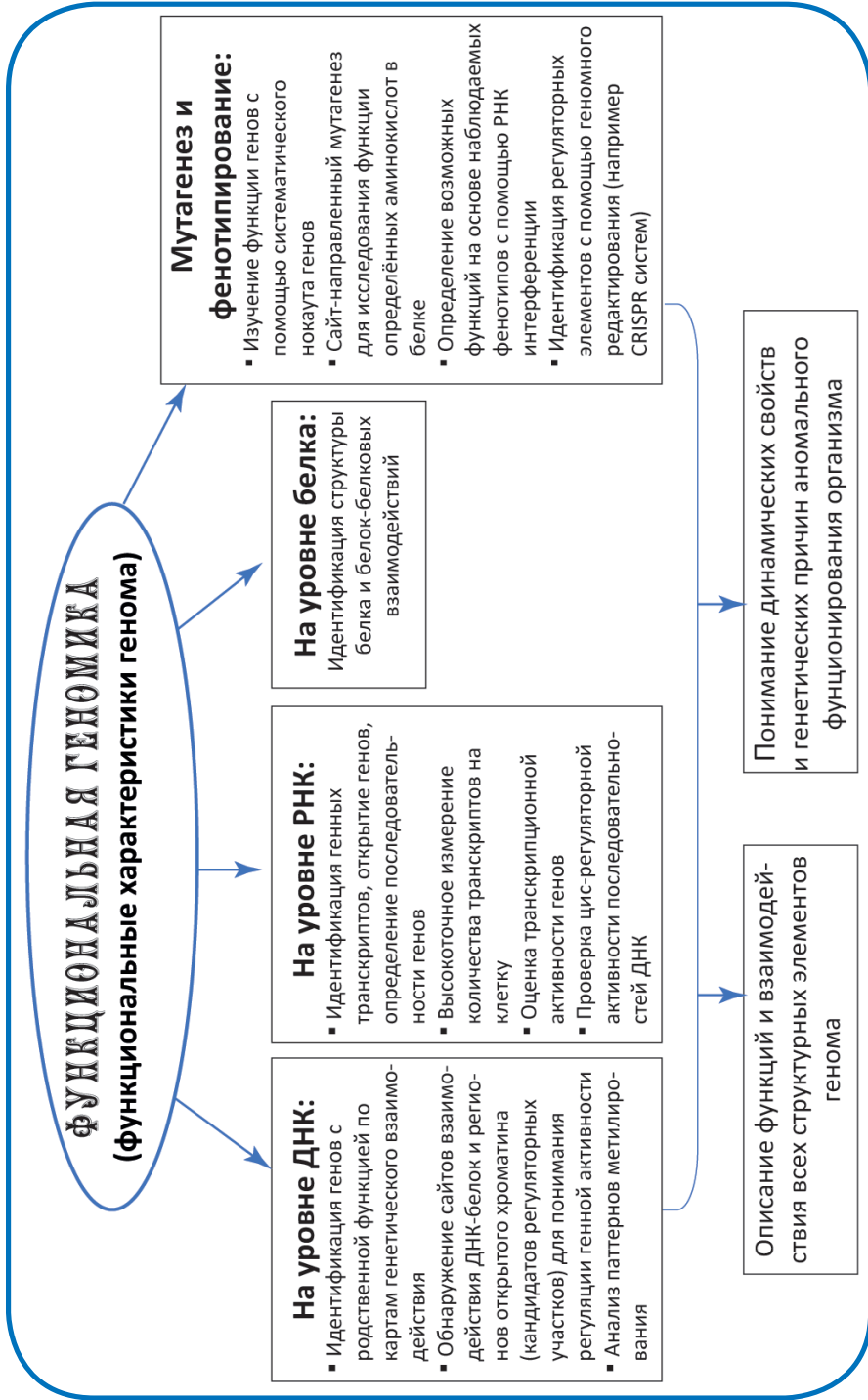


Рисунок 3.5. . Функциональная геномика

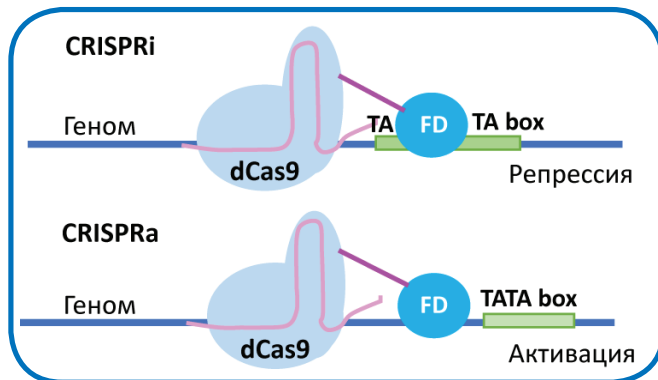
всех функциональных элементов, как кодирующих, так и некодирующих областей, в геноме человека была основной целью проекта ENCODE (Энциклопедия элементов ДНК) (Guigó et al., 2006).

На уровне РНК функциональная геномика использует различные подходы, такие как ESTs (фрагменты последовательностей мРНК, полученные в результате одиночных реакций секвенирования, проведённых на случайно выбранных клонах из библиотек кДНК), секвенирование РНК, SAGE (серийный анализ экспрессии генов), экспрессионные ДНК-чипы (в основе метода – молекулярная гибридизация) и MPSS (массово-параллельное опознавательное секвенирование; позволяет выявлять почти все кДНК-фрагменты в образце), с целью открытия новых генов и выявления генов, которые сильно или слабо экспрессируются при определённых условиях, либо являются активными и неактивными в различных тканях, что необходимо для понимания генетических причин аномального функционирования организма (Matsumura et al., 2005; Bilitewski, 2009; Aw et al., 2010).

Связывания белок–ДНК играют ключевую роль в процессах регуляции экспрессии генов. Секвенирование, сочетающее иммунопреципитацию хроматина с высокопроизводительным секвенированием (ChIP-seq), представляет собой метод анализа взаимодействий белка с ДНК. Его часто используют для картирования взаимодействий белок–ДНК *in vivo* в масштабе генома и для понимания того, как транскрипционные факторы (TF) и другие ассоциированные с хроматином белки воздействуют на механизмы, влияющие на фенотип (Jothi et al., 2008; Niu et al., 2011; Kumar et al., 2013; Muhammad et al., 2020).

Идентификация активных элементов регуляции генов является ключом к пониманию транскрипционного контроля биологических процессов, включая дифференцировку, развитие, пролиферацию, специфичность клеточного типа и реакцию на изменение окружающей среды. Сверхчувствительные к ДНКазе I сайты (DHS) являются маркерами открытого хроматина, которые обеспечивают очень эффективный метод нацеливания на все типы цис-регуляторных элементов (CREs), включая энхансеры, промоторы, инсуляторы, сайленсеры и области контроля локусов. Объединение полногеномных последовательностей и геномных карт регуляторной ДНК с отмеченными DHSs позволяет оценить гетерогенность уровня функциональных ограничений в регуляторной ДНК разных типов клеток, обнаружить изменчивость в функциональных ограничениях среди мотивов транскрипционных факторов в регуляторной ДНК, идентифицировать новые субстраты адаптивной регуляторной эволюции (Chen et al., 2018).

Новый и многообещающий способ изучения функции генов, а также потенциальных регуляторных элементов в масштабе всего генома заключается в модификации конкретных геномных локусов с помощью системы CRISPR–Cas9 (Shalem et al., 2014). CRISPR–Cas9 используется для удаления/репрессии генов (нокаут, CRISPRi) в клеточных линиях мультиплексным образом, или для их активации (CRISPRa) (рис. 3.6). Нокаут важного гена приводит к гибели клетки–хозяина и соответственно к истощению конкретной направляющей РНК. В эксперименте с CRISPR–Cas9 на кле-

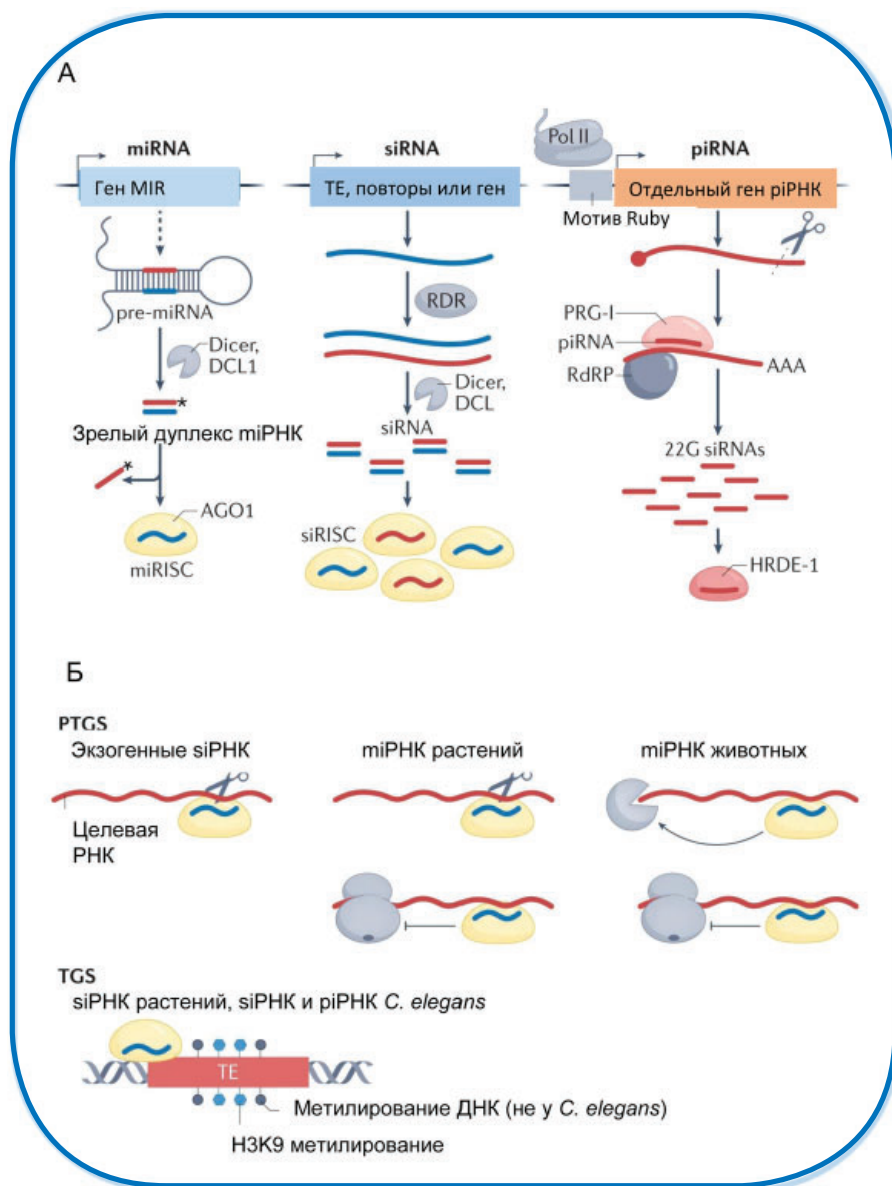


**Рисунок 3.6.** Репрессия и активация генов с помощью системы CRISPR–Cas9 (по: Rainha et al., 2021).

CRISPRi используется для репрессии генов: «мертвый» Cas9 (dCas9, т. е. способный только связываться с нужным участком ДНК) со слитым доменом (FD) воздействуют на область ТАТА-бокса, препятствуя образованию комплекса инициации транскрипции и, следовательно, подавляя экспрессию гена. CRISPRa используется для активации генов: dCas9, слитый с доменом активатора, взаимодействует с участком за пределами области ТАТА-бокса, что приводит к повышению уровня экспрессии гена.

точных линиях млекопитающих было показано, что около 2000 генов являются важными для нескольких клеточных линий, а некоторые гены – только для одной линии (Hart et al., 2015). Такая система, использующая пары направляющих РНК (gRNAs) для программирования тысяч делеций размером в тысячи пар оснований и глубокого сканирования целевой области мозаичным способом («ScanDel»), может прояснить роль мутаций в менделевских заболеваниях в разных локусах генома человека (Gasparini et al., 2017).

Другой подход, РНК–интерференция, часто используется для полногеномного скрининга потери функции генов, что позволяет идентифицировать компоненты клеточных сигнальных путей, и хорошо подходит для картирования и аннотации геномов. Например, масштабный скрининг клеток человека посредством РНК–интерференции привёл к открытию компонентов метаболического пути белка p53, который является наиболее широко действующим из известных ингибиторов апоптоза (Berns et al., 2004). Исторически РНК–интерференция была известна как метод посттранскрипционного сайленсинга генов. Методы РНК–интерференции можно использовать для временного подавления экспрессии генов (нокдаун) с помощью двухцепочечной РНК длиной ~20 пар оснований (обычно доставляемой путём трансфекции синтетических молекул короткой интерферирующей РНК, или вирусно-кодируемой короткой шпилечной РНК) (рис. 3.7). Возможные функции повреждённых генов определяются на основе наблюдаемых фенотипов. Открытие системы CRISPR–Cas9 и РНК–интерференции являются важнейшими событиями XXI века в области наук о жизни. Применение этих технологий для редактирования генов и геномов оказало существенное влияние на исследования в области биологии, в изучении болезней, развитии сельскохозяйственных культур и т. д. (например, Berns et al., 2004; Niemann, Petersen, 2016; Schiml, Puchta, 2016; Li et al., 2020a; Watters et al., 2021; Maf et al., 2022; Park, Baе, 2024). Появление редакторов оснований и первичных редакторов открыло новые возможности для точного преобразования оснований, а также для их вставки и удаления с высокой степенью точности. Кроме того, в настоящее время активно



**Рисунок 3.7.** Схемы механизмов биогенеза и способы действия малых РНК (из: Chen, Rechavi 2022).

А – Упрощённые механизмы биогенеза микроРНК (miRNA), коротких интерферирующих РНК (siRNA) и PIWI-взаимодействующих РНК (piRNA). Dicer/DCL1 – РНКазы у животных/растений; Argonaute – эффекторный белок; dsRNA – двухцепочечная РНК; siRISC – комплекс подавления; Ruby – мотив для экспрессии piRNA; HRDE-1 – белок Argonaute. Б – Посттранскрипционное (PTGS; расщепление/деградация РНК, репрессия трансляции) и транскрипционное (TGS; метилирование ДНК/гистонов) подавление генов.

продолжаются исследования, направленные на манипуляции с кольцевой ДНК в оргanelлах, а также на создание модифицированных мезенхимальных стволовых клеток, имеющих большой потенциал в качестве ресурса для регенеративной медицины. Ожидается, что будут успешно решены такие проблемы негативных последствий технологий редактирования генома, как создание крупных делеций ДНК, активация P53, хромосомные перестройки и нецелевые правки по всему геному (Park, Bae, 2024).

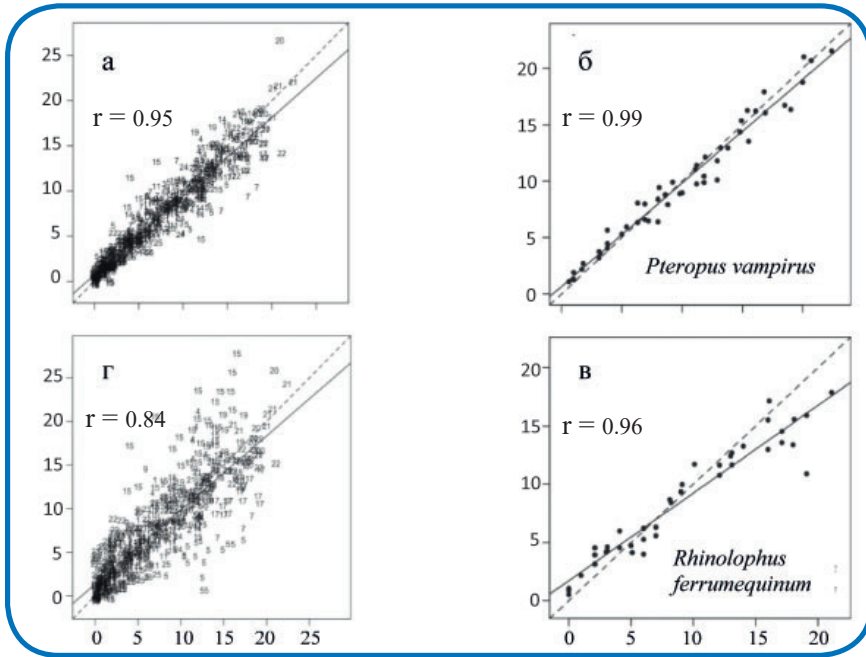
Быстрое получение результатов геномных исследований, чему способствовало высокопроизводительное секвенирование и развитие методов биоинформатики, значительно расширило возможности функциональной геномики в анализе «омических» данных и послужило основой создания технологий следующего поколения. Как следствие, появилась уникальная возможность использовать подходы глубокого обучения, например, для получения моделей метаболизма (Rai et al., 2019). Для обнаружения энхансерных и промоторных областей в геноме человека недавно был введён метод DECRES (глубокое обучение для выявления цис-регуляторных элементов), который основан на данных ENCODE и FANTOM (функциональная аннотация генома млекопитающих), а также на методе глубокого обучения для различения CRR (цис-регуляторные области, которые точно контролируют экспрессию генов). Модель DECRES показала высокий потенциал технологии глубокого обучения с перспективой применения в фундаментальных и прикладных исследованиях геномики (Li et al., 2018).

**3.2.1. Эпигеномика.** Открытие в конце 50-х годов немутационной генетической изменчивости (или эпигенетической изменчивости) показало, что *«мы нечто большее, чем просто сумма наших генов»* (Klar, 1998). В основе эпигенетической геновой регуляции лежат два основных механизма: метилирование ДНК и модификация гистонов. Эпигеномика изучает полный набор эпигенетических модификаций генетического материала клетки, которые влияют на экспрессию и регуляцию генов, участвуют во многих важных клеточных процессах (таких как дифференцировка, развитие, репарация ДНК и онкогенез), чтобы понять роль трансгенерационного эпигенетического наследования в адаптивной эволюции (Luikart et al., 2018). Анализ метилирования ДНК ранее был ограничен определёнными локусами, но с появлением новых методов секвенирования ДНК, таких как бисульфитное секвенирование, секвенирование одиночных молекул и нанопоровое секвенирование, стало возможным проводить исследования в масштабе генома и характеризовать целые метиломы с высоким разрешением (Rand et al., 2017). Набор посттрансляционных модификаций (метилирование, фосфорилирование, ацетилирование и др.) концевых доменов гистонов получил название «гистоновый код». Согласно гипотезе гистонового кода эпигенетические метки действуют как уникальный молекулярный язык, регулирующий структуру хроматина и транскрипцию генов. Модификации гистонов влияют на упаковку хроматина (компактного и транскрипционно неактивного гетерохроматина или транскрипционно активного эухроматина), что приводит к изменениям функционального состояния/активности генов (Jenuwein, Allis, 2001). Модифицированные

гистоны могут быть обнаружены с помощью метода иммунопреципитации хроматина и ДНК-микрочипов (технология ChIP-on-chip) (Barski et al., 2007). Для анализа данных полногеномного метилирования применяются методы искусственного интеллекта, включая глубокое обучение, а также анализ «путей изобретательности» (IPA), которые показали высокую эффективность в прогнозировании заболеваний человека (Bahado-Singh et al., 2021). Создан атлас метилома человека, позволяющий анализировать тысячи уникальных маркеров для 39 типов клеток, отобранных из 205 образцов здоровых тканей; это важный ресурс для изучения регуляции генов и генетических вариантов при разных заболеваниях и в различных тканях (Loyfer et al., 2023).

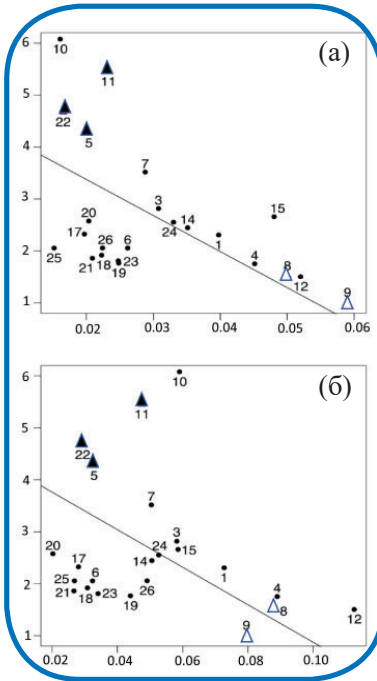
Эпигенетические модификации ДНК увеличивают регуляторную гибкость генома, способствуя эволюции фенотипической пластичности. Например, характер метилирования ДНК имеет большое значение в альтернативных «кастах» маток и рабочих у общественных насекомых, особенно у медоносных пчёл (для разных типов пчёл различия в паттернах метилирования обнаружены в нескольких десятках генов) (Weiner, Toth, 2012). Определяющие формирование пола факторы представляли одну из самых интригующих загадок в биологии со времён античности. Согласно недавним исследованиям, пол является «кульминацией комбинаторных эффектов генома, зародышевых клеток и окружающей среды с участием эпигенетических факторов, передающих биологический смысл этого взаимодействия» (Nagabhushana, Mishra, 2016), а закономерности метилирования ДНК связаны с фертильностью (Zhang et al., 2023c). Эпигеномика оказалась бесценной для выявления, профилактики и лечения различных заболеваний человека, включая онкологические (Huo et al., 2021; Dogan et al., 2021). Данные эпигенома также способствуют лучшему пониманию адаптации видов к изменчивым и суровым условиям окружающей среды (Miryeganeh et al., 2022). Возрастные изменения в метилировании ДНК можно использовать для прогнозирования возраста человека и других видов (рис. 3.8, 3.9). Причём эти преобразования зависят от процессов развития, в то время как изменения ДНК, связанные с долголетием, ассоциированы с врождённым иммунитетом или генами онкогенеза (Wilkinson et al., 2021). Продолжительность жизни вида выражается коэффициентом долголетия (LQ), который представляет собой отношение максимальной продолжительности жизни наблюдаемого вида к максимальной продолжительности жизни, прогнозируемой для нелетающих плацентарных млекопитающих с той же массой тела. Например, максимальная продолжительность жизни *Myotis lucifugus* более чем в шесть раз превышает ожидаемую, а максимальная продолжительность жизни *Molossus molossus* равна средней продолжительности жизни плацентарного млекопитающего того же размера (рис. 3.9).

Сравнительная эпигеномика – относительно новая область исследований, сочетающая эпигенетические признаки с филогенетическими отношениями и обеспечивающая богатый ресурс информации для многих областей, включая эволюционную биологию и биологию развития. Результаты исследования высококонсервативных цитозинон в геномах примерно 15000 образцов 348 видов показали, что метилирование



**Рисунок 3.8.** Зависимость метилирования ДНК от возраста летучих мышей (из: Wilkinson et al., 2021).

Зависимость показана для 26 видов летучих мышей (как на рис. 3.9); по оси Y – уровень метилирования ДНК, по оси X – возраст.



**Виды летучих мышей:**

- |                                  |                                      |
|----------------------------------|--------------------------------------|
| 1. <i>Antrozous pallidus</i>     | 14. <i>Phyllostomus discolor</i>     |
| 3. <i>Carollia perspicillata</i> | 15. <i>Phyllostomus hastatus</i>     |
| 4. <i>Cynopterus brachyotis</i>  | 17. <i>Pteropus hypomelanus</i>      |
| 5. <i>Desmodus rotundus</i>      | 18. <i>Pteropus poliocephalus</i>    |
| 6. <i>Eidolon helvum</i>         | 19. <i>Pteropus pumilus</i>          |
| 7. <i>Eptesicus fuscus</i>       | 20. <i>Pteropus rodricensis</i>      |
| 8. <i>Leptonycteris curasoae</i> | 21. <i>Pteropus vampyrus</i>         |
| 9. <i>Molossus molossus</i>      | 22. <i>Rhinolophus ferrumequinum</i> |
| 10. <i>Myotis lucifugus</i>      | 23. <i>Rhynchonycteris naso</i>      |
| 11. <i>Myotis myotis</i>         | 24. <i>Rousettus aegyptiacus</i>     |
| 12. <i>Myotis vivesi</i>         | 25. <i>Saccopteryx bilineata</i>     |
| 13. <i>Nyctalus noctule</i>      | 26. <i>Tadarida brasiliensis</i>     |

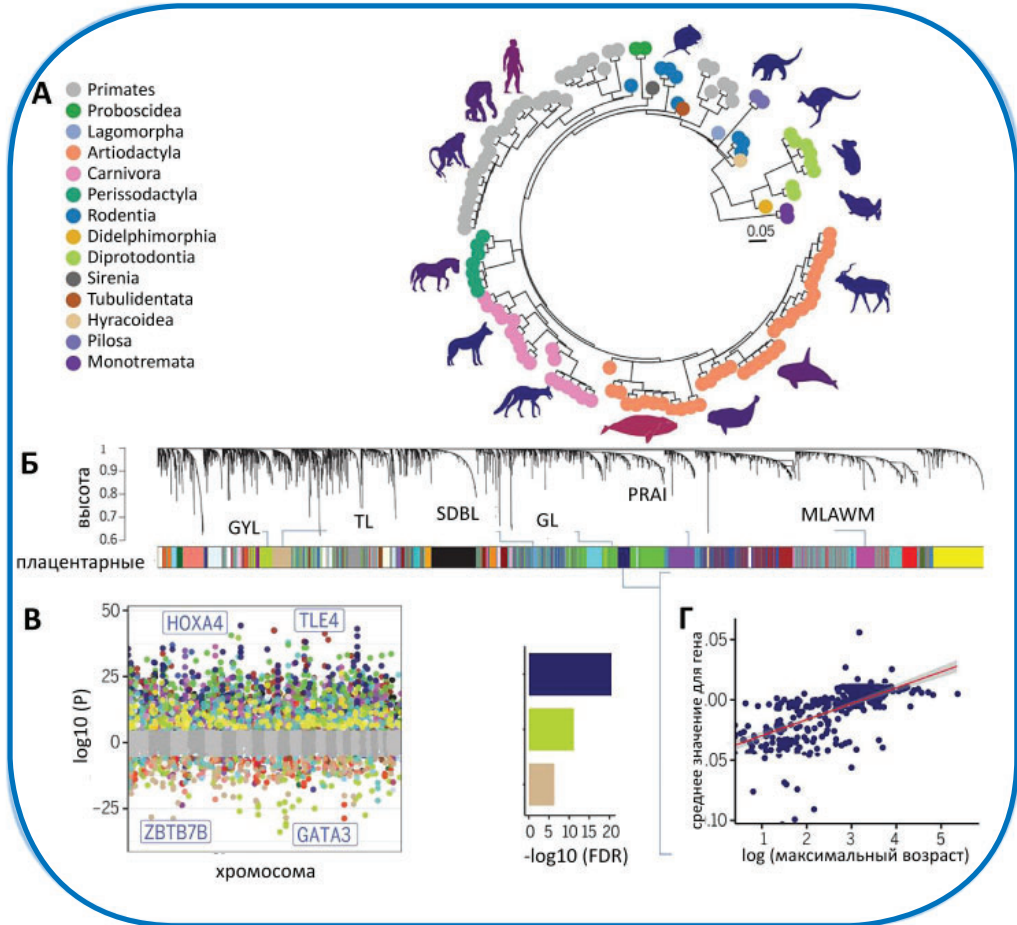
**Рисунок 3.9.** Средние скорости гиперметилирования (а) и гипометилирования ДНК (б) (из: Wilkinson et al., 2021). Средние скорости метилирования коррелируют с долголетием (у долгоживущих видов темпы изменения уровня метилирования ниже). Пять видов, используемых для идентификации DMP (возрастные дифференциально метилированные позиции) долголетия по разнице в скорости метилирования, обозначены чёрными (долгоживущие) и белыми (короткоживущие) треугольниками. LQ – по оси Y, средняя скорость метилирования – по оси X.

ДНК подвергается эволюционному давлению и отбору, а сети метилирования лежат в основе особенностей млекопитающих. Высокая филоэпигенетическая и филогенетическая конгруэнтность обусловлена различиями в уровнях метилирования, и расхождение профилей метилирования ДНК точно отражает генетическую эволюцию (рис. 3.10). Идентифицировано 30 цитозиновых модулей, связанных с максимальной продолжительностью жизни, возрастом, полом, массой тела и риском смертности человека. При этом максимальная продолжительность жизни связана с уровнями метилирования в генах гомеобокса подкласса *HOXL* и может регулироваться транскрипционными факторами плюрипотентности. В целом, модели метилирования у видов с большей максимальной продолжительностью жизни характеризуются уникальными пиками и падениями уровня метилирования, обычно отличаясь от ландшафта метилирования, связанного с возрастом или воздействием факторов, влияющих на риск смертности (Haghani et al., 2023).

### 3.3. Сравнительная геномика

Сравнительная или эволюционная геномика изучает содержание и организацию геномов у разных видов и помогает раскрыть механизмы их эволюции, приближая к разгадке происхождения эукариотического генома (рис. 3.11). Анализируя весь геном или большую его часть, сравнительная геномика оценивает биологическое сходство и различие между организмами и определяет их эволюционные взаимоотношения. Сравнительная геномика – мощный инструмент для быстрого поиска уникальных и консервативных участков генома, изучения эволюционных и функциональных особенностей целевых генов (например, связанных с наследственными или возникшими в результате повреждения ДНК заболеваниями), точной идентификации генов, необходимых для выживания в определённой среде, и установления минимального размера генома, необходимого для жизни (Hong et al., 2021).

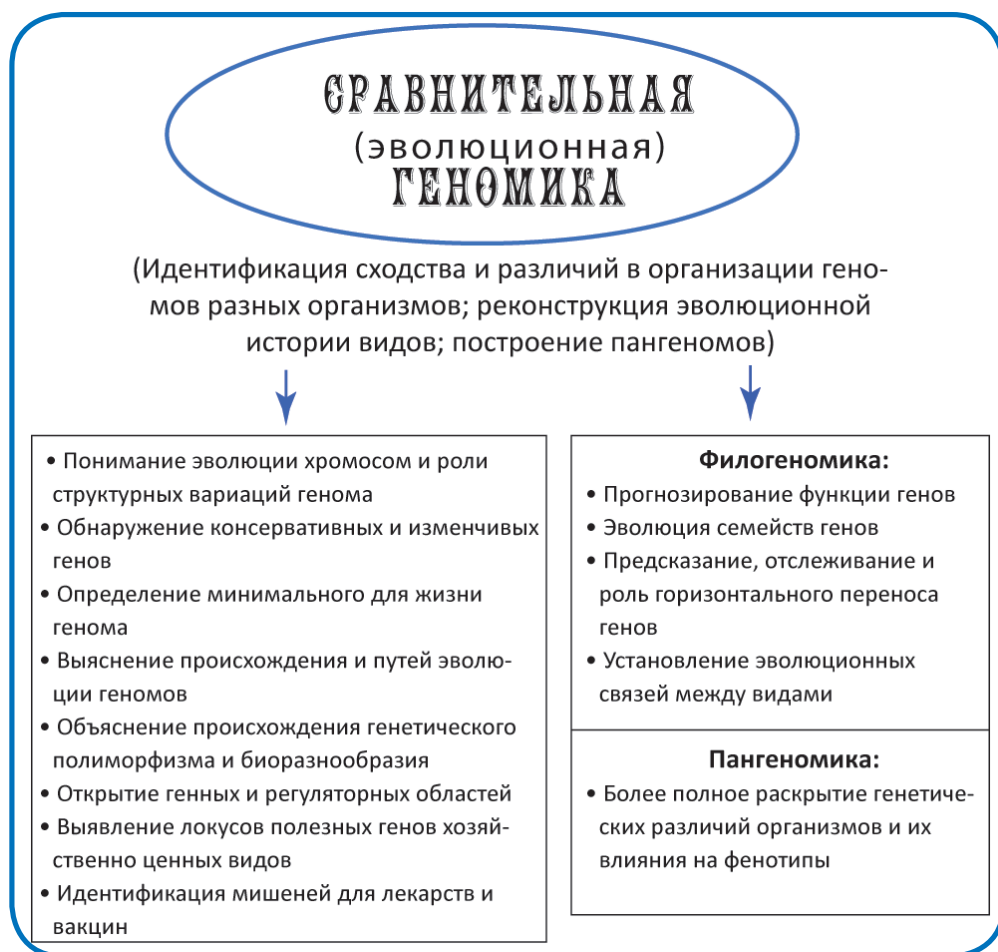
В качестве практического расширения эволюционной теории сравнительная геномика строится на предпосылке, что общие черты двух или более организмов закодированы в консервативных участках ДНК, и более близкие виды обладают похожими геномами. Элементы, ответственные за сходство между видами, должны сохраняться во времени, а элементы, ответственные за различия – дивергировать. Следовательно, гены, определяющие сходство или различия, подвергаются стабилизирующему или положительному отбору соответственно, а гены, не играющие значительной роли в эволюционном успехе, находятся под нейтральным отбором. Поэтому сравнительный геномный анализ начинается с выравнивания последовательностей (этот основной процесс реализуется для каждого нового генома) и далее осуществляется поиск ортологичных последовательностей, т.е. разошедшихся в результате видообразования и выполняющих одинаковые функции у разных видов, в отличие от паралогичных последовательностей, разделённых дубликацией генов (Lander et al., 2001; Hardison, 2003; Touchman, 2010).



**Рисунок 3.10.** Сеть метилирования ДНК связана с филогенией и особенностями млекопитающих (из: Naghani et al., 2023).

А – Филоэпигенетическое древо на основе данных метилирования ДНК из образцов крови. Б – Сети WGCNA идентифицировали 55 модулей кометилирования. В – EWAS данные максимальной продолжительности жизни. Каждая точка соответствует уровню метилирования высококонсервативного CpG. Показано логарифмическое преобразование значения P (ось y) в зависимости от координаты генома человека Hg19 (ось x). Справа – модульный анализ продолжительности жизни. Г – Модуль кометилирования коррелирует с максимальной продолжительностью жизни млекопитающих. Каждая точка соответствует отдельному виду.

WGCNA (взвешенный корреляционный сетевой анализ) – метод системной биологии для описания закономерностей корреляции между генами в образцах микрочипов; может использоваться для поиска кластеров высококоррелированных генов; EWAS (исследование ассоциаций в масштабе всего эпигенома) – исследование набора количественных эпигенетических меток по всему геному для установления ассоциаций между эпигенетическими вариантами и определённым фенотипом/признаком; GYL – жёлто-зелёный: продолжительность жизни; TL – жёлто-коричневый: продолжительность жизни; SDBL – голубой: продолжительность жизни пород собак; MLAWM – пурпурный: продолжительность жизни, масса тела, смертность.



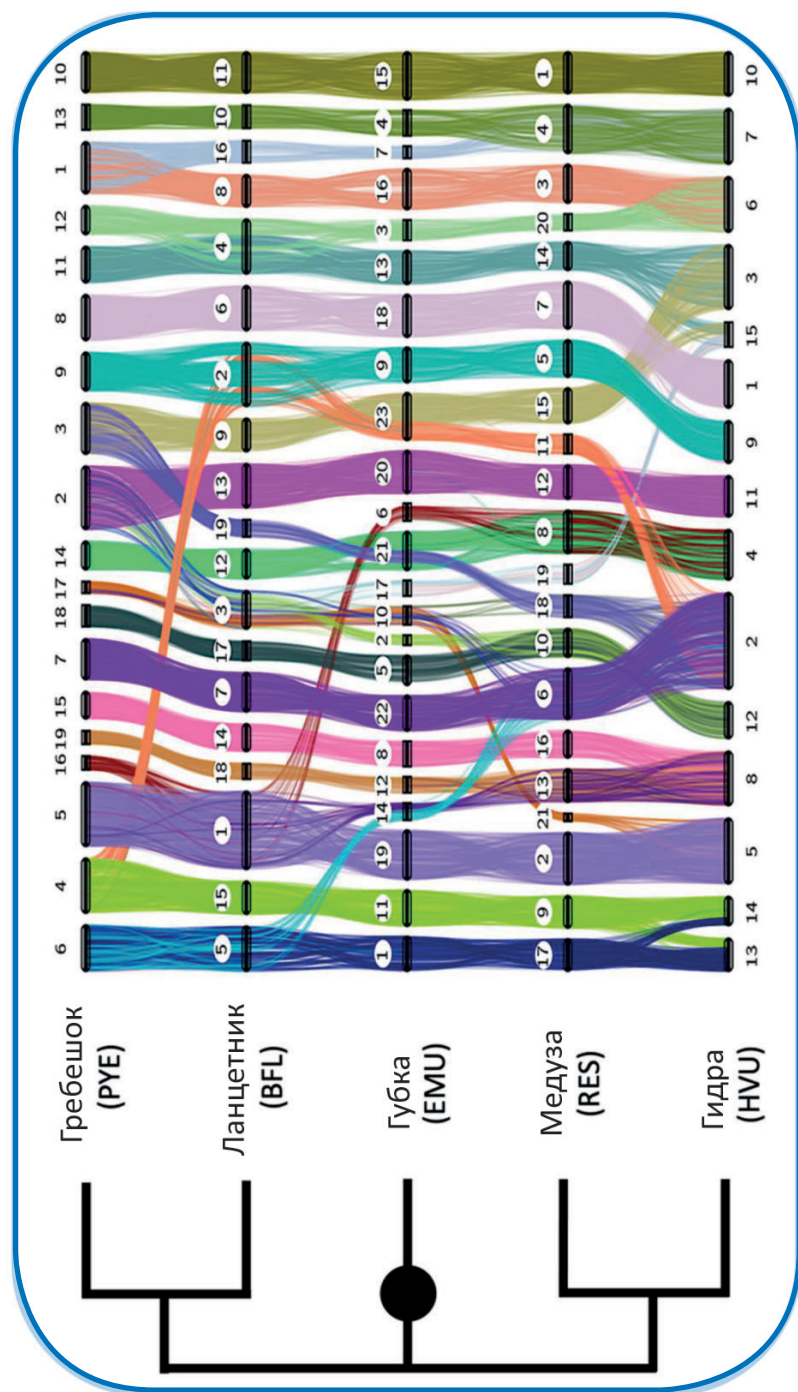
**Рисунок 3.11.** Сравнительная геномика

Прямое сравнение последовательностей ДНК между видами проводится для выявления более тонких различий, позволяющих оценить уровень синтении. В классической генетике под синтением понимается расположение каких-либо локусов на одной и той же хромосоме, наблюдаемое в разных наборах хромосом и у разных видов; если также совпадает и порядок локусов в хромосоме, это называется коллинеарностью. В сравнительной геномике синтением – это сохранённый порядок генов на хромосомах родственных видов, который является результатом происхождения от общего предка. Консервативная (или общая) синтением обозначает смежные выравненные области, которые сохраняют порядок и ориентацию выравнивания, допуская при этом микроперестройки внутри синтенных областей. Выделяют также макросинтением (коллинеарность порядка генов в масштабе всей хромосомы), микросинтением (коллинеарность порядка небольшого числа генов на участке хромосомы) и

мезосинтению (содержание генов в хромосоме сохраняется, но коллинеарность отсутствует). Эволюционные перестройки генома (инсерции, делеции, транслокации) могут привести как к приросту синтении, так и к её потере. Консервативная синтения считается одним из самых надёжных критериев для определения ортологии геномных областей у разных видов. Древняя консервированная синтения, сохраняющаяся на протяжении более полумиллиарда лет эволюции, обнаружена среди билатеральных животных, губок и книдарий (рис. 3.12). Несколько синтенных единиц Metazoa сохранились у одноклеточных родственников животных, подразумевая их существование до появления Animalia (Simakov et al., 2022).

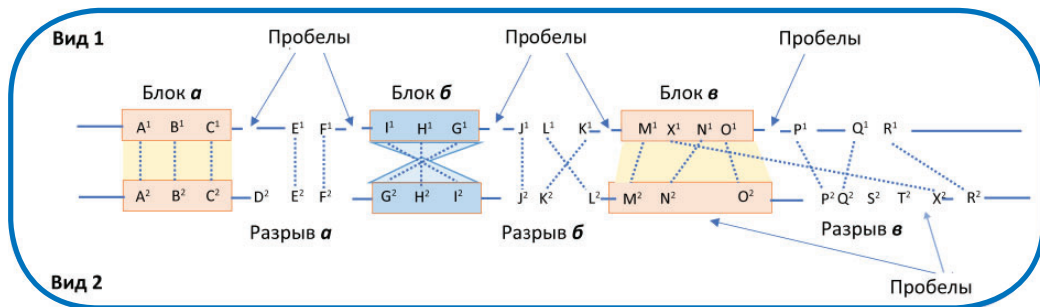
Высокая сохранность синтении может отражать важные функциональные взаимоотношения между генами, например, преимущество совместного наследования или общие регуляторные механизмы генного комплекса (Amores et al., 1998). Сравнительная геномика и анализ синтении дают определённое представление о механизмах геномных перестроек, происходящих после дивергенции видов. Анализ синтении широко используется для изучения сложных геномов с использованием простого модельного организма в качестве платформы. Например, геномы риса и гексаплоидной мягкой пшеницы, различающиеся по содержанию ДНК примерно в тридцать раз, содержат гомеологичные гены и фрагменты геномной ДНК в аналогичном порядке. Эта консервативность порядка генов в хромосомах может обеспечить основу для новых стратегий поиска генов у пшеницы (Kurata et al., 1994). Геномные исследования показали, что порядок генов не только различается между геномами, но и теряется он значительно быстрее, чем сходство белков. Тем не менее, удалось найти некоторые консервативные кластеры генов, которые либо взаимодействовали физически, либо обладали родственными функциями (Moreno-Hagelsieb et al., 2001). Сравнение последовательностей геномной ДНК также способствует локализации кодирующих и регуляторных генов. Благодаря такому сравнению, которое используется для анализа каждой новой геномной последовательности, было идентифицировано большинство функциональных областей генома человека (Lander et al., 2001; Hardison, 2003; Touchman, 2010). Интересно, что недавние исследования геномной синтении Metazoans выявили большое количество пар синтенных генов только в линиях билатеральных животных. При этом было показано, что у книдарий и первичноротых консервативные синтенные блоки включают множество высокоэкспрессируемых генов, а блоки синтении у позвоночных хотя и обширные, но имеют сравнительно небольшое число генов, часто с низкой коэкспрессией (Nehmeyer et al., 2023).

На результат сравнения геномных последовательностей существенно влияют филогенетические расстояния между видами. Когда они увеличиваются, сходство последовательностей уменьшается, а общие геномные признаки теряются. Поэтому, порядок как кодирующих генов, так и регуляторных последовательностей транскрипции редко сохраняется на больших филогенетических расстояниях (например, миллиард лет дивергенции), в то время как и кодирующие, и некодирующие последовательности ДНК обнаруживаются в консервативных элементах на меньших фило-



**Рисунок 3.12.** Древняя консервативная синтенция у билатерий, губок и книдарий (из: Simakov et al., 2022).

Синтенция (хромосомные связи без учета порядка генов) многих животных удивительно похожи даже после полуллиарда лет независимой эволюции. Пронумерованные горизонтальные полосы представляют хромосомы пяти видов: гребешок, ланцетник, губка, медуза, гидра. Губка находится в центральном положении для отображения консервативных синтенций как с билатериями (вверху), так и с книдариями (внизу). Цветные вертикальные линии соединяют ортологичные гены; каждый цвет представляет собой отдельную группу наследственного сцепления (ALG). Филогенетическое древо показано с корнем, отмеченным чёрным кругом.



**Рисунок 3.13.** Определение блоков и разрывов синтении (по: Liu et al., 2018).

Буквами обозначены гены, расположенные на хромосомах двух видов. Каждому гену присвоен номер, обозначающий вид, к которому он принадлежит (вид 1 или 2). Ортологичные гены соединены пунктирными линиями, а гены, не имеющие ортологического родства, рассматриваются как пробелы (гэпы) в программах синтении. По критерию не менее трёх ортологичных генов (якорей), синтениный блок может состоять из ортологов одинакового порядка (блок а), обратного порядка (блок б) или допускающих некоторые пробелы (блок в). Напротив, случаями нарушения синтении могут быть отсутствие ортологов (разрыв а), совмещенный порядок генов (разрыв б) или пробелы (разрыв в).

генетических расстояниях (50–200 млн л. дивергенции). Сравнение очень похожих геномов (около 5 млн л. дивергенции) позволяет выявить причины, ответственные за тонкие различия в биологических формах (Hardison, 2003; Touchman, 2010).

Опираясь на высококачественные полногеномные данные, можно реконструировать хромосомные перестройки, нарушающие консервативные синтениные блоки, и достигать максимального разрешения межвидовой гомологии на основе полногеномного выравнивания. Недавно разработанный метод halSynteny (часть пакета сравнительной геномики HAL, иерархическое выравнивание) позволяет точно и быстро обнаруживать синтению при множественных полногеномных сопоставлениях (Krashenninnikova et al., 2020). Существует много инструментов, предназначенных для поиска синтениных блоков (например, GRIMM-Synteny, BLASTZ, DAGChainer и др.), однако они требуют разных форматов данных, и для многих из них необходима строгая и надёжная аннотация ортологичных генов. При использовании halSynteny выравнивание является единственным необходимым входным параметром (Krashenninnikova et al., 2020). Блоки и разрывы синтении можно определить, как показано на рисунке 3.13, при этом большое значение имеет минимальное качество сборки генома (Liu et al., 2018).

Сравнительная геномика способствует пониманию структурных вариаций генома (SV), играющих ключевую роль в адаптивной эволюции и видообразовании. Структурные изменения могут трансформировать линейную организацию ДНК и повлиять на трёхмерную структуру генома, что неизбежно приводит к изменениям в транскриптоме и, в конечном итоге, – в свойствах организма. Накопление высококачественных полногеномных данных помогает каталогизировать весь спектр генетической изменчивости, получить представление о механизмах, которые создают «горячие точки» разнообразия в геноме, разделить эволюционные и демографи-

ческие влияния на эволюцию генома и лучше понять роль тандемных повторов в эволюционных процессах. Кроме того, синтез имеющихся данных даёт новое представление о роли структурной изменчивости в эволюции, её распределении между таксонами и т. д. Ожидается, что структурная изменчивость будет интегрирована в экологические и эволюционные приложения (например, в природоохранную геномику, сельскохозяйственную геномику и палеогеномику) (Mérot et al., 2020). Накопление полногеномных данных о видах из разных таксономических групп может повысить точность филогенетической и таксономической классификации (Yu et al., 2017). Данные сравнительной геномики также дают представление о метаболических путях таких заболеваний, как рак, об адаптации к окружающей среде и т. д. (Awasthi et al., 2021; Steindorff et al., 2021). Сравнительный подход анализа геномных данных оказался успешным для идентификации локусов полезных генов у видов сельскохозяйственных культур (Morrell et al., 2012), разработки вакцин (Seib et al., 2012) и др.

В международном проекте «The Great Ape Genome Project» методы сравнительной геномики использовались для исследования генетической изменчивости всех шести видов человекообразных обезьян. Неожиданно оказалось, что несмотря на значительное сокращение численности популяций, в их генофонде сохранились высокие уровни изменчивости (Prado-Martinez et al., 2013). В рамках сравнительных геномных исследований был также запущен проект зоономии (The Zoonomia Project), целью которого является понимание роли геномов в формировании огромного богатства разнообразия животных. Хотя проект отдаёт приоритет филогенетическому разнообразию, он также направлен на поддержание медицинских исследований и сохранение биоразнообразия (Zoonomia Consortium, 2020). Интересно, что термин «зоономия» впервые упоминается в книге «Zoonomia: The law of organic life», написанной в 1794 г. дедом Чарлза Дарвина – Эразмом Дарвиным, в которой он предугадал многие эволюционные идеи. Проект дарвиновского древа жизни (The Darwin Tree of Life Project) направлен на секвенирование геномов 70000 видов эукариот в Великобритании и Ирландии (биота Британских островов одна из наиболее изученных в мире). Он проводится в рамках проекта биогенома Земли (The Earth BioGenome Project), целью которого является создание геномного атласа всех эукариот планеты. Это необходимо для достижения понимания эволюции разнообразия и разработки практических приложений в области медицины, природоохраны и сельского хозяйства.

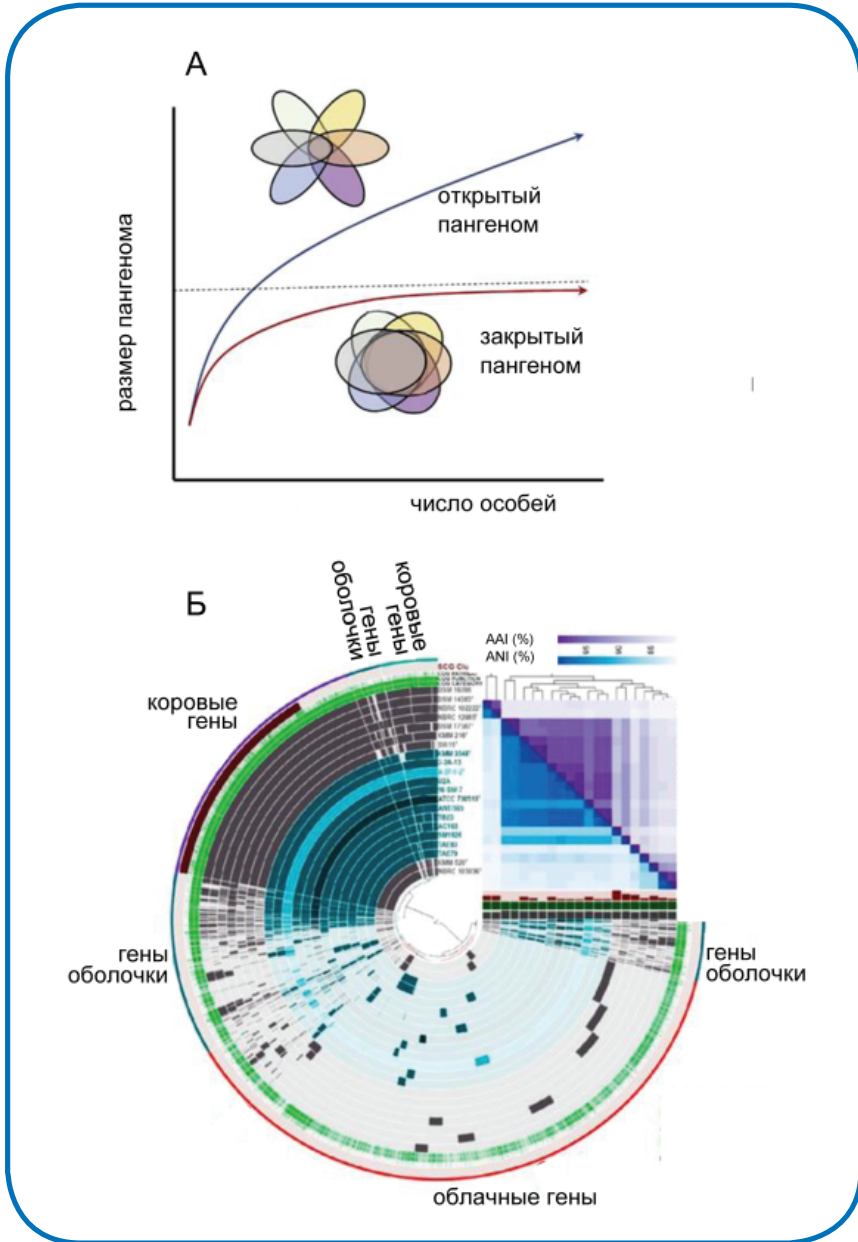
**3.3.1. Пангеном** имеет важнейшее значение в изучении эволюции методами сравнительной геномики; соответственно область изучения пангенома называется пангеномикой. Генетическая изменчивость, от однонуклеотидных полиморфизмов до больших структурных вариантов, может вызывать различия в содержании генов у особей одного и того же вида, и как следствие – привести к существенным фенотипическим изменениям. С развитием геномных исследований стало очевидным, что одного эталонного генома недостаточно для охвата всей полноты генетического разнообразия вида. Платформу для всесторонней оценки генетического разнообразия вида предоставляет пангеномный анализ (Tao et al., 2019). Термин «пангеном»

был впервые предложен Эрве Теттелином при изучении серотипов *Streptococcus agalactiae* Lehmann and Neumann, 1896 – основной причины неонатальных инфекций у человека (Tettelin et al., 2005). Было показано, что разные формы *S. agalactiae* могут быть описаны пангеномом, состоящим из основного генома, общего для всех изолятов, который составляет  $\approx 80\%$  любого отдельного генома, и дополнительного генома, состоящего из частично общих и штамм-специфичных генов.

В настоящее время различают структурные и функциональные пангеномы, а также пангеномы конкретного вида или таксона более высокого ранга. Пангеном можно разделить на «коровый пангеном / пангеном ядра», «пангеном оболочки» и «облачный пангеном», содержащий гены, присутствующие у всех представителей исследуемой группы, в двух или более, или только у отдельных из них соответственно (рис. 3.14). Иногда облачный пангеном также называют «дополнительным геномом», содержащим «необязательные» гены, хотя многие из них могут играть важную роль в эволюции генома и его взаимодействиях с окружающей средой (ассоциированы с адаптивными функциями) (Golicz et al., 2016; Tranchant-Dubreuil et al., 2018; Tao et al., 2019).

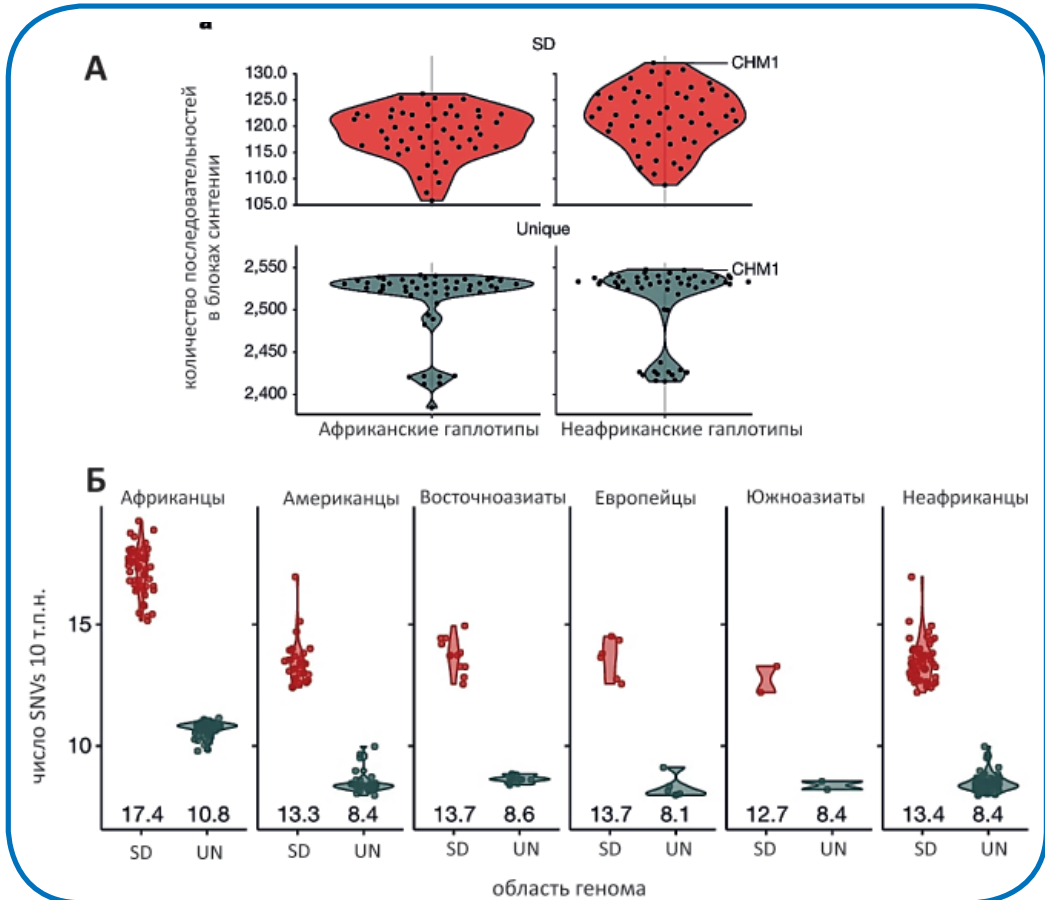
По мере увеличения числа секвенированных геномов количество коровых и новых генов уменьшается, а затем стабилизируется; такие пангеномы называются закрытыми (рис. 3.14). Пангеномы сильно различаются по размеру и соотношению основных и дополнительных генов. Предположительно, виды с большими открытыми пангеномами занимают более разнообразные ниши и имеют бóльший эффективный размер популяции по сравнению с видами, имеющими меньшие пангеномы. Выделяют несколько причин появления переменных геномов: (1) межвидовая гибридизация, (2) дупликации геномов, геномных сегментов и отдельных генов, а также горизонтальный перенос генов и (3) возникновение генов *de novo* (Richard, 2020). Кроме того, на формирование пангенома влияют отбор и дрейф, а также мобильные генетические элементы (Brockhurst et al., 2019). Однако, чтобы понять причины и механизмы динамики пангенома в пространстве и времени, вероятнее всего, потребуются новые теоретические и экспериментальные подходы (Brockhurst et al., 2019).

Большое внимание сейчас уделяется построению пангенома человека и его применению в различных аспектах (Liao et al., 2023; Vollgen et al., 2023; Hickey et al., 2024). Например, использование черного варианта пангенома человека (для данных с короткими прочтениями) снизило ошибку обнаружения минорных вариантов на 34% и увеличило количество структурных вариантов на 104% по сравнению с анализом на основе референсного генома GRCh38 (Liao et al., 2023). При исследовании дублицированных (труднодоступных) сегментов было установлено, что в них изменчивость вариантов одиночных нуклеотидов повышена на 60% по сравнению с уникальными областями, и имеются выраженные популяционные различия (рис. 3.15). Причём авторы считают, что не менее 23% этого увеличения связано с конверсией генов (Vollgen et al., 2023). Широко применяемый в исследованиях микроорганизмов, пангеномный анализ позволяет выявить у них новые кластеры генов биосинтеза, идентифицировать консервативные и уникальные природные продукты,



**Рисунок 3.14.** Пангеномы.

А – открытый и закрытый пангеном (из: Golicz al., 2020); Б – пангеном 21 штамма *Pseudoalteromonas* (из: Valabanova et al., 2023). Геномы организмов организованы в виде радиальных слоёв генных кластеров ядра, мягкого ядра, оболочки и облака, которые определяются филогенетическим деревом в центре диаграммы. Тепловая карта отображает попарные значения средней идентичности нуклеотидов (ANI) и аминокислот (AAI).



**Рисунок 3.15.** Повышенная изменчивость вариантов одиночных нуклеотидов (SNVs) в дуплицированных (SD) и уникальных (UN) областях (из: Vollger et al., 2023).

А – сглаженный скрипичный график; точками указаны отдельные гаплотипы.

Б – Среднее количество SNV (10 т. п. н. скользящее окно) в SD (красный) по сравнению с UN (серый) областями по суперпопуляциям и со средним значением (под каждой «скрипкой»).

лучше понять внутривидовые отношения и т. д. (Shi et al., 2022; Balabanova et al., 2023; Meesil et al., 2023).

К 2020 г. было сконструировано более 20 эукариотических пангеномов (таб. 3.2) для организмов, размер генома которых составляет от 12 Mb (пекарские дрожжи) до 17 Gb (мягкая пшеница) (Golicz et al., 2020). Ожидается, что в ближайшее десятилетие применение пангеномов станет нормой, заменив традиционный подход к геномному анализу с использованием единственной референсной последовательности. Дальнейшее распространение пангеномных исследований на более высокие таксономические группы создаст необходимые базы данных для изучения тонких геномных различий, облегчит идентификацию «минимального генома» и т. д. (Golicz et al., 2020).

**Таблица 3.2.** Хронология развития пангеномных исследований (по: Golicz al., 2020, с дополнениями).

Год	Событие
2005	Эрве Теттелин предложил первый пангеном
2006	Создан пангеном <i>Streptococcus pneumoniae</i>
2007	Мишель Морганте предложил концепцию пангенома
2008	Создан пангеном <i>Escherichia coli</i>
2009	На базе трёх геномов построен пангеном человека; по данным 573 геномов создан пангеном надцарства бактерий
2013	Создан пангеном одноклеточного организма фитопланктона ( <i>Emiliania huxleyi</i> )
2014	Создан пангеном дикорастущей сои ( <i>Glycine soja</i> )
2015	Опубликован обзор по аналитическим инструментам и моделям за 10 лет пангеномных исследований; построен пангеном <i>E. coli</i> по данным 2085 геномов
2016	Созданы пангеномы капусты огородной ( <i>Brassica oleracea</i> ) и трех видов тополя ( <i>Populus nigra</i> , <i>P. deltoides</i> и <i>P. trichocarpa</i> )
2017	Созданы пангеномы пшеницы и травянистого растения рода <i>Bracypodium</i>
2018	Созданы пангеномы риса ( <i>Oryza sativa</i> ), рапса ( <i>Brassica napus</i> ) и пекарских дрожжей (использовано 1011 изолятов)
2019	Созданы пангеномы человека африканского происхождения (с использованием 910 геномов) и китайцев ханьского происхождения (с использованием 275 геномов); Создан пангеном свиньи ( <i>Sus scrofa domesticus</i> ) Создан пангеном средиземноморской мидии ( <i>Mytilus galloprovincialis</i> )
2021	Созданы пангеномы земляники рода <i>Fragaria</i>
2023	Предложен конвейер “Minigraph-Cactus panggenome” для создания пангеномов непосредственно из выровненных геномных последовательностей (тестирован на данных 90 гаплотипов человека); Разработан алгоритм “seqwish” для получения несмещённых графов вариаций пангенома; Сконструирован первый пангеном овец ( <i>Ovis aries</i> ); Создан черновой референсный пангеном человека, опирающийся на 47 поэтапных диплоидных сборок из когорты генетически разнообразных индивидов

Разработаны несколько типов программного обеспечения для анализа пангеномов. К ним относятся программы для кластеризации гомологичных генов, определения однонуклеотидных полиморфизмов, построения пангеномных профилей, а также реконструкции филогенетических связей между ортологичными генами, семействами штаммов и изолятов. Кроме того, они включают программы для функционального поиска, аннотации и курирования данных, а также для визуализации результатов (Xiao et al., 2015). Создана мультимедийная платформа Anvi'o, которая включает пангеномный и метапангеномный анализы, а также рабочие процессы визуализации. В ней геномы представлены в виде концентрических кругов, где каждый радиус соответствует отдельному семейству генов. Это позволяет сравнивать более 100 геномов в интерактивном формате (Egen et al., 2021). Недавно для пангеномного анализа предложен конвейер Minigraph-Cactus pangenome, создающий пангеномы непосредственно из геномных выравниваний (может включать до 90 гаплотипов). Он позволяет отслеживать (строить графики) все формы генетической изменчивости, а использование эталона CHM13 от Консорциума теломер-к-теломеру повышает точность метода (Hickey et al., 2024). Сформировалась отдельная дисциплина – вычислительная пангеномика. Под пангеномом в ней понимается любая совокупность геномных последовательностей, которые должны быть проанализированы совместно, или использованы в качестве эталона (Marschall et al., 2018).

**3.3.2. Филогеномика.** Опирающаяся на методы сравнительного анализа геномов и эволюционный подход, филогеномика может рассматриваться, как самостоятельная ветвь сравнительной (эволюционной) геномики. Причём для филогеномного анализа полная последовательность генома необязательна. Филогеномика охватывает несколько областей исследований на стыке молекулярной и эволюционной биологии, преследуя две основные цели: (1) установить филогенетические отношения между таксонами и получить представления о механизмах молекулярной эволюции, а также (2) определить предполагаемые функции последовательностей ДНК (или белков), опираясь на данные филогенетических сравнений (Young, Gillung, 2020). В глобальном масштабе филогеномика – это наука о реконструкции эволюционной истории жизни на Земле. Термин «филогеномика» был предложен Джонатаном Эйзенем и первоначально предназначался для обозначения исследований в области предсказания функции генов у бактерий (Eisen et al., 1997), но вскоре его стали использовать в филогенетическом контексте (O'Brien, Stanyon, 1999).

Возможность получать данные о последовательностях в масштабе генома (чему способствовало удешевление методов полногеномного секвенирования) обеспечила переход от филогенетики (оперирующей одним или несколькими генами) к более эффективной филогеномике с новым более мощным вычислительным обеспечением для обработки и визуализации данных. Несмотря на то, что используемые в филогеномике маркеры часто составляют менее 5% генома (секвенирование с сокращённым представлением, RRS), они обеспечивают получение надёжных данных высокого разрешения, которые необходимы для построения

филогенетических гипотез (Burki et al., 2008; Young, Gillung, 2020; Carter et al., 2023). В частности, в качестве маркеров используются ультраконсервативные элементы (UCE, включая генетически изменчивые фланкирующие области; впервые описаны как «молекулярные окаменелости»), локусы закрепленного гибридного обогащения (АНЕ, метод для извлечения ортологичных участков ядерного генома, которые присутствуют в ограниченном количестве или в единственной копии в различных линиях), консервативные неэкзонные элементы (CNEE; отличаются от консервативных некодирующих элементов отсутствием последовательностей экзонов), нетранслируемые области (UTR), интроны, экзоны, полиморфизм одиночных нуклеотидов (SNP), анонимные области генома (RADseq; данный подход считали «самым важным научным прорывом» десятилетия 2010-х годов) и участки митохондриальной ДНК (мтДНК; как преимущество, имеет высокую скорость нуклеотидных замен и может сохранять (например, у позвоночных) высоко консервативную архитектурную структуру, однако размер генома небольшой и у большинства животных составляет всего 14–20 т.п.н.). Разные маркеры и подходы имеют свои преимущества, ограничения и недостатки. Например, экзоны и подходы по захвату цели (UCE, АНЕ и CNEE) ориентированы на более глубокие временные масштабы; для групп с недавней радиацией больше подходят мтДНК и некодирующие области; RADseq не используется, если необходимо исследовать конкретные области генома; анализ однонуклеотидных полиморфизмов удобен для изучения демографии, филогеографии, эволюции генома и др. (Carter et al., 2023).

В целом, филогеномика включает (1) прогнозирование функции генов (простой сравнительный анализ последовательностей может оказаться неэффективным, так как гены с разными функциями могут иметь похожие последовательности; в филогеномике функции неохарактеризованных генов предсказываются их филогенетическим положением по отношению к охарактеризованным генам); (2) выяснение особенностей эволюции семейств генов (например, сравнение геномов позволяет обнаруживать дубликации генов/геномов, что часто имеет большие эволюционные последствия); (3) предсказание и отслеживание горизонтального/латерального переноса генов (сравнение геномов позволяет уточнить, является ли сходство генов свидетельством их происхождения от общего предка, или нет, в чём простой филогенетический анализ значительно менее успешен); и (4) установление эволюционных отношений между таксонами (эффективные для близких видов, филогенетические реконструкции по данным отдельных генов часто вызывают проблемы при сравнении отдалённых таксонов из-за горизонтального переноса генов, конвергенции и разных скоростей эволюции, а геномные данные позволяют успешно справляться с этими проблемами) (Eisen et al., 1997; Eisen, 1998; O'Brien, Stanyon, 1999; Burki et al., 2008; Young, Gillung, 2020).

Выполнимость одновременного анализа сотен геномов предоставила беспрецедентные возможности для разрешения филогенетических противоречий, выявления геномных инноваций и генетической основы репродуктивной изоляции.

Благодаря усовершенствованиям в высокопроизводительном секвенировании и вычислительных технологиях в настоящее время достигнут значительный прогресс в создании полных филогений и временных шкал для многих таксономических групп эукариот. По оптимистичным прогнозам, геномы большинства ныне живущих видов будут секвенированы в ближайшие десять лет. Использование полных геномов со всеми кодирующими и некодирующими последовательностями поможет выявить и понять ранее скрытые геномные эффекты. В результате будет создан мощный ресурс для расшифровки механизмов эволюции эукариотического генома и реконструкции глобального древа жизни (Челомина, 2023).

Однако следует признать, что колоссальный приток геномных данных увеличивает вероятность систематических ошибок. Предложено несколько способов для их смягчения, включая тесты на композиционную неоднородность, апостериорные подходы к прогнозированию для оценки соответствия модели, а также методы для определения уровня чувствительности результатов к удалению сайтов или локусов и для проверки согласованности генеалогии между различными областями генома (Young, Gillung, 2020). Для датировки крупных эволюционных событий молекулярные часы, основанные на постоянных скоростях мутаций, не могут считаться надёжными из-за изменения во времени и между линиями. Поэтому был предложен подход калибровки молекулярных часов с использованием горизонтальных переносов генов, которые можно идентифицировать на филогенетических реконструкциях и которые соответствуют определённым геохимическим маркерам или точно датированным ископаемым (Wolfe, Fournier, 2018). Подход «перекрестного связывания», использующий аналогичные временные ограничения, основанные на ископаемых, использует дубликации генов/паралогичные белки путём формального связывания калибровочных дат узлов по всем паралогичным поддеревьям через иерархические байесовские модели (Moody et al., 2024).

### 3.4. Популяционная геномика

Популяционная геномика представляет собой крупномасштабное сравнение последовательностей ДНК популяций с использованием различных генетических маркеров для решения сложных вопросов (в том числе тех, на которые не может ответить классическая генетика), и она играет важнейшую роль в молекулярной и эволюционной биологии, в сохранении биоразнообразия, экологии и здоровья человека (рис. 3.16). Основываясь на современном высокопроизводительном секвенировании, инновационном программном обеспечении, статистических подходах и математических моделях, популяционная геномика изучает масштабы и причины геномной изменчивости в популяциях, включая совместные эффекты отбора и сцепления (Charlesworth, 2010).

В популяционной геномике часто применяется секвенирование с сокращённым представлением (RRS) (Luikart et al. 2018). Такая технология позволяет использовать массовое параллельное секвенирование для обнаружения и генотипирования тысяч



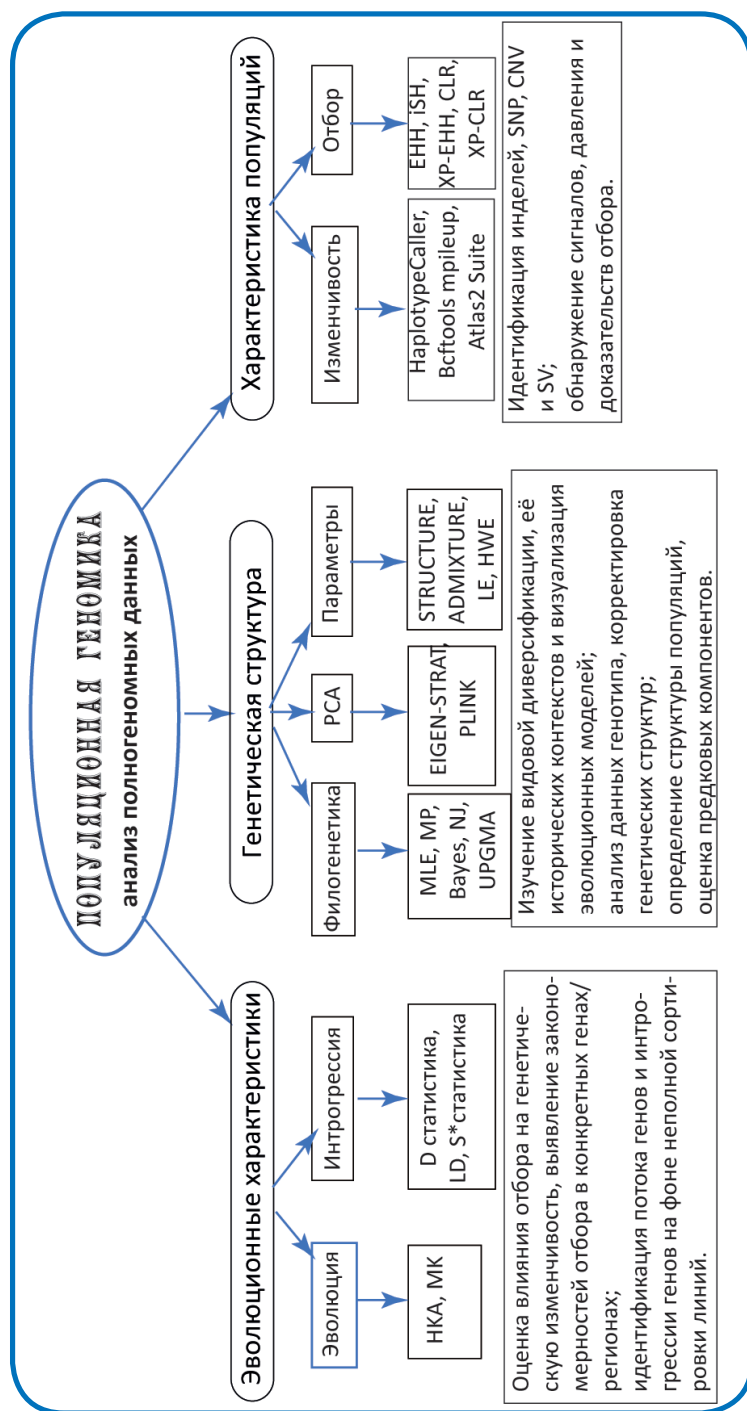
**Рисунок 3.16.** Популяционная геномика

локусов однонуклеотидного полиморфизма (SNP) у сотен особей одновременно и не требует полногеномных данных. В RRS для популяционного анализа обычно применяются три подхода: (1) секвенирование ДНК, связанной с сайтами рестрикции (RADseq); (2) целевой захват последовательности, позволяющий секвенировать любую область генома, например, экзоны; и (3) технология гибридизации на микрочипах (DArT), которая может обнаруживать и типировать вариации ДНК в нескольких сотнях геномных локусов параллельно, без предварительной информации о последовательностях (используется в основном при исследовании сельскохозяйственно важных видов растений) (Luikart et al., 2018).

Для изучения генетической структуры популяций используют анализ главных компонент (PCA), включая EIGENSTRAT 6.0.1 и PLINK 2.0; параметрические методы, такие как STRUCTURE 2.3.4 и ADMIXTURE 1.3.0, или основывающиеся на равновесии Харди–Вайнберга (HWE) и равновесии сцепления (LE); а также методы филогенетического анализа, опирающиеся на генетические расстояния или признаки, в таких программных обеспечениях, как MrBayes 3.2.2, MEGA 11.0 и PhyML 3.3. Для характеристики популяций применяют методы анализа генетической изменчивости и обнаружения сигналов отбора, например, F<sub>ST</sub> и XP-F<sub>ST</sub>, iHS, CLR и XP-CLR, а также VcfTools mpileup, HaplotypeCaller и Atlas2 Suite. Методы определения потока генов и интрогрессии включают статистику D Паттерсона, статистики S\* и D (тест AB-BABA), неравновесие по сцеплению (LD). Для обнаружения эволюционных изменений генома полезны тесты HKA и MK, которые оценивают давление отбора на генетическую изменчивость, выявляя закономерности отбора в конкретных генах или регионах (Lu et al., 2025) (рис. 3.17).

Популяционная геномика произвела революцию во многих областях биологии и предоставила концептуально новые подходы, значительно продвинувшие наше понимание роли таких эволюционных процессов, как генетический дрейф, поток генов, нейтральный отбор, а также геномных основ адаптации, инбридинга, акклиматизации и т. д. В популяционной геномике исследования, основанные на полногеномном секвенировании и ресеквенировании, могут использоваться для изучения генетической основы редких и тяжёлых заболеваний человека, таких как врождённая депрессия (включая идентификацию локусов, влияющих на инбредную депрессию и адаптацию к окружающей среде) и для развития прецизионной медицины. Геномные исследования также эффективны для оценки размера популяции и исторической демографии, идентификации локусов, оказывающих большое влияние на экологически важные фенотипические черты, для выявления адаптивного потока генов и понимания механизмов видообразования и т. д. (Charlesworth, 2010; Hendricks et al., 2018; Luikart et al., 2018; Hohenlohe et al., 2020; Челомина, 2024). Кроме того, знание внутривидовой генетической изменчивости на геномном уровне важно для эффективного управления сельскохозяйственными животными и растениями, а также природными ресурсами в целях сохранения экономически ценных, редких и исчезающих видов (Vernot et al., 2012; Lachance, Tishkoff, 2013; Hendricks et al., 2018; Longo et al., 2020; Pecoraro et al., 2020).

Результаты популяционной геномики используются для создания баз данных, которые в дальнейшем могут быть востребованы для более эффективного управления здоровьем населения. В процессе реализации международного проекта «1000 геномов» были созданы подробные карты генетических вариантов человека на уровне генома, включая однонуклеотидные полиморфизмы, индели и структурные изменения (включая вариации в числе копий генов), а также интерактивные карты генетического разнообразия более 1000 геномов жителей разных стран, этнических групп и т. д. (Abecasis et al., 2010; 2012). Как продолжение проекта «100000 геномов», пра-



**Рисунок 3.17.** Методы анализа популяций для данных секвенирования всего генома (по: Lu et al., 2025).  
 HWI – равновесие Харди-Вайнберга; LE – равновесие сцепления; iHS – показатель гомозиготности гаплотипа; EHH и XP-EHH – популяционный и кросс-популяционный тесты на гомозиготность гаплотипов; CLR и XP-CLR – популяционный и кросс-популяционный тесты отношения правдоподобия; PCA – анализ главных компонент; HKA – тест Хатгги–Крейтмана–Агуале; МК – тест Макдональда–Крейтмана

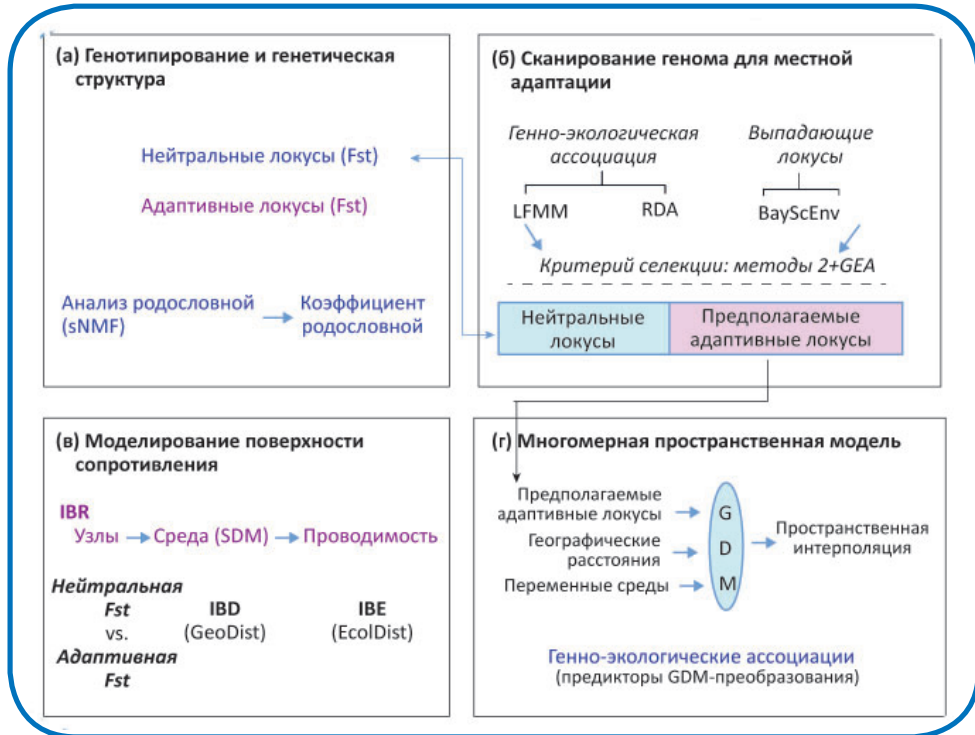
вительство Великобритании (UK Research and Innovation) недавно объявило о новом амбициозном проекте по секвенированию геномов 500000 человек с целью лучшего понимания, диагностики, лечения и профилактики жизненно важных заболеваний, включая карциному и деменцию.

### 3.5. Ландшафтная геномика

Ландшафтная геномика – это относительно новая ветвь геномики, возникшая на базе популяционной геномики и объединяющая некоторые аспекты популяционной и ландшафтной генетики и ландшафтной экологии. Целью ландшафтной геномики является идентификация изменений в геноме популяций диких видов, которые непосредственно вызваны местными факторами окружающей среды и вариантами генов, обеспечивающих локальную адаптацию. В отличие от классических генетических исследований, ландшафтная геномика использует десятки генов для одновременного мониторинга и обнаружения более тонких форм генетической изменчивости, сосредотачиваясь на адаптивных маркерах (Rellstab et al., 2015; Li et al., 2017). Другими словами, ландшафтная геномика предполагает *«использование геномных технологий для изучения нейтральных и адаптивных вариаций в масштабах всего генома экологически разнообразных популяций в гетерогенных ландшафтах для решения новых или ранее неразрешимых вопросов, таких как прогнозирование адаптивной способности при изменении окружающей среды»* (Grummer et al., 2019). Подобно другим исследованиям в области популяционной геномики, ландшафтная геномика основана на современных технологиях высокопроизводительного секвенирования и часто использует методы секвенирования с сокращённым представлением (см. выше) (Luikart et al., 2018).

В зависимости от цели исследования и типа генетических данных различают «нейтральную» и «адаптивную» ландшафтную геномику (Luikart et al., 2018; Balkenhol et al., 2017). Геномика нейтрального ландшафта основана на нейтральных генетических данных и фокусируется на процессах потока и дрейфа генов, изучая влияние окружающей среды на генетическую структуру популяций. Она включает в себя проверку четырёх основных ландшафтных гипотез: (1) изоляция расстоянием (IBD), т. е. как географические расстояния влияют на поток генов; (2) изоляция барьерами (IBB), т. е. каково влияние географических барьеров на поток генов; (3) изоляция посредством сопротивления (IBR), т. е. неоднородное сопротивление ландшафтной матрицы движению и потоку генов; и (4) изоляция окружающей средой (IBE), т. е. зависимость генетической дифференциации образцов от различий в их среде обитания (Balkenhol et al., 2017) (рис. 3.18). Статистическая увязка данных нейтральной геномики осуществляется различными методами, основанными на специальных аналитических подходах (Balkenhol, Fortin, 2016; Wagner, Fortin, 2016; Balkenhol et al., 2017).

Адаптивная ландшафтная геномика сначала идентифицирует локусы или геномные области, находящиеся в процессе отбора, а затем связывает эти адаптивные генетические данные с неоднородностью окружающей среды. Существует несколь-



**Рисунок 3.18.** Схема анализов, используемых для разделения нейтральных генетических локусов от потенциально адаптивных и описания моделей генетической дифференциации в ландшафтном масштабе (по: Shryock et al. 2021).

Сначала генотируют многие популяции и проводят анализ предков (а), затем используют сканированные геномы для отделения нейтральных локусов от потенциально адаптивных (б). Далее проводится причинно-следственное моделирование для выявления механистических процессов, лежащих в основе генетической изменчивости, включая изоляцию расстоянием (IBD), изоляцией сопротивлением среде обитания (IBR) и изоляцией окружающей средой (IBE) (в). В завершение, применяют многомерную обобщённую модель несходства (GDM) для прогнозирования пространственной изменчивости в ландшафтном масштабе в потенциально адаптивных локусах, что приводит к пространственной интерполяции (г).

ко подходов к достижению этих целей (Storfer et al., 2016; Balkenhol et al., 2017): (1) корреляционный (ЕАА); основан на выявлении выпадающих локусов (т.е. адаптивных локусов) и/или анализе ассоциации с окружающей средой; требует данные о нейтральной генетической структуре популяций; (2) фенотипический, основанный на анализе локусов качественных признаков (QTL) или полногеномном анализе ассоциаций (GWAS); (3) использование генов-кандидатов (основан на информации, доступной для модельных видов с известными адаптивными генами или областями генома); и (4) экзомный или транскриптомный подход, опирающийся на данные экзотов или транскриптомов (в реальном исследовании ландшафтной геномики ещё не применялся). Для количественной оценки уровня пространственной зависимости в распределении генотипов были разработаны высокопроизводительные вычисли-

тельные методы, такие как Samβada, позволяющий измерять глобальную пространственную автокорреляцию по всему набору данных (Balkenhol et al. 2017; Duruz et al. 2019).

Предметом изучения ландшафтной геномики служат как отдельные виды, так и группы видов, или сообщества наземных и водных экосистем. Ландшафтная геномика, нейтральная и адаптивная, применяется для решения большого разнообразия проблем, и сфера её применения постоянно расширяется. Данные ландшафтной геномики полезны для сохранения и восстановления биоразнообразия, управления ценными видами, понимания системы паразит–хозяин, прогнозирования последствий глобального изменения климата и урбанизации и т. п. (Глазко, Глазко, 2010; Grummer et al. 2019; Fernandez, Dodd, 2018; Luikart et al. 2018; Balkenhol et al. 2017; Тетушкин, 2013; Waldvogel et al., 2020).

### 3.6. Метагеномика

Мир микробов был открыт три столетия назад Антони ван Левенгуком с помощью изобретённого им микроскопа, а первый метод культивирования микроорганизмов разработан Робертом Кохом в конце XIX века. С тех пор были идентифицированы и исследованы (классификация, эволюция, применение) многие виды бактерий. В XX веке использование последовательностей 16S рибосомальной РНК (рРНК) привело к выделению архей в отдельную группу прокариот, а филогенетический маркер на основе 16S рРНК был предложен в качестве важнейшего инструмента в таксономических исследованиях микроорганизмов (Nam et al., 2023).

Метагеномика основана на идее, что набор генов из образцов окружающей среды можно анализировать подобно тому, как изучаются целые геномы. Термин «метагеномика» был введён Джо Хендельсманом с коллегами в 1998 г., а в 2005 году Кевин Чен и Лиор Пахтер определили его, как *«применение современных методов геномики к изучению сообществ микробных организмов непосредственно в их естественной среде, минуя необходимость выделения и лабораторного культивирования отдельных видов»* (Martellacci et al., 2019). В настоящее время под метагеномикой понимается изучение генетического материала, полученного непосредственно из образцов окружающей среды (вода, почва и др.) или клинических проб методом секвенирования ДНК; поэтому метагеномику также называют экогеномикой, геномикой сообществ или микробиомикой (Zhang et al., 2021; Nam et al., 2023).

Существуют разные классификации метагеномики. В зависимости от типа используемых данных, выделяют целевую метагеномику, основанную на секвенировании ампликонов с использованием генетических маркеров, а также метагеномику дробовика, которая основана на принципе случайного секвенирования всей ДНК, находящейся в образце, и обеспечивает полногеномную сборку (Nam et al., 2023). Целевая метагеномика позволяет идентифицировать все организмы на видовом уровне. В метагеномных исследованиях бактерий и архей для секвенирования обычно исполь-

зуют гены 16S рРНК, а эукариот – 18S рРНК, 26S рРНК и межгенные транскрибируемые спейсеры (ITS). Гены 16S рРНК помимо консервативных участков содержат девять гипервариабельных областей, поэтому различные регионы могут использоваться для изучения бактериальных сообществ на уровне семейства, рода и вида. Данные метагеномики дробовика включают все последовательности ДНК в образцах (хотя в конечном итоге доминируют геномы наиболее представленного вида/видов). Это даёт информацию как о таксономическом, включая новые виды, так и функциональном составе микробного сообщества. При достаточном выходе ДНК из образца, секвенирование может быть выполнено сразу, без использования библиотеки клонов (Luikart et al., 2018; Garlapati et al., 2019; Awasthi et al., 2020; Fadrosch et al., 2014).

Метагеномный анализ состоит из четырех основных этапов: отбора проб (должен по возможности исключать/ограничивать любые примеси) и выделения ДНК (с выбором оптимальных протоколов экстракции, которые зависят от типа используемых образцов), секвенирования (обычно с помощью методов NGS; в перспективе – технологии с длинными прочтениями), анализа и визуализации (с использованием специализированных программ) (Nam et al., 2023). Микробный геном на уровне вида, созданный на основе метагеномных данных сообщества из образца микробиома, называют метагеномно-сборным геномом (MAG). Сборку метагенома затрудняет наличие нескольких видов, неравномерное видовое обилие, общие для разных видов геномные области и внутривидовая изменчивость. Поэтому для получения MAG применяется биннинг – процесс группировки/сортировки собранных последовательностей ДНК в группы/бины, которые представляют отдельные геномы близкородственных организмов в микробном сообществе. Для биннинга разработаны различные алгоритмы, например, на основе композиционного состава, частоты использования кодонов или анализа глубины прочтений. Считается, что использование секвенирования HiFi (высокоточные длинные прочтения), улучшенных методов сборки метагенома и дополнительных стратегий биннинга обеспечит быструю и высокоэффективную каталогизацию микробных геномов в сложных микробиомах (Nam et al., 2023; Portik et al., 2024).

Поскольку данные метагеномного секвенирования содержат большое количество фоновой информации (в основном из-за загрязнения эукариотической ДНК), часто требуются специальные методы биоинформатики для сбора, обработки и извлечения полезной (целевой) информации и уменьшения «шума» (Wooley et al., 2010). В таблице 3.3 приводится список современных инструментов биоинформатики и платформ для метагеномного анализа. Их разнообразие и доступность привели к развитию конвейеров метагеномного анализа. Например, созданы эффективные конвейеры для контроля качества при окончательной визуализации метагеномных данных и для изучения взаимосвязи между микробиомами и патофизиологией заболевания. На веб-платформах конвейер анализа обычно состоит из контроля качества прочтений, сборки, прогнозирования функций, таксономического прогнозирования и визуализации результатов (Nam et al., 2023).

**Таблица 3.3.** Инструменты и платформы для метагеномного анализа, обновлённые в 2019–2023 гг. (из: Nam et al., 2023).

Инструменты/Платформы	Функции
MetaPhlAn 4	Комплексное метагеномное таксономическое профилирование
HUMAnN 3	Эффективное и точное функциональное профилирование
StrainPhlAn 3/PanPHLAn 3	Профилирование штаммов на основе нуклеотидных и генных вариантов
PhyloPhlAn 3	Филогенетическое размещение и возможное таксономическое присвоение
QIIME2	Анализ и визуализация данных метагеномики ампликонов
DIAMOND + MEGAN	Таксономический и функциональный анализы данных коротких и длинных метагеномных последовательностей
MEGAN6 и MeganServer	Таксономический и функциональный анализы и визуализация метагеномных данных
Kraken 2	Быстрая таксономическая идентификация

Как следует из определения, преимущество метагеномики состоит в том, что она позволяет идентифицировать не только культивируемые, но и некультивируемые микроорганизмы, существенно влияющие на видовое разнообразие микробных сообществ (Hugenholtz et al., 1998), а знание их разнообразия способствуют выявлению механизмов функционирования и определению метаболических взаимосвязей (Fremin, Bhatt, 2020). Это важно, поскольку несмотря на понимание решающей роли микробных сообществ в поддержании здоровья человека, животных и растений, а также окружающей среды, их состав и механизмы функционирования остаются недостаточно изученными. Метагеномика обладает значительным потенциалом для изучения микробных сообществ, населяющих почву, а также взаимодействия между растениями и микробами (Niu et al., 2017; Fadji et al., 2020), для реконструкции рациона травоядных животных (Chua et al., 2021), поддержания здоровья человека (Wallen et al., 2022; Wu et al., 2022), и который вносит вклад в ландшафтную геномику (Luikart et al., 2018). Помимо микробной ДНК, метагеномика изучает ДНК окружающей среды (eDNA), что помогает в мониторинге биоразнообразия, в управлении и охране окружающей среды, обнаружении антропогенного загрязнения, в оценке структуры прибрежных отложений, геопространственном отслеживании и реконструкции распространения патогенов, в исследовании экосистем и т. д. (Coward et al., 2018; Luikart et al., 2018; Garlapati et al., 2019) (рис. 3.19).

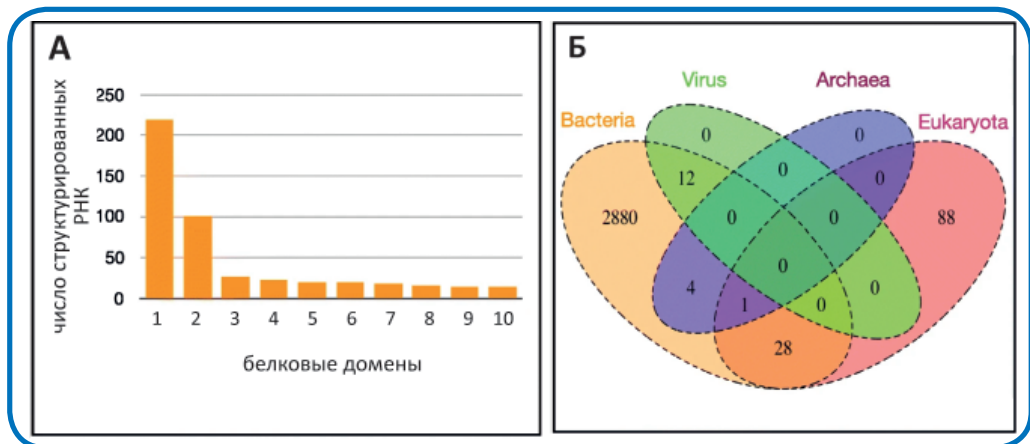
Предмет изучения метагеномики, сообщество микроорганизмов, населяющих конкретную среду обитания, или совокупность его генов, называется микробиомом.



Рисунок 3.19. Области применения метагеномики

Этот термин был впервые использован в 1952 г. при анализе загрязнения сточными водами (Mohr, 1952). С помощью метагеномного подхода можно выяснить, существует ли основной микробиом человека, и какие изменения в микробиоме могут быть связаны со здоровьем (Qin et al., 2010). Результаты метагеномного анализа показывают, что микробиом кишечника связан с различными заболеваниями, такими как неалкогольная жировая болезнь печени, аутоиммунные заболевания и злокачественные опухоли (Zhang et al., 2021). С целью лучшего понимания микробиома и его значения для здоровья человека в 2007 г. был инициирован проект «Микробиом человека» (HMP). Оказалось, что микробные гены, связанные с выживанием человека, более многочисленны, чем гены самого человека, причём бактериальные белок-кодирующие гены встречаются в 360 раз чаще, чем аналогичные им гены человека. В рамках выполнения данного проекта были, в частности, выявлены уникальные приспособления сегментированных нитчатых бактерий, которыми они пользуются в качестве кишечных симбионтов. Высокая значимость результата состоит в том, что эти бактерии стимулируют клетки Т-хелперов, играющих ключевую роль в аутоиммунных заболеваниях (Sczesnak et al., 2011).

В 2010 г. был запущен проект «Микробиом Земли» (EMP), направленный на обработку примерно 200000 проб в различных биомах и на создание полной базы данных о микробах Земли для характеристики окружающей среды и экосистем по микробному составу и взаимодействию. Полученная информация предназначена для хранения в специальных централизованных базах данных, таких как «Генный Атлас» (GA) и «Геномы, собранные микробиомом Земли» (EMAG), а полученные и аннотированные с помощью автоматизированного конвейера геномы будут храниться в открытых репозиториях. Предусматривается интерактивная визуализация данных



**Рисунок 3.20.** Предсказание и характеристика структурированных РНК-кандидатов (из: Fremin, Bhatt, 2021).

А – Гистограмма для десяти белковых доменов, чаще всего встречающихся вблизи структурированных РНК-кандидатов: 1 – cbiO, 2 – HTH\_XRE, 3 – AAA, 4 – Sigma70\_r4, 5 – Resolvase, 6 – fabG, 7 – PBP1\_LacI-like, 8 – HisKA, 9 – nitroreductase, 10 – LysM.

Б – Диаграмма Венна таксономического распределения структурированных РНК-кандидатов на уровне доменов.

через «Портал визуализации микробиома Земли» (EM-VIP) для установления взаимосвязи между микробным составом, параметрами окружающей среды и геномной функцией, а также предполагается реконструкция метаболических профилей в рамках «Метаболической реконструкции микробиома Земли» (EMMR).

Важной частью исследований микробного генома является идентификация структурированных РНК. У микробов такие РНК играют различные биорегуляторные роли. На сегодняшний день сотни структурированных РНК-кандидатов были предсказаны по геномным последовательностям с использованием биоинформационных подходов. Однако большая часть метагеномных данных микробиома человека остаётся неисследованной в отношении мотивов структурированной РНК, в основном из-за вычислительных ограничений. Крупномасштабный подход сравнительной геномики с тщательно разработанным автоматизированным анализом (требующим большой вычислительной мощности) идентифицировал более 3000 консервативных структурированных РНК-кандидатов в межгенных областях, а также около 2000 дополнительных структурированных РНК-кандидата в микробиоме человека, часть которых получила экспериментальное подтверждение (рис. 3.20). Чтобы надёжно классифицировать их на рибозимы, рибопереключателы, тРНК, антитоксины и т. д., необходимо сотрудничество многих лабораторий. В целом, этот ресурс из тысяч структурированных РНК позволит в дальнейшем функционально охарактеризовать и подразделить широкий спектр новых структурированных РНК из ранее неизученных областей микробного генома (Fremin, Bhatt, 2021).

В последние годы исследования микробиома быстро развиваются, и были разработаны методы его модификации *in situ*. Для этих целей используются системы CRISPR–Cas9, которые способны узнавать участки генома патогенных микробов, ответственные за устойчивость к антибиотикам, и вызывать их деградацию. Благодаря высокой специфичности, эти системы безопасны для нормальной микрофлоры и могут применяться, например, для профилактики дисбактериоза (Bikard et al., 2014).

### 3.7. Палеогеномика

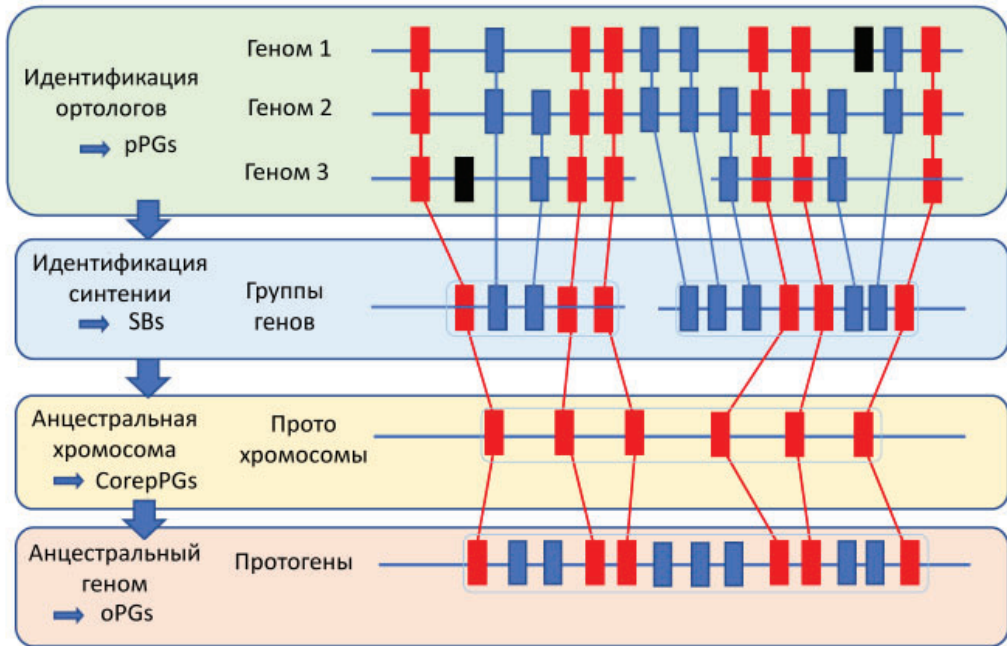
Исследования древней ДНК начались в середине 1980-х гг. одновременно в двух лабораториях, Сванте Паабо (Svante Pääbo) в Швеции и Аллана Вильсона (Allan Wilson) в США. Первая последовательность древней ДНК была выделена из шкуры вымершего вида семейства зебр *Equus quagga quagga* Boddaert, 1785, хранившейся в музее (Higuchi et al., 1984), а первая последовательность древней ДНК человека получена из тканей египетской мумии возрастом более 4400 л. (Pääbo, 1985). Исследование молекул ДНК, сохранившихся с течением времени, открыло доступ к доисторической генетической информации и нашло широкое применение, в том числе в таксономии, палеонтологических реконструкциях, популяционной генетике, гуманитарных аспектах и т. д. Однако низкая сохранность и посмертная модификация ДНК, наличие в образцах ингибиторов сильно усложняли работу с древней ДНК; поэтому первые исследования проводились в основном на последовательностях митохондриального генома (Челомина, 2006; 2022).

Достижения в области извлечения ДНК и новые технологии секвенирования значительно расширили возможности изучения древних геномов. Недавние усовершенствования методов обогащения (использование специальных «приманок» для захвата фрагментов древней ДНК) позволяют анализировать образцы, которые ранее считались непригодными. Древняя ДНК характеризуется такими повреждениями, как депуринизация (приводит к потере азотистого основания) и дезаминирование (в результате цитозин превращается в тимин), перекрёстные связи и высокая фрагментация, которые накапливаются с течением времени. Стратегии обогащения используются либо во время создания геномной библиотеки (предпочтительно для включения повреждённых фрагментов древней ДНК), либо после создания библиотеки (для разделения эндогенных и экзогенных фракций путём отжига с определённым набором зондов в растворе или на микрочипах, с предполагаемыми мишенями захвата размером примерно от 16 т. п. н. до 3 Gb) (Orlando et al., 2015). Другие подходы повышения эффективности извлечения древней ДНК включают целенаправленное расщепление ДНК микробов из окружающей среды, обогащение повреждённой матрицы (выборочное нацеливание на повреждённую ДНК во время подготовки библиотеки одноцепочечной ДНК), обогащение мишени в растворе без удлинения (на основе гибридизации мишень–зонд), твёрдофазное обогащение мишеней ДНК (прямое применение микрочипов) и обогащение всего генома (захват всего генома в растворе, WISC). В WISC специально разработанные биотинилированные РНК-

зонды отжигают с одноцепочечной ДНК (полученной через тепловую денатурацию), а несгибридовавшуюся ДНК элюируют. Связанная ДНК (обогащённая библиотечная фракция) высвобождается после обработки РНКазой, её амплифицируют и затем секвенируют (Orlando et al., 2015). Современные технологии позволяют успешно извлекать и анализировать геномы из окаменелостей, датированных сотнями тысяч лет и более. В настоящее время получены геномные данные для разных видов, от недавно вымерших, таких как тасманийский тигр (исчезнувший около 100 л. н.), до живших десятки или сотни тысяч лет назад, например, шерстистого носорога, пещерного медведя и древней лошади (палеобразцы датируются примерно 18000, 270000 и 560000 л. н. соответственно) (Mitchel, Rawlence, 2021).

Однако низкое содержание эндогенной ДНК, короткая длина фрагментов и ошибки, связанные с посмертными модификациями, требуют новых специализированных методов для успешной реконструкции древних геномов. Поскольку вычислительные инструменты имеют решающее значение при реконструкциях древних геномов, для их эффективной сборки используется специально разработанное для древней ДНК программное обеспечение, включая PALEOMIX (программный конвейер для быстрой обработки данных высокопроизводительного секвенирования) (Schubert et al., 2014) и EAGER (эффективная реконструкция древнего генома – система программирования на примерах для среды HyperCard) (Peltzer et al., 2016). Они упрощают анализ набора геномных данных, помогают предварительно обрабатывать, картировать, аутентифицировать, оценивать качество образцов древней ДНК и обеспечивают возможность для обнаружения, фильтрации и анализа вариантов (Peltzer et al., 2016; Lan, Lindqvist, 2018; 2019).

Использование геномных данных позволяет изучать эволюционную историю современных и вымерших видов, значительно расширяя наше понимание молекулярных основ видообразования, адаптации и восприимчивости к болезням. Эта область исследований опирается на два основных подхода, макро- и микроэволюционный (или синхронный и аллохронический соответственно). Первый подход сравнивает современные геномы с реконструкциями геномов предков в рамках глубоких временных масштабов (миллионы лет). Второй основан на прямом секвенировании генетического материала, извлечённого из антропологического и палеонтологического материала, сохранившегося в течение последних десятков и сотен тысяч лет. Геном предков – это «медианный» (или «промежуточный») геном, состоящий из эталонного порядка генов, общего для всех современных видов. Такие минимальные предковые геномы реконструируются в четыре этапа (рис. 3.21): (1) идентификация консервативных или дублированных генов; (2) выявление синтении в группах консервативных соседних генов; (3) реконструкция смежных наследственных регионов, содержащих гены, которые сохраняются у всех исследованных видов; и (4) реконструкция предкового генома с протохромосомами и переупорядоченными протогенами (Pont et al., 2019). В частности, такие исследования обнаружили высокую стабильность геномов плацентарных млекопитающих, позволили предположить, что изменения окру-

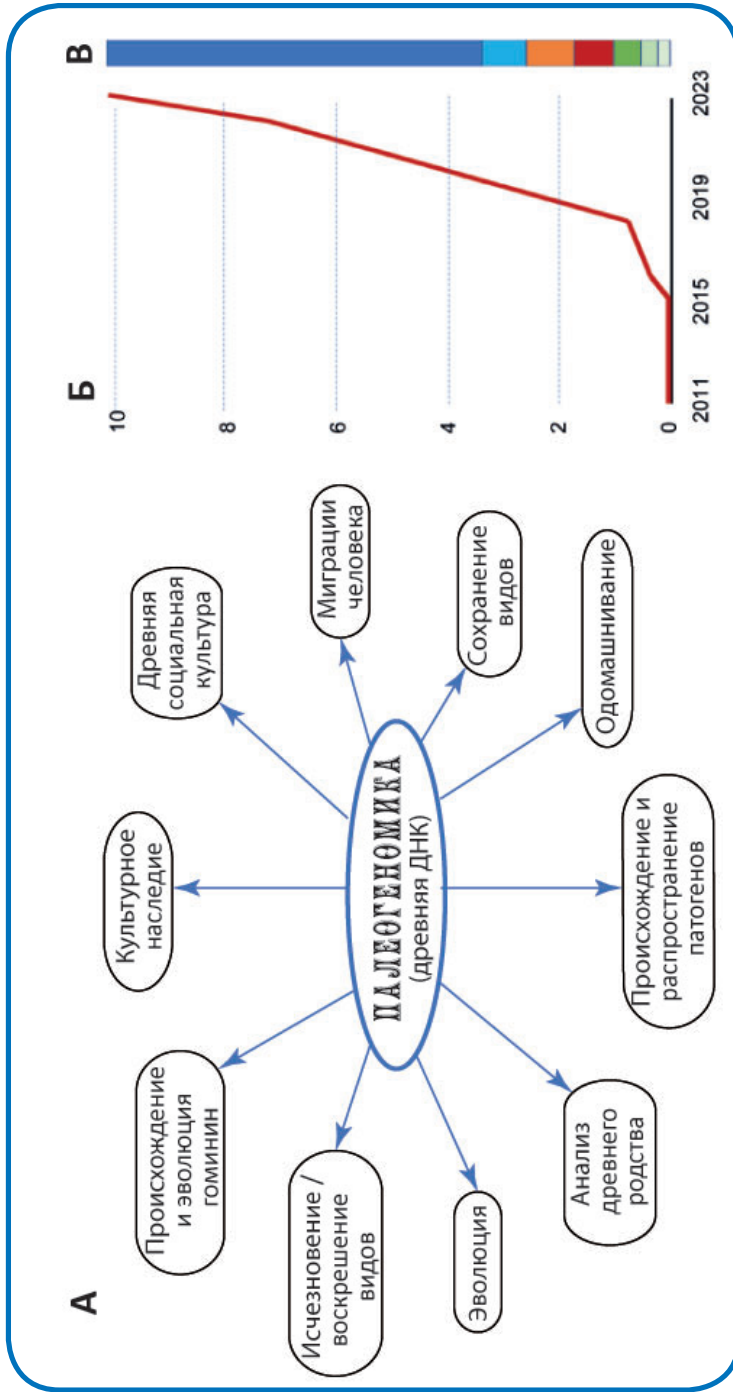


**Рисунок 3.21.** Методологические принципы реконструкции анцестрального генома (по: Pont et al. 2019).

pPGs – предполагаемые протогены, SBs – блоки синтении, Core-pPGs – коровые протогены, oPGs – упорядоченные протогены.

жающей среды могли влиять на пластичность генома, что привело к появлению ключевых адаптивных биологических функций (например, обонятельные рецепторы), показали поразительное сохранение синтении на протяжении примерно 320 млн л. эволюции позвоночных и сделали возможным реконструкцию анцестрального генома млекопитающих (Pont et al., 2019; Damas et al., 2022).

Палеогеномика (рис. 3.22), также известная как полногеномный анализ древней ДНК, оказала глубокое влияние на развитие современной науки, помогая в решении широкого спектра эволюционных, палеоэкологических и археологических вопросов, часто меняя прежние представления и открывая новые перспективы. Древняя ДНК позволяет отслеживать изменения генетического разнообразия во времени и документировать ход эволюционных событий. Благодаря палеогеномике стало возможным исследовать, «как отбор формирует геномы в течение продолжительных периодов времени, как события миграции и примеси порождают современные паттерны генетической изменчивости, как виды приобретают генетические варианты, обеспечивающие преимущества приспособленности в условиях новых ограничений отбора, и как популяции растений и животных реагируют на изменение окружающей среды, болезни или деятельность человека» (Brunson, Reich, 2019, с. 320). Таким образом, древняя ДНК может быть использована для тестирования связей между событиями окружающей среды и эволюционными изменениями в популяциях, уточ-



**Рисунок 3.22.** Палеогеномика.

А – направления исследований (по: Chen, Nedoluzhko, 2023); Б – рост данных (в тысячах) и В – географическое распределение исследований древнего генома человека (по: Callaway, 2023): синий – Европа и Россия (67%), голубой – Восточная Азия (8%), красный – Америка (7%), темно-зелёный – Южная и Центральная Азия (5%), зелёный – Африка (3%), светло-зелёный – Океания (2%).

нения эволюционных отношений между видами и калибровки молекулярных часов, выявления криптических взаимоотношений между исчезнувшими и современными популяциями, а также для изучения эволюции генома, включая эволюцию патогенности (Shapiro, Hofreiter, 2014). Геномные данные древних людей и архаичных гоминид дают новое представление об адаптации человека к изменяющейся окружающей среде, переходу к производящему хозяйствованию (земледелие и животноводство) и патогенам, об эволюции человека и её влиянии на поведение человека (Marciniak, Perry, 2017). Палеогеномика пересмотрела вымершее разнообразие, обнаружила скрытые миграции фауны, выявила прошлые события гибридизации и т. д. (Peltzer et al., 2016). Более того, новые технологии геномного анализа дают надежду на возрождение вымерших видов (Sandler, 2014). Примечательно, что хотя большая часть исследований сосредоточена на истории популяций человека, максимальное значение среднего количества цитирований имеют работы по применению древней ДНК в эволюционных исследованиях, а наиболее бурно развивается направление палеогеномики, связанное с реконструкцией экологической истории и генетикой сохранения видов (Pinakhina, Chekunova, 2020).

Развитие методов экстракции древней ДНК, технологий секвенирования и сборки геномов неизменно приведут к дальнейшему развитию палеогеномики и использованию её достижений во многих дисциплинах, значительно их обогащая и открывая новые горизонты. Можно ожидать открытия новых для науки видов. Например, вымершая линия гоминид, денисовцы, были обнаружены только на основе секвенирования генома (ранее найдены многочисленные скелетные останки неандертальца, и только два фрагмента костей, фаланги пальца из Денисовой пещеры на Алтае и нижней челюсти с Тибетского плато в Китае, принадлежащие денисовцам) (Meyer et al., 2012). Согласно данным палеогеномики, риск генетических заболеваний изменился за последние 10000 л., причём частота многих аллельных вариантов, связанных с риском заболевания, увеличилась в новейшей истории. Несмотря на то, что большинство этих локусов эволюционно нейтральны, некоторые находятся под сильным давлением отбора (например, локусы, предотвращающие развитие астмы и некоторые варианты генов гистосовместимости, МНС). Это направление является началом исследований эволюционной медицины, необходимых для лучшего понимания рисков наследственных заболеваний (Simonti, Laechance, 2021).

Анализ древней ДНК из отложений (sedaDNA) является относительно новым методом, который может предоставить информацию об окружающей среде и местах обитания в прошлом (Lan, Lindqvist, 2019). Основными подходами к изучению sedaDNA являются метабаркодирование (для микроорганизмов и растений), метагеномика (для экосистем) и гибридизационный захват (для доисторических людей) (Jia et al., 2022). Анализ древних геномов макрофоссилий растений может найти применение в реконструкции послеледниковых сдвигов видовых ареалов, отслеживании генетического разнообразия во времени и тестировании адаптивного потенциала растений к изменению климата (Schwörer et al., 2022). Древнюю ДНК из отложений можно использовать для интерпретации тенденций биоразнообразия, освещения

прошлой динамики пищевых сетей, прогнозирования долгосрочных экологических изменений в водных экосистемах, а также для прослеживания кратко- и долгосрочных трендов антропогенного воздействия (Nguyen, 2022).

Уникальные «окна», позволяющие наблюдать за генетическими и фенотипическими изменениями в процессе эволюции популяций, представляют «банки» покоящихся (эфиппиальных) яиц, которые способны выживать длительное время в самых неблагоприятных условиях. Например, дафния в ходе жизненного цикла производит спящие эмбрионы, которые приостанавливают своё развитие, сохраняя генетические последствия изменений окружающей среды и экологического стресса. Палеогеномные исследования такого материала пока ограничены, поскольку покоящиеся яйца не содержат достаточного количества ДНК для секвенирования всего генома. Однако уже разработан эффективный протокол для создания целых геномов одиночных эфиппиальных яиц и отдельных дафний, который может быть использован и для других мелких видов (Lack et al., 2018).

Палеоэпигеномные исследования – зарождающееся направление палеогеномики. Являясь основной движущей силой фундаментальных метаболических путей, эпигенетические изменения связаны с видообразованием, эмбриогенезом, развитием заболеваний и др. Поскольку эпигеном быстро реагирует на изменения окружающей среды, изучение характера метилирования древних геномов может прояснить, как виды адаптировались в изменяющемся мире, включая эпидемиологическую обстановку. Связь эпигенетических изменений с регуляцией транскрипции позволяет оценивать экспрессию генов вымерших видов, в то время как анализ дифференциально метилированных регионов (DMRs) даёт возможность успешно реконструировать их морфологические особенности. Например, использование карт метилирования позволило идентифицировать 56 морфологических особенностей, которые отличали денисовцев и неандертальцев от современных людей. Среди специфичных для неандертальцев DMRs было несколько гиперметилированных областей в кластере, содержащем гены, важные для развития конечностей у позвоночных (Liu et al., 2020; Niiranen et al., 2022).

Воскрешение (воссоздание) вымерших видов – новая область геномики, вызывающая большой научный и общественный интерес. В частности, новая технология редактирования генома системой CRISPR–Cas9 может быть использована для создания гибридного генома с чертами мамонта, запрограммированными в геноме эмбриона азиатского слона, что делает воскрешение мамонта реальностью. Однако такой эксперимент может столкнуться с определённой этической дилеммой (Sandler, 2014).

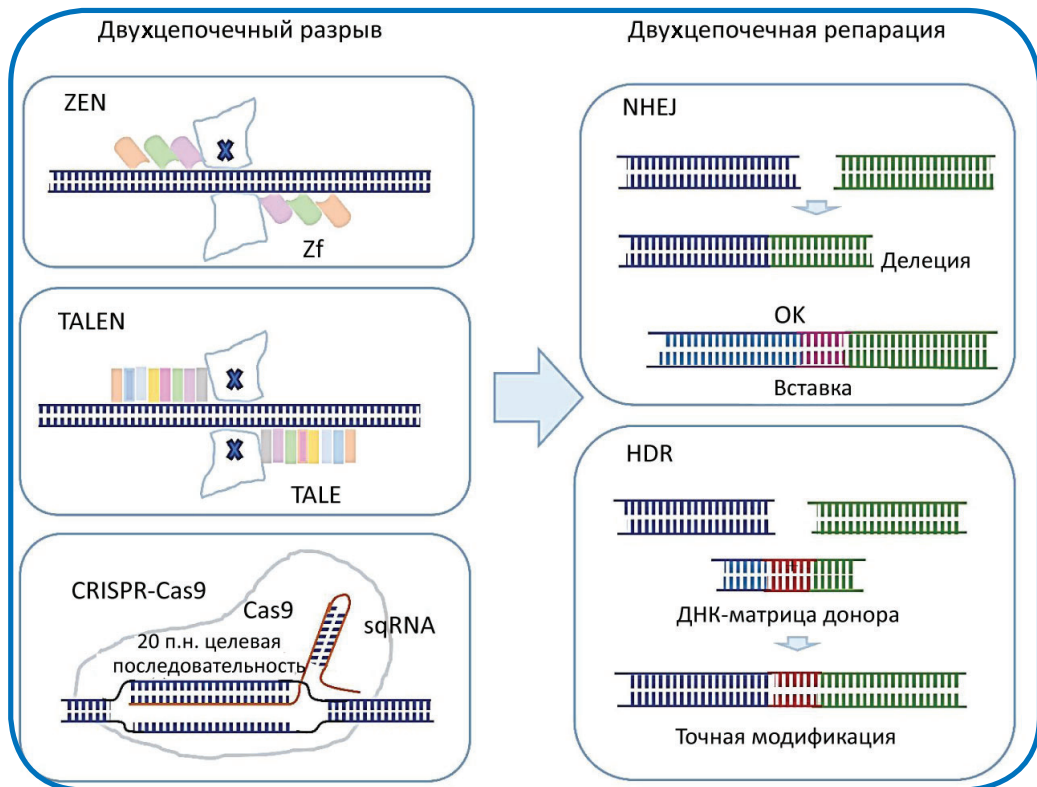
### 3.8. Синтетическая геномика

Определённая Дэниелем Шиндлером и его коллегами как «*новая область, в которой биология разрабатывается на уровне генома, и это является пересечением синтетической биологии и системной биологии*», синтетическая геномика (также известная как инженерная биология) направлена на создание химически синтези-

рованных полных геномов или их больших частей. Современные технологии позволяют вносить множественные инженерные изменения в генетическом материале организмов, что в итоге приводит к более глубокому пониманию биологических систем (Konig et al., 2013; Coradini et al., 2020; Thao et al., 2020; Yoneji et al., 2021). В то время, как только полностью синтезированные геномы (т.е. все блоки конечной молекулы ДНК создаются путём химического синтеза) являются синтетическими, не полностью синтезированные геномы считаются «инженерными» или «модифицированными» геномами (Schindler et al., 2018). Синтетическая геномика полезна для изучения многочисленных фундаментальных вопросов эволюции и генетики, таких как функциональные и фенотипические последствия изменений в структуре генома, функциональное и фенотипическое значение видовых пангеномов, экологическая интеграция организмов, оценка минимального генома и допустимые уровни редукции числа кодонов (Wang et al., 2018a; Schindler et al., 2018; Coradini et al., 2020). Как отмечают авторы, «нужно написать синтетические последовательности ДНК, чтобы понять грамматику жизни» (Schindler et al., 2018).

После синтеза первого искусственного гена (ген аланиновой тРНК дрожжей) Хар Гобинд Кораной в 1965 г. новые технологические разработки позволили синтезировать >10 т. п. н. за несколько недель, что сделало синтетические хромосомы доступными для функционального анализа *in vivo* (Schindler et al., 2018). Примером тому служит синтез первого РНК-генома вируса (7,5 т.п.н.) в 2002 г. (Cello et al., 2002), первого ДНК-генома (*Mycoplasma genitalium*, 590 т.п.н.) в 2008 г. (Gibson et al., 2008) и реконструкция генома SARS-CoV-2 в 2020 г. (Thao et al., 2020). Высшим достижением считается синтез бактериального генома – хромосомы *M. mycoides* (с кодовым названием JCVI-syn1.0) сотрудниками Института Крейга Вентера (JCVI). После переноса этой хромосомы в клетку родственного вида *M. capricolium* и последующего деления образовалась клетка, полностью управляемая искусственным геномом. Размер генома составил немногим более 1 млн п. н.; в процессе синтеза несколько генов были удалены, а в процессе сборки образовалось 19 мутаций (Gibson et al., 2010). Хотя полного синтетического эукариотического генома до настоящего времени не существует, 16 хромосом отдельных штаммов дрожжей синтезированы в рамках Консорциума Sc2.0 (проект синтеза генома *Saccharomyces cerevisiae* 2.0), и был получен синтетический митохондриальный геном мыши (первый и пока единственный) (Gibson et al., 2010; Schindler et al., 2018).

«Написание» генома (или индивидуальное построение генома) достигается за счёт синтеза ДНК и редактирования генома (Wang et al., 2018a). Синтез генома *de novo* обычно начинается с синтетических олигонуклеотидов одноцепочечной ДНК, за которыми следует ПЦР-амплификация и поэтапная сборка ДНК (в масштабах kb, Mb, хромосомы и генома). Когда нет необходимости в синтезированном *de novo* геноме, применяется редактирование существующего генома (рис. 3.23). Как правило, редактирование целевого генома осуществляется в два этапа. Во-первых, двухцепочечная ДНК разрезается одной из эндонуклеаз, такой как нуклеазы с цинковы-



**Рисунок 3.23.** Платформы редактирования генома и механизмы восстановления DSB – двухцепочечных разрывов эндогенной ДНК (по: Li et al., 2020a).

Нуклеазы редактирования генома (ZFN, TALEN и CRISPR/Cas9) индуцируют DSB в целевых сайтах. DSB можно восстановить с помощью негомологичного соединения концов (NHEJ) или, при наличии донорской матрицы, с помощью гомологичной репарации (HDR)

ми пальцами (ZFN) и эффекторные нуклеазы, подобные активаторам транскрипции (TALEN); обе системы представляют собой сконструированные белки, нацеленные на определённые последовательности ДНК через взаимодействия ДНК–белок. Кроме того, используется система коротких палиндромных повторов CRISPR/CRISPR-ассоциированного белка (Cas), основанная на простом дизайне и синтезе молекулы РНК, комплементарной сайтам–мишеням ДНК. Затем двухцепочечные разрывы ДНК восстанавливаются посредством негомологичного соединения концов (не требует донорской ДНК; InDels часто вводят в целевые сайты) или гомологичной прямой репарации (для облегчения точной модификации требуется донорская ДНК). Для сборки большого искусственного генома существующий геном можно отредактировать, а затем использовать для сборки генома *de novo* (Wang et al., 2018a).

Существует несколько конкретных классов генетических манипуляций, обеспечиваемых синтетической геномикой: реструктуризация (изменение структуры ге-

номных элементов или перегруппировка генетического материала внутри генома), перекодирование (замена аминокислотных и стоп-кодонов по всему геному, включая искусственные кодоны), минимизация (до минимального набора генов и физиологических функций, необходимых для жизни клетки), химеризм (объединение генетического материала разных штаммов или видов в единый геном), реинжиниринг органелл (митохондрии, хлоропласты) и воссоздание генома (Coradini et al., 2020). С помощью этих подходов были созданы полностью синтетические геномы *Escherichia coli* с семью элиминированными кодонами (обеспечивающими достаточную устойчивость к фагам), а также синтезированы хромосомы *S. cerevisiae* с удалением большинства интронов, транспозонов и повторяющихся элементов, с перемещением всех генов тРНК в независимую хромосому и с фланкированием всех несущественных генов сайтами loxPsym, позволяющими индуцировать крупномасштабные геномные перестройки (Schindler et al., 2018; Coradini et al., 2020; Schindler, 2020).

В 2016 году было официально объявлено о создании проекта «Genome». Этот проект, основанный на синтезе гена/генома, редактировании и других технологиях, направлен на понимание, проектирование и тестирование живых систем, а его глобальная цель – «понять план жизни, предоставленный Проектом генома человека» (Boeke et al., 2016; Schindler et al., 2018). Однако несмотря на свой огромный потенциал, синтетическая геномика поднимает важные этические вопросы, например о том, насколько генетические модификации приемлемы для существующих форм жизни (Coradini et al., 2020).

Манипуляции с геномами открывают новые перспективы в изучении эволюции и функционирования генома эукариот. Например, систематическая гуманизация генов дрожжей выявила консервативные функции и генетическую модульность генома (Kachroo et al., 2015, 2022). В этих исследованиях, чтобы определить, сохраняют ли гены свои наследственные функции на протяжении миллиарда лет эволюции, и выявить принципы глубокой эволюционной дивергенции, 414 основных генов дрожжей были заменены их человеческими ортологами. Оказалось, что почти половина (47%) генов дрожжей может быть успешно гуманизирована. Причём сходство последовательностей и экспрессия лишь частично предсказывали заменяемость, которая существенно зависела от генных модулей; т.е. гены, участвующие в одном и том же процессе, как правило, были одинаково заменяемы (например, биосинтез стиролов) или нет (например, инициация репликации ДНК). Моделирование также подтвердило, что выбор конкретной функции может поддерживать взаимозаменяемость, несмотря на значительную дивергенцию последовательностей. Ключевые наследственные функции многих важных генов, таким образом, сохраняются специфическим образом, устойчивым к дрейфу в последовательностях, сплайсинге и интерфейсах белков (Kachroo et al., 2015).

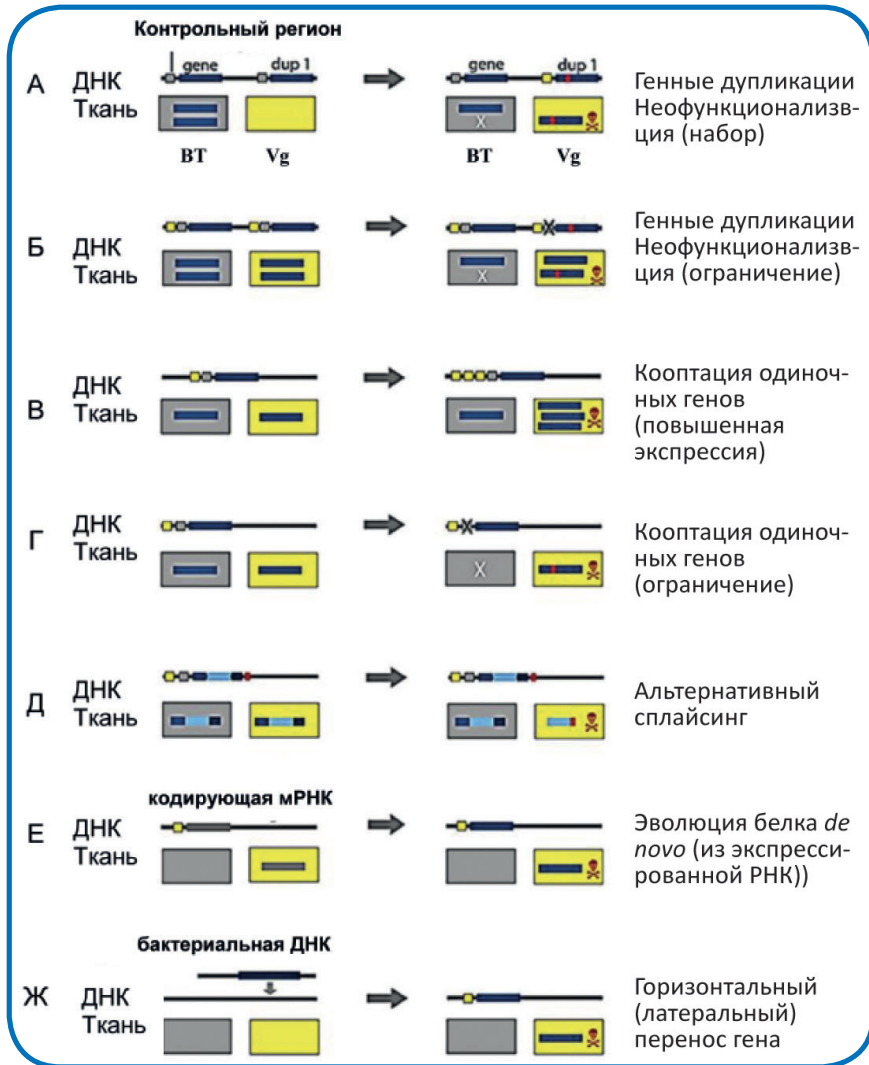
Большой интерес к созданию белков с заданными свойствами для применения в медицинских целях стимулировал развитие нового направления синтетической геномики, которое основано на использовании неприродных нуклеотидов (UB) и не-

канонических аминокислот (ncAA), и позволяющее создавать полусинтетические организмы (SSO) с дополнительными смысловыми кодонами. Синтезированы комплементарные друг другу нуклеотиды, такие как Z–P и Ds–Px, способные заменить классические пары A–T и G–C. Разработка UB пар (UBP) увеличила генетическую информационную ёмкость нуклеиновых кислот, придав им новые свойства и функциональные возможности, что открывает перспективу получения белков с нестандартными аминокислотами в определённых локусах (Fischer et al., 2020; Ostrov et al., 2020; Kimoto et al., 2022, 2023; Sun et al., 2022a; Romesberg, 2023). В 2020 г. Флойд Ромесберг и его коллеги опубликовали данные о девяти неприродных кодонах, способных производить неприродный белок с почти полным включением кодируемой ncAA, и показали, что не менее трех кодонов могут одновременно декодироваться в SSO, давая первый 67–кодоновый организм (Fischer et al., 2020). Эксперименты с UBП и ncAA пока проводятся на прокариотических системах, и геном кишечной палочки *Escherichia coli* уже превращён в первый SSO, а в перспективе – работы с эукариотическими системами *in vitro* и *in vivo*.

В целом, большинство проектов синтетической геномики используют стратегию разработки «сверху вниз», которая разделена на четыре категории: (1) разграничение спроектированной последовательности с исходной версией, (2) реструктуризация генома, позволяющая разделение или перемещение генов в функциональные модули, (3) уменьшение избыточности генома и его минимизация, и (4) введение конструктивных особенностей, обеспечивающих расширенные или новые функциональные возможности (Fu, Shen, 2024). Синтетическая геномика обладает потенциалом для создания инноваций, способных преобразовать производство продуктов питания, получать новые лекарства, производить альтернативные и экологически чистые материалы, а также разрабатывать новые методы защиты экосистем и решения экологических проблем (Fu, Shen, 2024). В 2023 г. свои дорожные карты и стратегии, которые позволят расширить возможности в этой новой научной области, опубликовали Великобритания (с планом инвестиций в размере двух миллиардов фунтов стерлингов), США, Китай и Австралия.

### 3.9. Веномика

В последнее время внимание исследователей было привлечено к геномам ядовитых видов, которые являются важными биоресурсами для создания новых терапевтических средств и биоинсектицидов. О производстве яда известно, как минимум, в 101 независимой линии животных (включая хорошо известных змей, пауков и скорпионов), который они используют для защиты, хищничества, внутривидовой конкуренции и т.п. (Farhat et al., 2023). Сравнительная геномика ядовитых видов раскрывает важные ключи к разгадке механизмов эволюции генов токсина и молекулярной основы морфологической адаптации «ядовитого аппарата». Геномные процессы, управляющие эволюцией белков яда, включают дубликации генов (с последующей неофункционализацией) и кооптации генов (повышающие или ограничивающие экс-



**Рисунок 3.24.** Геномные процессы, управляющие эволюцией белков яда (Drukewitz, Reumont, 2019). Контрольные и кодирующие участки ДНК отмечены разным цветом для генов, которые экспрессируются в тканях тела (серый) и ткани ядовитых желёз (жёлтый). Прямоугольники представляют место экспрессии: Vg – ядовитая железа; VT – ткань тела. X иллюстрирует устранение контрольных областей и отсутствие экспрессии в соответствующей ткани.

прессию), альтернативный сплайсинг, эволюцию белка *de novo*, и горизонтальный/латеральный перенос генов (рис. 3.24; Drukewitz, Reumont, 2019). Система ядов у разных видов изучается на уровне геномики, транскриптомики и протеомики. После характеристики состава ядов и отдельных токсинов с помощью испытаний на их синтез и активность, рассматриваются прикладные аспекты, такие как исследова-

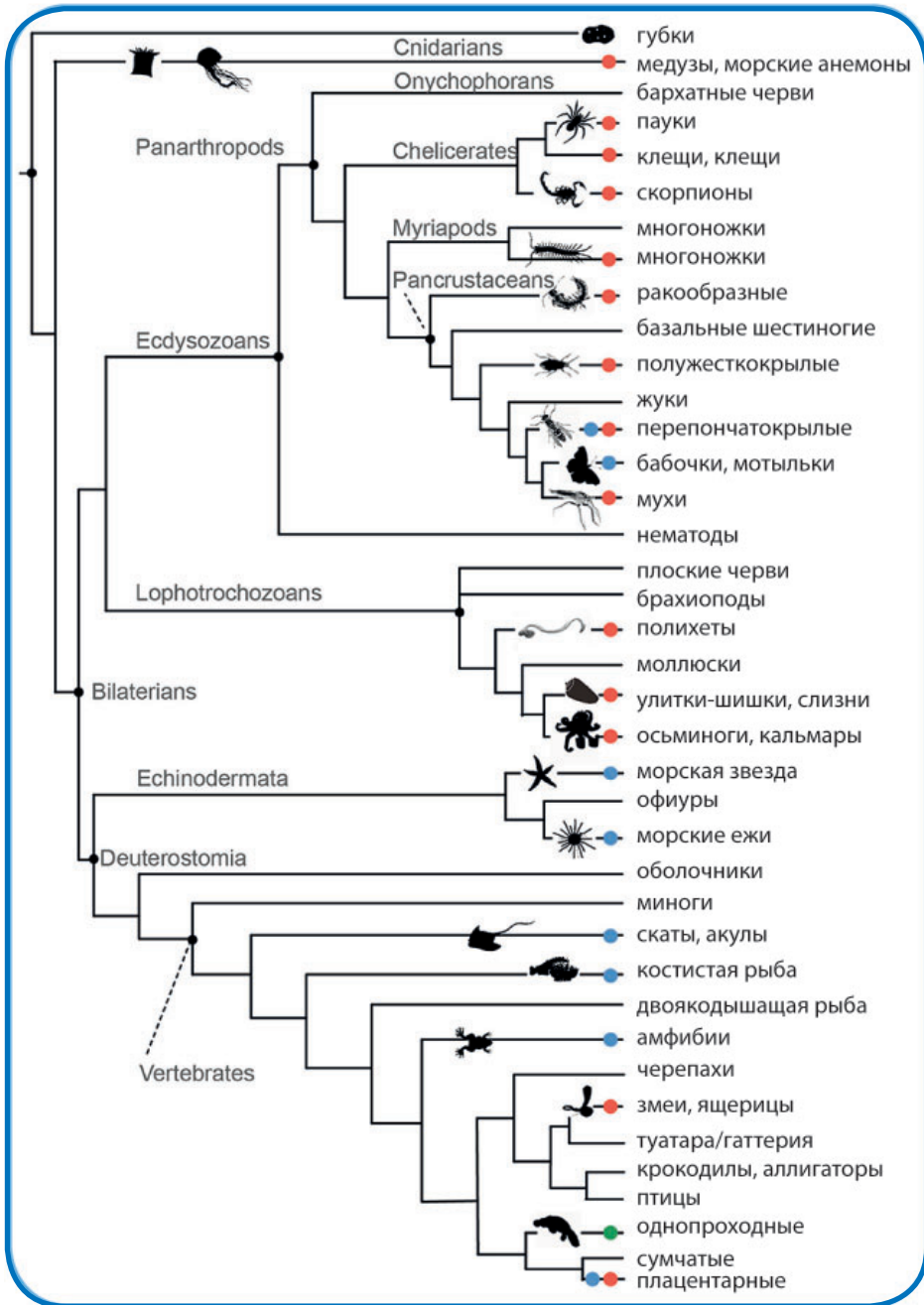


**Рисунок 3.25.** Веномика и её интегративные направления исследований (по: Drukewitz, Reumont, 2019).

СЗЭ – структурно-зависимая экспрессия, ЭОТ – эволюция отдельных токсинов, ЭСЯ – эволюция состава яда.

ния противоядий или разработка агрохимикатов и лекарств; вопросы эволюции состава яда или генов токсинов решаются эволюционной веномикой (рис. 3.25). Яды эволюционировали конвергентно в различных линиях животных в качестве ключевых приспособлений, повышающих эволюционную приспособленность видов. Они представляют собой сложные коктейли различных токсинов, обладающих широким спектром биологической активности. Из огромного разнообразия систем ядов в животном мире достаточно подробно изучена лишь небольшая часть. Описаны многие белки ядов, однако, эволюционные процессы и механизмы, которые способствовали их происхождению и вызывали рекрутирование в систему доставки яда, остаются дискуссионными. Высококачественные данные о последовательностях полных геномов ядовитых видов всё ещё немногочисленны и неравномерно распределены между таксонами (рис. 3.26) (Drukewitz, Reumont, 2019; Farhat et al., 2023).

Среди ядовитых видов перепончатокрылых на геномном уровне хорошо изучена паразитоидная оса *Nasonia vitripennis* (Walker, 1836). Чтобы понять, как у этого вида формируется явно целевой яд, было проведено сравнительное геномное исследование с анализом четырёх близкородственных паразитоидных ос из группы Pteromalidae. Насекомые оказались эволюционно довольно молодыми, с максимальным временем



**Рисунок 3.26.** Упрощённое филогенетическое древо эволюции яда в животном мире (из: Drukewitz, Reumont, 2019).

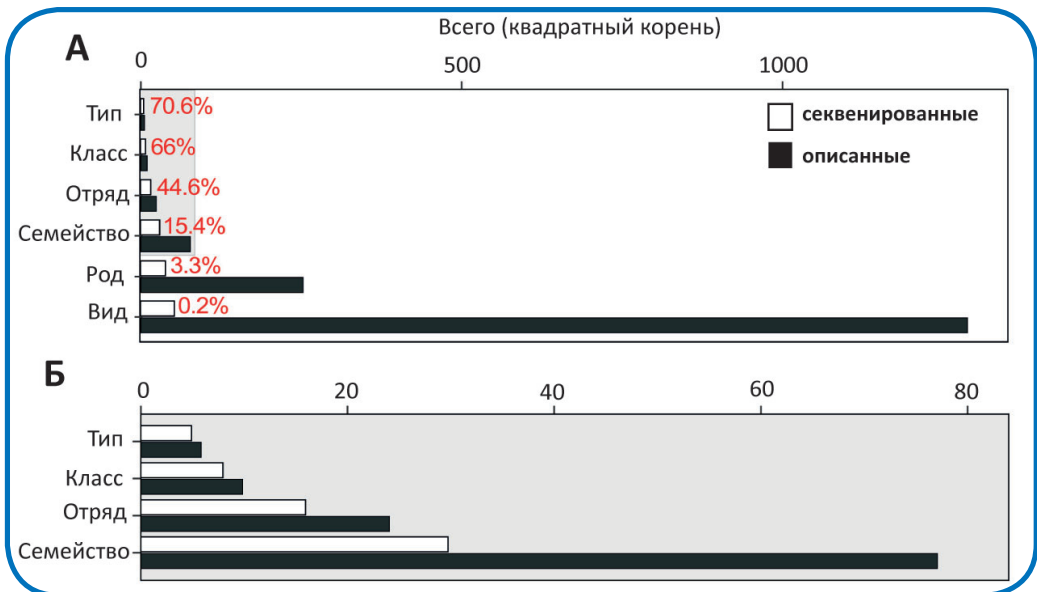
Таксоны, которые используют яды для хищничества, отмечены красными точками, для защиты – синими, для конкуренции – зелёными.

дивергенции около 4,9 млн л. н., но продемонстрировали закономерности специализации на разных хозяевах. В зависимости от вида хозяина в ядовитых железах у разных видов ос экспрессируются разные гены (Drukewitz, Reumont, 2019).

Neogastropoda представляют высокоразнообразную группу хищных морских моллюсков с разной стратегией питания (например, отравление, асфиксия и вампиризм), некоторые из которых вырабатывают специальные вещества для подавления своих жертв. Золотым стандартом для изучения яда этих моллюсков считаются улитки-конусы, каждый вид которых продуцирует около 200 токсинов. Некоторые из них уже прошли испытания, а один вышел на рынок (зиконотид, продаётся в качестве анестетика) (Farhat et al., 2023). Недавно были получены высококачественные геномы двух видов, которые могут служить эталоном для последующих исследований неогастропод и идентификации геномных процессов, определяющих их эволюционный успех и производство ядов (Farhat et al., 2023).

## ГЛАВА 4. ПРИМЕНЕНИЕ ГЕНОМИКИ В РЕШЕНИИ НАУЧНЫХ ЗАДАЧ

В 2021 г. самая обширная в мире генетическая база данных включала наилучшие доступные сборки генома в GenBank для 3278 видов животных из 24 типов (рис. 4.1), хотя и с резкими различиями в геномном представлении между разными таксономическими группами (Hotaling et al., 2021). Доступность полногеномной информации существенно преобразила научные исследования и обеспечила стремительный прогресс в решении широкого круга проблем (явных и скрытых), сформировав новую постгеномную биологию, основанную на фундаментальных знаниях молекулярных основ жизни на Земле во всём её разнообразии.



**Рисунок 4.1.** Доступность геномов царства Animalia по сравнению с таксономическими описаниями (из: Hotaling et al., 2021).

А – доля описанных таксономических групп по сравнению с количеством секвенированных геномных сборок от типов к видам (красным шрифтом указаны пропорции секвенированных геномов).

Б – увеличенное изображение категорий более высокого таксономического ранга.

### 4.1. Вариации размера геномов

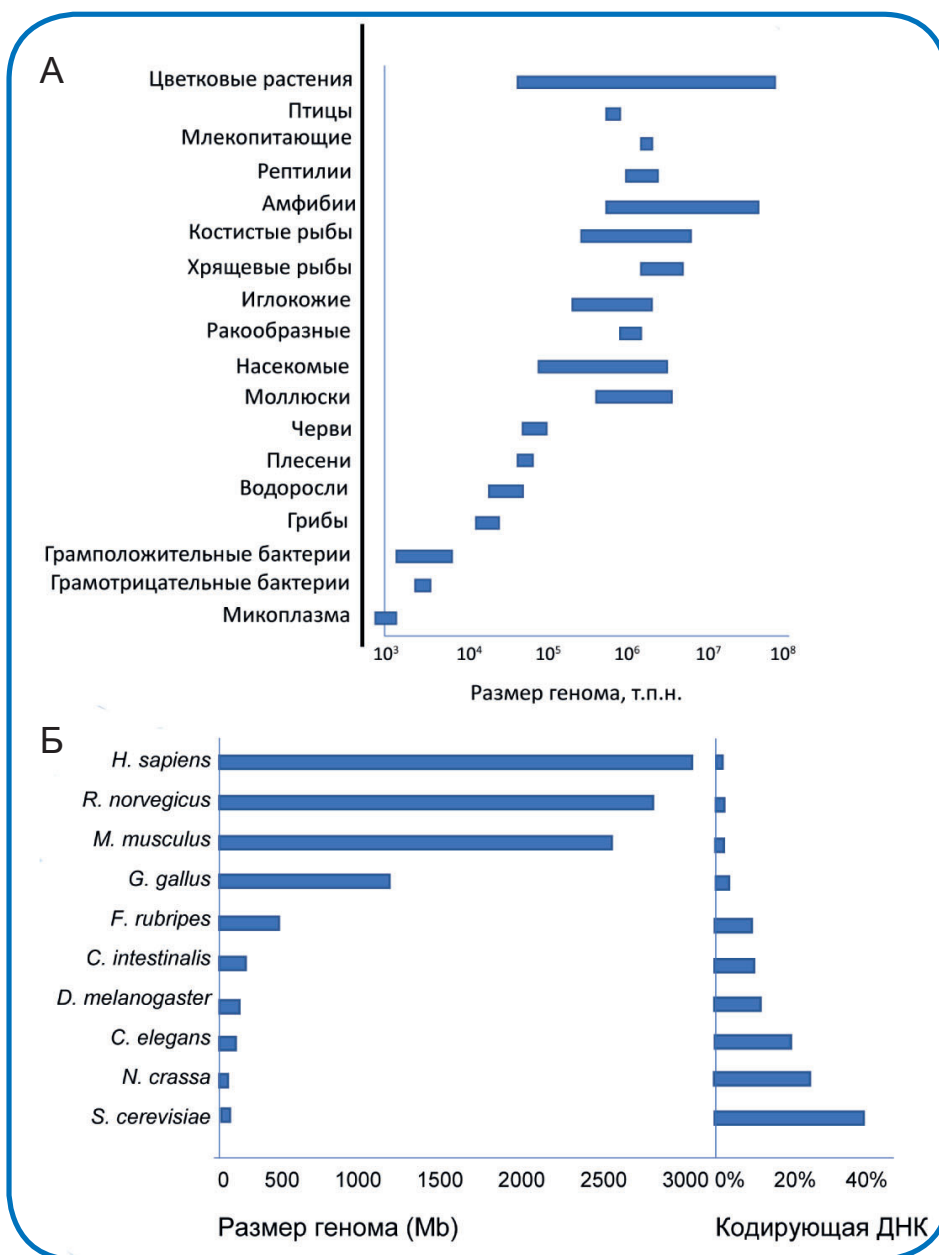
Размер генома является фундаментальной характеристикой биологического вида. Систематические исследования в этой области начались в 50-х годах прошлого столетия, однако вопросы, связанные с эволюционными причинами и последствиями изменений в размере генома, продолжают вызывать активные обсуждения. Главный вопрос заключается в том, является ли размер генома нейтральным признаком (Ellis et al., 2014). Проблему можно эффективно решить, сочетая классические эволюци-

онные подходы с современными технологиями секвенирования геномов, такими как MobiSeq. Этот метод, особенно полезный при анализе гигантских геномов, основан на создании библиотеки сокращённого представления с ориентацией на фланкирующие области мобильных элементов для идентификации однонуклеотидных полиморфизмов и генотипов (Rey-Iglesia et al., 2019). Считается, что естественные изменения размера генома имеют важнейшее значение в различных биологических процессах. Они оказывают влияние на геномную пластичность, динамику клеток, способность организмов адаптироваться к окружающей среде, а также на процессы диверсификации видов (Han et al., 2024).

Размеры генома у разных видов эукариот отличаются по меньшей мере на пять порядков (~0,002 Gb – ~150 Gb), и это не связано со степенью сложности организма (Fedoroff, 2012; Wang et al., 2023b) (рис. 4.2). Такое несоответствие, ранее известное как С-парадокс (название дал Чарльз Томас-младший в 1971 г.), в настоящее время называется «загадкой С значения» (термин введён Райаном Грегори в 2000 г.), поскольку непосредственной причиной вариаций размера генома установлена некодирующая ДНК (Blommaert et al., 2019; Blommaert, 2020). Опубликованные оценки размера геномов доступны в базах данных значений С (<https://cvalues.science.kew.org/> и AGSD; <https://genomesize.com/>); в них содержится информация для более 12000 видов растений и более 6000 видов животных (Kusgwaha et al., 2023).

Размеры геномов существенно различаются в разных группах организмов. Например, у членистоногих, рыб, покрытосеменных растений и водорослей наблюдаются 250, 350, 1000 и 5000 кратные отличия в размерах генома соответственно. В целом, самые крупные геномы имеют покрытосеменные растения, хвостатые амфибии и двоякодышащие рыбы (Kusgwaha et al., 2023). Например, размер генома (определён цитометрически) вороньего глаза *Paris japonica* Franchet (150 Gb) в 50 раз превышает размер генома человека (Pellicer et al., 2010). Однако рекордный размер эукариотического генома, который достигает 160,45 Gb, принадлежит эпифитному папоротнику *Tmesipteris oblaneolata* Copel. из Новой Каледонии (Fernández et al., 2024). Геномы двоякодышащих рыб – рогозуба *Neoceratodus forsteri* (Krefft, 1870) (43 Gb) и протоптеры *Protopterus aethiopicus* Heckel, 1851 (130 Gb) – в 14 и 40 раз соответственно больше размера генома человека (3,2 Gb). Основной причиной таких значительных различий являются протяжённые межгенные области и интроны, богатые повторами (около 90%), компоненты которых имеют большее сходство с аналогичными элементами генома четвероногих, чем лучепёрых рыб (Meyer et al., 2021).

Геномы всех двоякодышащих рыб продолжают увеличиваться, поскольку некоторые мобильные элементы (TE) у них всё ещё активны. В частности, геном американского чешуйчатника *Lepidosiren paradoxa* Fitzinger, 1837 демонстрирует паразитический рост на протяжении последних 100 млн л., добавляя эквивалент одного человеческого генома каждые 10 млн л. (Schartl et al., 2024). Интересно, что хромосомы *N. forsteri* сохраняют синтению с другими видами позвоночных, а микрохромосомы – консервативную гомологию с кариотипом предков позвоночных (Meyer et



**Рисунок 4.2.** Размеры геномов в разных таксономических группах (А) (по: Fedoroff, 2012) и у отдельных видов эукариот с долей кодирующей ДНК (Б) (по: Kumar et al., 2010). В целом, наиболее вариабельны размеры геномов у цветковых растений, а среди позвоночных – у амфибий; этим же группам принадлежат обладатели самых больших геномов. Доля кодирующей ДНК очень мала в более крупных геномах, т. е. большая часть генома, приобретенного этими организмами, является некодирующей.

al., 2021). У саламандр, геномы большинства видов которых диплоидны (т.е. размер генома, варьирующий от 8 до 120 Gb, не связан с полиплоидизацией), интроны также значительно длиннее, чем у других представителей позвоночных, и могут содержать большое количество регуляторных областей. Наряду с увеличением размера интронов, у саламандр обнаружили непропорционально большое количество ретротранспозонов с длинными концевыми повторами и предположили, что пролиферация этих элементов может быть движущим фактором гигантизма их геномов (Weisrock et al., 2021; Myers, Pyron, 2024).

Информация о гигантских геномах позволяет понять их структуру и функции в различных линиях эукариот, выявляя уникальные и универсальные геномные процессы для конкретных групп (Hidalgo et al., 2017). Значимость данных о крупных геномах также состоит в том, что они содержат больше информации об эволюционной истории видов. В таких геномах вероятность обнаружения свидетельств недавних популяционных событий значительно выше, и они могут содержать больше независимых маркеров, что способствует более точной реконструкции родственных связей (Weisrock et al., 2021). Возможность получения полных данных о гигантских геномах саламандр даёт надежду на выяснение тех особенностей их организации, которые обеспечивают беспрецедентную способность к регенерации повреждённых органов, представляющую исключительную ценность для регенеративной медицины человека. У нескольких видов амфибий идентифицированы некоторые гены (отсутствующие в других группах животных), которые экспрессируются в регенерирующей ткани конечностей. Имеющиеся данные позволяют предполагать, что способность к регенерации была широко распространена среди ранних четвероногих позвоночных, но была утрачена в процессе эволюции и в настоящее время сохранилась только у саламандр (Nowoshilow et al., 2018). Саламандры служат важными моделями для изучения эмбрионального развития и эволюции четвероногих. Секвенирование генома мексиканского аксолотля *Ambystoma mexicanum* (Shaw & Nodder, 1789) обнаружило отсутствие в нём важного гена развития *Pax3*, функцию которого взял на себя гомологичный ген *Pax7*. Примечательно, что в отличие от общего увеличения размера интронов, в локусе *НохА* аксолотля, который играет важную роль в развитии конечностей и реактивируется во время их регенерации, размеры интронов почти такие, как у других позвоночных (за исключением *AmНохА3*). Ограничение размера интронов в генах развития у аксолотля, возможно, происходит потому, что меньшие размеры генов способствуют быстрой транскрипции и, следовательно, более эффективной регуляции этих генов в определённых контекстах развития (Nowoshilow et al., 2018).

В целом процесс укрупнения генома и появление избыточных некодирующих блоков повторяющихся последовательностей является имманентным свойством, которое приводит к увеличению размера ядер и клеток, замедлению клеточных процессов и в итоге снижает скорость переработки энергии окружающей среды, что не позволяет считать его адаптивным признаком. Однако виды с большими геномами

оказываются эволюционно перспективными после периодов вымирания и геологических и экологических катастроф за счёт сброса избыточной факультативной ДНК или изменения её топографии в геноме (Голубовский, 2000). Юрий Александрович Филипченко отмечал, что «в эволюции играли главную роль какие-то внутренние силы, заложенные в самих организмах». Прежде всего, это разные фракции повторяющейся ДНК и мобильные элементы, обеспечивающие тенденцию к увеличению размера генома (Голубовский, 2000).

Изучение небольших геномов также предоставляет важную для эволюционной биологии информацию. Ортонектиды (Orthonectida, эндопаразиты беспозвоночных) имеют самый малый из известных геном (15,3 Mb) и самое низкое число генов (5120 кодирующих белки генов) среди многоклеточных животных. Геном чрезвычайно компактен из-за значительного сокращения количества генов, межгенных областей, длины интронов и повторяющихся элементов, что, вероятно, связано с паразитическим образом жизни (Slyusarev et al., 2020). Среди свободноживущих животных наименьшим геномом обладает почвенная нематода *Caenorhabditis elegans* (Маурас, 1900) Dougherty, 1955 (97 Mb, 19000 генов), при этом более 40% предсказанных белков этого вида находят существенные совпадения в других организмах (The *C. elegans* Sequencing Consortium, 1998). Среди растений наименьший геном (12,56 Mb) зарегистрирован у одноклеточной зелёной водоросли *Ostreococcus tauri* C. Courties & M.-J. Chrétiennot-Dinet (1995). Несмотря на высокую плотность генов (из-за редукции межгенных областей, слияния генов и т.п.), геном структурно сложен, с ненаблюдаемым ранее уровнем гетерогенности для эукариот. Две хромосомы, явно отличающиеся от остальных, имеют значительные отклонения в содержании G+C и несут большинство мобильных элементов, а гены одной из них характеризуются также уникальным использованием кодонов и сплайсинга. Как отмечают авторы, это делает *O. tauri* идеальной модельной системой для исследования эволюции генома эукариот (Derelle et al., 2006). Геном микроспоридии *Encephalitozoon cuniculi* C. Levaditi, Nicolau & R. Schoen, 1923 (2,9 Mbp, 2000 генов) широко известен как пример чрезвычайной редукции и уплотнения. Однако ядерный геном сестринского вида *E. intestinalis* (A. Cali, D. P. Kotler & J. M. Orenstein) оказался ещё более редуцированным (2,3 Mbp, т.е. сокращён ещё на 20%). Сравнение геномных последовательностей показало, что подавляющее большинство генов, отсутствующих у *E. intestinalis*, расположены на концах хромосом *E. cuniculi*. Потери генов и укорачивание межгенных областей в коровых участках хромосом объясняют большую часть уменьшения размера *E. intestinalis*, тогда как сокращение длины самих генов незначительно (рис. 4.3). Данные предполагают, что межгенные последовательности и интроны достигли пределов редукции для полностью функционального генома *E. intestinalis* (Corradi et al., 2010).

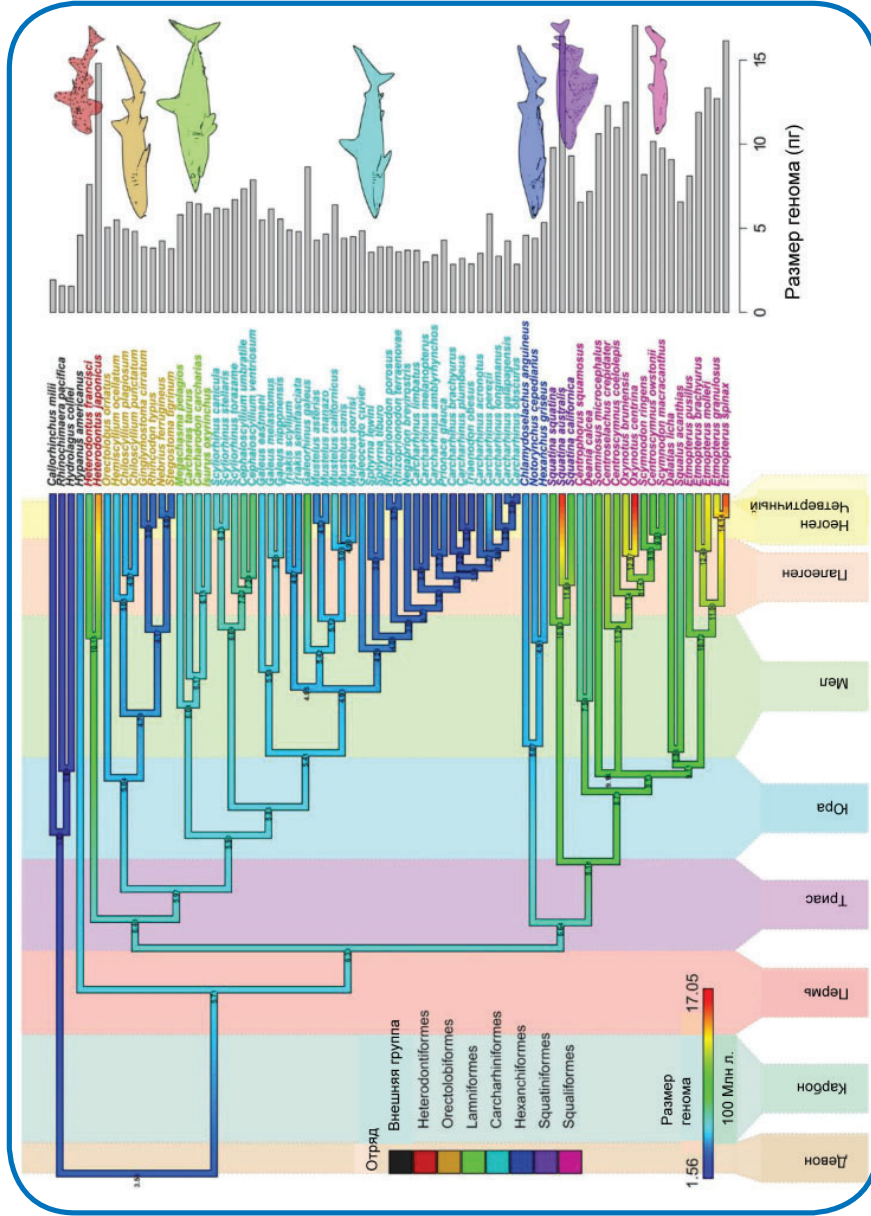
Эволюционной основой для понимания динамики и темпов изменений размера геномов во времени на различных таксономических уровнях служат филогенетические реконструкции (Kushwaha et al., 2023). Исследование геномов 71 вида акул (рис. 4.4), представляющих основные филогенетические линии, а также их исторические



**Рисунок 4.3.** Редукция генома у *Encephalitozoon intestinalis* (по: Corradi et al., 2010).

К – коровая часть, ORF – открытая рамка считывания.

и экологические особенности, обнаружило значительное увеличение геномов позд-непалеозойских предшественников акул, которое не было связано с динамичной эволюцией кариотипа этой клады. Размер генома общего предка для всех современных акул был оценен в 6,33 пг, а для всех современных хрящевых рыб Chondrichthyes – в 3,55 пг. На протяжении всего филогенеза прослеживался шестикратный диапазон вариаций в размере геномов. У современных акул надотряда Selachii размер генома эволюционирует по-разному. В то время как галеоморфные акулы в целом проявляют тенденцию к уменьшению размера генома (за исключением Heterodontiformes и Scyliorhinidae), большинство скваломорфных акул (особенно Squatiniformes и Squaliformes) демонстрируют его стабильный рост. Кроме того, большие размеры генома оказались связаны с «медленным» образом жизни и исключительно морской средой обитания, а также с неплацентарными способами размножения, что может рассматриваться, как метаболические ограничения на увеличение размера генома.



**Рисунок 4.4.** Эволюция размера генома у акул (из: Sáez et al., 2024). Филогения на уровне видов («максимальная достоверность клады» или гипотеза древа MCC), иллюстрирующая реконструкцию размера генома предков по наиболее подходящей модели максимального правдоподобия «раннего стада» (ЕВ) с подробностями по геологической шкале времени. Цветовая кодировка по среднему размеру генома каждого вида (столбчатая диаграмма) указана на врезке (внизу слева). Цвета внутренних узлов основаны на наиболее вероятных оценках размера генома предков. Названия видов имеют цветовую кодировку отрядов, к которым они принадлежат; виды внешней группы показаны чёрными цветом.

Однако в отличие от лучепёрых рыб, большой размер генома не был связан ни с таксономическим, ни с низким генетическим разнообразием (Sáez et al., 2024).

Нейтральная теория эволюции размера генома включает две гипотезы: мутационного риска и мутационного равновесия, предполагающие накопление ДНК за счёт дрейфа. Однако потеря ДНК в первом случае объясняется отбором (более эффективен по мере увеличения размера популяции) «слегка вредной избыточной ДНК», тогда как во втором случае к сокращению генома приводит постоянная скорость небольших делеций. Согласно другим гипотезам, размер генома может быть адаптивным признаком, влияющим на такие фенотипические характеристики, как размер тела, время развития, размер клетки и т. д. (Blommaert, 2020). Например, гипотеза «ограничения большого генома» предполагает наличие эволюционных ограничений для крупных геномов и селективное преимущество сокращения их размера. При этом, большие геномы могут отрицательно влиять на видовую диверсификацию (эволюционируемость). Гипотеза «отбора большого генома» признаёт связь между увеличением размера генома и адаптацией к специфическим экологическим условиям (например, к температурно-стабильным средам в тропиках). Примечательно, что эти тенденции отмечены в нескольких таксономических группах (некоторые растения, ракообразные, саламандры и т. д.), и они согласуются с различными (r и K) стратегиями воспроизводства (Saez et al., 2024). Гипотеза «экологически отобранного генома» подчёркивает важность оптимизации размера генома в ответ на экологические стрессы (засуха, изменение приливной зоны, ограничение питательных веществ и т. д.) (Han et al., 2024). Поскольку сокращение численности популяций происходит, когда скорость изменения окружающей среды превышает темпы адаптивных реакций, была выдвинута гипотеза о том, что размер генома может влиять на риск вымирания видов в быстро меняющихся условиях антропогенеза. Эта гипотеза «размера генома – вымирания» (не получившая поддержки при исследовании амфибий) опирается на концепцию, согласно которой увеличение генома приводит к снижению приспособленности (Pincheira-Donoso et al., 2023). В соответствии с гипотезой «нуклеотипа» размер генома у отдалённо родственных таксонов положительно коррелирует с размером клеток и ядер, но отрицательно – со скоростью деления клеток и скоростью основного обмена веществ. Действительно, в некоторых таксонах позвоночных летающие виды, для которых необходима высокая скорость метаболизма (что не могут обеспечить крупные геномы), обладают геномами минимального размера (Sáez et al., 2024). Считается, что предполагаемый направленный отбор на уменьшение размера генома может быть частично обусловлен биофизическим влиянием размера генома на клеточные и физиологические процессы, вне зависимости от последовательности ДНК (Gomez et al., 2024). Существует также гипотеза о том, что метаболические ресурсы (например, фосфор и азот) также важны для отбора на размер генома (Blommaert, 2020).

Размер генома может изменяться под влиянием различных внешних факторов, таких как экологический стресс и процесс одомашнивания (Zhang, Qiu, 2023). При-

чинами вариаций размера генома могут быть изменение его пloidности или дубликации отдельных участков (1), добавочные хромосомы (В-хромосомы) и гетерохроматические вставки (2), а также амплификация повторяющейся ДНК – tandemных повторов или мобильных элементов, TE (3) (Blommaert, 2020). Причём разные механизмы могут взаимодействовать между собой. Например, у некоторых растений и рыб вызванные полиплоидизацией ослабления селективных ограничений способствуют активности мобильных элементов (Falcon et al., 2023). В целом, размер генома отражает баланс между противоположными силами приобретения и потери ДНК. Анализ скоростей этих преобразований в геномах млекопитающих и птиц показал, что они имеют тесную взаимосвязь. Большинство потерь ДНК происходит в результате ~10 kb делеций, а скорость рекомбинаций в геноме отрицательно коррелирует с длиной интронов и содержанием TE. При этом считается, что aberrантная неаллельная рекомбинация между концами ретротранспозона с длинным терминальным повтором, видимо, является основным способом уменьшения размера генома (Falcon et al., 2023). Показательно, что у саламандр скорость геномных микроделеций значительно (в 3–10 раз) ниже, чем у бесхвостых амфибий, костистых рыб и Amniota, а также у них более редки крупные геномные делеции (Falcon et al., 2023). Предположительно, молекулярная основа эффективного удержания повторов может быть связана с появлением репрессорной системы белков цинковых пальцев, ассоциированных с боксом Круппеля (KRAB-ZFP; представляют собой большой класс факторов транскрипции четвероногих). По крайней мере, в целом отмечена аппроксимативная корреляция между размером генома и количеством доменов KRAB (Falcon et al., 2023). Мобильные элементы являются одним из основных факторов, определяющих общий размер генома, и у видов с очень большими геномами (>10 Gb) их содержание крайне высокое (Wang et al., 2023b). Вопрос о том, как связаны между собой сайленсинг мобильных элементов и геномный гигантизм, подробно рассматривается в работе Цзе Ванга с коллегами (Wang et al., 2023b).

Помимо глобальных для всех эукариот, выраженные различия в размере генома также наблюдаются на низких таксономических уровнях. В этом плане привлекательной моделью для изучения вариаций размера генома признаны коловратки (Rotifera, тип первичноротых животных). В комплексе видов коловраток *Brachionus plicatilis* (Müller, 1786) наблюдали семикратные отличия в размере генома. Причиной могли быть изменения в содержании повторяющихся (включая TE) элементов, но полиплоидия при этом не играла роли (Blommaert et al., 2019). Имеются примеры, предполагающие, что размер генома коррелирует с образом жизни. Например, у коловраток надотряда Gnesiotrocha размер генома у сидячих и одиночных видов существенно меньше, чем у подвижных и колониальных (Brown, Walsh, 2019). Хотя для биологического вида размер генома в основном считается стабильным признаком, существуют примеры его внутривидовой и даже внутриволюционной изменчивости. Например, у одного из видов коловраток *Brachionus asplanchnoidis* Charin, 1947 размер генома может отличаться почти в два раза в пределах одной географической популяции. Размер генома в популяциях этого вида положительно коррелирует с раз-

мером тела, размером яйца и продолжительностью эмбрионального развития. Дополнительная ДНК (присутствующая у одних особей, но отсутствующая у других) проявляет нуклеотипические эффекты и состоит преимущественно из тандемных повторов сателлитной ДНК с несколькими вкраплениями генов, что свидетельствует о наследовании и накоплении вариаций числа копий в поколениях (Stelzer et al., 2021a, 2021b).

Исследования на бделлоидных коловратках не выявили доказательств того, что ТЕ распространяются путём скрытой рекомбинации или сдерживаются особыми механизмами репарации ДНК. Напротив, обнаружено, что количество ТЕ в геномах меняется относительно медленно, в то время как семейства генов, участвующих в супрессии ТЕ посредством РНК-интерференции, претерпели значительное увеличение. Таким образом достигается снижение негативных последствий активности ТЕ и компенсируются неблагоприятные последствия длительной асексуальности, которая обычно рассматривается как эволюционный тупик (Nowell et al., 2021). Вместе с тем, для вида бделлоидных коловраток *Adineta vaga* (Davis, 1873), ранее считавшимся бесполом, получены свидетельства межиндивидуального генетического обмена и рекомбинации, хотя их механизмы остаются неясными (Vakhrusheva et al., 2020).

Логично предположить, что чем сложнее устроен организм, тем больше должен быть минимальный размер его генома, хотя геномы животных имеют удивительно стабильное количество генов, кодирующих белки (особенно среди билатерий и их ближайших филогенетических родственников) (Hehmeyer et al., 2023). Евгений Кунин, анализируя геномные данные, пришёл к заключению, что доминирующий способ эволюции генома сильно отличается от способа фенотипической эволюции, а архитектура генома, по-видимому, определяется балансом между давлением отбора, уровнем рекомбинации и активностью эгоистичных элементов. В результате, наблюдаемая архитектура генома в основном является продуктом нейтральных процессов и может быть связана с более общими селективными процессами (например, отбором на оптимизацию генома в успешных линиях с большими популяциями) (Koonin, 2009). Это соответствует взглядам Франсуа Жакоба, основанным на идее, что «эволюция обладает ограниченной способностью создавать совершенно новые компоненты, и что именно настройка и смешение ранее существовавших компонентов в значительной степени движет эволюционным процессом» (Falcon et al., 2023), а также представлениям о том, что «генетические локусы у эукариот устроены по модульному принципу...; это даёт возможность быстрой сборки новых конструкторов и создаёт мощные возможности регуляции генных ассамблей» (Голубовский, 2000). При этом, все геномы можно разделить на два класса: 1 – компактные, относительно небольшие геномы вирусов, архей, бактерий (обычно <10 Mb) и многих одноклеточных эукариот (обычно <20 Mb); большую часть в них занимают последовательности, кодирующие белки и РНК, и 2 – большие геномы многоклеточных и некоторых одноклеточных эукариот (обычно > 100 Mb), в которых большая часть нуклеотидной последовательности не является кодирующей (Koonin, 2009).

#### 4.2. Нуклеотидный состав и предпочтение кодонов

Геномы живых организмов часто имеют уникальные структуры, назначение которых в большинстве случаев неизвестно. Современные методы геномного анализа способствуют более глубокому изучению отдельных геномных элементов, их функционирования и роли в эволюции геномов и видов. GC-содержание (обычно применяется в качестве показателя плотности генов, хотя содержание GC и плотность генов не являются эквивалентными показателями) и использование кодонов являются простыми, но очень важными характеристиками всех эукариотических геномов. Современные исследования направлены на изучение моделей использования кодонов в целом, а не только в отдельных генах, а также на анализе роли содержания GC в организации и функционировании всего генома (Maldonado et al. 2018; Bouwman et al. 2023). Пропорции пар нуклеотидов АТ и GC часто оказываются неравными и могут существенно различаться как между видами, так и вдоль хромосом (хотя причины и последствия этой изменчивости до сих пор полностью не выяснены). Например, сравнение содержания GC в геномах различных растений показало, что злаки обладают самым высоким уровнем и гетерогенностью GC, у незлаковых однодольных этот уровень ниже, а у двудольных – самый низкий (Singh et al., 2016). Анализ содержания GC в геномах насекомых из четырёх отрядов (жесткокрылые Coleoptera, двукрылые Diptera, перепончатокрылые Hymenoptera и чешуекрылые Lepidoptera) обнаружил, что кодирующие белок последовательности имеют более высокое содержание GC, чем в среднем по геному, и что Lepidoptera, как правило, имеют более высокое содержание GC (Kyriacou et al., 2024). При межвидовых сравнениях были отмечены отдельные таксоны с необычно высоким или низким общим содержанием GC. В частности, содержание GC в геноме медоносной пчелы *Apis mellifera* Linnaeus, 1758 составляет 33%, обыкновенной гидры *Hydra vulgaris* Pallas, 1766 – 29%, а морской миноги *Petromyzon marinus* Linnaeus, 1758 – 46% (с увеличением до 75% в третьих позициях кодонов). Эти данные контрастируют со средним значением для выборочных геномов Metazoa (высококачественные сборки геномов для 1924 видов) в 39,9% GC (Kyriacou et al., 2024). Установление движущей силы, формирующей содержание GC, а также выяснение биологического смысла его вариаций необходимо для понимания эволюции генома (Singh et al., 2016). Имеющиеся в настоящее время данные позволяют предположить, что первичные процессы, лежащие в основе изменений содержания GC, не зависят от функции последовательности (Kyriacou et al., 2024).

От нуклеотидного состава зависят ключевые свойства нитей ДНК, такие как стабильность, гибкость и восприимчивость к повреждениям, а его изменения приводят к неоднородной структуре хромосом. Нуклеотидный состав связан с такими важными особенностями генома, как трёхмерная структура ДНК, или существование топологически ассоциированных доменов (Bernaola-Galván et al., 2023). Например, показано, что GC-богатые TAD имеют более высокую частоту петель и охватывают высокоэкспрессируемые гены и гены домашнего хозяйства, в отличие от TAD, характеризующихся более низким содержанием GC (Bouwman et al. 2023). GC-

ориентированный подход к изучению трёхмерной организации генома предполагает, что радиальный градиент содержания GC (от периферии ядра к его центру) является базисным принципом, управляющим структурой 3D-генома в отдельных клетках. Он может влиять на образование структурных доменов и формировать 3D-геном на наноуровне. Разные механизмы сплайсинга работают в разных радиальных положениях в ядре, и более 70% известных сайтов связывания транскрипционных факторов распределены вдоль радиальных градиентов GC содержания. Нарушение основного распределения содержания GC вдоль линейного генома, вероятно, может привести к дисфункциализации 3D-генома (Bouwman et al. 2023).

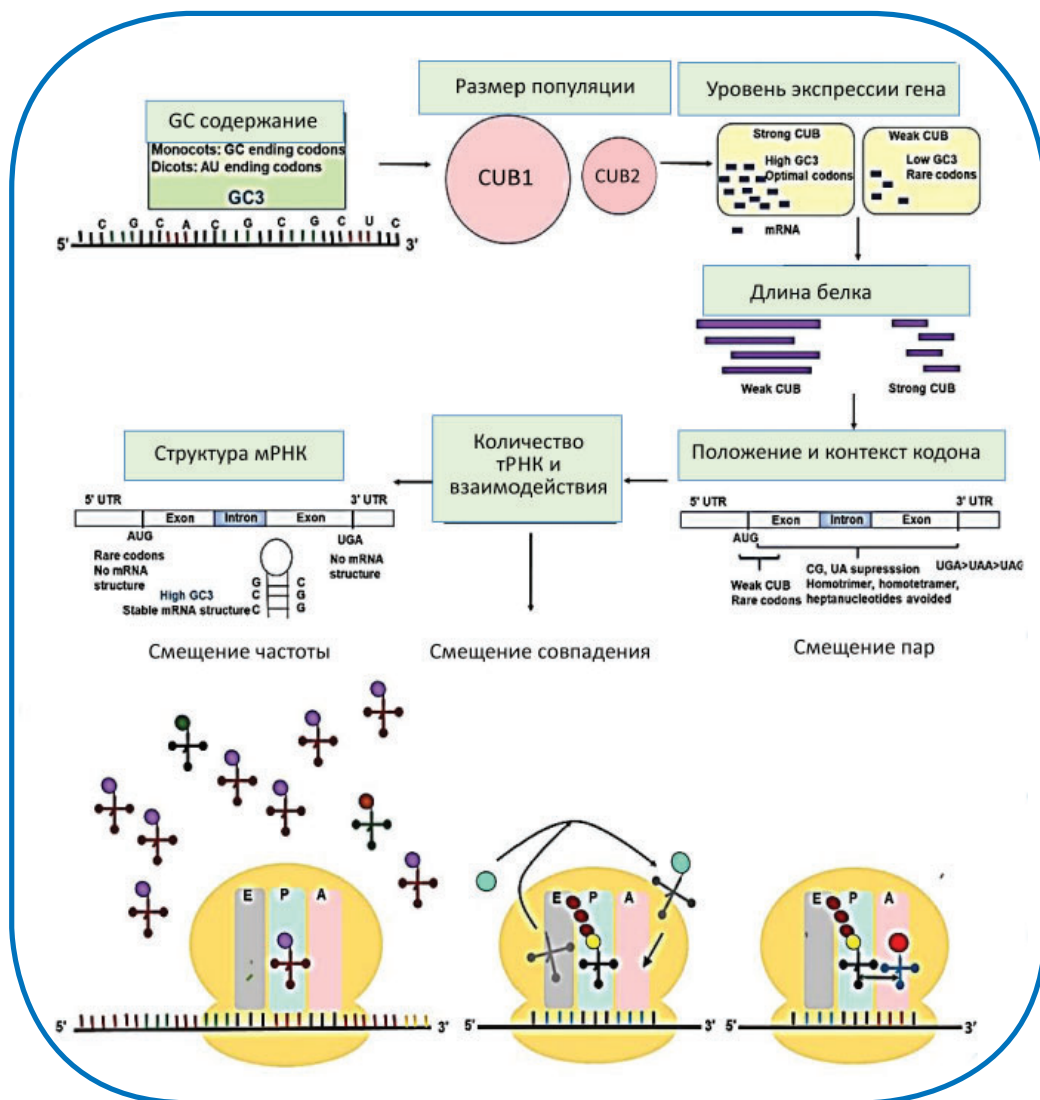
Интересно, что содержание GC оказывает влияние на формирование дифференцированных областей в континууме видообразования (Ke, Vasseur, 2024). Высокодифференцированные регионы генома, известные как геномные островки, устойчивы к потоку генов между зарождающимися линиями и предположительно играют роль в адаптации и репродуктивной изоляции. Изучение геномов 499 азиатских медоносных пчёл *Apis cerana* Fabricius, 1793 показало, что первоначально геномные островки, как правило, слабо дифференцированы и содержание GC в них ниже средних значений, но этот уровень возрастает при более глубокой дивергенции. При сравнении с регионами, имеющими высокое содержание GC, была обнаружена повышенная мутационная нагрузка, что указывает на неэффективный естественный отбор в регионах с низким содержанием GC. Таким образом, низкое содержание GC играет более значимую роль по сравнению с эволюционными процессами (например, таким как естественный отбор) в формировании геномных островков. Кроме того, регионы с низким содержанием GC демонстрируют более высокую долю полиморфизмов, характерных для отдельных линий, в отличие от областей с высоким содержанием GC, и это сглаживает несоответствия между митохондриальной и ядерной филогениями (Ke, Vasseur, 2024).

Являясь таксономическим признаком среди прокариот, GC-содержание в геномах эукариот, особенно у позвоночных, за некоторыми исключениями почти одинаково (40% – 45%; в геноме человека – 40,8%), но организация нуклеотидного состава до сих пор остается малоизученной, несмотря на наличие большого количества полногеномных данных (Vohnoutová et al., 2023). Тем не менее, геном позвоночных имеет гораздо более сложный композиционный состав, для объяснения которого около 40 лет назад в результате новаторских исследований группой Джорджио Бернарди была предложена теория изохор (согласно которой геномы представляют мозаику длинных однородных по нуклеотидному составу участков, а семейства изохор были разделены на АТ-богатые «лёгкие» и GC-богатые «тяжёлые»). Однако эта теория неоднократно подвергалась критике и в настоящее время не имеет всеобщего признания (Graur, 2022; Bernal-Galván et al., 2023; Vohnoutová et al., 2023). В частности, многими авторами оспаривалась концепция о том, что геном млекопитающих состоит из длинных однородных участков, различающихся по GC-содержанию; получены данные, опровергающие концепцию «композиционного перехода» между анимниотами

и амниотами; и, наконец, было показано, что изохоры не являются самыми длинными композиционными доменами. При этом более адекватной рассматривается концепция «композиционных однородных и неоднородных доменов» (Bernaola-Galván et al., 2023; Vohnoutová et al., 2023). Согласно современным представлениям, в геноме преобладает иерархическая композиционная структура, которую можно разделить на три основные категории: гетерогенности ближнего действия (тысячи пар оснований), изохоры размером от десятков до сотен тысяч пар оснований и суперструктуры, достигающие размера в десятки миллионов пар оснований. Определение композиционной структуры даёт возможность оценить меру композиционной сложности и применять её для сравнения геномов разных видов (Bernaola-Galván et al., 2023).

Хорошо известно, что во всех известных формах жизни частоты использования синонимичных кодонов отличаются. Оценка смещения традиционно проводится с помощью эвристических методов, таких как индексы использования кодонов, включая индекс адаптации кодонов (CAI), значения относительного использования синонимичных кодонов (RSCU) и содержание GC в третьей позиции синонимичных кодонов. Кроме того, используются другие подходы, основывающиеся на моделях популяционной генетики, которые в более полной мере учитывают влияние процессов мутации и отбора. Например, подход максимального правдоподобия (включает модель с несколькими аллелями для классификации кодонов по четырём классам), байесовские подходы (FMutSel и ROC SEMPPR, которые количественно определяют влияние отбора и мутации на предпочтение кодонов) и байесовская оценка DECUB (количественно оценивает совместные эффекты мутаций, отбора и GC-смещённой генной конверсии в масштабах всего генома) (Kotari et al., 2024).

Отклонение в использовании кодонов, известное как «codon usage bias» (CUB) или «codon preference», признано исключительно важным для генной экспрессии и клеточных функций (например таких, как связывание транскрипционных факторов, фолдинг белков, модификация убиквитина, сплайсинг мРНК, регулирование тканеспецифичных генных продуктов и т. п.) (Fox, Erill, 2010; Plotkin, Kudla, 2011; Mauro, Chappell, 2014 и т. д.). Для объяснения этого феномена предложены две гипотезы: мутационная и естественного отбора. Первая гипотеза, основанная на нейтральности мутаций, часто привлекается для объяснения межвидовой изменчивости в использовании кодонов. Гипотеза естественного отбора, предполагающая, что синонимичные замены могут влиять на жизнеспособность организма, обычно привлекается для объяснения CUB в геномах или генах (Plotkin, Kudla, 2011). CUB обычно указывает на баланс между мутационным смещением и естественным отбором для оптимизации трансляции и может прояснить эволюцию конкретного гена (Fox, Erill 2010; Maldonado et al. 2018; Moss et al., 2023; Wu et al., 2024 и т. д.). Возникает отклонение в силу различных факторов (уровень экспрессии генов, стабильность РНК, адаптация к условиям роста), но, тем не менее, нуклеотидный состав является сильнейшей детерминантой при межвидовых сравнениях (Parvathy et al., 2022; Tuagi et al., 2023; Khandia et al., 2024; Kotari et al., 2024) (рис. 4.5). В целом, предпочтение кодо-



**Рисунок 4.5.** Основные факторы, влияющие на смещение использования кодонов в организмах на примере цветковых растений (из: Parvathy et al., 2022). Содержание GC в геноме, размер популяции, уровень экспрессии генов, длина белка, положение и контекст кодона, количество тРНК и взаимодействия, а также структура мРНК показаны схематически. CUB1 – сильное смещение и естественный отбор, CUB2 – слабое смещение и генетический дрейф. Взаимодействия тРНК подразделяются на смещение частоты, смещение совпадения и смещение пар. E, P и A обозначают соответственно выход, пептидные и аминокислотные участки в рибосомах

нов представляет уникальную характеристику каждого организма (Fox, Erill, 2010; Plotkin, Kudla, 2011; Parvathy et al., 2022).

Для изменений в использовании кодонов установлен ряд закономерностей. Существует выраженная положительная корреляция между уровнем генной экспрессии и степенью отклонения в использовании кодонов, а также отрицательная корреляция между уровнем экспрессии гена и скоростью синонимичных замен между дивергентными видами (Plotkin, Kudla, 2011; Vicario et al., 2008). Гены с максимальной экспрессией в период быстрого роста организма и белкового синтеза имеют большие отклонения в использовании кодонов, чем гены с максимальной экспрессией в период медленного роста (Vicario et al., 2008). Кроме того, отклонения в использовании кодонов сильно влияют на экспрессию гетерологичных генов, причём использование определённых кодонов может увеличить уровень их транскрипции более, чем в тысячу раз (Plotkin, Kudla, 2011). Поэтому оптимизацию кодонов часто предлагают при создании высокоэкспрессирующих конструкций, пригодных для генной терапии и генетических вакцин. Однако при этом следует учитывать, что оптимизированные кодоны могут повлиять на конформацию и функции белка, повысить иммуногенность и в итоге снизить эффективность лечения (Mauro, Chappell, 2014). Учитывая важную роль конкретного синонимического кодонового состава гена на уровне его экспрессии, недавно предложены два новых простых статистических и эмпирических метода для прогнозирования уровня экспрессии белка по нуклеотидной последовательности соответствующего гена: индекс экспрессии кодонов (CEIS) и показатель продуктивности кодонов (CPS). Оба метода основаны на влиянии каждого отдельного кодона в гене на общий уровень экспрессии кодируемого белка и частоту изоакцепторов (Zaytsev et al., 2024).

Вначале считали, что синонимические кодоны обладают полностью эквивалентными функциями. Обнаружение неодинаковой частоты использования синонимических кодонов в мРНК позволило предположить, что положение определённых кодонов может быть не случайным и нести дополнительную (помимо кодирования аминокислотных остатков) функциональную нагрузку. Действительно, как показали дальнейшие исследования, *«генетический код имеет «несколько слоёв», несущих такую дополнительную информацию, и что синонимические кодоны расположены в мРНК не случайно, а стратегически – чтобы гарантировать определенную кинетику трансляции, которая, в свою очередь, обеспечивает тонкую настройку процесса сворачивания белка в клетке и способствует упорядоченному котрансляционному сворачиванию, которое происходит в ходе поэтапного и последовательного структурирования отдельных участков полипептидной цепи, выходящих из рибосомы в различные моменты времени»* (Комар, 2021). Наглядный пример – единственная синонимичная мутация (GTG→GTA) в гене *F9* (кодирует фактор свёртывания крови FIX) у пациентов с лёгкой формой гемофилии, которая оказалась причиной изменения кинетики синтеза и конформации FIX (Комар, 2019; 2021). Таким образом, использование кодонов существенно влияет не только на уровень экспрессии гена, но

также на структуру и функцию белков, что в конечном итоге формирует ландшафт протеома (Liu et al., 2021).

Исследования также показали, что гены с наиболее высокой экспрессией (HEG; например, кодирующие рибосомные белки) имеют тенденцию к увеличению количества специфических синонимических кодонов, которые известны как «оптимальные» или «мажорные» кодоны. Несмотря на то, что предпочтительные, оптимальные кодоны не всегда одинаковы у разных видов, они имеют тенденцию сохраняться в одних и тех же положениях геномной последовательности у близкородственных видов (что можно измерить как синонимическое расстояние), и эти кодоны могут «идти против» мутационного смещения, характерного для генома. Более того, у видов с действующим трансляционным отбором (согласно гипотезе колебания, допускающей возможность нестрогого соответствия первого нуклеотида антикодона тРНК с третьим нуклеотидом кодона мРНК при трансляции), всегда ожидается чрезмерная представленность некоторых кодонов в HEG: это UUC(Phe), UAC(Tyr), AUC(Ile) и AAC(Asn). Они признаны универсальными оптимальными кодонами и лежат в основе S-индекса трансляционного отбора, при этом замена синонимических часто используемых кодонов на редко используемые кодоны и наоборот влияет на уровень экспрессии генов (Parvathy et al., 2022; Комар, 2019; 2021). Редкие кодоны, как правило, собраны в кластеры, локализующиеся в определённых участках мРНК. Полногеномный анализ открытых рамок считывания (ORF) про- и эукариот показал, что кластеризация редких кодонов не ограничена определённым набором генов, не связана с общим GC содержанием и встречается чаще, чем при случайном выборе (Комар и др., 2019). Кроме одноклеточных видов, трансляционная селекция в синонимических сайтах также была обнаружена у растений (например, *Zea mays* L., 1753 и *Arabidopsis thaliana* (L.) Heunh., 1842). У некоторых гельминтов как мутационное смещение, так и трансляционный отбор являются основными силами, формирующими паттерны использования кодонов. Например, показано, что у трематоды *Schistosoma mansoni* Weinland, 1858 (кровавые сосальщики) уровень экспрессии является важным фактором, определяющим использование синонимических кодонов, причём отбор для трансляции, по крайней мере частично, ответственен за смещение использования синонимических кодонов (Lamolle et al., 2019; 2022).

Анализ особенностей использования синонимических кодонов в геномах вирусов подсемейства Coronavirinae (родов *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* и *Deltacoronavirus*), включая SARS-CoV-2, в открытых рамках считывания ORF1ab и ORF2 выявил достоверные различия в их использовании как разными родами, так и отдельными видами коронавирусов при значительном дистанцировании SARS-CoV-2 от большинства других бетакоронавирусов. Показательно, что коронавирусы летучих мышей и птиц часто занимают промежуточное положение по отношению к видам, относящимся к разным родам коронавирусов (Тюлько, Якименко, 2020).

Гены, кодирующие белок, заканчиваются одним из трёх стоп-кодонов (ТАА, ТГА или TAG), которые, как и синонимические кодоны, не используются одинаковым образом. Хотя в большинстве таксонов ТАА является оптимальным стоп-кодоном, ТГА широко распространён и консервативен в геномах млекопитающих. Эта исключительность млекопитающих хорошо объясняется тем, что рекомбинация GC (gBGC) имитирует очищающий отбор, и что ТАА является селективно оптимальным кодоном. Частое использование ТГА и высокие показатели ТАА→ТГА замен у млекопитающих согласуются с предсказанной способностью gBGC «управлять» вредными мутациями и подтверждают гипотезу о том, что сохранение последовательности не обязательно должно указывать на очищающий отбор (Ho, Hurst, 2022). Геномные исследования также показали, что для завершения этапа биосинтеза белка значительную роль играют не только белковые факторы терминации, но и ближайшее нуклеотидное окружение стоп-кодонов. Известно, например, что использование стоп-кодона UAA предпочтительно для АТ-богатых, а UAG и UGA – для GC-богатых генов, и это, в свою очередь, связано с уровнем их экспрессии. В геноме человека аналогичный нуклеотидный состав стоп-кодонов и их 3'-контекста стимулирует терминацию трансляции, а нуклеотидный состав 3'-контекста, отличающийся от последовательности стоп-кодона, снижает эффективность терминации трансляции (Сokolova и др., 2020).

Таким образом, использование кодонов является ранее недооценённым слоем генетической информации, который играет важную роль в регуляции экспрессии генов. Несмотря на достижения в этой области, по-прежнему остаётся множество нерешённых вопросов для дальнейших исследований (Liu et al., 2021).

#### 4.3. Малые интерферирующие РНК и РНК-интерференция

Впервые РНК-интерференцию наблюдали в 1928 г. в экспериментах с вирусами на растениях, но молекулярные механизмы были раскрыты только в 1989 г., когда Крейг Мелло и Эндрю Файер описали подавление экспрессии генов (сайленсинг) после введения двухцепочечной РНК в организм почвенной нематоды *C. elegans*. С тех пор РНК-интерференция стала широко использоваться для подавления экспрессии специфических генов *in vitro* в молекулярно-биологических исследованиях. Феномен ингибирования экспрессии генов комплементарными молекулами нуклеиновых кислот консервативен у большинства эукариот. Главная роль РНК-интерференции у растений – защита от патогенов, у млекопитающих – контроль развития, «судьбы» и гибели клетки при разных состояниях в физиологической норме и патологии (Isenmann et al., 2023). Леонид Иванович Корочкин (2006) рассматривал РНК-интерференцию как механизм формирования и наследственной передачи функционального состояния генома.

Малые некодирующие РНК, способные подавлять генную экспрессию (т.е. интерферирующие), и связанные с ними белки оказались одним из наиболее обильных рибонуклеопротеидных комплексов клетки. Однако их идентификация может быть

затруднена, если экспрессия ограничена немногочисленными клеточными типами или специфическими условиями окружающей среды. Поэтому для их выявления в дополнение к экспериментальным методам были разработаны вычислительные подходы *in silico* для геномных данных (Huang et al., 2010). Они основываются на разных принципах, включая поиск гомологии с известными генами и исследование их окружения, а также идентификацию консервативных генных сегментов, которые «выпадают» из предсказанных белок-кодирующих регионов. Потенциальные мРНК мишени могут быть идентифицированы с помощью таких программ, как MiRanda и PicTar, и экспериментальной биологии, включающих иммунопреципитацию мРНК в комплексах сайленсинга, а также секвенирование (Devaney et al., 2010).

Наиболее хорошо изучены три типа канонических малых некодирующих РНК с определённой структурой (Zhou, 2019). Короткие РНК (siRNA; размер 20–24 нуклеотида) являются продуктами как чужеродных нуклеиновых кислот, так и эндогенных геномных локусов. Они широко распространены у растений, одноклеточных и беспозвоночных и считаются эволюционно наиболее ранними. пиРНК – ассоциированные с Piwi (семейство регуляторных белков группы Argonaute) РНК (piRNA; размер 26–31 нуклеотид), являются самым большим классом некодирующих молекул РНК животных (обнаружено более 173 млн уникальных последовательностей пиРНК у 21 вида). пиРНК распределены по всему геному в виде высококонсервативных (в отличие от самих последовательностей генов) кластеров, которые проявляются только в ограниченном числе локусов.

микроРНК (microRNA; обычно 22 нуклеотида длиной) кодируются огромным набором генов, причём транскрипт может включать кластеры различных микроРНК, или микроРНК и белка (в этом случае последовательность микроРНК расположена внутри интрона). микроРНК обнаружены у растений, животных, некоторых вирусов, и в отличие от генов, кодирующих белки, они часто возникают *de novo*. микроРНК являются наиболее консервативными геномными элементами (Velandia-Huerto et al., 2022). В геноме человека обнаружено >2000 генов микроРНК, контролирующих работу более 60% генов, кодирующих белок (большинство из которых являются факторами развития и транскрипции) (Gupta et al., 2012). Аномальная экспрессия генов микроРНК может быть причиной болезней человека и сильных фенотипических изменений вплоть до летального исхода (Esau, Monia, 2007). Открыты стресс-индуцированные микроРНК, появляющиеся в ответ на специфические условия (Flynt, Lai, 2008). Показано, что микроРНК вовлечены в такие ключевые биологические процессы, как развитие, дифференцировка, апоптоз и пролиферация (Bartel, 2004; Harfe, 2005). Они играют важную роль в контроле генной экспрессии, эпигенетической модификации и регуляции активности гетерохроматина, а также в паразитохозяинных взаимодействиях (Obbard et al., 2009; Челомина, 2011).

Эволюция геномного набора микроРНК тесно коррелирует с важными морфологическими инновациями, и микроРНК использовались в качестве ценных маркеров в филогенетических анализах (Velandia-Huerto et al., 2022). Широкомасштабное срав-

нительное исследование (Hertel et al., 2006), включающее филогенетический анализ всех известных семейств микроРНК животных, позволило надёжно идентифицировать кластеры, которые произошли от общего транскрипта. Были выявлены три эпизода инноваций микроРНК, которые соответствуют основным инновациям развития животных. Первую волну инноваций представляет класс примерно из 20 микроРНК, который является общим для первичноротых и вторичноротых и может быть связан с появлением билатеральности. Вторая волна инноваций ассоциирована с позвоночными, а третья – с плацентарными млекопитающими. Кроме того, обнаружена ожидаемая экспансия микроРНК, обусловленная геномными дупликациями у ранних позвоночных и анцестральных костистых рыб. Нелокальные дупликации у предка позвоночных предшествовали локальным (тандемным) дупликациям, ведущим к образованию множества предковых кластеров микроРНК. Примечательно, что инновации микроРНК, тандемные и нелокальные дупликации генов микроРНК распределены в филогении Metazoa неравномерно. Основанное на геномных данных изучение микроРНК китайской печёночной двуустки *Clonorchis sinensis* (Cobbolt, 1875) Looss, 1907, возбудителя тяжёлых заболеваний человека и животных, привело к выводу, что у вида существует две основные стратегии микроРНК: хранение большого разнообразия семейств микроРНК различных животных, и поддержание строго консервативных областей и участков с высокой инновационной активностью, что идеально подходит паразиту для осуществления его сложного жизненного цикла (Xu et al., 2010).

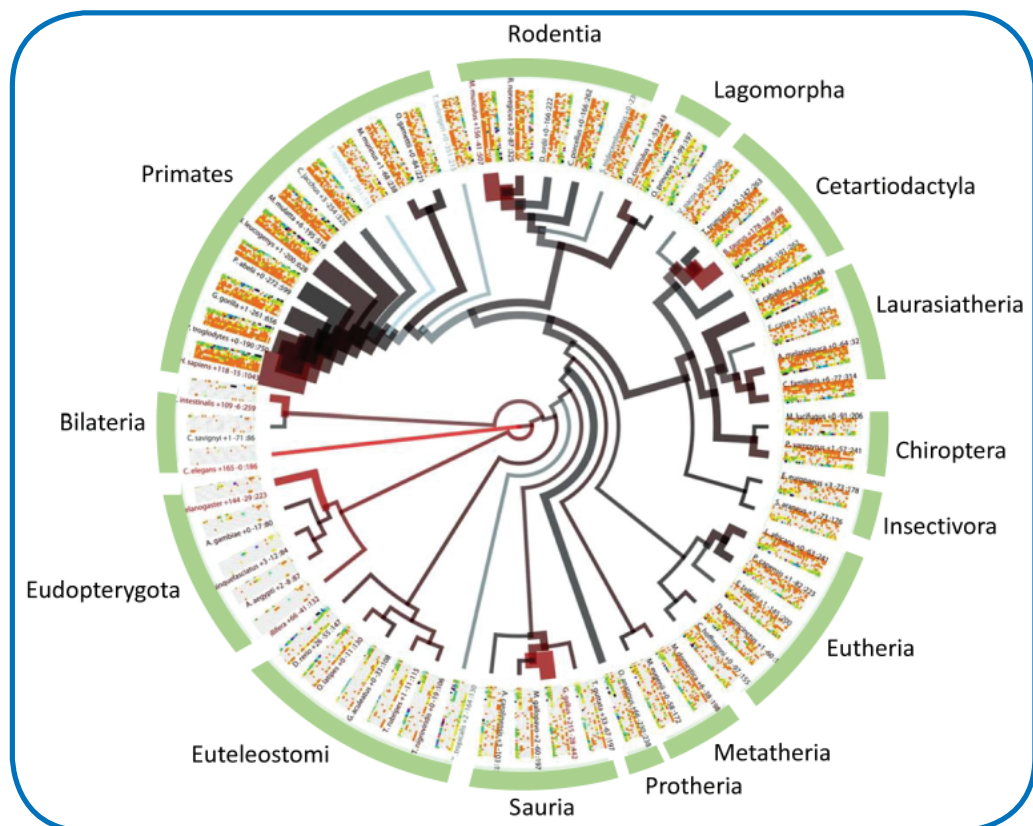
Полностью автоматизированный крупномасштабный анализ синтении, организации и эволюции микроРНК, основанный на геномных последовательностях и аннотациях семейств белков 80 видов, позволил сделать ряд важных выводов. В частности, был подтверждён недавний рост числа локусов микроРНК в ряде линий животных (например, приматов и грызунов) и выраженная потеря семейств микроРНК у других видов, таких как шпорцевые лягушки (*Xenopus tropicalis* (Gray, 1864) (1); обнаружено, что самые большие из выявленных экспансий включают микроРНК, которые участвуют в плюрипотентности и в переключении экспрессии генов у ранних эмбрионов с материнской на зиготную (2); установлен ряд новых ассоциаций между микроРНК и кодирующими белок генами, связанными с иммунным ответом (3), а также показано, что микроРНК имеют нетипичные паттерны синтении с предпочтением более длинных сгруппированных областей (4) (Guerra-Assunção, Enright, 2012).

Механизм подавления экспрессии генов с помощью малых РНК у животных был назван РНК-интерференцией, у грибов это явление получило название подавление транскрипции, а у растений – посттрансляционный сайленсинг генов. Термин РНК-интерференция относится к молекулярным процессам, в которых небольшой фрагмент РНК используется в качестве мишени для нацеливания на специфические нуклеотидные последовательности и регуляцию экспрессии генов (Torri et al., 2022). В зависимости от происхождения интерферирующих РНК, выделяют три категории таких процессов у животных: с участием микроРНК, коротких РНК и пиРНК. Короткие РНК препятствуют экспрессии специфических генов с комплементарными ну-

клеотидными последовательностями, разрушая мРНК-мишень после транскрипции и таким образом предотвращая трансляцию. Вместе с тем, существуют доказательства активации генов, индуцированной этими молекулами РНК. микроРНК связываются с комплементарными последовательностями соответствующей мРНК, затем сайленсинг генов происходит по одному из трёх сценариев: расщепление мРНК на две части (1), дестабилизация мРНК путём укорочения поли-А конца (2), и трансляция, требующая «помощи» рибосом (что, вероятно, малоэффективно) (3). пиРНК в основном участвуют в эпигенетическом и посттрансляционном сайленсинге мобильных элементов; сообщалось также, что индуцированный пиРНК сайленсинг ретротранспозонов необходим для развития эмбрионов млекопитающих и для сперматогенеза (Akgül et al., 2022; Torri et al., 2022; Velandia-Huerto et al., 2022; Zhou, 2019). Недавно выяснилось, что пути синтеза пиРНК существуют и в соматических клетках, хотя при этом их механизмы имеют отличия. Предполагается, что пиРНК могут играть важную роль в регенерации сердечной мышцы (участвуют в дифференцировке кардиомиоцитов), в старении (снижают экспрессию мобильных элементов, уменьшают повреждение ДНК и подавляют апоптоз), могут как стимулировать, так и ингибировать развитие онкологических заболеваний (Wang et al., 2020).

Развитие этого направления привело к созданию новой области исследований – микроРНК омики (miRNomics) (Akgül et al., 2022). Эволюционное распространение семейств микроРНК представлено на рис. 4.6. Происхождение РНК-интерференции обычно объясняется защитной гипотезой, согласно которой данная система развилась как защита от мобильных генетических элементов и РНК-вирусов и уже присутствовала у предков всех эукариот. В альтернативной, или нейтральной эволюционной гипотезе происхождения РНК-интерференции, переход от механизма регуляции генов прокариотической антисмысловой РНК к РНК-интерференции стал результатом конструктивной нейтральной эволюции (рис. 4.7). Согласно этой гипотезе РНК-интерференция уже существовала у последнего общего предка эукариот (до возникновения потребности в новой защитной системе), что помогло сформировать эукариотическую геномную архитектуру и стабильность (Torri et al., 2022).

Полногеномные дупликации (WGD) долгое время считались причинным механизмом, лежащим в основе резкого увеличения морфологической сложности из-за неофункционализации паралогов, генерируемых во время этих событий. Согласно альтернативной гипотезе, сохранение большинства паралогов происходит не из-за неофункционализации, а из-за степени взаимосвязанности предполагаемого генного продукта, а также самого способа WGD. Исследования распределения, экспрессии и молекулярной эволюции микроРНК у челюстноротых позвоночных и хелицеро-вых членистоногих привело к выводу, что WGD не приводят к созданию новизны микроРНК, и WGD не коррелируют с увеличением сложности. Напротив, именно количество начальных последовательностей микроРНК в самом геноме не только лучше коррелирует со случаями комплексообразования, но и механистически объясняет, почему сложность возрастает при возникновении новых семейств микроРНК (Peterson et al., 2022).

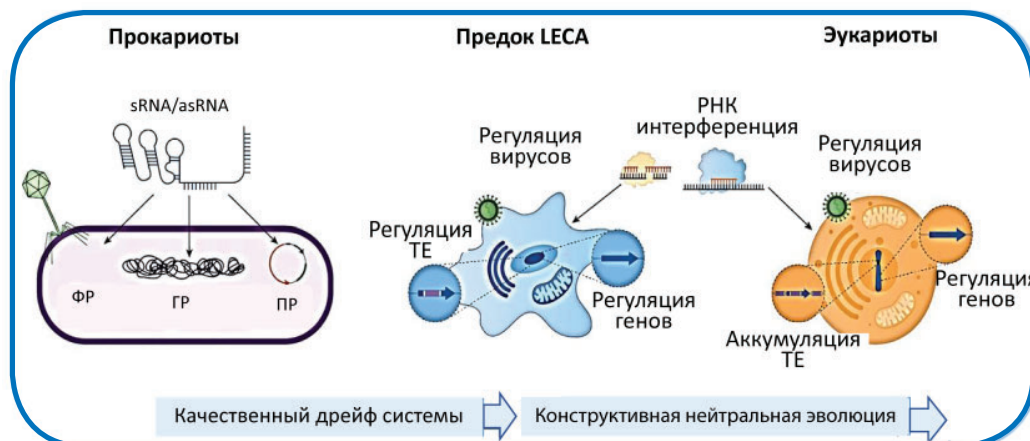


**Рисунок 4.6.** Филогенетическое древо, представляющее приобретения и потери семейств микроРНК (из: Guerra-Assunção, Enright, 2012).

Ширина ветви отображает количество семейств микроРНК; синий цвет означает значительную потерю семейства микроРНК, красный – увеличение. Каждый квадрат представляет определённое (из 408) семейство микроРНК: белый цвет указывает на то, что данный вид не относится к определённому семейству, а чёрный цвет – что данный вид содержит не менее 10 копий микроРНК в этом семействе. Копии от 1 до 10 обозначаются радужным градиентом (от красного до фиолетового). Группы видов обозначаются по названию предшествующей им эволюционной ветви.

#### 4.4. Мобильные генетические элементы и горизонтальный перенос генов

Мобильные генетические элементы (ГЭ) были впервые описаны Барборой Мак-клинтон в конце 40-х гг., обнаружившей в хромосомах кукурузы последовательности ДНК, способные перемещаться по геному. В настоящее время известно, что ГЭ составляют важную часть генома всех живых организмов, и в разной мере изучены у животных (впервые открыты отечественным биохимиком и молекулярным биологом академиком Георгием Павловичем Георгиевым с коллегами) и растений. У представителей типа плоских червей *Platyhelminthes* ретротранспозоны составляют до 40% генома, при этом многие из них экспрессируются на высоком уровне и являются



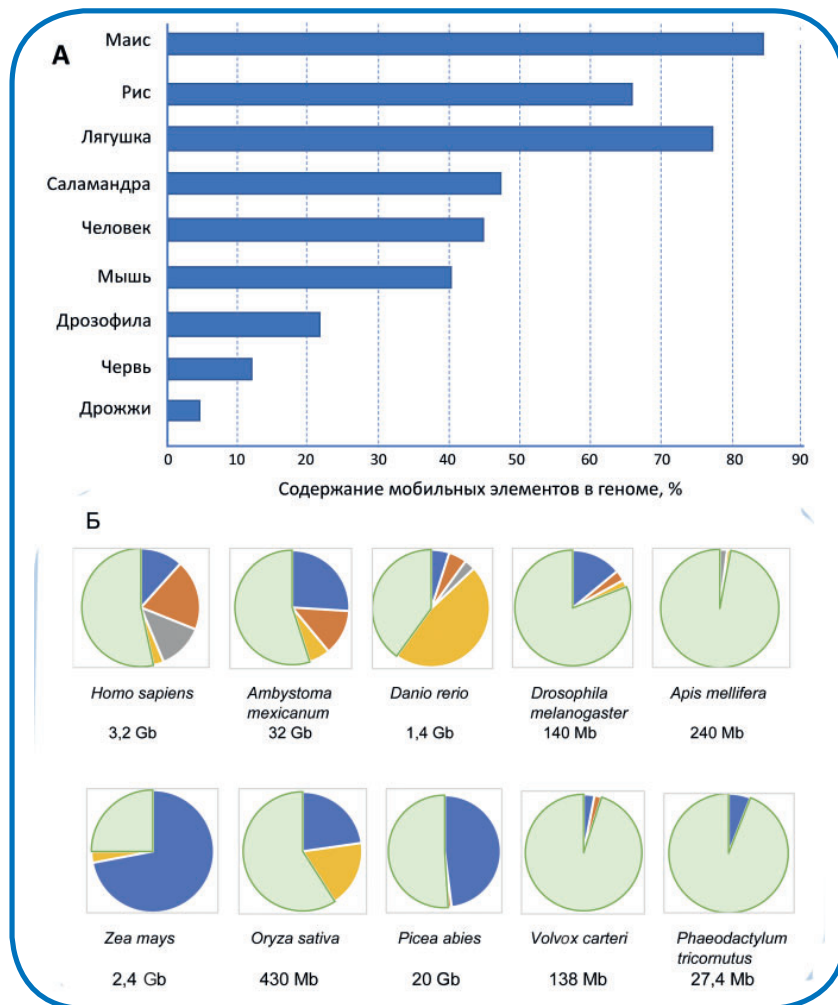
**Рисунок 4.7.** Эволюция от прокариотической РНК-опосредованной регуляции генов к эукариотической РНК-интерференции (из: Torri et al., 2022).

Предположительно, эволюционный путь от прокариотической РНК-опосредованной регуляции генов до эукариотической РНК-интерференции включал два отдельных эволюционных процесса. Первый связан с изменениями в молекулярном механизме без изменения конечных результатов процесса (качественный дрейф системы). Второй включал каскадный механизм, вызванный подавляющей ролью РНК-интерференции во вредных эффектах мобильных элементов (ТЕ), как постулирует конструктивная нейтральная эволюция.

sRNA – малая РНК; asRNA – антисмысловая РНК; LECA – последний общий эукариотический предок, ФР – регуляция фагов; ГР – регуляция генов; ПР – регуляция плазмид.

видо- или родоспецифичными (Johnston, 2006). Геном человека (как и большинства млекопитающих) почти наполовину состоит из мобильных элементов, и они играют важную роль во многих клеточных процессах (de Koning et al., 2011; Chiang et al., 2022), а максимальное содержание (около 85%) ТЕ отмечено для генома маиса (рис. 4.8) (Chénais et al., 2012). Быстрое накопление геномных данных позволяет прогнозировать обнаружение в ближайшем будущем новых семейств мобильных элементов в геномах немодельных видов (Nishihara, 2019).

Несмотря на то, что ТЕ традиционно рассматривались как нефункциональный «мусор» (эгоистичная ДНК), паразитические элементы или вредные мутагены, исследования показали, что они могут быть полезны для генома хозяина. ТЕ придают специфические признаки геномам, способствуют эволюционным изменениям размера и состава генома, влияют на эволюцию генных сетей и архитектуру генома, транскрипцию и процессинг пре-мРНК (Elbarbary et al., 2016; Nishihara, 2019). Помимо этого, ТЕ играют важную роль в организации, экспрессии и эволюции генов полового развития (Chiang et al., 2022). Выявление и характеристика ТЕ в данных секвенирования имеют большое значение для понимания их разнообразия и функции, а также для практического применения. Например, некоторые биотехнологические компании активно внедряют в клиническую практику новые стратегии генной инженерии, проявляя всё больший интерес к использованию возможностей ТЕ для модификации геномов (Arkhipova et al., 2024). Недавно разработана система классифика-



**Рисунок 4.8.** Вклад мобильных элементов в геномы разных видов эукариот (А) (по: Chénaïs et al., 2012) и пропорции разных ТЕ в геноме человека и некоторых видов животных и растений (Б) (по: Wells, Feschotte, 2020).

Размеры эталонного генома существенно отличаются у разных видов эукариот и слабо коррелируют с содержанием ТЕ. Синий – LTR (с включением ретроэлементов YR), коричневый – LINE, серый – SINE, жёлтый – элементы класса II; зелёный – не ТЕ.

ции и аннотации geNomad, которая объединяет геномную информацию и глубокую нейронную сеть для идентификации последовательностей плазмид и вирусов. Используя модель случайного поля, geNomad с высокой точностью обнаруживает про-вирусы, интегрированные в геном хозяина. Ожидается, что geNomad обеспечит дальнейший прогресс в изучении ТЕ (Camargo et al., 2024). Новые вычислительные инструменты, такие как Pantera, GraffiTE и Pannagram, идеально подходят для выяв-

ления активных транспозонов в пангеномах; причём эти элементы могут иметь как большой размер, так и малое количество копий (Arkhipova et al., 2024).

Все мобильные элементы по способу перемещения могут быть сгруппированы в два основных класса: 1 – ретротранспозоны (TE класса I), перемещающиеся с помощью РНК-посредника («copy and paste»), и 2 – ДНК транспозоны (TE класса II), перемещаются непосредственно ДНК («cut-and-paste»). Оба класса делятся последовательно на подклассы, суперсемейства и семейства, согласно механизму их интеграции. Ретротранспозоны обнаружены у всех эукариот, от грибов до млекопитающих, а ДНК транспозоны – самая древняя группа TE, распространённая среди про- и эукариот (Thomas et al., 2010; Bao et al., 2010). Например, только у гельминтов млекопитающих описано более 30 TE (ДНК транспозоны, LTR и non-LTR ретротранспозоны – содержащие и не содержащие длинные концевые повторы соответственно) (Thomas et al., 2010). LTR несодержащие ретротранспозоны могут быть автономными (LINEs – длинные диспергированные повторы), т.е. способными перемещаться по геному самостоятельно, и неавтономными (SINEs – короткие диспергированные повторы и SVA-SINE-VNTR-Alu элементы), требующими для перемещения присутствие других активных мобильных элементов (Klein, O’Neill, 2018). Мутации, создаваемые inserцией ретротранспозонов, более стабильны, чем мутации, вызванные ДНК транспозонами, поскольку в отличие от последних, ретротранспозоны, встроившись в геном уже никогда не исчезают (Georgiev, 1984). Для мобильных элементов установлен ряд общих закономерностей: ретротранспозоны с длинными концевыми последовательностями играют важную роль в изменении размера генома среди растений; у млекопитающих более распространены включения элементов LINEs и SINEs, чем LTR; в геномах птиц и чешуйчатых LINEs более распространены, чем SINEs и LTR; а в геномах амфибий практически отсутствуют SINEs (Dittrich et al., 2024). Различные типы TE могут иметь разные предпочтения относительно мест вставки. Например, в геноме человека SINEs обычно локализованы в областях, богатых генами, тогда как LINEs обогащены в межгенных областях (вероятно, действует отбор против вставки их длинных последовательностей в гены или рядом с ними) (Elbarbary et al., 2016). Оценки времени интеграции TE в геном также различаются в зависимости от вида и от типа TE (Dittrich et al., 2024). В целом, на распространение и распределение TE внутри и между геномами влияет уникальная биология различных TE, а экологические и генетические факторы, действующие на уровне вида-хозяина, дополнительно модулируют активность, диверсификацию и судьбу этих элементов (Wells, Feschotte, 2020).

Недавно у морских цианобактерий рода *Prochlorococcus* был обнаружен новый класс TE – транспозоны Tucherposons, названные в честь древнегреческой богини Тихе (в древнеримской мифологии ей соответствует Фортуна). Передающиеся от одной бактерии к другой, они переносят, в частности, гены для усвоения полезных метаболитов и для защиты от вирусов, в целом обеспечивая геномную пластичность. Tucherposons с высокой точностью встраиваются в определённые локусы генома (что

определяет интерес к изучению их рекомбиназ), не нарушая при этом его целостности. Интересно, что везикулы и фаговые частицы, собранные в морской воде, оказались обогащены тихепозонами и это, вероятно, обеспечивает для них возможность распространения в природе (Hackl et al., 2023).

Передвигаясь по геному и интегрируясь в новые сайты, ТЕ могут обеспечить потенциально вредную генетическую изменчивость, от полиморфизма простой последовательности, до кардинальных изменений в структуре хромосом, их целостности и инактивации соседних генов (Johnston, 2006). Например, все non-LTR ретротранспозоны или ретротранспозон-подобные последовательности одноклеточных паразитов (трипаносом *Trypanosoma brucei* Plimmer & Bradford, 1899 и *T. cruzi* Chagas, 1909, а также лейшмании *Leishmania major* Yakimoff and Schokhor, 1914) имеют тенденцию локализации в местах хромосомных инверсий, «strand-switch» регионов и хромосомных концов (Thomas et al., 2010). Вырезание и встраивание ТЕ индуцирует мутагенез, уровень которого может в тысячу раз превышать скорость спонтанных мутаций (Koga et al., 2006). Поэтому для подавления неконтролируемой пролиферации ТЕ сформировались специальные защитные механизмы против их транспозиции, включающие эпигенетические процессы, такие как система транскрипционного и посттранскрипционного сайленсинга генов. Однако различные виды физиологического стресса, в том числе изменения температуры, или колонизация новых мест обитания хозяевами, могут активировать ТЕ и привести к их пролиферации, разрушив эпигенетический сайленсинг (Wijayawardena et al., 2015).

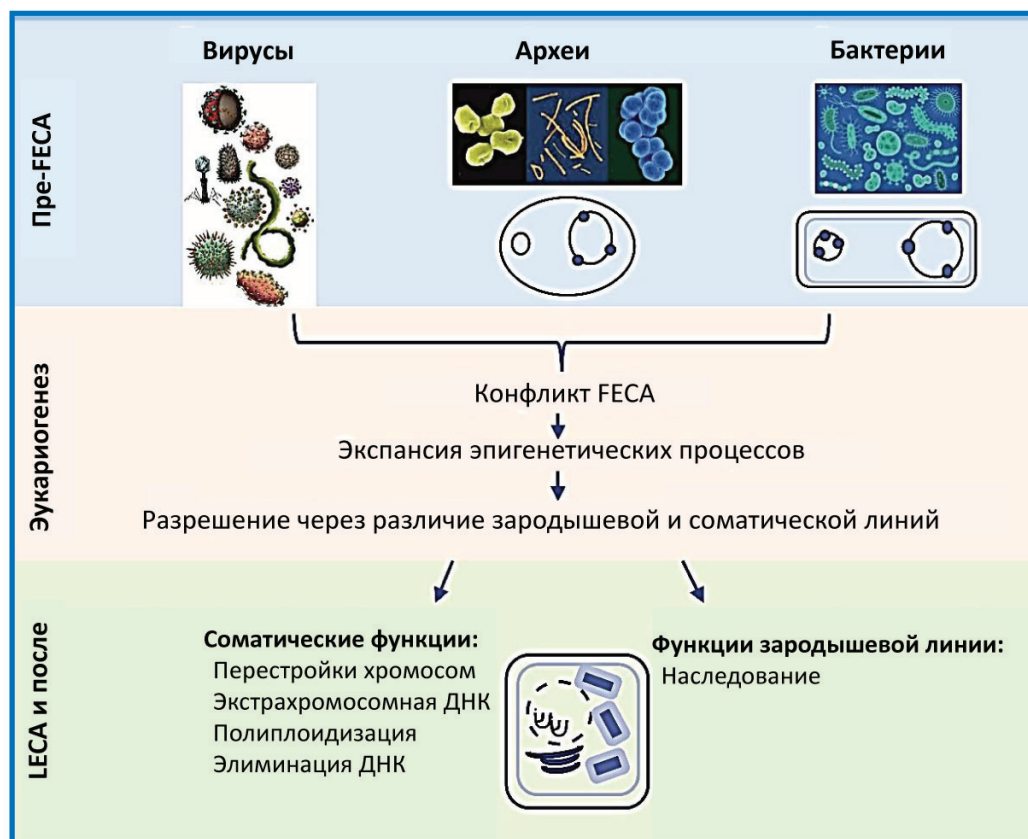
В то же время, получены данные, предполагающие коэволюцию ТЕ и геномов хозяев, положительное влияние и важную роль этих элементов в формировании и поддержании хромосом хозяина, модификации генной экспрессии и участии в разделении генома на хромосомные домены с эпигенетическими метками (Flutre et al., 2012; Thomas et al., 2010; Mashanov et al., 2012; Wijayawardena et al., 2015). Геномная доместикация может приводить к существенной функциональной эволюции ТЕ. В качестве примера можно отметить влияние этих элементов на эволюцию адаптивной иммунной системы челюстных позвоночных (Cooper, Alder, 2006) и нормального развития плаценты у млекопитающих (Kaneko-Ishino, Ishino, 2012), а также участие в репарации ДНК (Thomas et al., 2010). Описано большое число генов, произошедших от транспозонов, которые приобрели функцию регуляторов транскрипции, и получены данные, предполагающие происхождение теломеразы от обратной транскриптазы ретротранспозонов (Thomas et al., 2010). Иммунная система позвоночных постоянно эволюционировала благодаря интеграции транспозонов. Одним из недавних примеров этого процесса является кооптация экзона из ТЕ в рецептор интерферона типа I. В результате образовался рецептор-приманка, играющий важную роль в регуляции сигнальных путей интерферона. В более широком контексте «экзонизация» ТЕ создаёт значительный резервуар изоформ белков, который служит основой для естественного отбора у человека. Это явление требует глубокого экспериментального исследования, более доступного при изучении на других видах (Arkhipova et al., 2024).

Показано, что ТЕ связаны с тканеспецифичной экспрессией генов (Ito et al., 2017), управлением ранним эмбриогенезом (Wang et al., 2016) и с увеличением интронов у разных видов эукариот (Gozashti et al., 2022). Считается, что одним из факторов развития головного мозга человека являются события ретротраспозиций L1 и возникновение новых мобильных элементов (например, Alu и SVA) (Klein, O'Neill, 2018). Интересно, что взаимодействие последовательностей L1 человека с прилегающими к ним геномными участками, подверженными воздействию эпигенетических факторов, оказывает значительное влияние как на активность самих L1, так и на функции генома человека (т. е. L1, которые находятся под контролем хозяина, способны контролировать активность соседних генов) (Arkhipova et al., 2024).

Благодаря своему огромному потенциалу для создания генетической изменчивости, ретротранспозоны признаны основными средствами эволюции, которые приводят к фенотипической изменчивости, экологической адаптации, ремоделированию генома хозяина, а в долгосрочной перспективе – к видообразованию (Flutre et al., 2012; Thomas et al., 2010; Venancio et al., 2010). Существует гипотеза о том, что генетический конфликт с мобильными элементами способствовал эволюции генома у последнего общего предка эукариот и, вероятно, содействовал эукариогенезу (рис. 4.9) (Collens, Katz, 2021). ТЕ признаются одной из основных движущих сил регуляторной эволюции. Полагают, что они могут быть причиной радикального роста геномов у предков четвероногих (см. выше), а также эволюционных инноваций, которые позволили животным диверсифицироваться после перехода из воды на сушу (Faleon et al., 2023).

Считается, что транскрипционная активность ретротранспозонов может специфически контролироваться хозяином и играть важную роль в регенерации органов животных (Mashanov et al., 2012). Понимание механизмов регенерации, важное в теоретическом плане, может внести неоценимый вклад в регенеративную медицину человека (Moguel et al., 2015). Помимо фундаментального значения, практический интерес также может иметь способность ретроэлементов выполнять важные функции в эмбриогенезе, в частности, влиять на плюрипотентность клеток (Mashanov et al., 2012). Ретротранспозоны часто видо- или родоспецифичны, что открывает перспективу разработки чувствительных и специфичных молекулярных диагностикумов (Johnston, 2006). Изучение ТЕ представляет определённый интерес для эпидемиологических и медицинских исследований, поскольку они могут генерировать специфические генетические паттерны, ассоциированные с клиническими проявлениями заболеваний (Brindley et al., 2003; Челомина, 2015; Chénais, 2022).

ТЕ являются основными кандидатами на роль переносчиков в рамках гипотезы горизонтального/латерального переноса генов (HGT/LGT) между репродуктивно изолированными видами. Впервые HGT наблюдал Фредерик Гриффит в 1928 г. в эксперименте на бактериях, который продемонстрировал существование «трансформирующего принципа», и показал, что вирулентность может передаваться между штаммами пневмококков. Масштабное выявление HGT началось с наступлением эры секвенирования геномов и было показано, что HGT сыграл большую роль в эво-



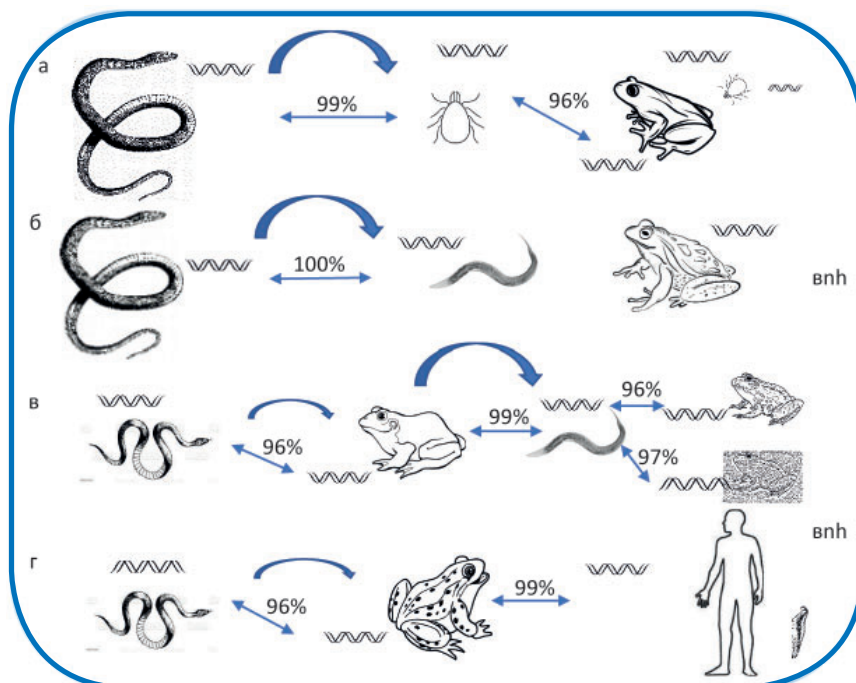
**Рисунок 4.9.** Происхождение эукариот (по: Collens, Katz, 2021).

В происхождении эукариот принимали участие вирусы, археи и бактерии (т.е., разнообразие вирусов и наличие ТЕ, интегрированных как в бактерии, включая предков митохондрий, так и в археи, включая вероятную клетку-хозяина FECA, первого эукариотического общего предка). Конфликт между этими геномами и ТЕ, вероятно, привёл к появлению эукариот.

люции не только бактерий, но и одноклеточных эукариот, растений и животных. События HGT, включая древний перенос генов между тремя доменами жизни (археями, бактериями и эукариотами), рассматриваются сейчас, как важная движущая сила эволюции эукариотического генома (Gladyshev et al., 2008; Prasad et al., 2021; Chen et al., 2021b; Undheim, Jenner, 2021). Причём считается, что система паразит-хозяин и симбиоз являются оптимальными моделями для изучения этого удивительного явления (Gilbert et al., 2010; Schaack et al., 2010; Wijayawardena et al., 2013). Например, одним из предполагаемых механизмов эволюции токсинов является горизонтальный перенос генов, а паразитоидные осы *Nasonia* Ashmead, 1904 представляют собой один из немногих примеров ядовитых видов, для которых механизм HGT описан достаточно подробно (Martinson et al., 2016). У *Nasonia* обнаружена хитиназа Gh19, которая происходит от одноклеточных микроспоридий и, вероятно, встречается также у других

паразитоидных ос. Ген порообразующих токсинов (PFTs) *Nematostella* Stephenson, 1935 с доменом аэролизина, очевидно, был перенесён горизонтально от патогенной бактерии *Aeromonas hydrophila* (Chester 1901) к актинии *Nematostella* (Moran et al., 2012), и эксперименты по нокдауну показали, что эти гены остаются функционирующими. События HGT были описаны и для других ядовитых видов, например, для генов латротоксинов пауков (Drukewitz, Reumont, 2019). Зерновка кофейных ягод *Hypothenemus hamperi* (Ferrari, 1867), вероятно, получила ген манназы *HhMAN1* от эндосимбиотических кишечных бактерий. Приобретение этого сложного гена даёт адаптивное преимущество, поскольку растения являются богатым источником сложных маннанов (Prasad et al., 2021). Геномный анализ многоножек обнаружил доказательства причастности HGT к эволюционному происхождению яда в линии животных, который может обеспечить «ускоренный канал для эволюции новшества за счёт превращения бактериального оружия в яды животных» (Undheim, Jenner, 2021).

Интересным примером HGT является фрагмент размером около 1,5 Mb бактерии вольбахии *Wolbachia* ssp. Hertig 1936 (род грамотрицательных плеоморфных бактерий из семейства Ehrlichiaeae класса альфа-протеобактерий), который, интегрировавшись в геном *Atmadillium vulgare* Latreille, 1804 (наземная мокрица-броненосец из класса высших раков), привёл к созданию новой половой хромосомы W6 (Prasad et al., 2021). Опосредованный вольбахиями HGT оказался связанным с цитоплазматической несовместимостью. У членистоногих цитоплазматическая несовместимость определяется как «эгоистичная репродуктивная манипуляция, вызванная эндосимбионтом *Wolbachia*». У самцов *Wolbachia* модифицирует сперму, что приводит к гибели эмбрионов при скрещивании с неинфицированными самками. У самок *Wolbachia* «спасает» скрещивание и позволяет яйцам нормально развиваться, что даёт им репродуктивное преимущество, позволяя передающемуся от матери симбионту быстро распространяться среди популяций хозяев. Сравнительный геномный анализ идентифицировал гомологи генов, лежащих в основе этого фенотипа, *cifA* и *cifB* (*cif* – фактор цитоплазматической несовместимости), тесно связанные с цитоплазматической несовместимостью. Оказалось, что в каждом геноме имеется до семи копий этих генов, и согласно филогенетическому анализу, *Wolbachia* часто приобретает новые копии из-за повсеместного горизонтального переноса между штаммами. Во многих случаях гены впоследствии приобретают мутации с потерей функции, превращаясь в псевдогены. Вначале это происходит в *cifB*, единственной функцией которого является модификация сперматозоидов, а затем – в *cifA*, который необходим для сохранения скрещивания у самок. Несмотря на то, что гены *cif* рекомбинируют, рекомбинация в значительной степени ограничивается близкородственными гомологами. Такой сценарий предсказывается в контексте модели коэволюции между модификацией сперматозоидов и защитой эмбрионов, когда рекомбинация между отдалённо родственными парами генов ведёт к созданию самонесовместимого штамма. Вместе эти паттерны приобретения, потери и рекомбинации генов способствуют поддержанию эволюционных моделей цитоплазматической несовместимости (Martinez et al., 2021).



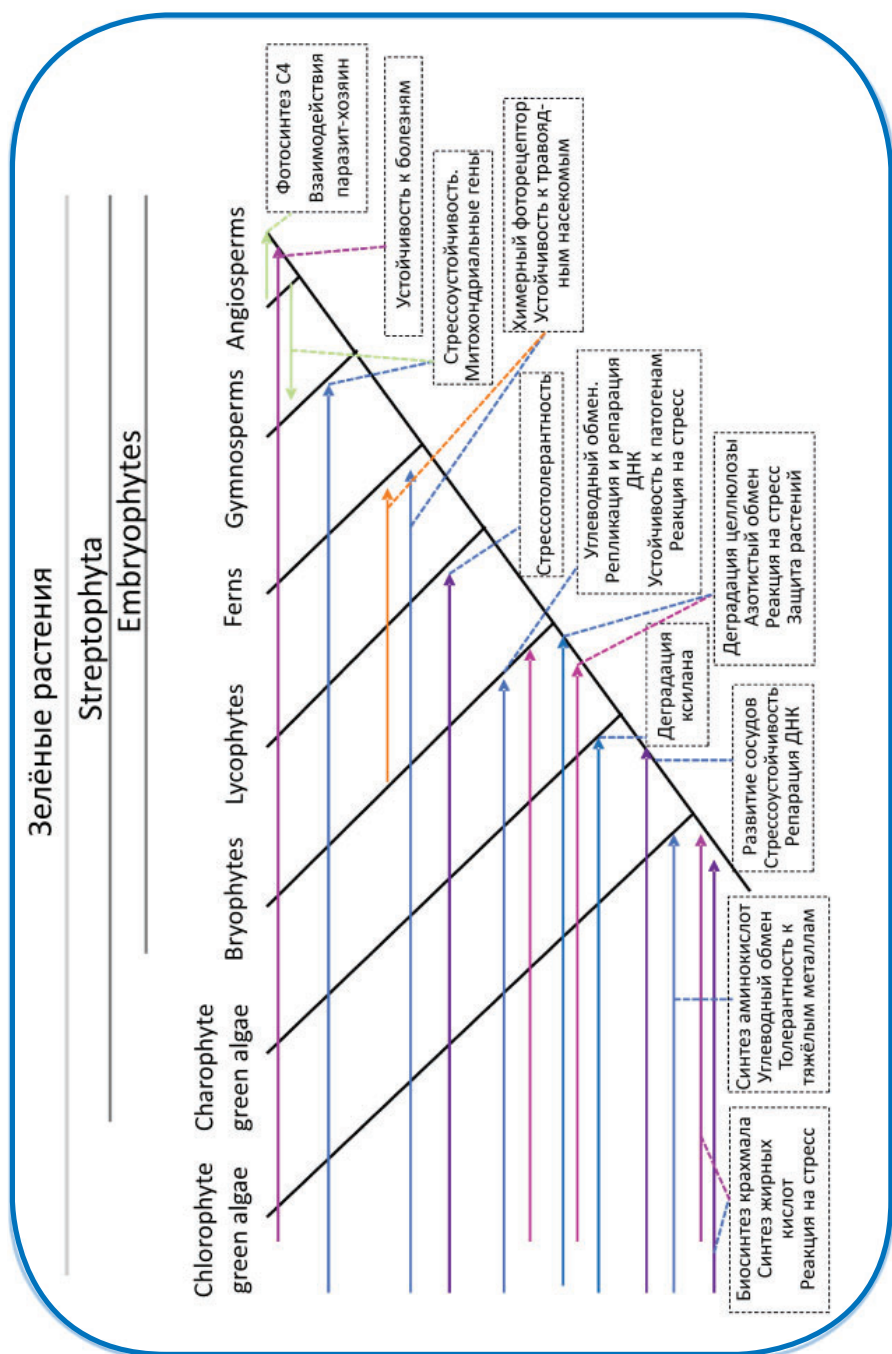
**Рисунок 4.10.** Пути передачи мобильного элемента BovV змей через паразитов (по: Kambayashi et al., 2022).

Толстые и тонкие стрелки показывают направление горизонтального переноса и сходство последовательностей BovV между таксонами соответственно; vnh – BovV negative host.

Благодаря новым геномным технологиям, горизонтальный перенос генов (широко распространённый среди микробных организмов) между многоклеточными животными (считавшийся крайне редким) обнаруживается всё чаще, но его глобальная географическая тенденция и механизмы переноса не изучены в должной мере. Недавно геномный скрининг обнаружил фрагменты мобильного генетического элемента BovV в геномах лягушек Мадагаскара, имеющие высокое сходство (> 94%) с аналогичными последовательностями, выявленными ранее у змей. Это удивительный пример передачи геномного элемента у позвоночных от хищника к жертве (рис. 4.10). Элемент BovV длиной около 3,2 т. п. н. содержится также в геномах жвачных, сумчатых, чешуйчатых, однопроходных и африканских млекопитающих. Интересно, что последовательности BovV являются основным геномным компонентом некоторых из этих видов. При этом, согласно геномным данным, не менее 54 случаев горизонтального переноса BovV, по-видимому, имели место 85–1,3 млн л. н. (Kambayashi et al., 2022). Использование комплексной трансконтинентальной выборки показало, что горизонтальный перенос BovV широко распространён только в одном из исследованных географических регионов, на Мадагаскаре (что подразумевает важные региональные различия в распространённости HGT), где также высока доля BovV-

положительных особей среди переносчиков (паразитов). Примечательно, что в двух линиях лягушек последовательности BovB появились после миграции в «горячую» точку (Мадагаскар) из Африки. Эти результаты дают новое представление о том, как распространённость паразитов может влиять на возникновение событий HGT в регионе (т. е. аналогично ситуации с патогенами и их переносчиками при некоторых эндемических заболеваниях) (Walsh et al., 2013; Kambayashi et al., 2022).

Зелёные растения (Viridiplantae), включающие около 500000 видов, играют огромную роль как в наземных, так и в водных экосистемах и являются источником пищи для большинства живых существ на Земле. Их эволюция представляет собой процесс непрерывных инноваций (морфологических и физиологических), которые позволили растениям адаптироваться к новым условиям. С увеличением геномных данных, доступных в последние годы, был достигнут огромный прогресс в исследованиях HGT у растений, и было обнаружено по крайней мере 200 генов или семейств генов, горизонтально приобретённых гораздо большим числом таксонов, чем предполагалось ранее (Chen et al., 2021b). Внёсший огромный вклад в пластичность генома и адаптивную эволюцию прокариот и некоторых одноклеточных эукариот, HGT значительно способствовал эволюции и адаптации к разнообразным условиям зелёных растений, включая хлорофитные и харофитные зелёные водоросли, мохообразные, ликофиты, папоротники и семенные растения (рис. 4.11). Зелёные растения приобрели множество генов, связанных с реакцией на стресс, а также с устойчивостью к патогенам и травоядным насекомым, толерантностью к тяжёлым металлам, синтезом аминокислот и жирных кислот, репарацией и репликацией ДНК и др. Некоторые гены, приобретённые харофитами, в дальнейшем превратились в функционально важные для экспансии на сушу. Большинство приобретённых посредством горизонтального переноса генов у несеманных растений происходят от бактерий, грибов или вирусов, в то время как у семенных растений часто встречается горизонтальный перенос генов от растения к растению (Chen et al., 2021b). Кроме того, горизонтально приобретённые гены сыграли важную роль в адаптивных инновациях (в углеродном и энергетическом обменах) растений, что можно использовать в селекции и улучшении качества сельскохозяйственных культур. Действительно, некоторые из таких генов продемонстрировали огромный потенциал для улучшения урожая. Например, ген грибов *Fhb7* (кодирует глутатион-S-трансферазу, GST, и придаёт широкую устойчивость к видам *Fusarium* Link, 1809) нашёл применение для селекции устойчивости к фузариозу. Фузариоз колоса – заболевание растений, вызываемое грибами рода *Fusarium*, которые выделяют пищевые токсины и наносят значительный ущерб производству пшеницы (в зародышевой плазме которой было обнаружено мало ресурсов устойчивости) во всём мире (Wang et al., 2020). Некоторые некодирующие молекулы РНК, в том числе малые интерферирующие РНК, играющие ключевую роль во многих биологических процессах, включая реакции на биотический и абиотический стрессы, также могут переноситься горизонтально среди животных, микробов и растений. Дальнейшие исследования механизмов и функций горизонтально приоб-



**Рисунок 4.11.** Упрощённое филогенетическое древо и ключевые функции генов, происходящих в результате НГТ, в основных линиях зелёных растений (по: Chen et al., 2021b).

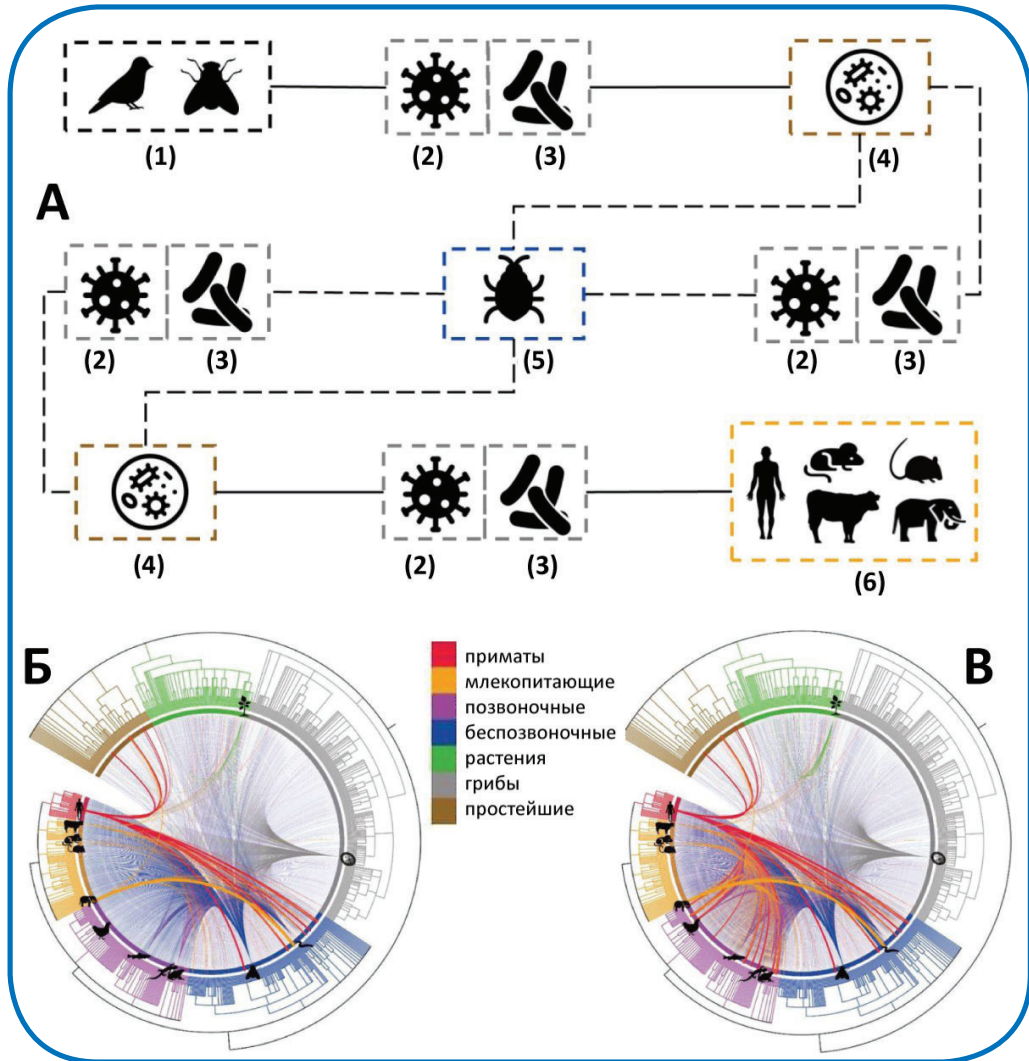
Синие и пурпурные линии указывают на происхождение от бактерий и низших грибов соответственно; оранжевые и зелёные линии обозначают ген, перенесённый от мохообразных и покрытосеменных соответственно. Стрелки указывают линии растений, которые приобрели генетический материал, а пунктирные прямоугольники – на усиление функций для разных линий растений

ретённых РНК способны внести значительный вклад в развитие сельского хозяйства (Wang et al., 2020; Chen et al., 2021b).

Считается, что в среднем 81% генов прокариот оказались вовлечены в HGT. Анализ ~128000 геномов прокариот и вирусов выявил несколько потенциальных путей горизонтального переноса генов с участием кровососущих паразитов, внутриклеточных патогенов и бактерий (рис. 4.12А) (Li et al., 2022b). Чтобы оценить масштабы HGT среди эукариот, недавно было проведено исследование по идентификации областей HGT у 13 модельных видов эукариот путём сравнения их геномов с геномами других 824 эукариотических видов. В геномах первых было обнаружено от 4 до 358 уникальных HGT на вид, многие из которых ранее были неизвестны. Большинство из 824 эукариот с полноразмерными геномными последовательностями также содержали горизонтально перенесённые гены; они трансформировали геномы своих хозяев тысячами копий и повлияли на сотни, и даже тысячи генов. Таким образом, результаты свидетельствуют о широком распространении HGT среди эукариот, являющегося «вездесущим» двигателем эволюции их генома (рис. 4.12Б, В) (Li et al., 2022b).

Широко признано, что у прокариот HGT, опосредованный трансформацией и трансдукцией, является движущей силой эволюции за счёт приобретения адаптивных признаков, таких как устойчивость к антибиотикам, вирулентность и т.п. Как подчёркивает Роман Бениаминович Хесин в своей книге «Непостоянство генома», роль плазмидной миграции настолько велика, что не будь плазмид, не было бы и многих видов бактерий и возможно – человека, «*поскольку мы бы уже отравились продуктами своей мощной химической промышленности, которые бы не обезвреживались бактериями*» (Хесин, 1984). Несмотря на то, что благодаря перемещающимся элементам генофонды всех живых организмов потенциально образуют общий генофонд всего живого мира, частота и важность HGT у эукариот всё же остаются спорными. Причиной во многом является отсутствие ясности его механизмов у эукариот, за исключением переноса фрагментов ДНК вирусными носителями, а также некоторые методологические особенности, не позволяющие надёжно подтвердить наличие этого феномена (Wijayawardena et al., 2013; Chen et al., 2021b).

Общепринято считать, что приобретение ТЕ *de novo* подтверждается, если гомология нуклеотидных последовательностей в геномах неродственных организмов составляет более 90%. Также считается, что для HGT не требуется долговременной физической связи между видами, но физический контакт может увеличить вероятность HGT (1); смена хозяев может затруднить идентификацию события HGT, ассоциированного с паразитизмом (2); обнаружение ТЕ часто интерпретируют как HGT, однако необходимо проверять, не существует ли первичного механизма, обусловившего его распространение (3); наконец, HGT нередко подвергается сомнению из-за потенциально возможной контаминации, и эту проблему следует рассматривать с должным вниманием (4). Дальнейшие исследования горизонтального переноса генов, прежде всего с помощью геномного секвенирования с длинными прочтениями, могут существенно пополнить наши представления о геномной коэволюции и, в частности, дать



**Рисунок 4.12.** Горизонтальный перенос генов (HGT) (из: Li et al., 2022b).

А – предполагаемый путь горизонтального переноса генов между млекопитающими и дальними родственниками эукариот. 1 – филогенетически далёкие виды эукариот, такие как *Drosophila willistoni* и *Serinus canaria*; 2 – вирусный генофонд; 3 – бактериальный генофонд; 4 – внутриклеточные паразиты, такие как *Toxoplasma gondii* ME49; 5 – кровососущие паразиты, такие как *Cimex lectularius*; и 6 – млекопитающие, включая *Homo sapiens*, *Bos taurus*, *Loxodonta Africana*, *Mus musculus*, *Rattus norvegicus*. Сплошные линии обозначают хорошо подтвержденные события HGT, а пунктирные линии – непроверенную гипотезу.

Б, В – HGT у эукариот. 824 вида эукариот были сгруппированы в семь подгрупп: приматы, млекопитающие не приматы, не млекопитающие позвоночные, беспозвоночные, простейшие, грибы и растения

Межтиповые (Б) и межклассовые (В) HGT показаны линиями, соединяющими родственные виды, а толщина линии соответствует числу HGT.

эволюционное понимание эпидемиологии паразитарных инфекций (Wijayawardena et al., 2013; Челомина, 2017).

#### 4.5. Эпистаз, плейотропия и гетерозис

Высокопроизводительная геномика открывает новые перспективы в широко-масштабном изучении разнообразия взаимодействий между геномными локусами и аллелями. В настоящее время рутинной стратегией расшифровки ассоциации генотип-фенотип является полногеномное исследование ассоциаций (GWAS), т.е. исследование связи генотипа с различными фенотипическими признаками в масштабе всего генома.

**Эпистаз** представляет собой такой тип взаимодействия генов, при котором на экспрессию одного гена влияет другой (неаллельный) ген, и, таким образом, мутация одного гена может маскировать фенотипические эффекты мутации в другом локусе (Phillips, 2008; Roth et al., 2009). В глобальном масштабе паттерны взаимодействия генов имеют отношение к системной биологии. Эпистаз *«признан фундаментально важным для понимания как структуры и функции генетических путей, так и эволюционной динамики сложных генетических систем»* (Phillips, 2008). Несмотря на важную роль эпистаза, его обнаружение представляет собой сложную задачу, хотя искусственные популяции со сбалансированными частотами аллелей упрощают её и делают более выполнимой (Liu, Yan, 2019). Исследования показывают, что дисперсия эпистаза может оказывать значительное влияние на генетическую изменчивость роста и экофизиологических признаков, и существует вероятность проявления эпистатического эффекта для количественных признаков. Однако его точная оценка иногда затруднена из-за отсутствия подходящих моделей, или слишком низкой дисперсии эпистаза (Garel, Philippe, 2022). С помощью масштабного компьютерного моделирования было обнаружено, что полигенный параллелизм согласуется с положительным синергетическим эпистазом между аллелями сильнее, чем другие механизмы (Stern et al., 2022). Также имеются свидетельства, что гибридная стерильность возникает в результате множественных сложных эпистатических взаимодействий, часто с участием половой хромосомы (Rosser et al., 2022).

**Плейотропия**, или явление множественного действия гена, описывается состоянием, когда один ген или одна мутация влияют на два или более различных фенотипических признака. Плейотропия была подробно исследована для некоторых генов, однако её геномные паттерны остаются в значительной степени неизвестными (Wang et al., 2010; Brown, Lad, 2019). Учитывая роль плейотропии в развитии сложности, полногеномное изучение этого явления представляется важным для понимания функций генов (особенно потому, что функции большинства генов остаются неизвестными – так называемый «тёмный геном»), роли генов в заболеваниях, а также в разработке новых терапевтических подходов. Таким образом, плейотропия является фундаментальной областью генетики человека, имеющей решающее значение для будущего развития геномной медицины (Brown, Lad, 2019). Анализ полногеномных

последовательностей из базы данных геномного проекта фонда муковисцидоза (содержит информацию о WGS от более чем 5000 человек с муковисцидозом в США) обнаружил плеiotропные модификаторы возрастного диабета и неонатальной кишечной непроходимости (заболевания со слабо коррелирующими признаками) при муковисцидозе (Aksit et al., 2022). Полногеномные исследования (выборка из 88250 пациентов с возрастной дегенерацией жёлтого пятна, диабетической ретинопатией, глаукомой, отслоением сетчатки и миопатией) выявили 23 плеiotропных локуса, связанных как минимум с двумя глазными заболеваниями (Xue et al., 2022). Исследование плеiotропии может также иметь большое значение для развития сельского хозяйства. С помощью полногеномных ассоциативных исследований и анализа однонуклеотидных полиморфизмов (5,6 миллионов предполагаемых вариаций последовательностей) идентифицированы плеiotропные варианты и гены–кандидаты на признаки фертильности и воспроизводства у крупного рогатого скота голштинской породы (Chen et al., 2022c). Исследование истории селекции пшеницы с помощью GWAS указало на синергетический отбор плеiotропных геномных участков для архитектуры растений и урожайности зерна (Li et al., 2022a).

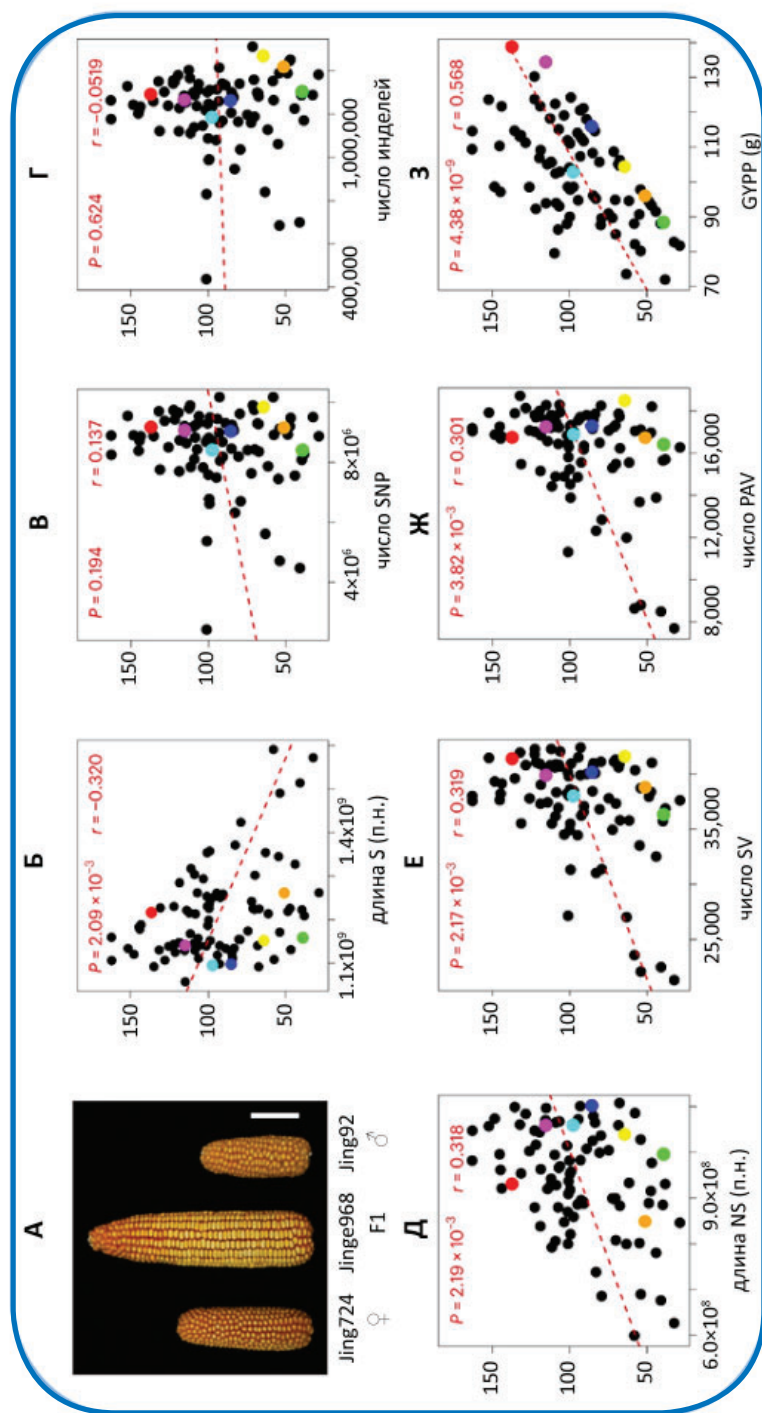
*Гетерозис*, известный как превосходство гибридов по сравнению с родителями, или «гибридная сила», изучен значительно лучше (Yang et al., 2017; Chen et al., 2019; Liu et al., 2022; Wang et al., 2023a; Saxena et al., 2024). Эффект зависит от степени родства родительских особей (хотя эта зависимость нелинейна) и может наблюдаться только у гибридов F1 на протяжении всего или почти всего жизненного цикла. В последующих потомствах, когда некоторые области генома становятся гомозиготными, он «затухает». Явление гетерозиса было известно ещё до открытия законов Грегора Менделя. Позднее шведский ботаник Оке Густафссон пришёл к выводу, что у растений могут быть выделены три формы гетерозиса: репродуктивная (повышается плодородность и урожайность гибридов), соматическая (увеличиваются линейные размеры и масса гибридов) и адаптивная (повышается приспособленность гибридов к неблагоприятным факторам) (Gustafsson, 1946).

Являясь одним из наиболее значимых феноменов биологии, гетерозис эффективно используется в сельскохозяйственной практике, хотя его генетические причины до последнего времени оставались неясными из-за генетической и фенотипической сложности гетерозиса и недостатка консенсуса относительно механизмов, лежащих в его основе (Yang et al., 2017; Chen et al., 2019; Liu et al., 2022). Для объяснения гетерозиса с точки зрения взаимодействия внутри гибридного генома были предложены три основные гипотезы: доминирования (возникновение гибридной силы из-за полногеномного маскирующего эффекта вредных рецессивных генов одного родителя доминантными аллелями другого родителя), сверхдоминирования (результат взаимодействия аллелей родителей) и эпистаза (взаимодействие между аллелями в разных локусах генов от двух родителей, приводящее к гибридной силе) (Liu et al., 2022). Результаты многочисленных исследований свидетельствуют, что *«гетерозис должен рассматриваться как следствие сложных взаимодействий, которые при-*

водят к комплексу изменений на уровне генных, эпигенетических, биохимических и регуляторных сетей», а гетеротическим ответом в гибридах F1 совместно управляют три типа генного действия: аддитивность, доминирование и эпистаз. При этом «омические» исследования способствуют выявлению ключевых механизмов регуляции гетерозиса и определению перспективных направлений его использования в сельском хозяйстве (Хотылева и др., 2016; Шаптуренко, Хотылева, 2016).

Молекулярные механизмы гетерозиса изучаются в двух основных направлениях: вариации последовательностей нуклеотидов в геноме и изменения экспрессии генов (Sun et al., 2023b). Высококачественные сборки геномов облегчили идентификацию различий родительских геномов у гибридов, включая вариации числа копий (CNV) и отсутствия/присутствия (PAV), однонуклеотидные полиморфизмы, индели, а также их роли в формировании гетерозиса. Например, PAVs составляют важную часть геномной изменчивости среди различных инбредных линий кукурузы и могут способствовать гетерозису за счёт функциональной или экспрессионной комплементации включённых в них генов (Liu et al., 2022). Тем не менее, консенсусного принципа связи между степенью дивергенции родительских геномов, оцениваемой по SNP, и уровнем гетерозиса у гибридов не существует, и не все гетерозиготные области в гибридном геноме играют важную роль в гетеротическом фенотипе. В то же время, аллельные варианты в некоторых ключевых локусах могут способствовать гетерозису по определённым признакам конкретных гибридов (Liu et al., 2022). Исследования молекулярных основ гетерозиса у кукурузы (Wang et al., 2023a) впервые показали, что длина синтенных областей имеет значительную отрицательную, а несинтенных областей – существенную положительную корреляцию с гетерозисом лучшего родителя по урожайности зерна с растения. Было также обнаружено, что количество структурных вариантов (SV) и PAV между родительскими линиями сильно коррелировали с уровнем гетерозиса, тогда как однонуклеотидные полиморфизмы и индели связаны с гетерозисом слабо (рис. 4.13). Кроме того, были получены убедительные данные, подтверждающие вклад как доминирования, так и сверхдоминирования в формирование гетерозиса у кукурузы (Wang et al., 2023a).

Геномные подходы обеспечили эффективное исследование многих локусов, лежащих в основе преимуществ гибридного риса (для 1495 гибридных сортов риса идентифицировало 130 локусов, связанных с гетерозисом), выявили геномную архитектуру гетерозиса риса и обнаружили ряд генетических вариантов, важных для дальнейшего улучшения селекции. Показано, что индели в промоторных областях могут приводить к появлению полиморфных цис-регуляторных элементов, которые связаны с экспрессией гетеротических генов у гибридов через взаимодействия с активаторами/репрессорами транскрипции (Chen et al., 2019; Liu et al., 2022). Создана полная нуклеотидная карта количественных признаков (QTN) риса и обобщены наиболее полные ключевые функциональные вариации генов количественных признаков и их роль в гетерозисе (Wei et al., 2021). Обнаружено перепрограммирование метилирования ДНК в гибридной сое, связанное с фенотипической изменчивостью



**Рисунок.4.13.** Идентификация связи между гетерозисом и генетическими вариантами родителей при диаллельном скрещивании кукурузы (из: Wang et al., 2023a).

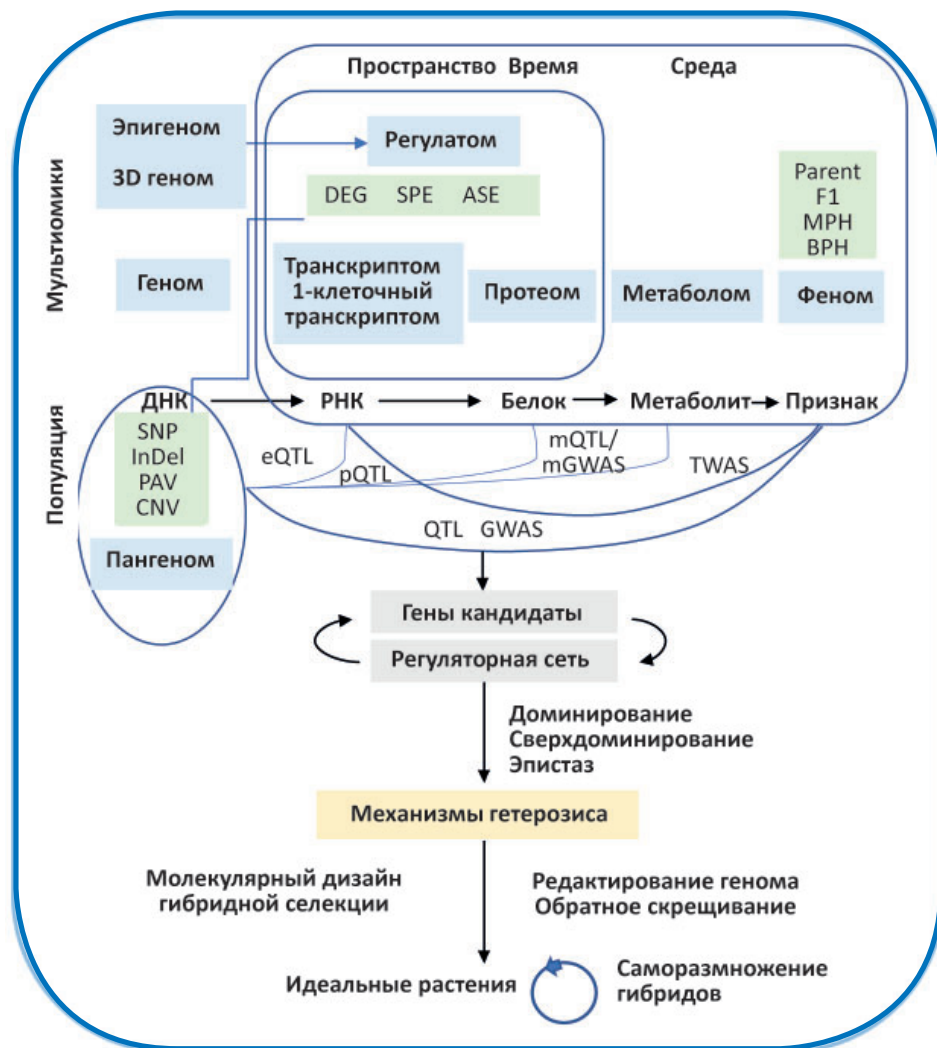
А – морфология початков родителей и их гибрида F1. Б–Ж – анализ корреляции между гетерозисом лучшего родителя (ВРН) по урожайности зерна с растения (GYPP) (%; ось Y) и различными типами генетических вариантов между родителями (ось X); З – ВРН по GYPP существенно коррелирует с соответствующим признаком у F1. S – синтенные области, NS – несинтенные области, SNP – однонуклеотидный полиморфизм, SV – структурные варианты, RAV – варианты присутствия/отсутствия, п.н. – пар нуклеотидов. Значения P и коэффициенты ранговой корреляции Спирмена (r) показаны в верхней части каждого графика. Семь зарегистрированных сортов в Китае или США показаны цветными точками.

и инициацией гетерозиса. Сайты неаддитивного дифференциального метилирования генома идентифицированы в незрелых гибридных семенах и обнаружены в большом количестве у гибридов. Причём по сравнению с нормальным распределением, большинство метилированных сайтов были более сконцентрированы в областях генов, участвующих в различных биологических процессах, и особенно – в регуляции транскрипции и гормональной функции (Liu et al., 2022). Примечательно, что активности некоторых эпигенетически модифицированных генов могут вызывать каскадные эффекты в генных сетях, способствуя развитию гетерозиса (Шаптуренко, Хотылева, 2016).

Изменения экспрессии генов в масштабе всего генома являются связующим звеном между генетической/эпигенетической изменчивостью и лучшими свойствами гибридов по сравнению с их родителями по многим биологическим признакам (Liu et al., 2022; Sun et al., 2023b). В последнее время появилось много свидетельств комплементации экспрессии генов у гибридов, например, при гетерозисе у кукурузы (обнаружена с помощью исследований транскриптома) (Liu et al., 2022). Выявлены различия между гибридом и родителями в регуляторных сетях генов (GRN; описывают взаимодействия и регуляторные отношения между генами) у *Arabidopsis*. Идентификация «генов-концентраторов» сети (hub genes, взаимодействуют со многими другими генами и обычно играют решающую роль в биологических процессах и регуляции генов) помогла выявить основные биологические пути формирования гетерозиса (Liu et al., 2021, 2022). Показано, что многие регуляторные сети подчинены действию циркадных часов (т.е., координируют функциональные процессы в течение суточного цикла). Изменение пиков экспрессии у циркадных генов может влиять на такие важные биологические механизмы, как стрессоустойчивость, регулирование иммунитета растений, работа генов сигнальных путей фитогормонов и т.п. (Шаптуренко, Хотылева, 2016; Liu et al., 2022). Дальнейшее развитие геномных технологий с использованием мультиомик и популяционных подходов открывает большие возможности для более полного понимания генетических и молекулярных механизмов, лежащих в основе гетерозиса (рис. 4.14).

#### 4.6. Гибридизация и гибридные зоны

Гибридизация представляет собой вполне обычное явление, широко распространённое на древе жизни, с большим диапазоном расхождений в частоте между таксонами. Она охватывает разные временные масштабы, и есть убедительные свидетельства того, что гибриды существовали миллионы лет назад (Moran et al., 2021). Александр Сергеевич Серебровский рассматривал естественную гибридизацию как скрещивание между особями, принадлежащими разным видам. Ботаники вавилонской школы выделяли микро- и макро- (отдалённую) гибридизацию с промежуточными вариантами между ними. В современной литературе даются очень разные определения, от гибридизации между генетически несходными особями одной популяции до гибридизации между видами, в том числе принадлежащими к разным



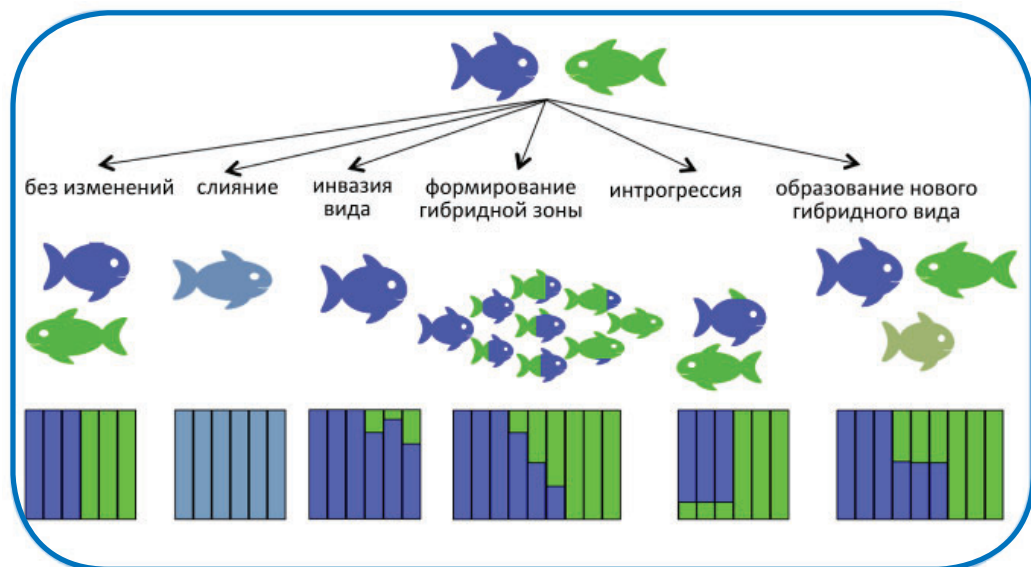
**Рисунок 4.14.** Перспективы стратегий мультиомики и популяционной генетики в научных исследованиях и в практическом использовании гетерозиса (по: Liu et al., 2022).

ASE (аллель-специфическая экспрессия) количественно определяет относительную экспрессию двух аллелей; TWAS (полнотранскриптомное исследование ассоциаций) выявляет значимые ассоциации генов и признаков с упором на моделирование регуляций на уровне транскрипции; eQTL (локусы экспрессии количественных признаков) – геномная область, содержащая варианты последовательностей ДНК, влияющие на уровень экспрессии одного или нескольких генов; объясняет различия в уровнях экспрессии мРНК; pQTL (локусы количественных признаков белка) влияют на количество конкретного белка; mGWAS (полногеномные исследования метаболических ассоциаций) опирается на фенотипическую информацию, основанную на накоплении метаболитов, и может идентифицировать гены, поддерживающие содержание первичных и вторичных метаболитов; mQTL (локусы количественных признаков метаболома) способствуют более широкому пониманию генетических регуляторных сетей.

таксонам более высокого ранга. Согласно Международному кодексу зоологической номенклатуры, гибрид – потомок двух особей, относящихся к разным таксонам (Боркин, Литвинчук, 2013).

У человека около 2–5% генома произошло в результате древней гибридизации с неандертальцами и денисовцами, а у других таксонов, таких как некоторые виды бабочек, воробьёв, цихлид и др., более 10% геномов образовались в результате древней гибридизации, при этом частично такой генетический обмен отражает процесс адаптивной интрогрессии (Moran et al., 2021). В целом, гибридизация чаще встречается у растений, где она отмечена у 25% видов, тогда как у животных она обнаружена только у 10% видов (Runemark et al., 2019). Открытие и понимание роли гибридизации, когда в одном организме происходит объединение генов, принадлежащих разным таксонам, привело к огромному успеху в изучении проблем эволюции и видообразования. В отличие от обычного, такой тип видообразования называют сетчатым (ретикулярным), или гибридогенным, когда родственные отношения между таксонами могут быть представлены в виде сети, а не дерева. Эволюционная значимость естественной гибридизации чаще всего обсуждается в контексте филогенетического распространения и адаптивной природы, а одним из ключевых вопросов является её влияние на специфические фенотипы. Формулируя эволюционные последствия естественной гибридизации, Майкл Арнольд (Arnold, 1992) отмечает, что, как крайние варианты, возможны либо слияние гибридизующих форм, либо усиление репродуктивных барьеров за счёт ассортативного скрещивания (1); возможно получение относительно приспособленных генотипов, позволяющих интрогрессивным формам распространяться в новые места обитания (2); гибридные особи могут оказаться более привлекательными для паразитов, что ограничивает их адаптацию к родительским особям (3); и, наконец, естественная гибридизация может приводить к формированию гибридных видов, включая партеногенез (партеногенетические виды/расы часто являются полиплоидами, возникшими в результате отдалённой гибридизации), гибридогенез или гомоплоидное (диплоидное) видообразование (4). Таким образом, генетический обмен между видами может рассматриваться как весомый фактор увеличения разнообразия с внедрением адаптивных вариантов через границы видов, включая быстрое видообразование, либо как угроза природному биоразнообразию, вплоть до поглощения одного вида другим (рис. 4.15).

Существует много подходов к выявлению гибридов, оптимальных к определённым моделям исследований. Для близкородственных таксонов рекомендуются подходы геномной кластеризации в таких программных обеспечениях, как STRUCTURE (Pritchard et al., 2000) и ADMIXTURE (Alexander et al., 2009). Они определяют количество генетических групп анализируемого набора данных и место в них для каждой особи. Например, ADMIXTURE использовали при изучении гибридной зоны между домашним и португальским воробьями в Италии (Elgvin et al., 2017). Для определения интрогрессивных областей генома, направления и количественной оценки потока генов используют филогенетические методы (Pease, Hahn, 2015). Недавно с



**Рисунок 4.15.** Потенциально возможные эволюционные результаты гибридизации (из: Runemark et al., 2019).

помощью этого подхода были получены убедительные доказательства гибридного происхождения азиатских чёрных медведей (*Ursus thibetanus* G. Cuvier, 1823) (Zou et al., 2022), а также обнаружено древнее гибридное происхождение группы макак (Macaca Lacépède, 1799) (Zhang et al., 2023a). Для набора данных с большим количеством таксонов хорошо подходят методы построения графов, дающие оценки дрейфа и интрогрессии; из-за межпопуляционного обмена генами простые дихотомические деревья могут давать неверное представление об истории популяций (Pickrell, Pritchard, 2012). Метод моделирования графов успешно применялся для определения событий генетической примеси в древних и современных популяциях тигра (*Panthera tigris* (Linnaeus, 1758)) (Sun et al., 2023a). По отношению к небольшим наборам таксонов подходит демографическое моделирование (Pinho, Heu, 2010). С помощью такого подхода было подтверждено гибридное происхождение перуанского морского котика (*Arctocephalus australis* (Zimmermann, 1783), перуанские и северо-чилийские популяции) (Lopes et al., 2023).

Среди других методов, часто используется статистика Паттерсона (Patterson et al., 2012; Martin et al., 2015), широко известная как тест АВВА-ВАВА, которая позволяет сделать вывод об интрогрессии, несмотря на неполную сортировку линий (ILS). Однако устойчивая в широком диапазоне генетических расстояний, эта статистика может быть чувствительна к размеру популяции. Программа NewHybrids (Anderson, Thompson 2002), дающая оценки апостериорной вероятности попадания генетически отобранных особей в определённую категорию, эффективна в обнаружении недавней гибридизации. HyDe – это пакет программного обеспечения, кото-

рый выявляет гибридизацию в наборах филогеномных данных (Blischak et al., 2018). Сравнение методов показало, что оценки для разных пропорций родительского вклада с помощью HyDe являются исключительно высокоточными и надёжными, тогда как STRUCTURE и ADMIXTURE не всегда позволяют идентифицировать гибриды, особенно если родительские вклады асимметричны. HyDe и статистика Паттерсона эффективны для обнаружения гибридизации во всех сценариях, кроме случаев с высоким показателем ILS (Kong, Kubatko, 2021).

Для анализа гибридных зон выделяют три типа современных категорий геномных данных (Gompert et al., 2017): (1) большой (от тысяч до сотен тысяч) набор однонуклеотидных полиморфизмов, который анализируется без эталонного генома, или с использованием генома низкого качества; в эту категорию входит секвенирование с сокращённым представлением; (2) большие наборы SNP, которые были сопоставлены с эталонным геномом; и (3) данные ресеквенирования всего генома, приведённые в соответствие с высококачественным эталонным геномом. Использование разных категорий данных позволяет точно оценить не только такие ключевые показатели, как гибридные индексы, пропорции примесей, среднюю степень интрогрессии, но также границы предковых блоков и распределение сигналов интрогрессии по геному, и такую дополнительную информацию, как скорости рекомбинации и модели генов (Gompert et al., 2017).

**Формирование гибридного генома: аллополиплоиды и гомоплоиды.** Аллополиплоидию, когда гибридные геномы имеют по одному полному набору хромосом от каждого родителя, и гомоплоидию, при которой гибридные геномы представляют мозаику геномов родительских видов без увеличения числа хромосом, признают основными вариантами геномов гибридных видов (Runemark et al., 2019). Доля интрогрессивного материала, унаследованного гибридами, сильно отличается между видами, составляя 50% во многих полиплоидных таксонах после первоначального события гибридизации. У некоторых гомоплоидных гибридов также обнаружен равный вклад родительских геномов, но в других случаях он крайне неодинаков и у некоторых видов (например, у рыб-меченосцев (*Xiphophorus* Heckel, 1848) и итальянских воробьёв (*Passer italiae* Vieillot, 1817)) существуют популяции с очень разными пропорциями родительских геномов (Runemark et al., 2018, 2019; Elgvin et al., 2017).

Критериями гомоплоидного гибридного видообразования считаются (1) наличие репродуктивной изоляции между двумя родительскими таксонами, (2) свидетельства генетической смеси двух родительских линий, а также (3) доказательства возникновения репродуктивных барьеров в результате гибридизации (Wang et al., 2022c; Vock et al., 2023). От интрогрессии такой тип видообразования отличается тем, что он приводит к стабильной линии как отдельной таксономической единицы. Гомоплоидия является важным механизмом образования новых видов и всё чаще признаётся распространённым явлением как у растений, так и у животных. Вместе с тем, эволюционные последствия гомоплоидной гибридизации часто остаются малоизученными, и не всегда удаётся подтвердить гомоплоидное видообразование. Например, ге-

номные исследования продемонстрировали гибридное происхождение оксфордской амброзии (крестовик весенний или крестовник оксфордский) *Senecio squalidus* L., медуницы *Pulmonaria helvetica* Bolliger, двух из трёх хорошо изученных гомоплоидных гибридных видов подсолнечника (*Helianthus annuus* L., 1753), но не смогли подтвердить предполагаемого гомоплоидного видообразования у ириса *Iris hexagona* Walter и флоксов *Phlox* L. (1753) (Bock et al., 2023). Геномное секвенирование и популяционный анализ выявили также «тупиковую», т.е. без инициации видообразования, гомоплоидную гибридизацию у актинидий *Actinidia* Lindl., 1836. Результаты показали, что *A. zhejiangensis* C. F. Liang в основном представляют собой гибриды F1 между *A. hemsleyana* Dunn и *A. eriantha* Benth., а межвидовой поток генов существует около 1 млн л. н., что подразумевает как сильные генетические барьеры, так и продолжающуюся гибридизацию (Yu et al., 2023b). Роль гомоплоидного видообразования у животных широко обсуждается, но подтверждённых геномными данными случаев немного. Например, секвенирование геномов с сокращённым представлением обнаружило, что перуанский морской котик является видом со смешанным геномом, возникшим около 400000 л. н. в результате гибридизации двух видов морских котиков, Южноамериканского *Arctocephalus australis* (Zimmermann, 1783) и галапагосского *A. galapagoensis* Heller, 1904, путём гомоплоидного гибридного видообразования (Lopes et al., 2023). Этот же тип видообразования выявлен при ресеквенировании геномов 44 особей родов *Carpinus* L., 1753 и *Ostrya* Scop. Исследование подчёркивает вероятность событий гомоплоидного гибридного видообразования между предками современных родов на начальных этапах их расхождения, что, вероятно, привело к сетчатой филогении на более высоких таксономических уровнях (Wang et al., 2022c).

Аллополиплоидное видообразование чаще встречается у растений, чем у животных, при этом из-за различий в числе хромосом аллополиплоиды могут быть сразу изолированы от родительских видов, а для изучения ранних стадий аллополиплоидизации часто используются синтетические аллополиплоиды растений. Как и при аутополиплоидии, аллополиплоидный геном после дубликации подвергается диплоидизации, в результате которой теряется большая часть последовательностей (преимущественно некодирующая ДНК, в основном – тандемные повторы) (Runemark et al., 2019). В кодирующей части генома крупные реорганизации недопустимы, а со временем каждый ген помимо консервации (сохранения функций) имеет возможность реализовать несколько эволюционных сценариев, включая два варианта функциональной дивергенции: субфункционализацию, когда каждый паралог сохраняет определённую часть функции родительского гена (как следствие – расширение адаптивного потенциала) и неофункционализацию – приобретение принципиально новых функций, а также псевдогенезацию, т.е. полную утрату функций и превращение в псевдоген, часто из-за эпигенетических изменений генома (Chen, Ni, 2006; Chaudhary et al., 2009; Akhunov et al., 2013; Yoo et al., 2013). Скорость структурно-функциональной дивергенции значительно варьирует между видами, но в целом субфункционализацию рассматривают как наиболее раннее событие, а со временем возрастает вероятность неофункционализации и псевдогенезации (Adams, Wendel, 2005). Согласно

геномным данным, у гексаплоидной пшеницы (вероятный результат перекрёстного опыления *Triticum aestivum* L. (1753) дикими травами) псевдогенизации подверглись 38% экспрессирующихся генов (Akhunov et al., 2013). Для сравнения, у резуховидки Таля *Arabidopsis thaliana* (L.) Heynh., 1842, Brassicaceae), долгое время считавшейся классическим диплоидом (является полиплоидом, сохранившим только 30% дублированных генов после последнего события дубликации примерно 86 млн л. н.) (Bowers et al., 2003), около 62% пар генов подверглись функциональной диверсификации (Blanc, Wolfe, 2004). Генетические исследования аллотетраплоида *Tragopogon mirus* Ownbey семейства астровых (Asteraceae) показали, что субфункционализация ряда генов произошла всего за 40 поколений (Buggs et al., 2010). Примечательно, что потеря дубликатов не является случайной, и гены, сохранившиеся после одного цикла дубликации, имеют большую вероятность сохраниться при последующем цикле дубликации (Seoighe, Gehring, 2004).

Дубликация генома, связанная с аллополиплоидией, позволяет наборам хромосом от разных видов сохраняться как отдельные «субгеномы» во время последующих мейозов, а хромосомное распределение и время вставок транспозонов служат историческим индикатором хромосом. Новые разработанные на основе геномных данных подходы позволяют надёжно идентифицировать субгеномы у аллоплоидов растений и животных, а также определять порядок гибридизации диплоидов–предшественников с организмами более высокой ploidy (Session, Rokhsar, 2023). Считается, что в более высоком полиплоиде (который может возникнуть в результате гибридизации предшественников различной ploidy) субгеном, добавленный последним, должен (1) обладать более высоким сохранением генов (смешённое фракционирование), поскольку другой полиплоидный предшественник уже утратит избыточные гены, и (2) более устойчивую экспрессию генов (доминирование генома), поскольку у другого, полиплоидного предшественника, экспрессия оставшихся избыточных генов будет ниже, чем у диплоидного (Session, Rokhsar, 2023). Субгеномы могут эволюционировать ассиметрично с образованием «доминантных» и «подчинённых» субгеномов, причём последние имеют более высокую плотность мобильных элементов (Runemark et al., 2019; Session, Rokhsar, 2023). У октоплоидной земляники один из четырёх геномов является доминантным и его гены экспрессируются чаще. Кроме того, определённые черты этого растения, как например устойчивость к болезням, контролируются доминантным субгеномом (Edger et al., 2019).

Аллополиплоидные и гомоплоидные гибриды часто имеют нарушение регуляции фертильности. Аллополиплоидия влияет на прохождение хромосомных спариваний и поэтому может вызвать бесплодие из-за дефектных гамет у гибридов (Bashir et al., 2018). Интересный пример (без получения полногеномных данных) был недавно описан для водяных лягушек рода *Pelophylax* Fitzinger, 1843, известным признаком которых является гибридизация. Обитающие в долине р. Днестр лягушки комплекса *P. esculentus* (Linnaeus, 1758) состоят из двух родительских видов, прудовой *P. lessae* (Camerano, 1882) и озёрной *P. ridibundus* (Pallas, 1771), а также их гемиклонального

гибрида *P. esculentus*, и все они являются диплоидами. Оказалось, что в популяциях *P. esculentus* численность самцов сильно (90%) превышала численность самок, половина изученных гибридных самцов была стерильной, а большинство фертильных гибридных самцов продуцировали сперму с геномом *P. lessae*. Более того, несколько гибридных самцов дали смесь сперматозоидов с геномами *P. lessae* и *P. ridibundus* (амфиспермия) и только один самец – с геномом *P. ridibundus* (Litvinchuk et al., 2023).

В гибридном геноме одновременно взаимодействуют многие факторы, поэтому его эволюция необычайно динамична. Наблюдения показывают, что в целом у гибридов действует отбор против второстепенного родителя, т.е. от которого гибриды получают меньшую часть своего генома в наиболее функционально важных геномных областях (Moran et al., 2021). Гибридная дисфункция, известная как «несовместимость Добржанского–Мюллера», часто возникает в результате эпистатических взаимодействий между генами, когда ранние комбинации аллелей в разных локусах объединяются и взаимодействуют, снижая приспособленность. Согласно этой модели, локусы, полученные от второстепенного родителя, с большей вероятностью обнаружат несовместимости в других частях генома, что приводит к глобальному отбору против происхождения от второстепенного родителя (Moran et al., 2021; Rosser et al., 2022).

**Изменения гибридного генома: геномный и транскриптомный «шок».** Гибридизация, при которой происходит слияние дивергировавших геномов с различными адаптациями и несовместимыми комбинациями генов, инициирует цепь событий, которые Барбара МакКлинток назвала «геномным шоком» (McClintock, 1984). Согласно гипотезе геномного шока, к последствиям гибридизации относится увеличение активности мобильных генетических элементов, нарушение эпигенетического ландшафта, вариации в структуре хромосом и повышенная нуклеотидная вариабельность, изменение числа копий генов и количества повторяющейся ДНК, а также трансформация профилей экспрессии генома. Из них, в качестве основной причины дестабилизации гибридного генома рассматривают нарушение эпигенетического контроля, сопровождающегося активацией мобильных элементов (Shapiro, 2023).

Нестабильность генома у гибридов за счёт всплесков транспозиций продемонстрирована на многих видах. Например, показано, что у дрозофил (*Drosophila* Fallén, 1823) гибридизация реактивирует мобильные элементы, которые эффективно подавлялись в геномах родительских видов (Guerreiro, 2014). При этом нарушается эпигенетический контроль, управляемый пиРНК, которые обычно подавляют передвижение мобильных элементов в зародышевой линии (Gamez-Visairas et al., 2020). Исследования геномов лягушек рода *Xenopus* Wagler, 1827 также позволили предположить, что мобильные элементы (ДНК-транспозоны) и их пиРНК могут сильно влиять на эволюцию аллотетраплоидного генома после межвидовой гибридизации (Suda et al., 2022). Данные секвенирования генома дрожжей с использованием длинных ДНК-прочтений показали, что большинство ретротранспозонов с длинными концами (LTR-RT) были вставлены в геном после гибридизации, и их количество широко

варьирует (от 0 до 87), увеличиваясь с повышением уровня геномной примеси (Tusso et al., 2022). Во вновь синтезированных алло- и аутотетраплоидных геномах картофеля (*Solanum tuberosum* L., 1753) копии LTR-RT активировались преимущественно в гибриде (29 копий), в аллотетраплоидных линиях активации не наблюдалось, а в аутотетраплоидных линиях активировалась только одна копия (Gantuz et al., 2022). Вместе с тем, есть данные, что после аллополиплоидизации пролиферация может быть ограничена всего несколькими мобильными элементами, а в некоторых случаях она вообще не обнаруживается. Таким образом, геномный ответ на полиплоидизацию является высокоспецифичным, причём геномный дисбаланс мобильных элементов может влиять на взаимодействия субгеномов и их эпигенетический контроль (Gantuz et al., 2022). Показательно, что у межродовых гибридов (встречаются значительно реже межвидовых) некоторых рыб вблизи мобильных элементов располагались сайты, подвергшиеся интрогрессии (Nedoluzhko et al., 2022).

Эпигенетические модификации, прежде всего метилирование ДНК, являются важным фактором регуляции многих клеточных функций, включая фертильность. Среди межвидовых гибридов морских гребешков – *Argopecten purpuratus* (Lamarck, 1819) и гермафродитного *A. irradians* (Lamarck, 1819) – более высокие уровни метилирования были выявлены у стерильных особей по сравнению с фертильными, особенно в одной из хромосом отцовского родителя. Всего было обнаружено 3619 дифференциально метилированных генов, которые, как полагают авторы, могут вызывать гибридную стерильность различными путями, например, через ингибирование оогенеза и созревания яиц, а также индукцию апоптоза (Yu et al., 2023a). У тихоокеанской устрицы *Crassostrea gigas* Thunberg, 1793 общий уровень метилирования ДНК был значительно ниже в гибридных скрещиваниях, чем в родительских, и отрицательно коррелировал с ростом (Yang, Li, 2022); у гибридных форм сои (*Glycine max* (L.) Merr., 1917) перепрограммирование метилирования ДНК оказалось связанным с фенотипической изменчивостью и гетерозисом (Chen et al., 2022b), а сбалансированная геномная диверсификация у аллотетраплоидов резуховидки *Arabidopsis Heynh.* (1842) сопровождалась конвергентными и согласованными изменениями в метилировании ДНК между двумя субгеномами (Jiang et al., 2021).

В гибридных геномах часто возникают структурные варианты (SV), которые могут непосредственно вызывать репродуктивную изоляцию с последующим видообразованием. Например, инверсии могут способствовать видообразованию за счёт уменьшения рекомбинаций между локусами, в частности, отвечающими за репродуктивную изоляцию, а крупные транслокации напрямую вызывают гибридную стерильность из-за неправильной сегрегации во время мейоза (Bock et al., 2023). Полногеномный анализ с использованием технологии секвенирования длинными ДНК-прочтениями позволил выявить более 100000 SV для двух гибридирующих видов бабочек рода *Lycaeides* Hübner, 1819 (Lycaenidae) с достоверными отличиями в моделях интрогрессии в зоне гибридизации. В целом, гибриды продемонстрировали направленный сдвиг в сторону одного из предков, особенно выраженный для Z-сцепленных SV

и инверсий, содержащих множество генов. Таким образом, SV могут иметь очень сильное влияние на приспособленность гибридов, способствуя репродуктивной изоляции (Zhang et al., 2023a).

Гибридизация, помимо крупных геномных перестроек, может изменять частоту соматических мутаций и вызвать их «взрывное» накопление. Для объяснения этого феномена предложены разные теории, например, что причиной может быть глобальное разрушение у гибридов коадаптированных в родительских геномах белковых комплексов, или локально повышенная частота мутаций вокруг гетерозиготных локусов (Bashir et al., 2023). В любом случае, повышенная генетическая изменчивость полезна для создания адаптивных инноваций новых гибридных видов (Shapiro, 2023). Возможность существования различий в скоростях мутаций между гибридами и родительскими видами была впервые продемонстрирована на примере табака (*Nicotiana L.*) (Bashir et al., 2023).

Активация мобильных элементов может также привести к крупным изменениям в функционировании генома с нарушением паттернов экспрессии у гибридных видов – так называемый «транскрипционный шок». Экспрессия генов менее строго регулируется непосредственно после гибридизации, новые модели экспрессии возникают и стабилизируются в течение десятков поколений (Runemark et al., 2019), а для стабилизации генома и совместимости транскрипции требуется корректировка эпигеномного ландшафта (Shin et al., 2022). Однако несмотря на важность понимания механизмов эволюции полиплоидов, знания о полногеномных эффектах экспрессии у зарождающихся аллополиплоидов весьма ограничены. Известно, например, что в геноме  $\times$ *Brassicoraphanus* Sageret, синтетического аллотетраплоида репы *Brassica rapa L.* и редьки *Raphanus sativus L.*, 1753, родительские хромосомы сохраняются без перестройки, но происходит реконфигурация транскрипционной сети и, в частности, мобильные элементы от *B. rapa* транскрипционно подавляются из-за повышенного метилирования ДНК (Shin et al., 2022). Интересно, что изучение экспрессии у реципрокных межвидовых гибридов лучепёрых рыб рода *Xiphophorus* обнаружило увеличение семейств генов и положительно отобранные гены, связанные с живорождением; при этом положительно отобранные семейства генов были значительно обогащены непалиморфными мобильными элементами (Lu et al., 2023b).

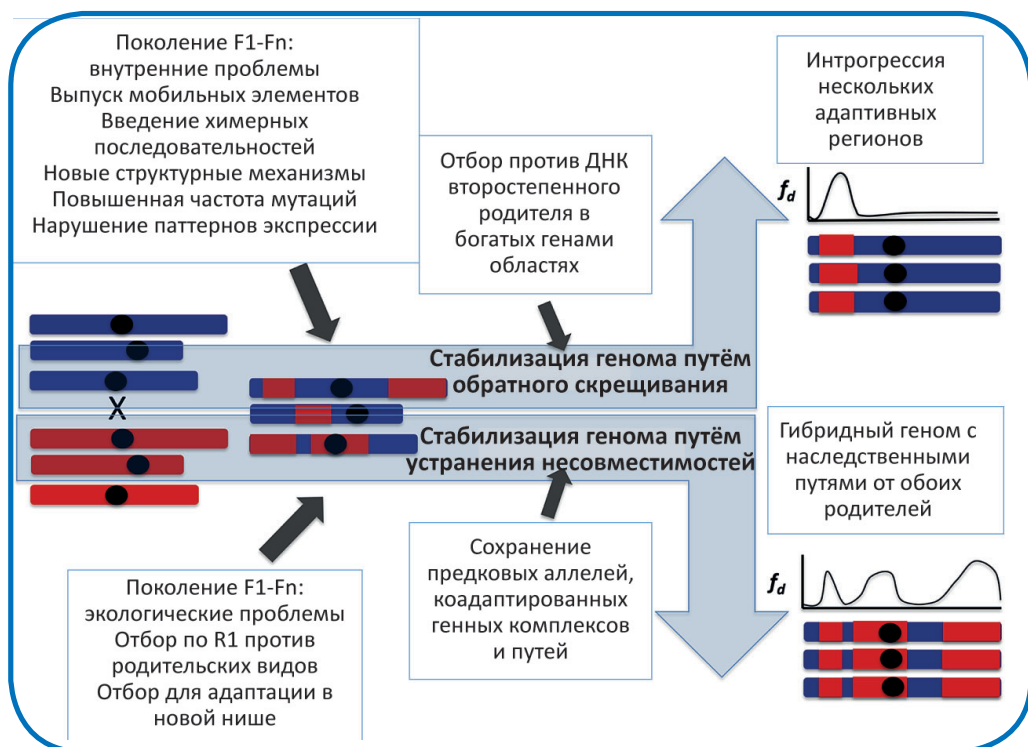
При аллополиплоидизации часто наблюдается отклонение от аддитивной экспрессии, при которой эффекты тождественных генов суммируются. Неаддитивная регуляция генов контролируется в процессе развития, что может приводить к субфункционализации дублирующих генов (Wang et al., 2023a). Кардинальные изменения в неаддитивной регуляции генов наблюдали у аллотетраплоидов резуховидки. Отклонение экспрессии (для примерно 5–6% генов) от среднего значения для родителей были обнаружены в двух независимо полученных синтетических аллотетраплоидах. Оказалось, что >65% неаддитивно экспрессируемых генов в аллотетраплоидах репрессированы, а >94% репрессированных генов в аллотетраплоидах соответствуют генам, которые экспрессируются на более высоких уровнях у *Arabidopsis thaliana*, по

сравнению с *A. arenosa* (L.) Lawalrée, 1960. В целом, комбинация двух дивергентных геномов индуцирует неаддитивную регуляцию генов по всему геному, обеспечивая тем самым молекулярную основу для изменчивости *de novo* и для эволюции аллополиплоидов (Wang et al., 2006). Неаддитивная экспрессия стресс-чувствительных белков может способствовать гетерозису у гибридов и, возможно, гибридных культур резуховидки (June et al., 2023). У гибридной сои были в большом количестве обнаружены неаддитивные сайты дифференциального метилирования (в основном были гиперметилированы), которые принимали участие в разных биологических процессах, включая регуляцию транскрипции и гормональной функции (Chen et al., 2022b).

**Стабилизация гибридного генома.** В целом, относительно генетических последствий гибридизации выделены новые принципы, применимые к большому количеству случаев, когда отбор в среднем действует против гибридизации, но которые могут не соответствовать системам, где гибридизация глобально нейтральна или выгодна. Они включают (1) сочетание быстрого (десятки поколений) и медленного (сотни и тысячи поколений, в зависимости от общей скорости рекомбинации) удаления чужеродных предков (генетических примесей), стабилизирующего смешанные геномы; (2) снижение скорости интрогрессии в функционально важных областях генома; и (3) стабилизацию генома, в которой ключевую роль играет рекомбинационный ландшафт (например, различные корреляции между скоростью рекомбинации и плотностью генов) (Moran et al., 2021).

Сразу после прекращения гибридизации между гибридным и родительскими таксонами начинается процесс стабилизации генома, когда некоторые пути интрогрессии удаляются отбором против несовместимостей, а другие, напротив, фиксируются (рис. 4.16). Время стабилизации геномов варьирует в широких пределах, охватывая сотни и тысячи поколений; например, 400 поколений в геномах грибов *Zymoseptoria Quaedvlieg & Crous, 2011*, 2500 в гибридных геномах рыб-меченосцев *Xiphophorus* и 5000 у гибридного итальянского воробья *Passer italiae* (Runemark et al., 2019). Скорость локальной рекомбинации важна для вероятности события интрогрессии, потому что в случае широко распространённых несовместимостей интрогрессированные аллели с большей вероятностью рекомбинируют вдали от несовместимостей в областях с частой рекомбинацией. Процесс фиксации родительских аллелей может замедлить ассоциативное сверхдоминирование (когда гибриды с гаплотипами от обоих родителей имеют более высокую приспособленность по сравнению с гибридами, имеющими гаплотип одного родителя), наиболее сильного в областях низкой рекомбинации (Runemark et al., 2019). Считается, что в природе полная геномная несовместимость, определяющая невозможность скрещивания, достигается в среднем за 3 млн л., а решающее значение для понимания факторов, определяющих геномную архитектуру репродуктивной изоляции в процессе видообразования, имеет изучение геномной дивергенции (Caeiro-Dias et al., 2023).

Репродуктивная изоляция формируется постепенно, по мере накопления вредных эпистатических взаимодействий между независимыми локусами (так называемая несовместимость Бейтсона–Добжанского–Мюллера), либо из-за возникающих по



**Рисунок 4.16.** Процесс стабилизации генома при гибридном видообразовании и интрогрессии (из: Runemark et al., 2019).  
RI – репродуктивная изоляция.

всему геному несовместимостей генов многих несвязных локусов, в то время как барьерные локусы появляются позднее в процессе видообразования (Caeiro-Dias et al., 2023). Отсутствие гетероморфных половых хромосом приводит к более медленному накоплению репродуктивной изоляции. Правило Джона Холдейна гласит, что, когда в потомстве F1 двух разных рас животных один пол отсутствует, редок или бесплоден, этот пол является гетерозиготным (Runemark et al., 2019). У чешуекрылых, например, гибридная стерильность возникает в результате множественных и сложных эпистатических взаимодействий, часто (согласно правилу Холдейна) с участием половой хромосомы (Rosser et al., 2022), а предотвращение потока генов между родственными таксонами при гибридной стерильности способствует видообразованию (Blanckaert, Payseur, 2021; Valiskova et al., 2022). Несмотря на то, что идентификация локусов репродуктивной изоляции по-прежнему остаётся сложной задачей, геномные исследования ясно показали, что постоянная изменчивость играет огромную роль в видообразовании, особенно на ранних этапах, тогда как новые мутации приобретают значимость на более поздних стадиях процесса (Bock et al., 2023).

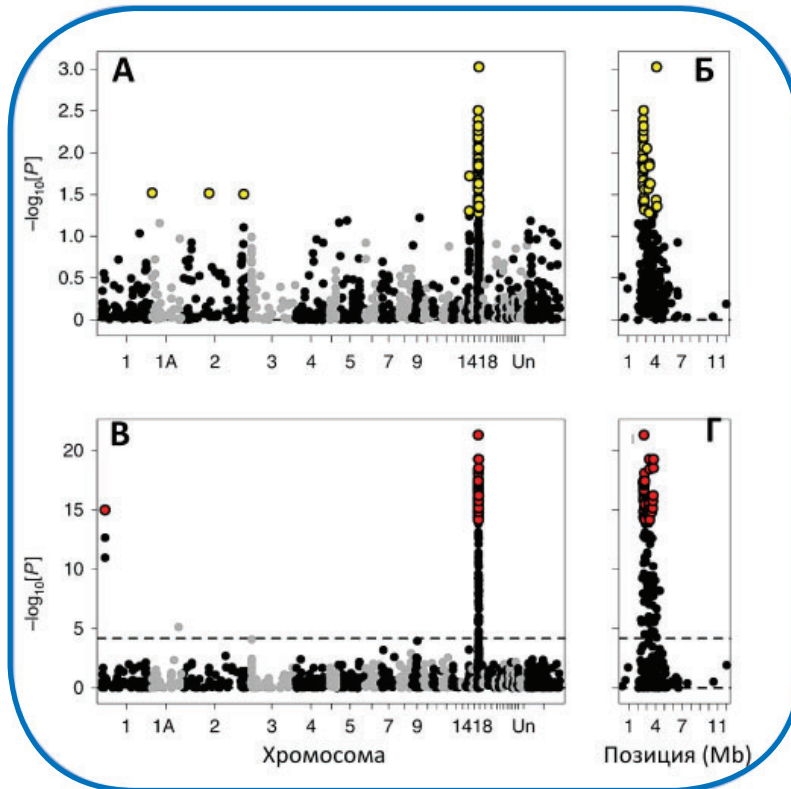
**Популяционная геномика гибридных зон.** Гибридные зоны – географические регионы, где происходит гибридизация и генетическое смешение расходящихся групп – давно рассматриваются, как природные лаборатории эволюционных исследований и окна в эволюционные процессы (Barton, Hewitt, 1985; Harrison, 1993). Поскольку гибридные зоны содержат больше гетерозиготных особей, их отличительной чертой является атипичное неравновесие по сцеплению (LD) между локусами, в том числе физически не связанных (Gompert et al., 2017). Последняя ледниковая эпоха создала очаги гибридных зон, а климат и география формировали их динамику в пространстве и времени. Полагают, что чаще гибридные зоны возникают в результате вторичного контакта разошедшихся линий с формированием как широких, так и узких регионов контакта и гибридизации (Volte et al., 2024). Гибридные зоны могут сильно различаться по возрасту и стабильности, а состав современных гибридных зон может представлять собой смесь локусов и особей, отражающую события как недавней, так и более древней гибридизации. При этом гибридные зоны с балансом между расселением особей и отбором против гибридов (зоны напряжения) перемещаются в районы с низкой популяционной плотностью (Barton, Hewitt, 1985; Gompert et al., 2017). Растущее число примеров перемещения гибридных зон подразумевает, что это явление, вероятно, гораздо более распространено, чем предполагалось раньше (Kalaentzis et al., 2023). Предложено множество моделей для объяснения гибридных зон и управляющих ими эволюционных процессов (Nicolakis et al., 2022). Например, модель зоны напряжения (почти не зависящая от окружающей среды) предполагает, что гибридизация прежде всего приводит к вредной несовместимости и далее к снижению приспособленности или к нежизнеспособности гибридов; согласно модели ограниченного превосходства гибридов и экотональной модели (гибридная зона, возникающая над градиентом окружающей среды), приспособленность гибридов зависит от окружающей среды; а геноцентрические модели рассматривают локус/ локусы, которые могут способствовать репродуктивной изоляции (так называемые «барьерные локусы») с акцентом на одном гене или признаке. Однако в любом случае в пределах одной гибридной зоны действуют несколько селективных процессов, которые могут взаимодействовать между собой, снижая или усиливая эффекты каждого из них (Nicolakis et al., 2022). Таким образом, изучение гибридных зон имеет важное значение для понимания генетических основ видовых отличий, которые являются ключевыми в поддержании репродуктивной изоляции, а также может дать представление о реакции видов на природные и антропогенные изменения (Wagner et al., 2020).

Одним из наиболее известных примеров естественной гибридизации является скрещивание серой *Corvus cornix* (Linnaeus, 1758) и чёрной и *C. corone* (Linnaeus, 1758) ворон, которые имеют протяжённые зоны вторичного контакта и гибридизации в Западной Европе и Восточной Азии. Широко распространённые и хорошо известные, эти виды считаются одними из самых умных животных в мире, с коэффициентом энцефализации, как у приматов (Emery, Clayton, 2004). До недавнего времени у них не было известно видоспецифичных генов и экологической дифференцировки,

они легко скрещивались между собой, но хорошо различались по окраске и поведению. В результате сравнительного геномного анализа было обнаружено, что практически все генетические различия между серой и чёрной воронами сконцентрированы на небольшом участке хромосомы 18. Эта часть хромосомы длиной 1,95 млн п. н. в своё время подверглась инверсии, поэтому при скрещивании она не рекомбинирует с гомологичными участками близкородственных видов. На инвертированном участке сконцентрированы гены окраски оперения туловища, гены зрительного восприятия и гены гормональной регуляции поведения (идентифицирован 81 сайт с фиксированными заменами примерно в 40 генах), чего оказалось достаточным для формирования и поддержания видовой самостоятельности (Poelstra et al., 2014).

Геномные исследования гибридной зоны чёрной и серой ворон обнаружили также, что ассортативное спаривание и эпистатическая архитектура признаков спаривания вызывают сложное движение гибридной зоны (Metzler et al., 2021; Knief et al., 2019). При этом умеренная степень ассортативного спаривания вызывала пре- и постзиготическую изоляцию посредством положительного частотно-зависимого отбора, а эпистаз между двумя локусами индуцировал перемещение гибридной зоны в пользу аллелей, передающих тёмное оперение, с последующим сдвигом в противоположном направлении в пользу фенотипов серого окраса поколений после вторичного контакта (Metzler et al., 2021). Более 68% изменчивости цвета оперения было связано с эпистазом между геном NDP (белок болезни Норри) и областью ~ 2,8 Mb на хромосоме 18 с подавленной рекомбинацией, при этом оба локуса пигментации показали признаки дивергентного отбора против интрогрессии. В целом, лишь малая часть локусов оказалась способной эффективно управлять предзиготической изоляцией и защищать фенотипическую дивергенцию от потока генов. Интересно, что скрининг локусов геномных аутсайдеров в независимых контактных зонах в Азии (рис. 4.17) указал на некоторые из тех же самых генов, которые действуют в системах, модифицирующих паттерны пигментации (Knief et al., 2019).

В Северной Америке известна гибридная зона черношапочных *Poecile atricapillus* Linnaeus, 1766 и каролинских *Poecile carolinensis* (Audubon, 1834) синиц, смещающаяся на север в ответ на повышение зимних температур, но которая не увеличивается в ширину, что указывает на сильный отбор против синиц со смешанными геномами. Согласно геномным данным, гибриды птиц имеют более высокий уровень основного метаболизма, чем любой из родительских видов, а также испытывают дефицит пространственной памяти и способности к решению задач. Таким образом, метаболические и когнитивные нарушения у гибридных синиц могут действовать как сильные механизмы постзиготической изоляции в этой движущейся гибридной зоне, т.е. могут быть вовлечены в отбор против гибридов (Wagner et al., 2020). В основе репродуктивной изоляции в гибридной зоне североамериканской иволги Балтимора *Icterus galbula* (Linnaeus, 1758) и иволги Буллока *I. bullockii* (Swainson, 1827) лежит сложная геномная архитектура. В масштабе всего генома выявлено несколько предполагаемых барьеров для репродукции, включая повышенные



**Рисунок 4.17.** Анализ геномных клинов центральной и южной гибридной зоны представителей рода *Corvus* (чёрной и серой ворон) для 1111 SNP в геноме (А и В) и для области выброса на хромосоме 18 (chr18) (Б и Г) (из: Knief et al., 2019). Пунктирные линии представляют порог полногеномной значимости

пики дивергенции, большую предполагаемую инверсию на Z-хромосоме и сложные взаимодействия между генами-кандидатами путей меланогенеза (рис. 4.18). В целом данные предполагают относительно быстрое накопление барьеров между видами и демонстрируют сложный путь процесса видообразования (Walsh et al., 2023). Исследование шести гибридных зон, образованных семью видами ящериц рода *Podarcis* Wagler, 1830, обитающими на Пиренейском полуострове, показывает, что адаптивная интрогрессия является ключевым механизмом, формирующим разнообразие главного комплекса гистосовместимости (МНС). Поскольку адаптивная интрогрессия включает вновь приобретённые признаки, дающие явные преимущества в приспособленности, МНС представляется особенно интересным в плане изучения из-за его необычайного полиморфизма, роли в адаптивном иммунном ответе, а также важности для биологии сохранения (Gaczorek et al., 2023).

Геномные данные, полученные на многочисленных примерах, свидетельствуют о том, что гибридизация между разными линиями происходит регулярно и приводит к



некоторой степени интрогрессии, причём в процессе видообразования отдельные области генома становятся устойчивыми к потоку генов раньше других – образование так называемых «островков» дифференциации. К таким давно признанным областям относятся половые хромосомы, играющие непропорционально большую роль в процессе видообразования птиц. Например, использование геномного подхода для изучения ландшафта дифференциации между гибридизирующими подвидами острохвостой амадины *Poephila acuticauda* (Gould, 1840) из семейства вьюрковых ткачиков (Estrildidae) обнаружило, что репродуктивную изоляцию в зоне гибридизации птиц обеспечивают инверсии половых хромосом, тогда как аутосомная дифференциация между подвидами минимальна. Таким образом, Z-хромосома и цвет клюва (большая часть генов-кандидатов находится на хромосоме 8, а остальные – на Z-хромосоме) способствуют репродуктивной изоляции, установленной между этой парой подвидов (Hooper et al., 2019). Важная роль Z-хромосомы в репродуктивной изоляции предполагается также для зоны гибридизации ящериц *Podarcis bocagei* (Seoane[es], 1884) и *P. carbonelli* Pérez-Mellado, 1981 (Caeiro-Dias et al., 2023).

При изучении гибридной зоны между бабочками парусниками *Papilio syfanius* Oberthür, 1886 и *Papilio maackii* Ménétriés, 1859 (Papilionidae) было обнаружено смешение темпов эволюции. В этом случае геномные области, наиболее подверженные генным потокам, эволюционируют с одинаковой скоростью между видами, а области генома с высокой степенью репродуктивной изоляции – с видоспецифичной скоростью. Таким образом, репродуктивная изоляция поддерживает не только геномную дифференциацию, но и скорость накопления дифференциации, т.е. асимметричные темпы эволюции обеспечивают дополнительные «метки» локусов, вовлечённых в репродуктивную изоляцию (Xiong et al., 2022). В гибридной зоне между двумя таксонами бабочек голубянок из рода *Lycaeides* с помощью полногеномного секвенирования с использованием длинных ДНК-прочтений было показано, что структурные варианты, особенно делеции, могут способствовать видообразованию, непосредственно вызывая репродуктивную изоляцию или подавляя рекомбинацию в больших областях генома (Zhang et al., 2023).

Интересную модель для изучения представляет семейство гремучих змей (Crotalinae), поскольку у них в природе распространена гибридизация между отдалённо родственными видами, а лабораторные скрещивания дают жизнеспособное потомство. В совокупности это предполагает отсутствие полной изоляции, поэтому изучение геномных интрогрессий может дать ценную информацию о механизмах и процессах, лежащих в основе частичной репродуктивной изоляции. Исследование интрогрессии в гибридной зоне между двумя видами гремучих змей – степной *Crotalus viridis* (Rafinesque, 1818) и карликовой *Crotalus oreganus concolor* Woodbury, 1929 – нашли доказательства того, что геномная несовместимость и другие мультилокусные процессы влияют на приспособленность гибридов. В частности, обнаружены доказательства геномных паттернов интрогрессии, которые различаются в зависимости от участка генома; выявлены закономерности, согласующиеся с таким

моделями, как зоны напряжения, или такими, где мультилокусные процессы управляются внешними факторами. В целом система находится на стадии накопления мультилокусных эффектов и, по-видимому, имеют место антагонистические геномные взаимодействия, т.е. как за, так и против смешанных генотипов внутри гибридной зоны (Nikolakis et al., 2022). Примечательно, что яды этих видов (которые разошлись в плиоцене, около 3–5,2 млн л. н.) хорошо отличаются; яд карликовой змеи характеризуется наличием гетеродимерного нейротоксического компонента, а яд степной змеи состоит в основном из литических веществ. Недавно было показано, что яды гибридов этих гремучих змей представляют собой уникальную комбинацию признаков, проявление которых, по-видимому, ограничено гибридной зоной, и высказано предположение, что скрещивание видов может привести к появлению новых фенотипов яда с уникальными действиями (Smith et al., 2023).

Когда мутации, возникшие в геноме каждого родительского вида, при взаимодействии вызывают вредные последствия, у гибридов формируется гибридная несовместимость Добржанского–Мюллера. Причём управляют появлением гибридной несовместимости, вероятно, более глобальные эволюционные силы, чем выделенные на сегодняшний день отдельные функционально разнообразные локусы (Moran et al., 2021). Гибридизация, особенно отдалённая, может вызывать нарушение ядерно-цитоплазматических взаимодействий и приводить к ядерно-цитоплазматической несовместимости. Вместе с тем, при изучении гибридизации и видообразования внимание на ядерно-цитоплазматических взаимодействиях обычно не акцентируется, хотя имеются исключения. Например, в гибридной зоне колокольчика *Hyacinthoides hispanica* (Mill.) Rothm. и *H. non-scripta* (L.) Chouard ex Rothm на севере Испании обнаружили, что ядерно-цитоплазматическая несовместимость не предотвращает интрогрессию нейтральных маркеров по всему геному, но, по-видимому, исключает перемещение некоторых аллелей и скоординированность между клеточными компартментами, а высокая приспособленность межвидовых скрещиваний согласуется с продолжающимся потоком генов (Marquardt et al., 2022). Как показали исследования геномов цитрусовых, эволюционные конфликты, возникающие из-за ядерно-цитоплазматических взаимодействий, способны оказывать стойкое воздействие на генетику и селекцию растений и могут служить основой цитоплазматической мужской стерильности. Полногеномное исследование ассоциаций (GWAS) выявило три ядерных гена, кодирующих белки пентатрикопептидного повтора (PPR), которые способствуют ядерно-цитоплазматическим взаимодействиям у представителей рода *Citrus* L. (Wang et al., 2022c).

Изучение гибридных зон полезно для понимания видообразования и того, как виды реагируют на окружающую среду, позволяя иногда делать прогностические выводы. При этом гибридизация и видообразование в морских экосистемах менее изучены по сравнению с пресноводными или наземными экосистемами. Морские звёзды являются ключевыми видами в приливно-отливной зоне Северной Атлантики. Исследование двух близкородственных видов морских звёзд, *Asterias rubens* Linnaeus, 1758 и

*A. forbesi* (Desor, 1848), различающихся по ширине своей экологической ниши, предполагает важную роль экологического отбора в поддержании гибридной зоны. Показано, что на распределение геномных вариантов морских звёзд влияет окружающая среда, которую будет крайне важно учитывать по мере изменения климата (Giakoumis et al., 2023). Комплекс *Mytilus edulis* включает три близкородственных вида, обитающих в северном полушарии: *Mytilus edulis* Linnaeus, 1758 и *Mytilus trossulus* Gould, 1850 (регионы с холодным и умеренным климатом), а также *Mytilus galloprovincialis* Lamarck, 1819 (более тёплые регионы). В местах, где их ареалы перекрываются, происходит гибридизация и образуются гибридные зоны, которые характеризуются разными экологическими условиями и степенью межвидового потока генов и примесей. Исследования мозаичной гибридной зоны *M. galloprovincialis* и *M. edulis* на юго-западе Англии подтвердили ее долгосрочную (в течение 35 лет наблюдений) стабильность, несмотря на более высокую жизнеспособность *M. galloprovincialis*. Стабильность совпадает с резким изменением температуры, позволяя предположить, что эти гибридные популяции могут противостоять последствиям глобального потепления. Однако, один локус SNP, связанный с трансмембранным сигнальным белком Notch (принимает участие в латеральном ингибировании и эмбриогенезе), продемонстрировал заметно отличающийся характер изменчивости, возможно связанный с адаптацией *M. galloprovincialis* к более холодным северным температурам (Diz et al., 2023).

Гибридизацию обычно изучают в контексте пар видов, но в природе существуют гораздо более сложные системы межвидовой гибридизации и интрогрессии. В некоторых случаях один вид может выступать в качестве «канала/моста» для потока генов между двумя другими видами (например, у некоторых карпообразных (Cypriniformes), бабочек семейства нимфалид (Nymphalidae), дарвиновских вьюрков (Geospizinae)), в других случаях некоторые виды могут функционировать как «интрогрессивные центры», скрещиваясь со многими не гибридизующими между собой видами (например, кряква (*Anas platyrhynchos* Linnaeus, 1758), обыкновенный фазан (*Phasianus colchicus* Linnaeus, 1758) и серебристая чайка (*Larus argentatus* Pontoppidan, 1763)) и, наконец, иногда несколько видов могут гибридизовать между собой, как, например, это происходит в роде *Bos* Linnaeus, 1758 (Natola et al., 2022). С помощью методов популяционной геномики изучалась зарождающаяся гибридная зона, которая включает две сталкивающиеся гибридные зоны трёх видов дятлов: красногрудого *Sphyrapicus ruber* (Gmelin, JF, 1788), красношейного *S. nuchalis* Baird, 1858 и желтобрюхого *S. varius* (Linnaeus, 1766) сапсака. Гибридизация в этой трёхвидовой гибридной зоне оказалась более обширной, чем в обеих межвидовых гибридных зонах. Поток генов между видами предполагает возникновение интрогрессии до образования трёхвидовой гибридной зоны, что может быть результатом «мостовой гибридизации», бродяжничества или других трёхвидовых взаимодействий. Таким образом, история взаимодействия трёх видов дятлов, видимо, значительно сложнее, чем предполагали ранее, и, кроме того, она может представлять распространённую биологическую закономерность, присущую другим таксонам (Natola et al., 2022).

Геномная революция привела к стремительному росту количества полногеномных данных, демонстрирующих огромное криптическое биоразнообразие. Правильное понимание криптических видов имеет важное научное значение и необходимо для практического применения в сфере охраны природы, биологического контроля и т.д. (Kalaentzis et al., 2023). Одним из перспективных направлений геномных исследований является использование гибридных зон для проверки статуса криптических видов. Такой анализ позволяет непосредственно проверить, возникли ли барьеры для потока генов между предполагаемыми криптическими видами, и насколько они сильны. Перспективность подхода была недавно продемонстрирована при анализе гибридной зоны полосатых тритонов (род *Ommatotriton* Gray, 1850) в Анатолии: полученные данные убедительно подтвердили ранее предполагаемый видовой статус *O. nesterovi* (Litvinchuk, Zuiderwijk, Borkin, and Rosanov, 2005) и *O. ophryticus* (Berthold, 1846) (Kalaentzis et al., 2023).

Необычный пример гибридизации описан у общественных насекомых. Репродуктивное разделение труда у них обычно определяется сигналами окружающей среды, однако у муравьёв было обнаружено генетическое влияние на касту. Крайней формой генетической кастовой детерминации является «социальный гибридогенез», при котором сосуществующие генетические линии обязательно скрещиваются, чтобы произвести рабочих (дочери-королевы имеют чистое/негибридное происхождение). Результаты проверки гипотезы о том, что социальный гибридогенез в роде муравьёв *Pogonomyrmex* Mayr, 1868 является результатом одного или нескольких событий межвидовой гибридизации, предполагают, что сдвиги в детерминации репродуктивной касты могут усилить репродуктивную несовместимость, подобно эволюции гибридогенеза в несоциальных системах (Cahan et al., 2022).

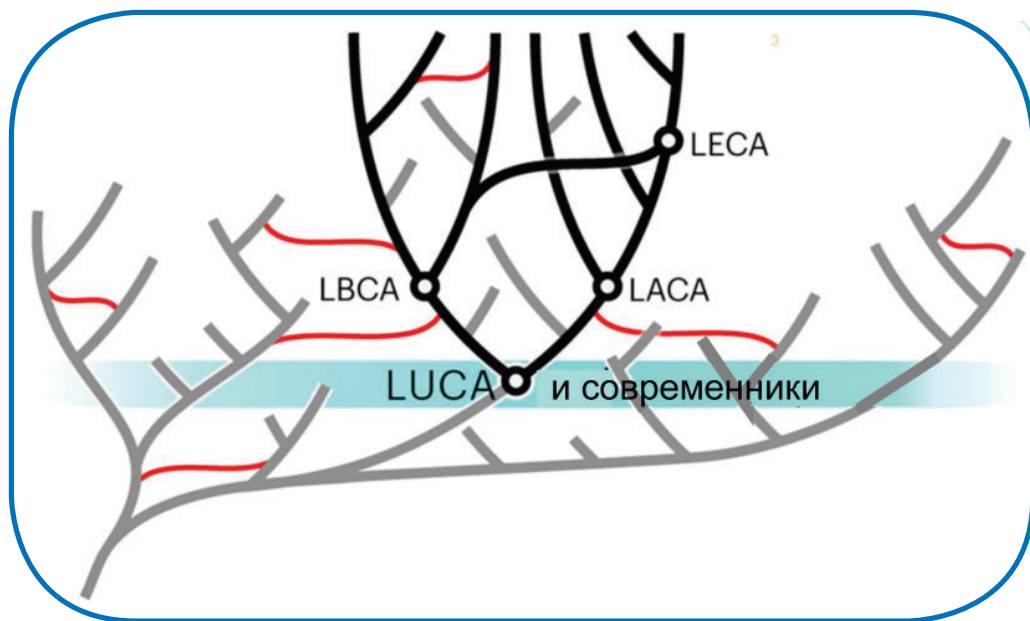
Однако, несмотря на постоянный рост объёма полногеномных данных, многие вопросы о свойствах гибридного генома и последствиях гибридизации по-прежнему остаются нерешёнными. В задачи будущих исследований входит установление продолжительности процесса стабилизации гибридного генома (одинакова ли она у гомоплоидных и аллополиплоидных гибридов) и существования/отсутствия тенденции к возврату в сторону одного из родителей у гомоплоидных гибридов; выявление отличий в создании новых фенотипов у аллополиплоидов по сравнению с автополиплоидами; определение влияния отбора на совместимость при формировании гибридного генома и т.д. (Runemark et al., 2019). Согласно недавно проведенным исследованиям, анализ гибридных зон может быть использован для проверки статуса форм, классифицируемых как один вид из-за высокого морфологического сходства, т.е. криптических видов или видов-двойников (Kalaentzis et al., 2023). Наименее изученным вопросом, решение которого ускорит как фундаментальные, так и прикладные исследования, признают молекулярные механизмы, задействованные в формировании гибридного генома (Bashir et al., 2018). Таким образом, помимо фундаментальной биологии, дальнейшие исследования гибридных геномов необходимы для лучшего понимания потенциала гибридизации в целях сохранения биоразноо-

бразия и эффективного управления природными ресурсами, а также оптимизации сельскохозяйственного производства.

#### 4.7. Эволюция, филогенез и биология видов

Ранние филогенетические исследования с использованием многолокусного анализа внесли большой вклад в понимание истории **филогенеза** и поставили под сомнение некоторые устоявшиеся представления о взаимоотношениях между разными группами растений и животных. Однако для решения ряда проблем они оказались бессильны. Развитие геномных технологий превратило область молекулярной филогенетики в более мощную и эффективную филогеномику, открывая новые и ранее невообразимые возможности для эволюционных исследований (Young, Gillung, 2020). Например, филогения из 10575 равномерно отобранных бактериальных и архейных геномов на основе полного набора из 381 маркера с использованием нескольких стратегий анализа данных позволила по-новому взглянуть на древо жизни бактерий и архей. Филогеномные реконструкции указали на значительно более тесную эволюционную близость между археями и бактериями, чем предыдущие оценки, опирающиеся на меньшее количество «основных» генов (Zhu et al., 2019).

Последний универсальный общий предок, LUCA, представляет собой одну из самых ранних стадий эволюции, которую можно исследовать с помощью стандартных филогенетических методов (рис. 4.19). В последние десятилетия было проведено множество исследований, направленных на выяснение природы LUCA, его возраста, метаболизма и экологического контекста. Однако результаты этих работ, часто основанные на фрагментарных данных и разрозненных методах, были достаточно противоречивыми и вызывающими споры. Крупным достижением в этой области стала недавно опубликованная работа международного коллектива авторов (Moody et al., 2024). В ней для датировки возраста LUCA, его геномном составе, метаболизме и влиянии на экосистему была использована методология молекулярных часов, филогенетическое согласование с учётом горизонтального переноса генов и имеющиеся биогеохимические модели. Используя данные для 350 видов бактерий и 350 видов архей, авторы провели анализ 10000 семейств генов, чтобы сконструировать эволюционные древа и определить гены, с наибольшей вероятностью принадлежащие LUCA. В итоге, было выделено 399 генов, входивших в состав генома древнего предка, и рассчитаны вероятности для тысяч других генов, чтобы получить максимально точную картину. Филогенетические реконструкции, для которых использовали универсальные семейства паралогичных белков, позволяют предполагать, что LUCA имел размер генома не менее 2,5 Mb, и в нём было закодировано 2600 белков, что сопоставимо с современными прокариотами. Универсальные паралоги – это семейства белков, которые дуплицировались до последнего LUCA. На филогенетических реконструкциях каждый паралог образует отдельную кладу с собственным узлом LUCA; сравнивая оценку времени узла по обоим поддеревьям паралогов одного и того же семейства, оценку времени можно перекрёстно откалибровать. Возраст



**Рисунок 4.19.** LUCA в контексте древа жизни (из: Moody et al., 2024).

Ветви древа жизни, на которых сохранились отобранные потомки, окрашены в чёрный цвет, а не оставившие потомков – в серый. Как общий предок всех существующих форм клеточной жизни, LUCA представляет собой самый древний узел, который можно реконструировать с помощью филогенетических методов. Он существовал на ранней Земле вместе с другими линиями (выделены бирюзовым цветом), которые не оставили потомков среди отобранных клеточных форм. Однако эти линии могли оставить след в современных организмах, перенеся гены в отобранное древо жизни (красные линии) до своего исчезновения. LUCA – последний универсальный общий предок, LBCA – последний общий предок бактерий, LACA – последний общий предок архей и LECA – последний общий предок эукариот.

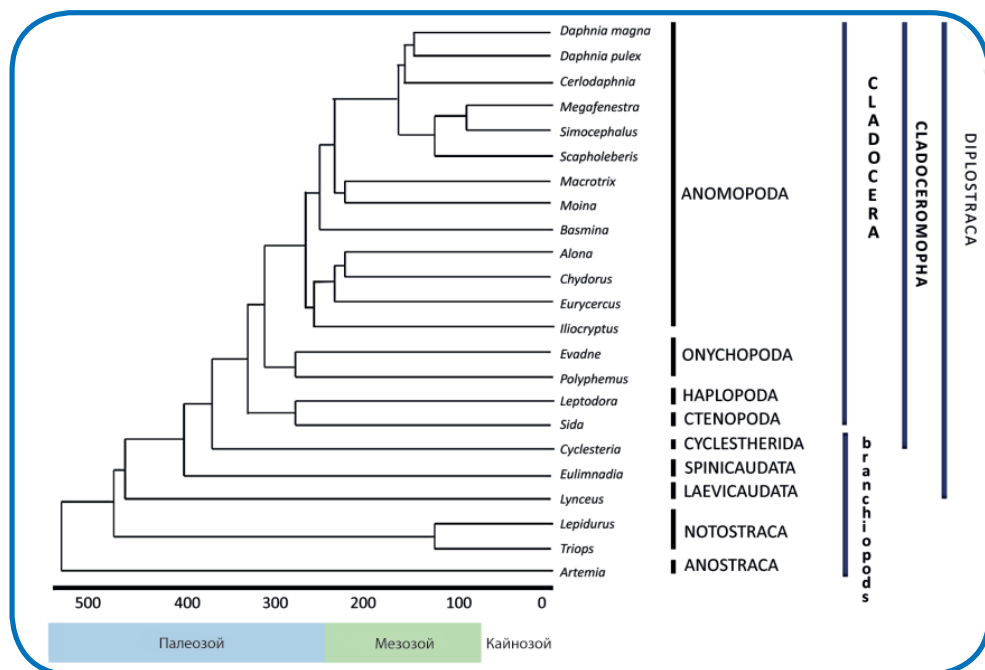
LUCA был оценен в 4,2 млрд л., что совпадает с периодом активных астероидных бомбардировок Земли, и существует предположение, что в таких сложных условиях эволюция могла происходить значительно быстрее, чем предполагали ранее. Согласно результатам, LUCA был анаэробным ацетогеном прокариотного уровня, который обладал ранней иммунной системой и являлся частью устоявшейся экологической системы, а его метаболизм мог обеспечить нишу для других членов микробного сообщества (Goldman, Vecerra, 2024; Moody et al., 2024).

Построение высоконадёжного глобального древа эукариот – одна из важнейших задач эволюционной биологии. Эволюция эукариот, которая привела к наблюдаемому сегодня разнообразию видов, началась более миллиарда лет назад, и воссоздать её чрезвычайно сложно. Чтобы исследовать раннюю эволюцию среди эукариотических супергрупп (к ним относятся Opisthokonta и Amoebozoa, Plantae/Archaeplastida, Excavata, Chromalveolata и Rhizaria), был собран обширный набор данных (65 видов, 135 генов, представляющих 31921 аминокислоту), анализ которого показал, что с высокой степенью вероятности эукариот можно разделить на две монофилетические

мегагруппы и несколько менее разнообразных линий (Burki et al., 2008). Возросшие возможности филогеномики помогли разрешить отношения между супергруппами, обеспечили создание новых гипотез для более глубокого понимания древа эукариот, позволили разместить таксоны «сироты», которые ранее не поддавались надлежащей классификации. В то же время, возникли новые проблемы, и осталось несколько ключевых линий неизвестного эволюционного происхождения. Очевидно, что для полного решения древа эукариот потребуется продолжение исследований с интегративным подходом, сочетающим морфологию, филогенез одного гена и филогеномику. Считается, что в дальнейшем необходимо проводить систематическую оценку распространённости горизонтального переноса генов, способного (по мнению многих исследователей) глобально влиять на эволюцию эукариот и как следствие – иметь самые негативные последствия для разрешающей способности филогеномики (Burki, 2014).

Одним из примеров изучения филогенеза является филогеномная реконструкция древа жизни амфибий с использованием 220 локусов для 286 видов (представляющих 94% семейств и 44% родов). Результаты этих исследований прояснили несколько спорных вопросов, которые в сочетании с набором проверенных ископаемых калибровок подтверждают удивительно более молодые временные рамки для диверсификации кроновых и порядковых амфибий, чем сообщалось ранее (Hime et al., 2021). Были получены надёжные подтверждения монофилии амфибий и сестринских отношений между лягушками и саламандрами (что согласуется с гипотезой *Batrachia*). Ожидается, что дальнейшие исследования по систематике амфибий будут сосредоточены на более глубоком понимании основных движущих сил их диверсификации в палеогене и неогене, способствовавших радиации, последствия которой наблюдаются сегодня (Hime et al., 2021).

Хотя оценка филогенеза сложна для изучения событий, которые происходили несколько сотен миллионов лет назад, филогеномные подходы предлагают новые способы изучения взаимоотношений между древними линиями и оценки гипотез, которые являются ключевыми для эволюционной биологии. Реконструкция взаимоотношений между сохранившимися видами одной из древнейших ветвей членистоногих – жаброногих ракообразных – с применением калейдоскопического подхода однозначно разрешила давно обсуждаемые связи между существующими отрядами ветвистоусых *Cladocera* – водяными блохами (экологически значимая группа зоопланктона в глобальных водных и морских экосистемах). Филогеномный анализ (рис. 4.20) подтвердил монофилию *Cladocera* наряду с обнаружением гомоплазии в их строении, а также выявил большие филогенетические расстояния между линиями с похожей экологической специализацией. Последнее предполагает независимую эволюцию в основных планах строения тела, например, у пелагических хищных отрядов *Narloroda* и *Onychoroda* (*Gymnomera*). Таким образом, результаты поднимают новые вопросы о роли гомоплазии и быстрой радиации в диверсификации кладоцер и помогают изучить эволюцию признаков с геномной точки зрения в функционально хорошо изученной древней группе членистоногих (Van Damme et al., 2022).



**Рисунок 4.20.** Молекулярная филогения Branchiopoda, калиброванная по ископаемым находкам (по: Van Damme et al., 2021). Значения апостериорной вероятности для всех узлов равны 1, кроме внутреннего узла Chydoridae (0,96)

Palaeognathae представляют собой одну из двух основных линий современных птиц и включают летающих (Tinamiformes) и нелетающих (Apterygiformes, Casuariiformes, Rheiformes и Struthioniformes) бескилевых. Разрешение филогенетических отношений палеогнатов исторически считалось трудной задачей, и более ранние молекулярные филогении предполагают, что обширная неполная сортировка линий могла сопровождать их быструю древнюю диверсификацию. Используя полногеномные наборы трёх типов некодирующих ядерных маркеров (в общей сложности 20850 локусов и более 41 млн п. н.), была реконструирована полностью разрешённая топология, с помещением нанду как сестринского вида киви и эму+казуар, и подтверждаемая паттернами вставок для ретроэлементов CR1. Вместе с тем, анализ суперматриц объединённых данных позиционирует нанду как сестринский вид всех других палеогнатов, но эта альтернатива не имеет поддержки данными по ретроэлементам. Отличие наиболее распространённой топологии генного древа для каждого типа маркеров от древа вида указывает на существование «эмпирической зоны аномалий» у палеогнатов. Авторы подчеркивают, что не только высокая бутстреп поддержка, но также конгруэнтность между типами маркеров и подтверждение редких геномных изменений (таких как вставки ретропозонов), а также стратегии филогеномной подвыборки для оценки основного филогенетического сигнала важны для получения надёжных топологий (Cloutier et al., 2019).

Современные виды отряда *Crocodylia* являются остатками древней линии крупных рептилий-архозавров, филогенетические отношения которых изучались на протяжении десятилетий. Филогеномные исследования *Crocodylus Laurenti*, 1768 с использованием генотипических данных из 17538 SNP, собранных для 33 особей шести видов, выявили новые связи для всех неотропических видов. Впервые острорылый крокодил *C. acutus* (Cuvier, 1807) был определён как монофилетический вид с двумя отдельными линиями (анtilьской и континентальной) и сестринским центральноамериканским крокодиллом *C. moreletii* (Duméril et Bibron, 1851) (Milián-García et al., 2020).

Китообразные, многие из которых представляют самых крупных обитателей Земли, претерпели наиболее радикальную морфологическую трансформацию среди млекопитающих, задокументированную многочисленными окаменелостями эоцена (56–34 млн л. н.). Несмотря на то, что происхождение и эволюция китообразных, от их раннего перехода к водному образу жизни до последующей диверсификации, была предметом многочисленных исследований, некоторые аспекты систематики внутри отдельных групп оставались невыясненными. Филогеномный анализ *Cetacea*, охватывающий 6527596 выровненных п. н. для 89 таксонов, позволил полностью разрешить отношения между кловорылыми китами, а также спорные отношения между океаническими дельфинами – особенно проблематичным подсемейством *Delphinidae* (McGowen et al., 2020).

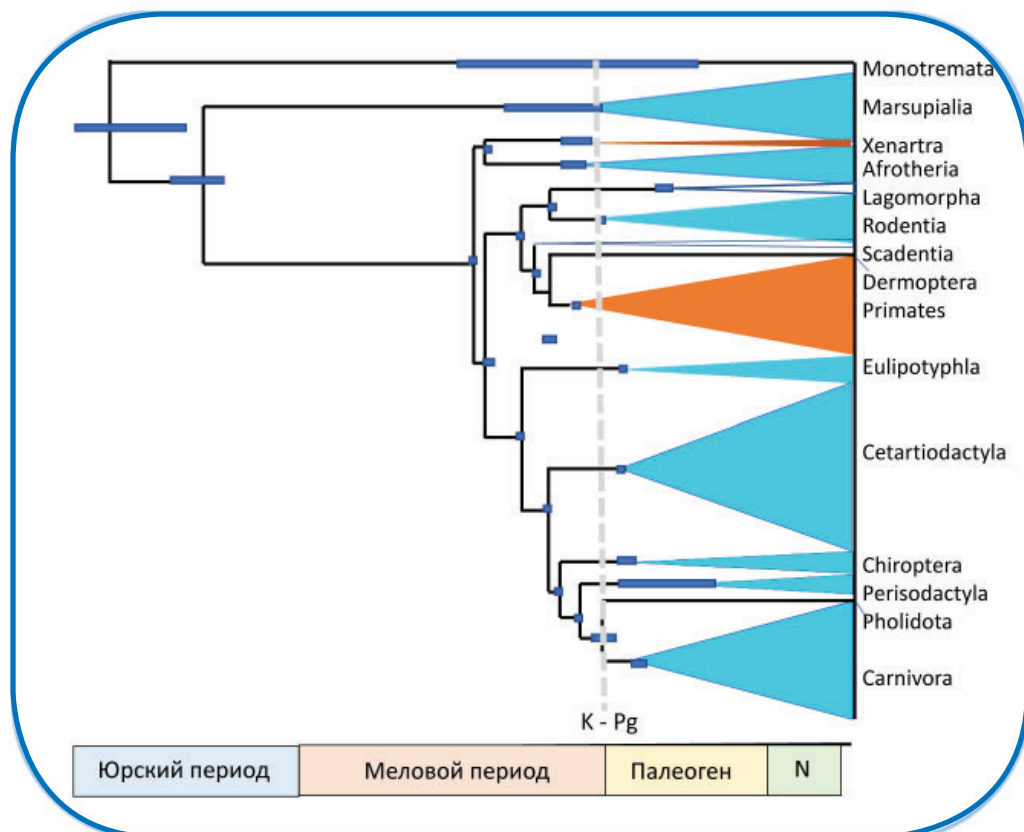
Другой пример – жуки, представляющие собой отряд животных с наиболее высоким биоразнообразием, насчитывающий свыше 380000 только описанных видов. Анализ набора данных по 68 однокопийным генам, кодирующим ядерные белки, в выборке 129 из 193 признанных существующих семейств, воссоздал топологию (противоречивую в предшествующих исследованиях), совпадающую с морфологическими данными, и позволяющую сделать несколько формальных изменений в классификации *Coleoptera* (Cai et al. 2022).

Филогеномный анализ также успешно применялся для изучения групп животных и растений более низкого таксономического ранга. Семейство Ореховые (*Juglandaceae*) включает коммерчески важные древесные растения, обычно называемые грецким орехом, крылатым орехом, орехом пекан и гикори. Филогенетические связи и разнообразие среди *Juglandaceae* долгое время были популярными научными вопросами, которые удалось частично решить благодаря новым данным об ископаемых остатках, а также результатам морфологических, молекулярных и палеоэкологических исследований. Однако сохранялась необходимость дальнейшего раскрытия внутри- и межродовых отношений. Поэтому были реконструированы основные филогенетические отношения *Juglandaceae*, с использованием данных о ДНК органелл (хлоропласты) и ядерном геноме 27 видов. В результате анализа было выделено три подсемейства (*Juglandoideae*, *Engelhardioideae* и *Rhoipteleoideae*), а пять родов получили высокую статистическую поддержку валидности. Кроме того, филогеномный анализ обнаружил хорошо подтверждённые позиции неконгруэнтности для *Juglans cinerea* L., *J. hopeiensis* Hu, и *Platycarya strobilacea* Siebold & Zucc (Zhou et al., 2021).

Пример филогеномного исследования на родовом уровне – катальпа *Catalpa* Scop, небольшой род древесных растений семейства бигнониевых (Bignoniaceae), дизъюнктно распространенный в Восточной Азии, восточной части США и Вест-Индии. Растение имеет красивые соцветия и выращивается как важное декоративное дерево для ландшафтного дизайна, садоводства и как источник лесоматериалов. Однако филогенетические отношения и биогеографическая история рода до недавнего времени оставались невыясненными. Исследования с использованием большого набора геномных данных показали, что катальпа является монофилетическим родом, содержащим две основные клады: секцию *Catalpa* с тремя субкладами и секцию *Macrocatalpa* Griseb. Важным результатом является обнаружение события межвидового потока генов, включающего *C. bungei* s. str. с векторами наследования от *C. duclouxii* Dode. и *C. fargesii* Vireau. Неполная сортировка родословных, вероятно, сыграла важную роль в эволюционной истории катальпы. Опираясь на филогеномные и биогеографические данные, были предложены три основных события расселения катальпы, два из которых произошли в олигоцене (Dong et al., 2022).

Галлюциногенные грибы рода *Psilocybe* веками использовались в Мезоамерике. Несмотря на большой интерес к этой группе как объекту культуры, науки и медицины, сведения о таксономии и филогенетических отношениях оставались неполными. По данным геномного секвенирования 71 образца с использованием 2983 семейств генов была создана полностью подтверждённая филогения и показано, что биосинтез псилоцибина впервые возник у *Psilocybe* (Fr.) P. Kumm., возможно, с несколькими событиями горизонтального переноса к другим грибам 40–22 млн л. н. В целом, результаты дали новое представление об эволюционном происхождении биосинтеза псилоцибина, что очень важно для понимания функциональной роли и применения этого многообещающего психотропного средства для лечения различных психических заболеваний (Bradshaw et al. 2022).

Филогеномный подход успешно применяется для **датировки важнейших эволюционных событий**. В отношении событий мела–палеогена (К–Pg) и их связи с летописью окаменелостей, известно несколько моделей диверсификации плацентарных млекопитающих. Согласно «взрывной» модели общий предок Placentalia возник и диверсифицировался в палеогене; модель «длительной выплавки» предполагает возникновение предка Placentalia в меловом периоде, но с внутривидовой диверсификацией в палеогене; в то время как в модели «быстрой выплавки» несколько отрядов плацентарных диверсифицировались в меловом периоде вскоре после возникновения Placentalia, а другие – в палеогене (dos Reis et al., 2012). Байесовский анализ 36 ядерных и 274 митохондриальных геномов (в сумме 20,6 млн п. н.) в сочетании с надёжной калибровкой ископаемых находок показал (рис. 4.21), что сумчатые отделились от плацентарных 168–178 млн л. н., а кроновые сумчатые дивергировали 64–84 млн л. н., Placentalia разошлись 88–90 млн л. н., а современные отряды плацентарных (за исключением Primates и Xenarthra) появились в промежутке примерно 20 млн л. (45–65 млн л.) после массового вымирания, на границе мела и палеогена (dos Reis et al., 2012).



**Рисунок 4.21.** Временное древо млекопитающих (по: dos Reis et al., 2012).

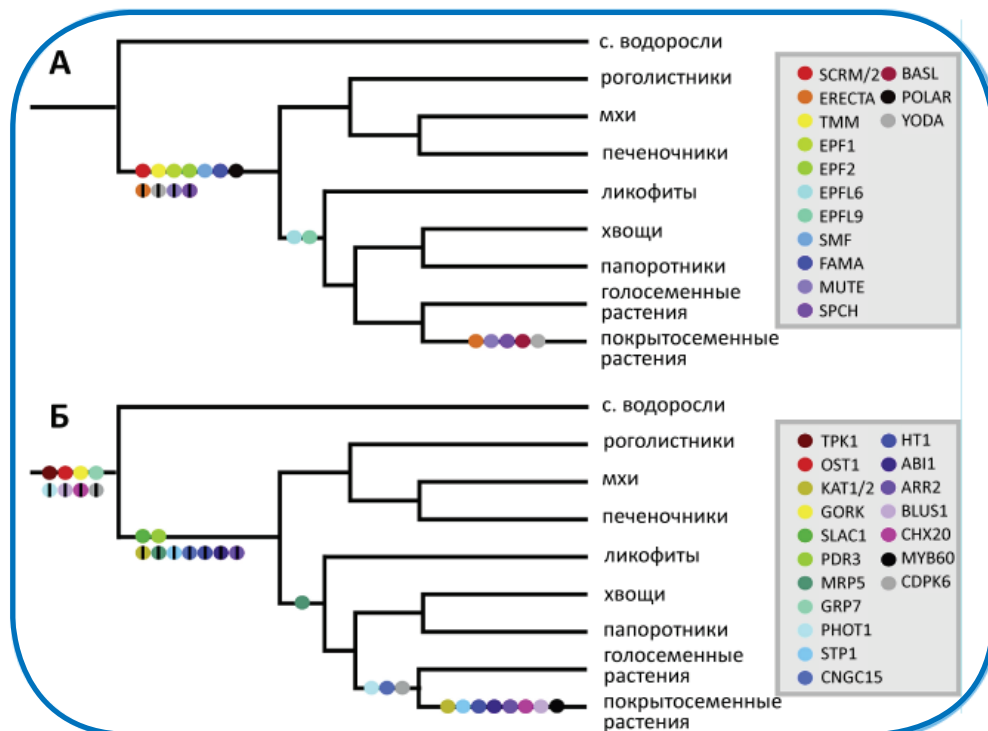
Синие горизонтальные полосы представляют 95% доверительный интервал для возраста узлов. Два отряда млекопитающих диверсифицировались до события вымирания на рубеже мела и палеогена (K–Pg; показаны оранжевым цветом), а все отряды внутри Afrotheria и Marsupialia диверсифицировались после этого события (показаны голубым цветом); N – Неоген.

Анализ филогеномных данных амфибий отодвинули дату происхождения и первоначальной диверсификации лягушек, червяг и саламандр, что сделало их на несколько миллионов лет моложе (Hime et al., 2021). Временные оценки филогеномного древа Cleoptera указали на позднекаменноугольное происхождение жесткокрылых, позднепалеозойское происхождение всех современных подотрядов жуков и триасово-юрское происхождение большинства существующих семейств; однако фундаментальные расхождения в филогении жуков не совпали с гипотезой меловой земной революции (Cai et al., 2022). Полученные результаты относительных сроков дивергенции (опирающиеся на надёжные палеонтологические данные) в основных линиях древа жаброногих, предполагают палеозойское происхождение около 325 млн л. н. для предка Cladosceta и древнюю быструю радиацию около 252 млн л. н. на границе перми и триаса (Van Damme et al. 2022). Датированные филогеномные дере-

вья с интеграцией новых данных об окаменелостях указывают на то, что диверсификация кроновых китообразных Cetacea началась до позднего эоцена, а дивергенция кроновых дельфиновых Delphinidae – уже в среднем миоцене (McGowen et al., 2020). Время дивергенции семейства грецких орехов было оценено в 78,7 млн л. н., причём основные линии диверсифицировались в тёплых и сухих местообитаниях в середине палеоцена и раннем эоцене. Предположительно, Juglandaceae произошли в конце мелового периода, а Juglandoideae – в палеоцене, что свидетельствует о быстрой диверсификации в течение нескольких ледниковых периодов (Zhou et al., 2021). Отряд насекомых Psocodea, представляющий собой разнообразную филогенетическую линию, включает как паразитических (Phthiraptera), так и непаразитических представителей (Psocoptera). Филогеномный анализ с использованием данных для 2370 ортологичных генов указал на единое происхождение паразитизма внутри отряда Psocodea, разрешая противоречивые результаты предыдущих исследований. Согласно датировкам, паразитизм мог возникнуть в любое время между 115 млн л. н. и начальной диверсификацией паразитических видов 100 млн л. н. (De Moya et al., 2021).

Объединение филогеномики с летописью окаменелостей позволяет не только датировать важные эволюционные события, но может также выявить давно утерянные **биогеографические связи**. Анализ геномов (1105 экзонов для семи ныне живущих видов) с использованием обновлённых морфологических данных всех существующих и вымерших таксонов показал, что два современных рода семейства лучепёрых рыб, короткорылые *Atractosteus* (Rafinesque, 1820) и длиннорылые *Lepisosteus* Lacépède, 1803 панцирники, разошлись примерно 105 млн л. н., и многие из предполагаемых расхождений в их филогении тесно связаны с крупными мезозойскими тектоническими событиями, включая разделение Америки, раннее расширение Атлантики и меловую реорганизацию речных систем Северной Америки. Кроновые клады *Atractosteus* и *Lepisosteus* возникли в кайнозое на востоке Северной Америки, т.е. этот регион служил как источником существующего разнообразия, так и рефугиумом этой древней линии (Brownstein et al., 2022).

**Глобальная смена среды обитания и формирование адаптаций.** Филогеномный анализ проясняет важные аспекты биологической эволюции разных таксономических групп, в том числе при их переходе в другую среду обитания. «*Колонизация суши растениями явилась основополагающим событием в истории жизни на Земле, открыв совершенно новые ниши для диверсификации земной жизни и навсегда изменив углеродный цикл*» (Harris et al., 2020). Выход растений из воды на сушу (400–500 млн л.н.) признаётся ключевым событием в развитии наземной растительной флоры. Общий предок эмбриофитов и водорослей Zygnematomphuseae имел гены (например, семейства белков GRAS и рецепторов абсцизовой кислоты PYR/PYL/RCAR), повышающие устойчивость к биотическим и абиотическим стрессам у наземных растений. Приобретённые через горизонтальный перенос от почвенных бактерий, они сыграли важную роль в эволюции и радиации эмбриофитов (Cheng et al. 2019). Возникновение наземных растений сопровождалось новыми приспособлениями к жизни на суше, в том числе эволюцией устьиц (специализированных пор на поверхности



**Рисунок 4.22.** Наличие и отсутствие ортологичных генов, участвующих в развитии и функционировании устьиц (из: Harris et al., 2020).

Однокопийные ортологи (сплошные кружки) сохраняются как один ген у современных растений и, вероятно, выполняют ту же функцию. Ортологи предудупликации (полосатые кружки) представляют собой гены, которые присутствовали в единственной копии у предка-эмбриофита, но позже были продублированы в некоторых линиях. Однокопийные ортологи генов, участвующих в устьичной функции (Б) и в развитии устьиц (А), уже присутствовали у общего предка эмбриофита; с. водоросли – стрептофитовые водоросли

растений, регулирующих газообмен и транспирацию). Основные гены, ответственные за развитие и функции устьиц были хорошо изучены у модельных растений (например, *A. thaliana*). Однако этого было недостаточно, чтобы понять эволюционное происхождение и предковые функции этих приспособлений. Расширенный филогенетический анализ предоставил данные о наличии и отсутствии ортологичных генов, участвующих в развитии и функционировании устьиц у мохообразных (древняя обособленная ветвь в эволюции высших растений) и убедительные доказательства их монофилии. Оказалось, что ряд генов основных замыкающих клеток, включая гены для белков SPCH/MUTE, SMF и FAMA семейств, произошли от общего предка эмбриофитов (рис. 4.22). Следовательно, у первых эмбриофитов устьица были более сложными, чем предполагалось ранее, а устьица мохообразных претерпели редуцированную эволюцию, включая их полную утрату у предка печёночников (но возникли позже в процессе эволюции) (Harris et al., 2020).

Переход к наземному образу жизни был связан с основными видами радиации у животных, включая наземных улиток и слизней. Для прояснения эволюционной истории радиации лёгочных брюхоногих моллюсков *Ranpulmonata*, был проведён филогеномный анализ последовательностей ~1160 генов, кодирующих ядерные белки, и данных транскриптомов для 110 видов надотряда *Sacoglossa* (включает мелких морских слизней и морских улиток, морских брюхоногих моллюсков) и надсемейства *Siphonarioidea* (представленно лёгочными морскими улитками в подотряде сидячеглазых). Результаты показали, что *Sacoglossa* является сестринской группой клады, *Pneumopulmonata*, внутри которой *Siphonarioidea* была сестринской остальных линий в большинстве анализов. Кроме того, сдвиги в маргинальную среду обитания (эстуарии, мангровые заросли и приливные зоны) предшествовали и ускорили эволюцию пневмостом, подтверждая гипотезу о том, что адаптация к маргинальным зонам сыграла важную роль в основных переходах с моря на сушу (Krug et al., 2022).

В девонский период саркоптерингии, или лопастепёрые рыбы (в том числе двоякодышащие), «заоевали» сушу и в конечном итоге дали начало всем наземным позвоночным, включая человека. Филогеномный анализ подтвердил, что двоякодышащие рыбы с гигантскими геномами являются ближайшими из ныне живущих родственниками четвероногих. Это крупное эволюционное событие сопровождалось рядом эволюционных инноваций (например, органов чувств, кровообращения, нервной системы и т. д.) и прежде всего – обеспечивающих дыхание атмосферным кислородом и передвижение по твёрдой поверхности (Meyer et al., 2021). Примечательно, что в то время как происхождение позвоночных сопровождалось двумя раундами полногеномной дупликации, в результате которых увеличилось число генов и размер их геномов, переход от воды к суше не сопровождался какими-либо событиями WGD (Falcon et al., 2023). Секвенирование геномов саркоптеригий и амфибий показало, что существующие линии, непосредственно предшествующие или следующие за этим переходом, содержат исключительно большое количество мобильных элементов. Таким образом, мобильные элементы могли способствовать адаптации, которая позволила позвоночным выйти на сушу (Falcon et al., 2023).

Геномные подходы позволили лучше понять эволюцию и адаптацию к среде обитания цианобактерий – фотосинтезирующих прокариот, населяющих разнообразную водную и наземную среду. Филогенетический и сравнительный анализ 650 геномов цианобактерий показал, что увеличение семейств генов является обычной стратегией адаптации наземных цианобактерий к изменчивой среде, тогда как геномы многих морских штаммов, напротив, подверглись сокращению, чтобы адаптироваться к бедным питательными веществами условиям. Оказалось, что сотни генов, важных для выживания и адаптации этих организмов в конкретных местообитаниях, по-разному распространены в геномах морских, пресноводных и наземных цианобактерий. Они участвуют в восприятии и поглощении света, в хемотаксисе, переносе питательных веществ, реакциях на осмотический стресс и т. д. Интересно, что значительная часть таких генов вносится в результате горизонтального переноса, и эти генетические обмены более часто встречаются у наземных цианобактерий (Chen et al. 2021a).

**Происхождение и эволюция новшеств (приспособлений).** Каждый вид обладает приспособлениями (морфофизиологическими и поведенческими), которые помогают ему существовать в определённых условиях; любое приспособление возникает в результате длительного отбора случайных мутаций. Известные своими необычными биоматериалами, включая яды и шелка, пауки имеют древнее хищническое происхождение. Несмотря на большую практическую важность, отношения между основными родословными пауков оставались долгое время нерешёнными или требующими новых надёжных подтверждений. Филогеномный анализ пауков, включая таксоны, представляющие все основные линии, восстановил некоторые базисные и бесспорные клады, но отверг преобладающую парадигму эволюции паутиных сфер монофилетической *Orbiculariæ*, самой разнообразной линии, содержащей пауков-кругопрядов. Согласно геномным данным, круговая паутина либо возникла намного раньше, чем предполагали прежде, и появилась у предков большинства пауков, либо имела несколько независимых источников происхождения. Примечательно, что ткачи кругов из семейства *Deinopoidea*, использующие механически клейкий шёлк, оказались более тесно связанными с разнообразной кладой пауков (в основном беспутинных), чем с пауками семейства *Arachnoidea*, плетущими круги с использованием клейких капель на шёлке. Оценки времени расхождения филогенетических линий относят происхождение кругов паутины к нижнеюрскому периоду. Эти данные имеют большое значение для интерпретации эволюции пауков, их биоматериалов и паутины (Bond et al. 2014).

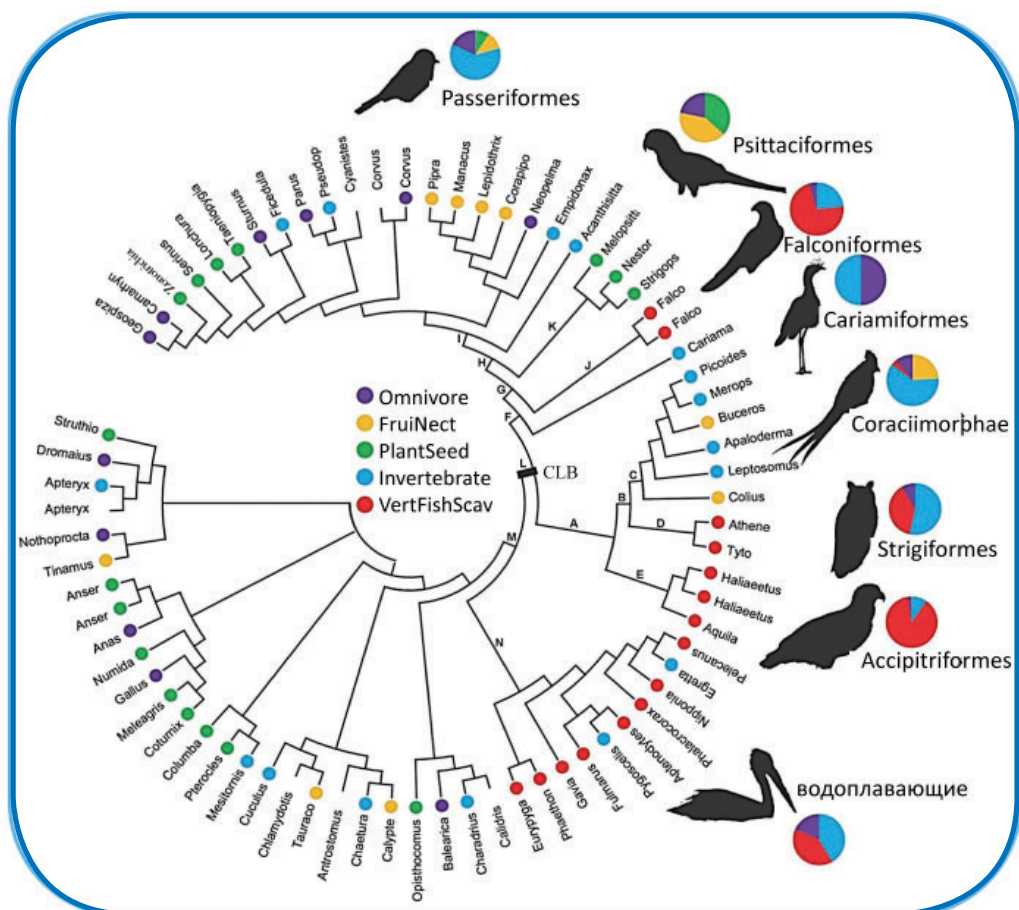
Акустическая коммуникация – один из самых заметных способов передачи сигналов у животных, хорошо задокументированный у насекомых, костистых рыб, лягушек, птиц, китообразных и наземных млекопитающих. Используя филогеномные данные, была проведена реконструкция эволюции слуховых и звукопроизводящих органов у прямокрылых, дающая представление о том, как акустическая коммуникация возникла и разнообразилась на протяжении нескольких сотен миллионов лет эволюции. В частности, авторы пришли к заключению, что в подотряде *Ensifera* (длинноусые прямокрылые) стридуляция на основе передних крыльев и большеберцовые тимпанальные уши эволюционировали одновременно, в то время как в подотряде *Caelifera* (короткоусые прямокрылые) брюшные тимпанальные уши сначала развились в несексуальном контексте, а позднее были кооптированы для передачи половых сигналов. Однако, свидетельства того, что эволюция слуха и органов, производящих звук, увеличила скорость диверсификации в тех линиях, где была известна акустическая коммуникация, ограничены (Song et al., 2020).

Для увеличения эффективности расселения на новые территории, многие растения в процессе эволюции приобрели специфические приспособления, например такие, как крыловидные выросты у семян. Филогеномный анализ семейства Ореховых показал, что плоды-крылатки были наследственным состоянием подсемейства *Juglandoideae*, но адаптация к новым режимам расселения и регенерации после рубяжа мела и палеогена привела к независимой эволюции зоохории среди нескольких родов *Juglandaceae* (Zhou et al., 2021).

**Биология видов.** Геномные исследования способствуют пониманию биологии вида, включая морфофизиологические особенности, диетические предпочтения, образ жизни и адаптацию к условиям среды обитания. В частности, геномные данные являются ценным ресурсом для изучения генетических механизмов адаптации птиц-хищников к окружающей среде, особенно обоняния, ночного образа жизни и реакции на УФ-излучение. Чтобы изучить эволюционную основу охотничьего поведения и адаптации ястребиных и сов к разным световым циклам, *de novo* были собраны черновые геномы индийского перепелятника *Accipiter virgatus* Temmink, 1822 и восточной сплюшки *Otus sunia* (Hodgson, 1836). Сравнение новых и имеющихся в базе данных геномов птиц обнаружило четыре гена, находящихся под действием положительного отбора: *XRCC5*, *PRIMPOL*, *MDM2* и *SIRT1*, связанных с реакцией ястребиных на ультрафиолетовое излучение, и один ген, *ALCAM*, связанный с развитием сетчатки у сов, что согласуется с их соответственно дневной и ночной хищнической активностью. Кроме того, была выявлена диверсификация репертуаров обонятельных рецепторов хищников, что может отражать важную роль обоняния в их образе жизни. Сравнение числа генов *TAS2R* (связаны с горечью вкуса) показало, что пищевые токсины были основной селективной силой, формирующей разнообразие репертуаров *TAS2R* у птиц. Меньшее количество генов *TAS2R* у хищников отражало их плотоядную диету, поскольку ткани животных с меньшей вероятностью содержат токсины, чем растительная пища (Zhou et al., 2019).

Исследования основных наземных птиц (Telluraves – хищные птицы, дятлы, попугаи и певчие птицы) с помощью секвенирования большого набора генов, связанных с формой клюва, зрением, слухом, языком, ощущением температуры, вкуса, а также перевариванием и поглощением углеводов, белков и жиров, позволили обнаружить полногеномные адаптивные дифференциации, лежащие в основе их экологической диверсификации. Сравнительный молекулярно-филоэкологический анализ (рис. 4.23) показал, что разные наземные птицы проявляют усиление адаптаций в разных аспектах, с формированием двух общих закономерностей. Хищные демонстрируют конвергентное адаптивное усиление переваривания и всасывания жиров, в то время как нехищные птицы, как правило, проявляют повышенную способность к перевариванию и всасыванию белков и углеводов. Результаты также показали относительно сильную поддержку хищнического образа жизни общего предка основных наземных птиц, предполагая единое происхождение хищников, за которым последовали две вторичные потери хищничества у основных наземных птиц. Кроме того, в соответствии с предыдущими результатами (см. выше) у дневных птиц наблюдалась тенденция к адаптивному усилению зрения при ярком свете, а у ночных – к зрению при тусклом освещении (Wu et al. 2021).

Относительно простая архитектура геномов птиц способствует изучению молекулярной основы их специфических черт. Например, сравнительный геномный анализ выявил конвергентный положительный отбор по 227 генам, из которых 73% экспрессируются в головном мозге, а также увеличение скорости замен в регуляторных областях в 2–3,5 раза при обучении вокалу. У птиц семейство генов  $\beta$ -кератина



**Рисунок 4.23.** Филогенез птиц и их диетические категории (из: Wu et al., 2021).

Категории диеты показаны разным цветом; пробелы означают, что информация о диете недоступна. Диаграммы показывают соотношение числа видов с различными типами питания для целевых групп птиц: PlantSeed (растения и семена), FruiNect (плоды и нектар), Invertebrate (беспозвоночные), VertFishScav (позвоночные, рыба и падаль) и Omnivore (оценка  $\leq 50$  во всех четырёх категориях). CLB – основные наземные птицы; буквы А–N показывают ветви, прошедшие анализ положительного отбора на основе всего набора данных о птицах в этом исследовании.

увеличилось почти в два раза, при этом кластер генов пера содержит 56% генов бета-кератина, а 55% генов окостенения находятся под положительным отбором. Деляция экзонов со сдвигом рамки считывания у генов эмали и дентина ответственны за потерю зубов, ген белка гомеобокса ALX1 (специфическая функция у человека не определена) связан с формой клюва, а ген негистонового белка хроматина HMGA2 (у человека связан с ростом и образованием опухолей) – с размером клюва. Зрение у птиц преимущественно тетрахроматическое, с четырьмя классами основных генов опсинов (Stiller, Zhang, 2019).

Диета считается самой важной силой отбора для всех видов, когда-либо живших на Земле, и с особенностями питания связан целый комплекс морфофизиологических особенностей. В летописи окаменелостей специализация на плотоядности часто связана с относительно коротким временем вымирания, поэтому особый интерес представляет изучение молекулярной адаптации к хищничеству (Kim et al., 2016). Углублённый анализ всего генома представителей трёх диетических групп млекопитающих – плотоядных, всеядных и травоядных – значительно улучшил понимание эволюции их диетических специализаций. Сравнивали 18 репрезентативных геномов млекопитающих, сосредоточив внимание на геномах кошачьих (домашняя кошка, тигр, лев, гепард и леопард), гоминидах и полорогих. Помимо явного сокращения в семействах генов, отвечающих за метаболизм крахмала и сахарозы, геномы плотоядных животных имели свидетельства общих эволюционных адаптаций в генах, ответственных за успешную охоту и потребление мяса. Причём анализ высококонсервативных регионов на уровне семейства выявил молекулярные признаки специфической пищевой адаптации у кошачьих, гоминид и полорогих. Однако, в отличие от плотоядных, всеядные и травоядные имели меньше общих адаптивных признаков, вероятно потому, что плотоядные животные находятся под сильным селективным давлением, обусловленным диетой. Наблюдаемое сокращение генетического разнообразия у кошачьих, связанное с уменьшением размеров популяции, может быть следствием негибкого характера строгой диеты, что подчёркивает их уязвимость и критический статус охраняемых видов (Kim et al., 2016). Согласно сравнительному геномному анализу, в основе биологии и одомашнивания кошек лежит положительный отбор генов, участвующих в метаболизме липидов (основа в адаптации к гиперхищной диете), сенсорных процессах (особенно влияющих на зрение и слух у хищников) и связанных с нервным гребнем (синдром одомашнивания). Кроме того, существует эволюционный компромисс между функциональным репертуаром генов обонятельных и вомероназальных рецепторов в геномах кошек и собак с расширением хемосенсорной системы кошек для обнаружения феромонов через запахи (Montague et al., 2014). В целом, эти результаты создают основу для будущего картирования болезней и филогеномных исследований всех представителей семейства кошачьих, а также предоставляют полезный ресурс для исследований в области генетики и здоровья, связанных с питанием (Montague et al., 2014; Kim et al., 2016).

Филогеномный анализ оказался успешным в выявлении таких важных эволюционных событий, как *геномные дупликации и гибридизация*. Полногеномные дупликации происходят либо вследствие гибридизации двух разных родительских видов (аллополиплоидизация, см. выше), либо в результате удвоения одного и того же генома (автополиплоидизация), которые имеют разные цитогенетические последствия (Redmond et al., 2023). В 1970 г. Сусуму Оно в своей книге “Evolution by gene duplication” впервые высказал предположение, что полногеномные дупликации могут быть двигателями эволюционного процесса (Ohno, 1970). Распределённые по всему древу жизни, события WGD особенно хорошо изучены у растений, дрожжей и позвоночных (Farhat et al., 2023; Redmond et al., 2023). Считается, что у цветковых

растений имеются следы четырёх полногеномных дупликаций ~150–200 млн л. н., а дупликации генома позвоночных происходили на ранних этапах их эволюции ~500–600 млн л. н. (Байрамов и др., 2021). Геномы позвоночных, включая млекопитающих, подверглись двум раундам WGD, а виды костистых рыб испытали затем ещё 1–3 раунда WGD (Lu et al., 2023a).

Являясь драматическим эволюционным событием в генетическом аспекте, WGD обычно сопровождается массивной потерей или дупликацией генов и другими геномными перестройками. Эти геномные преобразования подразумевают появление новшеств (включая новые гены и гены с подмножеством исходной наследственной функции) у полиплоидных видов, в том числе способствующих выживанию в условиях массовых вымираний (Farhat et al., 2023; Redmond et al., 2023). Будучи также крупномасштабным событием, WGD изменяют геномы значительно сильнее, чем мелкомасштабные геномные дупликации или дупликации единичных генов, и у позвоночных они вносят значительный вклад в видообразование и видовое разнообразие (Lu et al., 2023a). Принципиально важно, что у аутополиплоидов процессы дупликации генома и генов разделены процессом редиплоидизации, которая может быть асинхронной по всему геному и растянутой по времени на десятки миллионов лет. При этом редиплоидизация может происходить независимо в линиях потомков одной и той же WGD, если видообразование произошло до завершения редиплоидизации у предка, уровень которой также имеет важные эволюционные последствия (Redmond et al., 2023).

Для предсказания и визуализации геномных дупликаций создана база данных “DupScan”, в которую интегрировали 106 геномов на уровне хромосомной сборки. Она позволяет анализировать и визуализировать синтении в масштабе генома и отдельных генов, а также показывать значения Ka/Ks (соотношение частоты несинонимичных и синонимичных замен) и 4DTV (четырёхкратно вырожденный синонимичный сайт), и просматривать геномы. Являясь универсальной аналитической платформой для исследования синтений и WGD для 106 видов с полногеномными последовательностями, DupScan помогла также выявить закономерности эволюции генома у более 60000 видов позвоночных с событиями WGD и синтенией (Lu et al., 2023a). Более ранние базы данных, связанные с синтенией и WGD, включают PGDD, PhyloFish и OHNOLOG, предназначенные для растений, костистых рыб и позвоночных соответственно (Lu et al., 2023a).

Возросшая экономическая доступность и повышение качества секвенирования (посредством технологий с использованием длинных ДНК-прочтений в сочетании с Hi-C – позволяет представить пространственную структуру хромосомных контактов в масштабе всего генома – или Omni-C, обеспечивающих более широкий и равномерный охват генома, увеличивая потенциал обнаружения особенностей генома, от SNP до крупномасштабных конформационных изменений) существенно упростили получение сборки на хромосомном уровне для разных организмов (Farhat et al., 2023). Например, сравнительный геномный анализ выявил два раунда недавних событий полногеномных дупликаций у циклокарии болотной *Cyclocarya paliurus* (Batalin)

Pijnsk – реликтового вида растений, листья которого традиционно использовались для лечения ожирения и диабета. Были обнаружены сайты, связанные с событиями изменения окружающей среды, и множество генов, участвующих в критически важных биологических функциях (таких как производство вторичных метаболитов и защита растений) (Qu et al., 2023).

Астериды – одна из самых успешных линий покрытосеменных растений, имеющая большое морфологическое разнообразие и включающая ряд важных сельскохозяйственных культур. Предок астерид был древесным наземным растением с простыми листьями, обоеполыми и актиноморфными цветками со свободными лепестками и свободными пыльниками, верхней завязью со столбиком и косточковыми плодами. Филогеномика и филотранскриптомика астерид раскрыли морфологическую эволюционную историю и предоставили неопровержимые доказательства 33 полногеномных дупликаций у Asteridae и одного – у Berberidopsidales, включая четыре надсемейных и семь семейных/подсемейных полногеномных дупликаций. Полученные результаты улучшают понимание филогении астерид и дают многочисленные новые эволюционные представления об их разнообразии и морфологической эволюции (Zhang et al., 2020). Филогеномный и сравнительный геномный анализы обнаружили клонально-специфическую WGD у многолетней полыни *Artemisia argyi* Lévl. Et Vant (широко используется в народной медицине многих азиатских стран), значительно улучшив представления об эволюции этого лекарственного растения. Недавняя WGD привела к появлению множества дублицированных генов, включая увеличение семейства генов терпинсинтазы (что могло способствовать разнообразию летучих терпеноидов), и потери функции гомологов аморфа-4,11-диен синтазы, ADS (что могло привести к отсутствию синтеза артемизинина) (Chen et al., 2023).

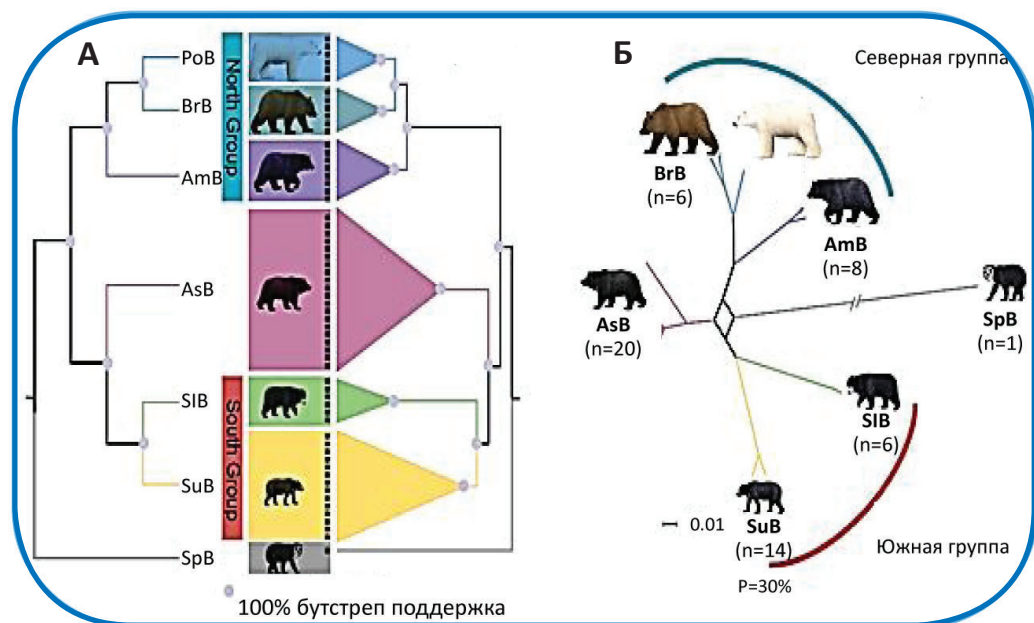
Привлекательными моделями для изучения полногеномных дупликаций и эволюции плоидности у рыб являются виды отряда Acipenseriformes, представляющие собой «живые ископаемые» со сложными геномами. Применение филогеномного подхода позволило пересмотреть время WGD по отношению к дивергенции осетров (Acipenseridae) и веслоносов (Polyodontidae) и привести убедительные свидетельства единственной наследственной аутополиплоидии (несмотря на существование множества независимых дупликаций генов), имевшей место во время, близкое к пермско-триасовому вымиранию, т.е. более 200 млн л.н. (Redmond et al., 2023; Wang et al., 2024).

WGD существенно повлияли на формирование различий между геномами растений и животных, которые наблюдаются в структуре семейств генов, скорости рекомбинации, вариантах сплайсинга, консервативных некодирующих последовательностях, в размере мультигенных семейств и других признаках (Murat et al., 2012). Примерно 20–60% наших генов совпадают с генами растений; еще меньше совпадений с геномом растений в некодирующих участках. В течение 20 млн л. после разделения у животных в среднем сохранилось в три раза больше генов (40–50%), кодирующих белки, чем у растений (10–20%). Это может быть напрямую связано с массивной потерей дублированных генов после WGD, и потому порядок генов у

растений более нарушен, чем у животных за аналогичные периоды времени. Например, отчётливую картину сохранения коллинеарности при 70–80% сохранении генов наблюдали после предполагаемых 65 млн л. эволюции у однодольных растений и ~450 млн л. у животных. Перетасовка генов после дупликации у растений стимулирует ремоделирование генома, а у животных – увеличение вариантов сплайсинга. В целом, из-за различных эволюционных историй современные геномы животных (не имеющие или испытавшие лишь несколько раундов WGD) в два раза более консервативны, чем геномы растений (все диплоидные виды покрытосеменных являются диплоидизированными древними полиплоидами) в отношении содержания генов. Эволюция цветковых растений через ряд WGD объясняет специфические взаимодействия между miRNA и связанными с ними мишенями. Причём модель появления miRNA у растений включает в себя генные дупликации на основе шпилек, а у млекопитающих вновь сформировавшиеся miRNA предположительно происходят из шпилек РНК. Кроме того, древние и короткие элементы TE класса I характерны для современных геномов животных, в то время как длинные ретротранспозоны LTR оказались особенно успешными колонизаторами хромосом растений. При этом плотность TE однородна в геноме человека, тогда как у растений вставки и редукции путём незаконной рекомбинации привели к появлению горячих точек TE. Полиплоидия, более частая у растений, чем у животных, может вызывать генетические и эпигенетические изменения, приводящие к постоянной реструктуризации и перепрограммированию генома. Таким образом, вследствие различных палеоэволюционных моделей и процессов, геномы растений представляются значительно более динамичными и быстро эволюционирующими, а геномы млекопитающих – более консервативными и стабильными (Murat et al., 2012).

Понимание масштабов гибридизации и роли гибридизации в эволюции является важным направлением современных эволюционных исследований. Полногеномные данные имеют решающее значение для определения видов с ретикулярной эволюцией, поскольку позволяют учитывать неполную сортировку по линиям и гибридизацию при построении видовых деревьев. Многочисленные филогенетические исследования не смогли прийти к консенсусу по многим аспектам эволюционных отношений меченосцев и пятитыпцилий (Poeciliidae, род *Xiphophorus* Heckel, 1848, подроды *Xiphophorus* и *Platypoecilus* Günther, 1866). Используя геномные данные, были получены надёжные реконструкции филогенетических связей видов *Xiphophorus*, которые с высокой достоверностью объясняют как неполную сортировку линий, так и гибридизацию. Эти результаты позволяют пересмотреть давние разногласия относительно эволюции отбираемого половым путём меча у *Xiphophorus* и показать, что гибридизация была поразительно широко распространена в эволюционной истории этого рода. Примечательность примера состоит в том, что исследуемая группа видов характеризуется не только значительным морфологическим разнообразием, но также поведенческими барьерами для межвидового скрещивания (Cui et al., 2013).

Медведи имеют сложную модель видообразования и быструю эволюцию различных фенотипов. Обычная для медведей межвидовая гибридизация сформировала их



**Рисунок 4.24.** Филогенетические отношения шести видов Ursinae (А) и филогенетический сетевой анализ с пороговым уровнем 30% (Б) (из: Zou et al., 2022).

А – имеет 100% поддержку для всех ветвей межвидового уровня; в качестве внешней группы использован очковый медведь; Б – показывает сетчатую эволюционную историю азиатского чёрного медведя.

PoB – белый медведь (*Ursus maritimus* Phipps, 1774); BrB – бурый медведь (*Ursus arctos* Linnaeus, 1758); AmB – американский чёрный медведь (*Ursus americanus* Pallas, 1780); AsB – азиатский чёрный медведь (*Ursus thibetanus* G. Cuvier, 1823); SuB – солнечный медведь (*Helarctos malayanus* Raffles, 1821); SIB – медведь-ленивец (*Melursus ursinus* Shaw, 1791).

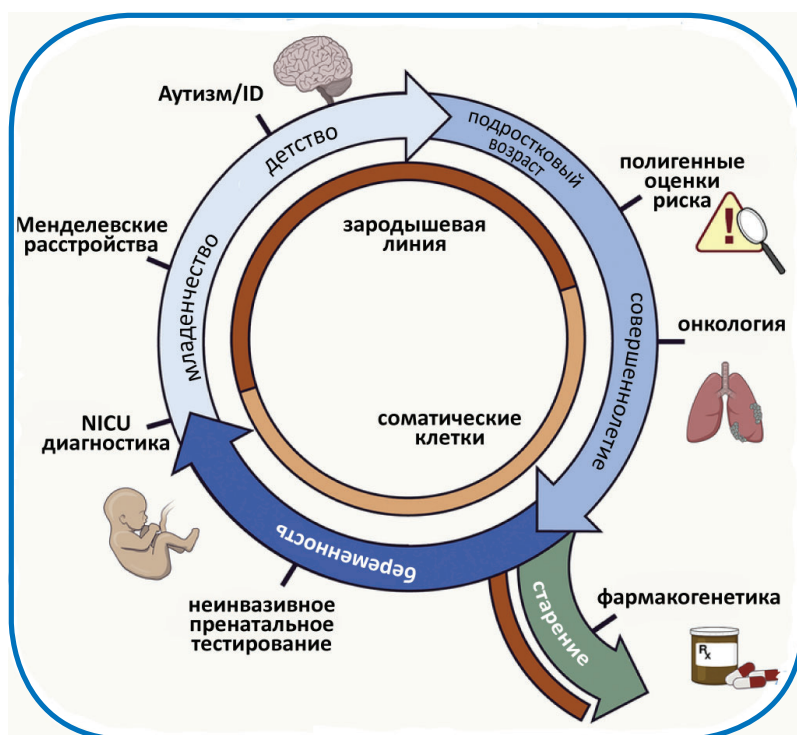
сложную эволюционную историю со следами интрогрессии в палеогеномах. Недавние исследования филогенетических отношений (рис. 4.24), основанные на большом наборе геномных данных и новых методах обнаружения гибридного видообразования, обнаружили весомые доказательства гибридного происхождения азиатских чёрных медведей. Хотя считается, что этот тип видообразования редко встречается у млекопитающих, в отличие от растений, которые могут приспосабливаться к полиплоидии. Выяснилось, что чёрные азиатские медведи произошли в результате исторической гибридизации между предками двух групп животных с контрастными размерами тела: (1) белый медведь, бурый медведь и американские чёрные медведи и (2) солнечный медведь и медведь-ленивец. Полагают, что репродуктивная изоляция от родительских линий, вероятнее всего, произошла за счёт альтернативного наследования дивергентных локусов. Кроме того, были идентифицированы ключевые гены, участвующие в фенотипе промежуточного размера тела азиатского чёрного медведя. Это исследование подчеркивает роль гибридизации как «катализатора формирования новых видов и фенотипической эволюции млекопитающих» (Zou et al., 2022).

## ГЛАВА 5 ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ ГЕНОМИКИ

Значительные достижения в области технологий секвенирования и сборки геномов стимулировали большой интерес к геномным исследованиям, включая создание множества новых международных проектов по изучению геномов разных организмов (табл. 5.1). В результате, за последнее время накоплен обширный материал о геномах и получены ответы на многие актуальные теоретические и практические вопросы о видах и их взаимодействии с окружающей средой, о происхождении, эволюции и болезнях человека. Некоторые из данных находят практическое приложение уже сейчас, другие создают прочную основу для дальнейших усовершенствований и новых разработок.

### 5.1. Медицина

Востребованная в течение всей жизни человека, от пренатального развития до глубокой старости, геномная медицина (рис. 5.1) известна как предиктивная, прецизионная и персонализированная медицина. Предиктивная (прогностическая) медицина



**Рисунок 5.1.** Геномная медицина на протяжении всего жизненного цикла человека (из: Shendure et al., 2019).

NICU – отделение интенсивной терапии новорожденных; ID – интеллектуальная инвалидность.

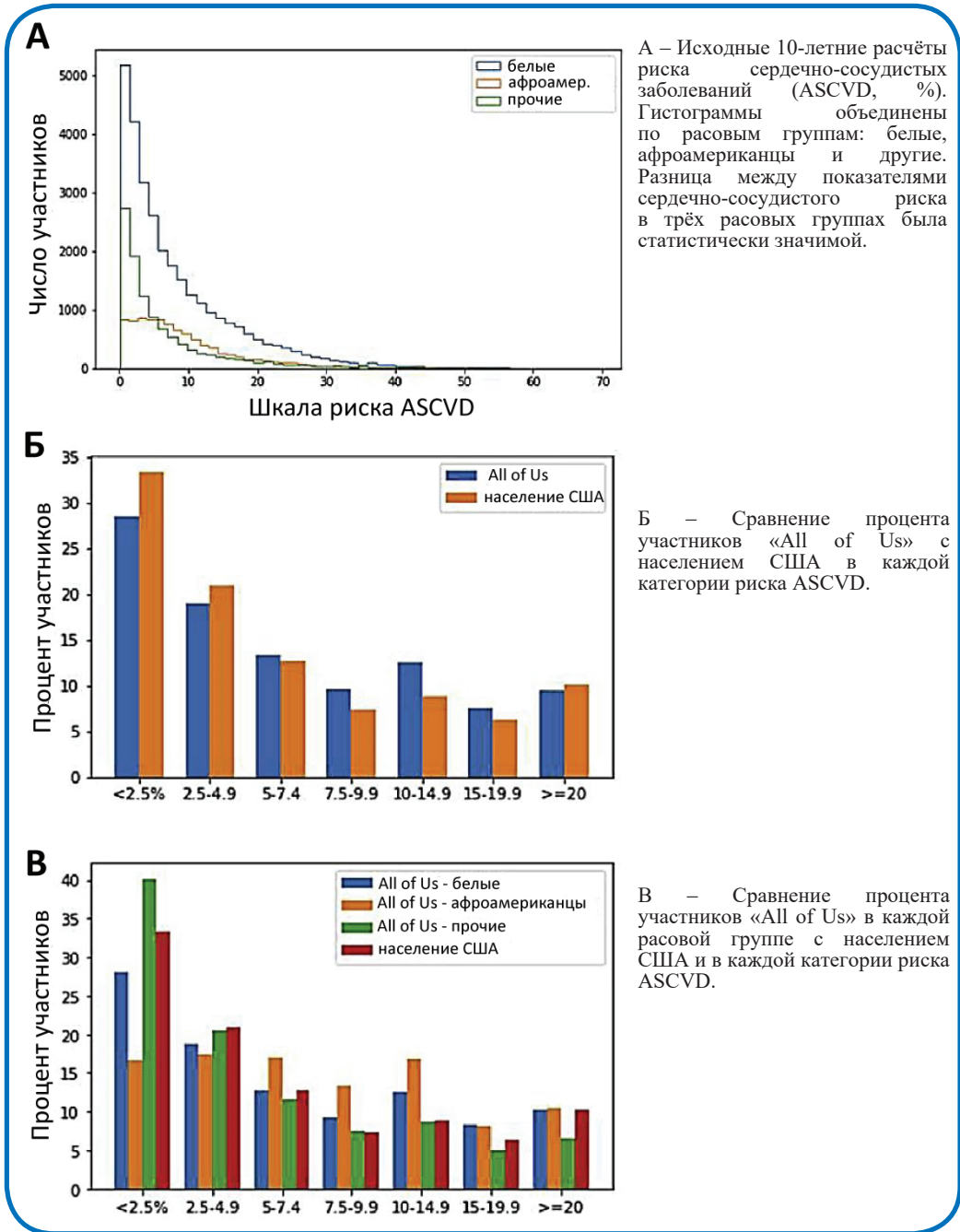
**Таблица 5.1.** Примеры масштабных геномных проектов

Дата	Название	Цели
1990	The International Human Genome Project	Расшифровка полного генома человека
2002	The International HapMap Project	Выявление общих генетических вариаций среди народов и вариантов, влияющих на распространённые заболевания, для улучшения диагностических инструментов и выбора наиболее подходящих терапевтических мишеней
2008	1,000 Human Genome Project	Лучшее понимание генетической изменчивости человека
	1,000 Plant Genome Project	Секвенирование генома около 1000 видов растений по всему миру
2009	10,000 Genome Project	Обеспечение основы для разнообразных исследований генома человека в будущем
2011	5,000 Insect Genome Project	Лучшее понимание биологии насекомых; совершенствование управления членистоногими, угрожающими здоровью человека и экономической безопасности
2012	100,000 Genome Project	Улучшение знаний о случаях и лечении редких, онкологических и инфекционных заболеваний
	100,000 Pathogen Genome Project	Секвенирование патогенных микроорганизмов для использования полученных данных в здравоохранении
2015	Bird 10,000 Genomes Project	Создание черновиков геномов всех существующих видов птиц, чтобы понять их эволюцию, разнообразие, поведение и т. д.
2017	Vertebrate Genome Project	Создание эталонного генома для каждого существующего вида позвоночных
	Bat 1,000 Genome Project	Секвенирование генома всех живых видов летучих мышей для выявления молекулярной основы их адаптации (например, уникальный иммунитет и сокращённый геном)
	1,000 Fungal Genome Project	Широкомасштабное исследование биологии грибов
2018	Earth BioGenome Project	Секвенирование и каталогизация геномов всех описанных видов эукариот
	Darwin Tree of Life Project	Секвенирование всех видов эукариот в Великобритании и Ирландии.

использует информацию, предоставляемую персональной геномикой при выборе схемы лечения для конкретного человека. Прецизионная (точная) медицина описывает диагноз заболевания с более высокой точностью, используя геномные и другие технологии, т. е. представляет новую систему болезней человека, основанную на молекулярной биологии. В 2018 г. в США началось масштабное исследование в области прецизионной медицины под названием «All of Us» с целью привлечь один миллион участников. Полученные к настоящему времени результаты в рамках этого проекта включают различия в использовании лекарств в зависимости от расы при депрессии и диабете второго типа, подтверждение известных ассоциаций онкологических заболеваний с курением, а также расчёт показателей риска сердечно-сосудистых заболеваний, учитывающий расовые особенности (рис. 5.2) (All of Us Research Program Investigators, 2019). Современная персонализированная медицина, отождествляемая с прецизионной и предиктивной, использует методы профилактики, диагностики и лечения заболеваний, основанные на индивидуальных генетических особенностях пациента и таких факторах, как история здоровья семьи, род занятий и окружающая среда. Персонализированная медицина имеет древнюю историю – выдающиеся врачи прошлого рассматривали каждого пациента как уникальную комбинацию факторов наследственности и приобретений, а её развитие тесно связано с достижениями в области геномики (Ginsburg, Willard, 2009; Roth, 2019).

По имеющимся на сегодня оценкам, в геноме человека насчитывается от 20000 до 25000 генов с размером от нескольких сотен до более 2 млн п. н., что немного больше, чем у червей, но меньше, чем у таких видов, как рис или пшеница. Следовательно, геном отражает не только количество генов, но и сложность генных сетей (Roth, 2019). По образному выражению Сиддхартха Мукерджи, *«геном человека яростно изобретателен, он динамичен, его части удивительно красивы, он пронизан историей, он непостижим, уязвим, устойчив, адаптируем, повторяем и уникален. Он готов развиваться. Он завален обломками своего прошлого. Ему суждено выжить. Он похож на нас.»* (Roth, 2019).

Фармакогеномика является частью персонализированной медицины, которая изучает влияние индивидуальных генетических различий человека в его ответной реакции на лекарственные препараты, обеспечивая тем самым максимальную эффективность лечения при минимальных побочных эффектах. База знаний по фармакогеномике (PharmGKB) – это ресурс, который собирает, обрабатывает и распространяет клинически значимую информацию о влиянии генетической изменчивости человека на лекарственную реакцию. Важным аспектом работы этой базы является присвоение уровня доказательности для ассоциаций между генетическими вариантами и препаратами. Таким образом, PharmGKB является источником надёжной информации, поддерживающей проекты по внедрению персонализированной медицины (Whirl-Carrillo et al., 2012). Для лучшего понимания патогенеза аутоиммунных заболеваний, иммунного отторжения после трансплантации органов, реакции «трансплантат против хозяина» и реакции «трансплантат против лейкемии» после



**Рисунок 5.2.** Расчёты риска сердечно-сосудистых заболеваний (ASCVD) на основании геномных данных пациентов (из: Ramirez et al., 2022)

пересадки костного мозга, а также после трансплантации гемопоэтических стволовых клеток, при пищевой аллергии и иммунных реакциях после вакцинации и лечения онкологических заболеваний могут применяться подходы иммуногеномики/иммунофармакогеномики, включая иммунотерапию. Эти подходы также важны для создания методов прогнозирования лекарственной эффективности или нежелательных побочных эффектов после лечения (Zewde et al., 2018).

Применение полногеномного секвенирования в клинической медицине изменило диагностику, лечение и профилактику заболеваний человека (Van et al., 2013). В частности, WGS помогает выявлять генетические варианты, требующие раннего медицинского вмешательства (Dewey et al., 2014), и обеспечивает неинвазивный пренатальный скрининг ДНК на хромосомные анеуплоидии и другие генетические нарушения (Гольцов и др., 2019). Функциональная геномика способствует выяснению механизмов, лежащих в основе таких заболеваний, как саркопения (возрастная потеря мышечной массы и силы, этиология которой остаётся неясной), и открытию новых терапевтических мишеней. Например, известно, что структурные перестройки генома (дупликации, делеции, инверсии и транслокации) способны вызывать изменения числа копий (CNV) определённых сегментов ДНК, и это может привести к аномальному числу одного или нескольких генов. Полногеномный анализ структурных вариантов у пациентов с саркопенией помог идентифицировать ген *gremlin1* и два новых CNV (CNV1191 и CNV2580), а также родственных генов, потенциально важных для изменения мышечной массы аппендикулярного отростка и, возможно, лежащих в основе фенотипов саркопении (Melouanea et al., 2019).

Геномные исследования показали, что факторы окружающей среды (включая токсины, стресс и необычное питание) могут приводить к эпигеномным изменениям зародышевой линии, и существует отчётливая связь между эпигеномными вариациями и фенотипическими, экологическими и патологическими признаками (Hjort et al., 2021; Lu et al., 2021). Полногеномное, индуцированное окружающей средой (воздействие ДДТ или винклозолина), трансгенерационное эпигенетическое наследование болезней наблюдали у мелких млекопитающих (Luikart et al., 2018). Дети большинства женщин с посттравматическими стрессовыми расстройствами, как правило, имели более высокий уровень кортизола и дифференциальное метилирование в генах-кандидатах *NR3C1*, *HTR3A* и *BDNF157*, участвующих в регуляции глюкокортикоидов, серотонина и окситоцина соответственно (Hjort et al., 2021). Исследование пар монозиготных и дизиготных близнецов выявило связь между потреблением алкоголя и метилированием ДНК крови в масштабе всего генома. При этом оказалось, что гиперметилирование CpG сайта cg07326074, расположенного в способствующем развитию опухоли гене *C16orf59*, связано с потреблением алкоголя (Lu et al., 2021).

**Онкологические заболевания.** Рак занимает второе место среди основных причин смертности в мире, и, по прогнозам, заболеваемость онкологическими заболеваниями со временем будет только возрастать. Для усовершенствования диагностики, лечения и предотвращения развития злокачественных опухолей были инициированы

проекты «Раковый геном» (CGP) и «Атлас ракового генома» (TCGA), направленные на изучение основ молекулярных механизмов развития рака. Однако разработка новых методов лечения осложняется высокой меж- и внутриопухолевой гетерогенностью, что является основным препятствием для эффективной терапии. Новые данные свидетельствуют о том, что трёхмерная архитектура генома (см. выше) является ещё одним уровнем межпациентной гетерогенности (Johnston et al., 2019; The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020). Мета-анализ геномных особенностей разных типов опухолей (для 2658 геномов, представляющих 38 типов рака) привёл к созданию базы международного обмена данными, известной как «Панраковый анализ полных геномов» (PCAWG). С помощью этого ресурса было обнаружено, что (1) в среднем геномы клеток опухолей содержат 4–5 управляемых мутаций, (2) онкологические заболевания с аномальным сохранением теломер часто происходят из тканей с низкой репликативной активностью, а (3) редкие варианты зародыша влияют на характер соматических мутаций. Результаты исследований в рамках геномных проектов по изучению рака легли в основу созданных онлайн баз данных о соматических мутациях для разных типов опухолевых заболеваний – «Каталог соматических мутаций при раке» (COSMIC), а также чувствительности опухолевых клеточных линий к противораковым препаратам – «Геномика чувствительности к лекарственным средствам при раке» (GDSC). Использование технологий РНК-интерференции помогло идентифицировать более десяти потенциальных супрессоров опухоли, а скрининг с помощью CRISPR-Cas9 выделил ген *Ptpn2* (кодирующий протеинтирофосфатазу нерецепторного типа 2) в качестве перспективной мишени для иммунотерапии рака (Manguso et al., 2017; Li, Kasinski, 2020). Технология CRISPR-Cas9 также важна для разработки лекарств, а благодаря способности вырезать и заменять последовательности ДНК, она подаёт большие надежды на лечение некоторых типов рака, ВИЧ, β-талассемии, серповидно-клеточной анемии и других заболеваний. Клинические применения редактирования генов соматических клеток (т. е., которые не передают наследственные мутации будущим поколениям) человека уже разрешены, но запрещено генетически модифицировать ДНК человеческих эмбрионов, используемых для имплантации, поскольку такие изменения в клетках зародышевой линии остаются постоянными, и их последствия в настоящее время непредсказуемы (Roth, 2019). Особый научный и практический интерес представляет вопрос о влиянии полногеномных дупликаций и полиплоидии (факторы хромосомной нестабильности) на приспособленность клеток и прогрессирование раковых заболеваний. Исследования показали, что WGD происходит на ранних стадиях трансформации клеток, позволяя клеткам впоследствии стать анеуплоидными, что приводит к прогрессированию рака. Согласно другим данным, полиплоидия играет роль супрессора опухолей, вызывая остановку клеточного цикла, старение клеток, апоптоз и даже стимулируя дифференциацию клеток в зависимости от типа клеток ткани (Sharma, 2023).

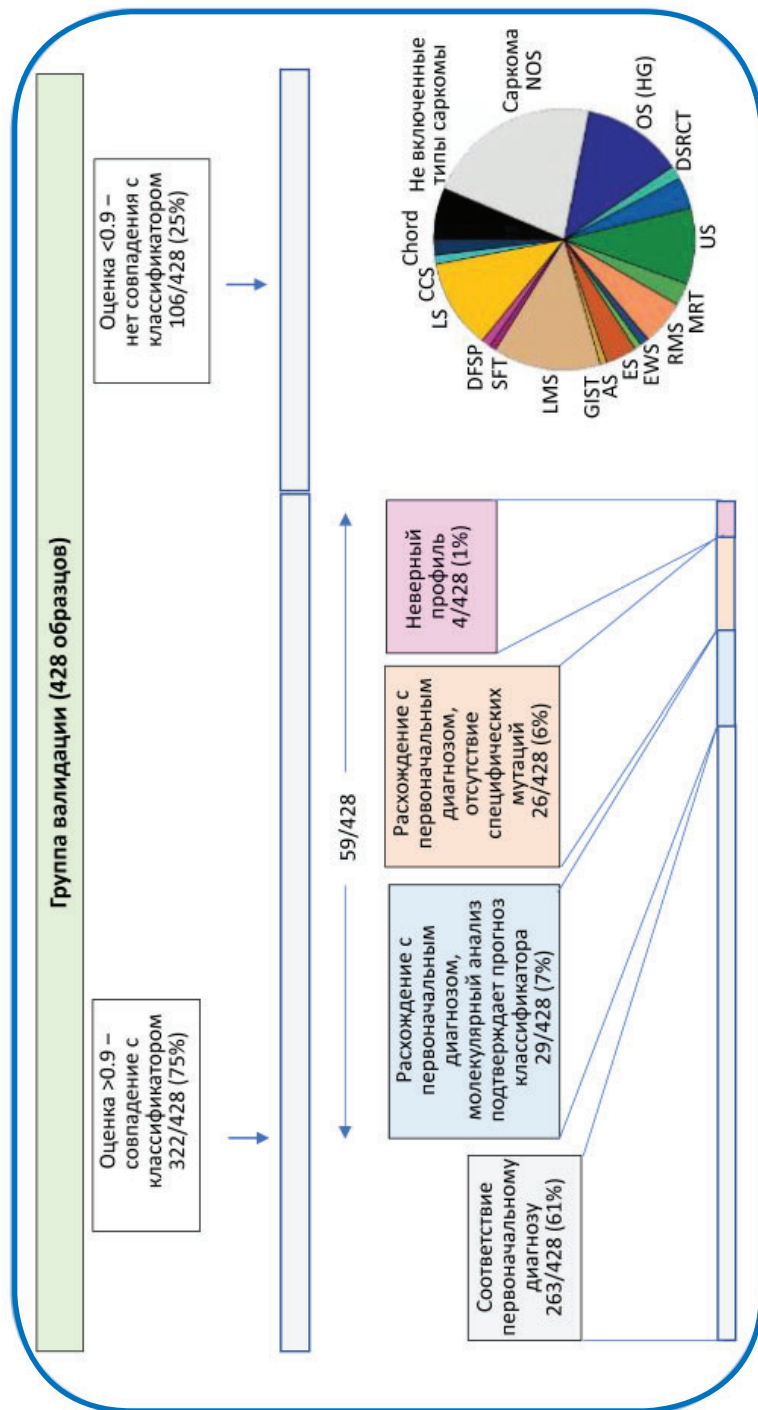
Саркомы – это злокачественные опухоли мягких и костных тканей, поражающие взрослых, подростков и детей. Они представляют собой морфологически гетероген-

ный класс новообразований, у некоторых из которых отсутствуют определяющие гистопатологические признаки, что часто приводит к ошибочному диагнозу. Недавно был разработан новый метод их систематизации с помощью алгоритма классификатора машинного обучения на основе данных о метилировании ДНК. Использовали набор из 1077 профилей (включающих 62 класса) метилирования генома опухолей у всесторонне изученных клинических случаев (составляющих широкий спектр подтипов сарком в разных типах тканей по всему возрастному спектру). Результаты продемонстрировали высокий потенциал классификации сарком на основе метилирования ДНК для исследований и будущих диагностических приложений. Эффективность подтверждена в когорте из 428 саркоматозных опухолей, из которых 322 случая были категорированы классификатором саркомы (рис. 5.3) (Koelsche et al., 2021).

Большинство типов рака являются заболеваниями с глобальной эпигенетической дисрегуляцией и представляют серьёзную проблему для диагностики и клинического лечения. Сосредоточив внимание на эпигеномных (помимо геномных и транскриптомных) изменениях при раке, онкогеномика помогает идентифицировать новые онкогены и гены-супрессоры опухолей, на которые можно ориентироваться для улучшения диагностики, лечения и прогнозирования клинических исходов. Изменения эпигенома могут быть связаны и с другими заболеваниями, например, болезнью Альцгеймера, болезнью Паркинсона и диабетом второго типа (Holtzman et al., 2021). Последние технологические достижения, позволяющие изучать эпигеном на уровне одной клетки, открывают перспективу разработки новых и эффективных стратегий борьбы с такими заболеваниями (Berkel, Casan, 2019; Bond et al., 2020). Анализ структуры генома с высоким разрешением выявляет новые терапевтические мишени при ряде заболеваний, в том числе глиобластоме, наиболее распространённой злокачественной опухоли головного мозга у взрослых (Shergalis et al., 2018).

**Психические заболевания.** У человека примерно 70% генов экспрессируются в головном мозге, а генетические вариации этих генов связаны с 40% клинических состояний (Plomin, Spinath, 2004). Психические расстройства, поражающие более 25% населения планеты, лидируют среди причин инвалидности во всём мире. Поскольку сотни генов вовлечены в когнитивные функции, развитие мозга и межнейронную коммуникацию, использование инструментов когнитивной геномики существенно способствует раскрытию молекулярных основ психических и нейродегенеративных расстройств, улучшению диагностики, разработке лекарств и прогнозированию риска заболеваний (Smoller et al., 2019; Fitzgerald et al., 2020).

Несмотря на то, что многие психические расстройства имеют глубокие генетические корни, лежащие в их основе молекулярные механизмы выяснить полностью ещё предстоит. Комплексный функциональный геномный ресурс и интегративная модель человеческого мозга прояснили их возможные молекулярные механизмы, включая основные гены и пути прогнозирования расстройств (Wang et al., 2018b). На основе полногеномной оценки риска *de novo* было установлено, что мутации как в кодирующих, так и в некодирующих областях генома способствуют риску расстройства аути-



**Рисунок 5.3.** Валидация классификатора саркомы (по: Koelsche et al., 2021).

Всего проанализировано 426 независимых образцов саркомы; 75% соответствовали установленному классу метилирования ДНК с пороговым значением предсказания классификатора  $\geq 0.9$ , а 25% достигли порогового значения предсказания классификатора. AS – ангиосаркома, CCS – светлоклеточная саркома мягких частей, Chord – хордома, DFSP – дерматофибросаркома, ES – эпителиоидная саркома, EWS – саркома Юинга, GIST – желудочно-кишечная стромальная опухоль, LMS – лейомиосаркома, LS – липосаркома, MLS – миксоидная липосаркома, MRT – злокачественная рабдомиосаркома, NOS – точно не идентифицированная саркома, OS (HG) – конвенциональная остеосаркома высокой степени злокачественности, RMS – альвеолярная рабдомиосаркома, US – недифференцированная саркома

стического спектра (ASD). Причём мутации в некодирующих регионах генома, возникающие *de novo*, вероятно, могут нарушать транскрипцию генов (An et al., 2018). Полногеномные исследования восьми психических расстройств (таких как шизофрения, большое депрессивное расстройство, нервная анорексия и др.) выявили три группы взаимосвязанных нарушений и обнаружили 109 локусов, ассоциированных как минимум с двумя психическими заболеваниями (в том числе 23 локуса с плейотропными эффектами и 11 – с антагонистическими). Примечательно, что плейотропные локусы в генах, которые проявляют повышенную экспрессию в головном мозге на протяжении всей жизни, могут играть важную роль в процессах развития нервной системы (Smoller et al., 2019). Таким образом, высокопроизводительное геномное секвенирование и геномика с клеточным разрешением имеют решающее значение для понимания когнитивных расстройств и их эволюции (Berto et al., 2020). Наиболее перспективным в геномной медицине представляется интегративный подход, объединяющий данные геномики, протеомики и метаболомики, так как он позволяет полно описать весь патологический процесс (Van et al., 2013).

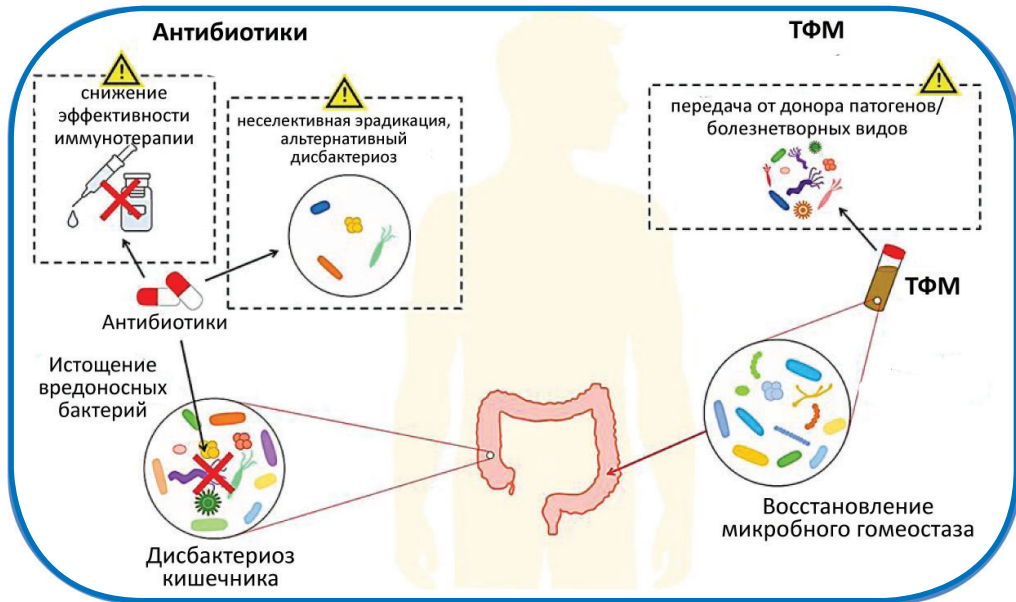
**Микробиомы.** Микробные сообщества очень персонализированы и считаются уникальными для каждого человека, как и отпечатки пальцев; они даже уникальны для каждого участка тела. На микробное разнообразие и численность влияют многие факторы, а их совокупность приводит либо к сбалансированной и разнообразной здоровой микробиоте, либо к несбалансированному и дисбиотическому составу. По-видимому, микробные токсины могут влиять на микробные взаимодействия и секретируемые метаболиты, которые, в свою очередь, способны изменить метаболизм хозяина и другие функции его организма. Взаимодействию микробиома и хозяина способствуют различные факторы, включая экологические, эпигенетические и генетические (Link, 2021; Mousa et al., 2022), а одним из самых спорных вопросов является определение «здорового» микробиома человека (Joos et al., 2024). Интересно, что согласно недавно проведённым исследованиям, микробиота человека является резервуаром мутаций SARS-CoV-2 (Cao et al., 2024).

Выделяют несколько ключевых концепций сложных взаимоотношений между микробами и человеком: (1) концепция «врождённых и адаптивных геномов» (способствует более глубокому пониманию генетических и эволюционных аспектов генома человека) и (2) «синдрома отсутствия микробов» (ставит под сомнение традиционное представление о микробах как исключительно патогенах, подчёркивая их важность для поддержания здоровья), (3) концепция «рабской ткани» (акцентирует внимание на сложных симбиотических отношениях и отмечает динамичное влияние микробных взаимодействий на здоровье) и (4) «приобретённого микробного иммунитета» (рассматривает микробиом как важный компонент иммунной системы человека, что служит обоснованием для применения пробиотической терапии и рационального использования антибиотиков), (5) «гипотеза гомеостатического перепрограммирования» (связывает микробиом с теорией внутренней среды, что открывает возможности для объяснения изменений в гомеостатических показателях в период постиндустриализации), (6) «модель клеточно-микробной коэкологии» (объ-

ясняет, как симбиотическая регуляция влияет на клеточный баланс), 7) концепция «мета-хозяина» (позволяет расширить определение хозяина, включая в него симбиотические микробы) и 8) «модель преобразования здоровья в болезнь» (объединяет врождённое и адаптивное взаимодействие геномов с закономерностями дисбактериоза). Задача концепций заключается в том, чтобы обеспечить более глубокое понимание микробиома, а также выделить перспективные направления исследований, которые могут привести к созданию более эффективной и результативной системы здравоохранения (Ma et al., 2024).

Уже в ходе ранних метагеномных исследований было обнаружено множество интересных фактов. Например, метагеномный анализ 124 человек с избыточным весом (здоровых и страдающих кишечными заболеваниями) из Дании и Италии показали, что микробиомы представителей группы с избыточным весом и кишечными заболеваниями имеют на 25% меньше генов и более низкое бактериальное разнообразие (Qin et al., 2010). Идентифицированы маркерные гены, например, значительно коррелирующие с возрастом, полом, или с индексом массы тела, т.е. имеющие диагностическую ценность (Arumugam et al., 2011). Установлено, что анализ перимплантного микробиома на основе метагеномного подхода также помогает прояснить некоторые важные аспекты микробиома полости рта; такая информация способствует правильному выбору стратегий лечения стоматологических заболеваний (Martellacci et al., 2019). Определены факторы, влияющие на силу вирулентности штаммов бактерии *Gardnerella vaginalis* Greenwood and Pickett 1980 при вагинозе у женщин (Yeoman et al., 2011), и связи между оральной микробиотой и атеросклерозом (Koren et al., 2011), а также продемонстрирован обмен факторами вирулентности между патогенными представителями рода бактерий *Neisseria* Trevisan 1885 (участвующими в развитии менингита, сепсиса и передающихся половым путём заболеваний) с их симбионтами (Marri et al., 2010).

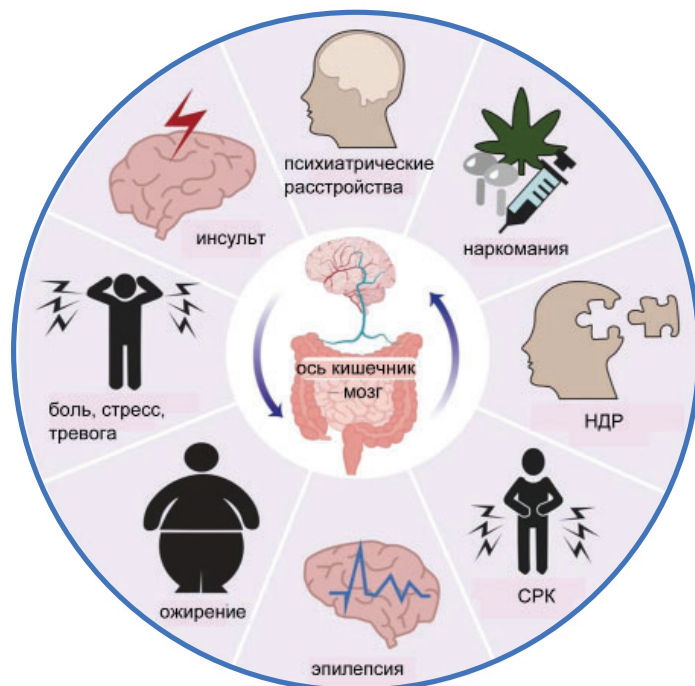
В настоящее время появились «микробно-центрированные терапевтические стратегии, которые обеспечивают или воссоздают желаемую функцию интересующего микробного сообщества, либо подавляют или устраняют определённый нежелательный или патогенный элемент» (Whitfill, Oh, 2019). Например, известны методики применения CRISPR–Cas9 системы для нормализации микрофлоры и предотвращения дисбактериоза у человека (Bikard et al., 2014). Пересадки фекальных микробиомов от здоровых доноров пациенту могут восстановить нормальную микробиоту при инфекциях *Clostridium difficile* Lawson and Rainey 2016 (Austin et al., 2014). Исследование роли микробиома кишечника при колоректальном раке является новой областью исследований, которая показала эффективность трансплантации фекальной микробиоты для купирования дисбактериоза и восстановления гомеостаза. Однако из-за присутствия множества неизвестных компонентов в образцах донора, такой подход также несёт в себе риск передачи реципиенту патогенов или генов, вызывающих заболевания (рис. 5.4) (Fong et al., 2020). Примечательно, что ещё в IV веке китайский даосский учёный и алхимик Ге Хун опубликовал книгу, в которой



**Рисунок 5.4.** Предполагаемые механизмы действия антибиотиков, трансплантация фекальной микробиоты (ТФМ) и связанные с ними проблемы безопасности (из: Fong et al., 2020).

впервые было описано лечение фекальной жидкостью пищевых отравлений и тяжёлой формы диареи (Yan et al., 2023).

Микробиом кишечника человека содержит наибольшее количество бактерий и может оказывать существенное влияние на метаболизм не только локально, но и системно. Микробный дисбиоз связан с возникновением и прогрессированием многих заболеваний, включая воспалительные заболевания кишечника, ожирение, нарушения обмена веществ, а также психические расстройства. Это так называемая «ось кишечник–мозг» или «ось кишечник–печень–мозг» (рис. 5.5). Наличие тесной связи между дисбактериозом в кишечнике и нейродегенеративными и нейроразвивающимися заболеваниями предполагает, что состав микробиома кишечника может быть использован для ранней диагностики, а его изменение – для лечения таких заболеваний, как болезнь Альцгеймера, болезнь Паркинсона, рассеянный склероз и т. д. (Mousa et al., 2022; Bicknell et al., 2023; Pezzino et al., 2023; Yan et al., 2023). Кишечник, печень и мозг образуют «трёхстороннее шоссе», обеспечивающее информационное взаимодействие между желудочно-кишечным трактом, печенью и нервной системой. В настоящее время применяются как общие терапевтические подходы, так и специализированные методы лечения, направленные на данную ось. Дальнейшие исследования взаимодействий внутри этой оси будут способствовать разработке новых профилактических стратегий и открытию точных терапевтических мишеней при многих заболеваниях (Yan et al., 2023).



**Рисунок 5.5.** Общая карта различных заболеваний оси кишечник–мозг  
(из: Yan et al., 2023).

НДР – нейродегенеративные расстройства, СРК – синдром раздражённого кишечника.

В последнее время активно обсуждается вопрос о наличии микробиома в мозге здорового человека, и высказано предположение, что резидентные микробы способны оказывать значительное влияние на многие нейропсихиатрические и нейродегенеративные заболевания. Кроме того, существует вероятность, что в мозге могут находиться ранее неизвестные микробы, которые не поддаются обнаружению с помощью стандартных методов. В качестве аргумента приводятся убедительные данные о новых микробиомных последовательностях, полученных в результате *de novo* сборки ДНК бесклеточной крови человека (Link, 2021; Bicknell et al., 2023). Однако если наличие микробиома у больного мозга имеет веские доказательства, то установление факта существования микробиома в здоровом мозге сталкивается с большими трудностями, начиная с возможной контаминации реактивов и неизбежным низкоуровневым загрязнением при извлечении самого образца (Link, 2021). Хотя в этом вопросе пока нет единого мнения, в недавно опубликованном исследовании, в котором был разработан и применён новый метод «электронного древа жизни» (eToL; основан на использовании зондов рибосомальной РНК малой субъединицы рибосом и дополнительным анализом рРНК большой субъединицы), сообщается о значительном разнообразии микробиома мозга как у пациентов с болезнью Альцгеймера, так и у здоровых людей (Hu et al., 2022; 2023).

Активно развивающуюся область исследований также представляет микрогендер, или половой диморфизм в микробиоме человека. Эксперименты сосредоточены на изучении двусторонних взаимодействий между микробиотой, половыми гормонами и иммунной системой, а также их влиянии на восприимчивость к заболеваниям и эффективность терапии (Ma, 2024). В частности, анализ структуры микробиомов показал, что у мужчин в кишечнике, по-видимому, содержится больше уникальных видов микробов, чем у женщин, в то время как в дыхательных путях, ротовой полости и на коже женщин, вероятно, больше уникальных микробных видов по сравнению с мужчинами. Наблюдаемые различия, очевидно, обусловлены половым диморфизмом в гормональной и иммунной системах. При этом важную роль в выделении и циркуляции половых гормонов, таких как эстроген и андрогены, играет микробиом кишечника, а половые гормоны также могут оказывать влияние на микробиом (Ma, 2024).

Микробиом человека представляет собой богатый источник разнообразных биоактивных натуральных продуктов, обладающих значительным потенциалом для терапевтического использования. В целях упрощения систематического исследования и анализа биосинтетического ландшафта человеческого микробиома был создан Атлас кластеров биосинтетических генов (ABC-HuMi). Эта база данных способствует глубокому пониманию молекулярного репертуара, закодированного в микробиоме человека, и представляет собой масштабный ресурс для открытия и характеристики новых биоактивных соединений (Hirsch et al., 2024).

**5.1.1. Инфекционные и паразитарные болезни.** Геномика и биоинформатика также имеют важнейшее значение для улучшения нашего понимания инфекционных заболеваний, включая патогенез, механизмы устойчивости к противомикробным препаратам, прогностические и диагностические маркеры, рекомендации по лечению и возможному взаимодействию лекарств и вакцин (Bah et al., 2018). Микробы сопровождали человека с момента его появления в Африке 200000–300000 л. н. и последующего расселения по миру, а эпидемии, вероятно, были частью нашей истории со времён первых предков. Однако частота эпидемических заболеваний увеличилась после того, как наши прародители занялись сельским хозяйством около 10000–12000 л. н. Этому способствовал ряд обстоятельств, таких как повышенные численность и плотность населения, плохая санитария и диетическая зависимость. Таким образом, бронзовый век и неолит были важным периодом биологической адаптации человека к патогенам (Kerner et al., 2021a; Larsen, Crespo, 2022). Геномы представляют собой молекулярную запись о происхождении и эволюции, филогеографии и адаптации древних патогенов. Палеогеномика существенно расширила знания о палеопатологии, так как не все и не всегда инфекционные заболевания оставляют следы повреждения костей (результат инфекций). Новые данные позволили идентифицировать возбудителей прошлых пандемий, открыть вымершие линии микробов, дать хронологию появления патогенов, реконструировать эволюционную историю некоторых современных эпидемиологически значимых видов (Spyrou et al., 2019).

Все эти знания необходимы для успешной борьбы с современными инфекционными заболеваниями.

Один из наиболее вирулентных патогенов, чумная палочка *Yersinia pestis* van Loghem 1944, стала причиной трёх крупных пандемий чумы: Юстиниановой чумы (VI–VIII вв. н. э.), Чёрной смерти (XIV–XVIII вв. н. э.) и третьей пандемии (XIX–XX вв.). Палеогеномные исследования возбудителя обнаружили, что генеалогии пандемий отличаются, и что современная родословная, вероятно, произошла от родословной, связанной со второй волной пандемии (Wagner et al., 2014; Feldman et al., 2016). Изучение семи евразийских штаммов *Y. pestis* возрастом 5000–2800 л. показало, что самый последний общий предок (MRCA) всех линий существовал более 5000 л.н. Древняя менее патогенная линия приобрела генетические изменения и стала высоковирулентной около 3000 л.н. У всех штаммов в период до 3600 л.н. отсутствовал ген *umt*, который кодирует белок, необходимый для поддержания жизнеспособности патогена в кишечнике его переносчика, блохи, в то время как все штаммы *Y. pestis* примерно с 2900 л.н. содержали интактную копию гена *umt*. Вместе с тем, последовательность генома *Y. pestis* с датировкой около 660 л.н., полученная от жертв Великой эпидемии чумы в Лондоне, не имела каких-либо нуклеотидных изменений, не характерных современным геномам. Эти данные предполагают, что потенциально важную роль в Пандемии Чёрной смерти могут играть поведение человека и экология, а не генетические изменения самого возбудителя (Rasmussen et al., 2015; Marciniak, Perry, 2017).

Широко распространённая в прошлом, инфекция палочки Коха *Mycobacterium tuberculosis* Lehmann and Neumann 1896 до сих пор остаётся глобальной угрозой. Относительно эволюционной истории заболевания существовали две гипотезы. Согласно одной из них, туберкулёз был завезён в Новый Свет постконтактным путём, согласно другой – туберкулёз появился в Новом Свете ещё до контакта с европейцами. Исследование трёх геномов туберкулёзной микобактерии возрастом 1000 л. показало, что патоген инфицировал население Южной и Северной Америки в доколумбову эпоху. Однако, эти микобактерии обнаружили наибольшее сходство с современными штаммами туберкулёза не человека, а мигрирующих тюленей южного полушария (Bos et al., 2014). Интересно, что анализ древней ДНК также подтвердил наличие генетического материала *M. tuberculosis* в останках бизона возрастом около 17870 л. (Joseph, Lindo, 2023). Данные филогенетических исследований позволяют предположить, что существовала предконтактная зоонозная линия, которую вытеснила европейская форма, принесённая колонистами (Brynildsrud et al., 2018). Недавно обнаружили, что гомозиготы по *P1104A* варианту гена тирозинкиназы (аутоиммунный защитный вариант, ограничивающий координирующие сигналы, необходимые для генерации специализированных субпопуляций Т-клеток) подвержены более высокому риску развития клинических форм туберкулёза. Исследования 1013 геномов древнего человека показали, что *P1104A* возник у общих предков западных евразийцев примерно 30000 л.н. Частота его заметно варьировала в последние 10000 л. (на фоне крупномасштабных миграций населения) с резким снижением по-

сле окончания бронзового века. Это падение частоты связано с сильным негативным отбором, который начался примерно 2000 л. н., с относительным снижением приспособленности гомозигот на 20%, и это один из самых высоких показателей для генома человека. Таким образом, палеогеномика установила, что туберкулёз стал тяжёлым бременем для европейцев за последние два тысячелетия (Kerner et al., 2021a).

Учитывая важность туберкулёза как глобальной угрозы для здоровья, Консорциум структурной геномики туберкулёза сосредоточил внимание на *M. tuberculosis* с целью охарактеризовать потенциальные мишени для лекарств (существующие методы лечения длительны и имеют множественную лекарственную устойчивость) и получить представление о биологии туберкулёза. Полное секвенирование генома позволило клонировать около одной трети генов *M. tuberculosis* и определить более 100 белковых структур, часть из которых уже является проверенными мишенями для лекарств. Кроме того, исследование мутагенеза транспозонов в масштабе генома выявило множество генов, необходимых для роста *M. tuberculosis* (Baker, 2007).

Другой вид микобактерий, палочка (бациллы) Хансена *Mycobacterium leprae* Lehmann and Neumann 1896, также является возбудителем заболевания, которое поражало человечество на протяжении всей его истории. Первые исторические записи симптомов проказы в Китае, Индии и Египте датируются примерно 600 г. д.н.э. Лепра была широко распространена в Европе, но в XIV–XVI вв. неожиданно исчезла. Исследование геномной ДНК археологических образцов возрастом от римского периода до XIII в. обнаружило коинфекцию в 40% анализируемого материала. Это позволило предположить, что увеличение смертности от туберкулёза привело к тандемному снижению заболеваемости проказой – так называемый эффект перекрёстного иммунитета. Геном *M. leprae* оказался значительно консервативным в последние 1000 лет, что делает маловероятным исчезновение лепры из-за потери вирулентности её возбудителя. Филогенетические исследования указывают на европейское, в отличие от туберкулёза, происхождение лепры в Америке, а также существование последнего общего предка всех штаммов *M. leprae* около 3000 л.н. (Schuenemann et al., 2013; Mendum et al., 2014).

*Streptococcus pyogenes* Rosenbach 1884 является ещё одним важным патогеном для человека, с 700 млн случаев заражения ежегодно во всём мире, но без лицензированной вакцины. Интегративный анализ популяционной геномики, транскриптома и вирулентности не только раскрывает новые детали патогенеза этой бактерии, но также с высокой достоверностью идентифицирует однонуклеотидную вставку в межгенной области, которая значительно изменяет глобальные профили транскриптов и, в конечном итоге, вирулентность (Kachroo et al., 2019).

Тяжёлыми патогенами для человека являются некоторые виды вирусов. Вирус натуральной оспы *Variola virus* (класс I по Балтимору – вирусы, содержащие двухцепочечную ДНК) является возбудителем болезни, которая в XX в. унесла жизни от 300 до 500 млн человек на планете. Как декларировала ВОЗ, человечество победило натуральную оспу благодаря всеобщей вакцинации. Исторические источники сви-

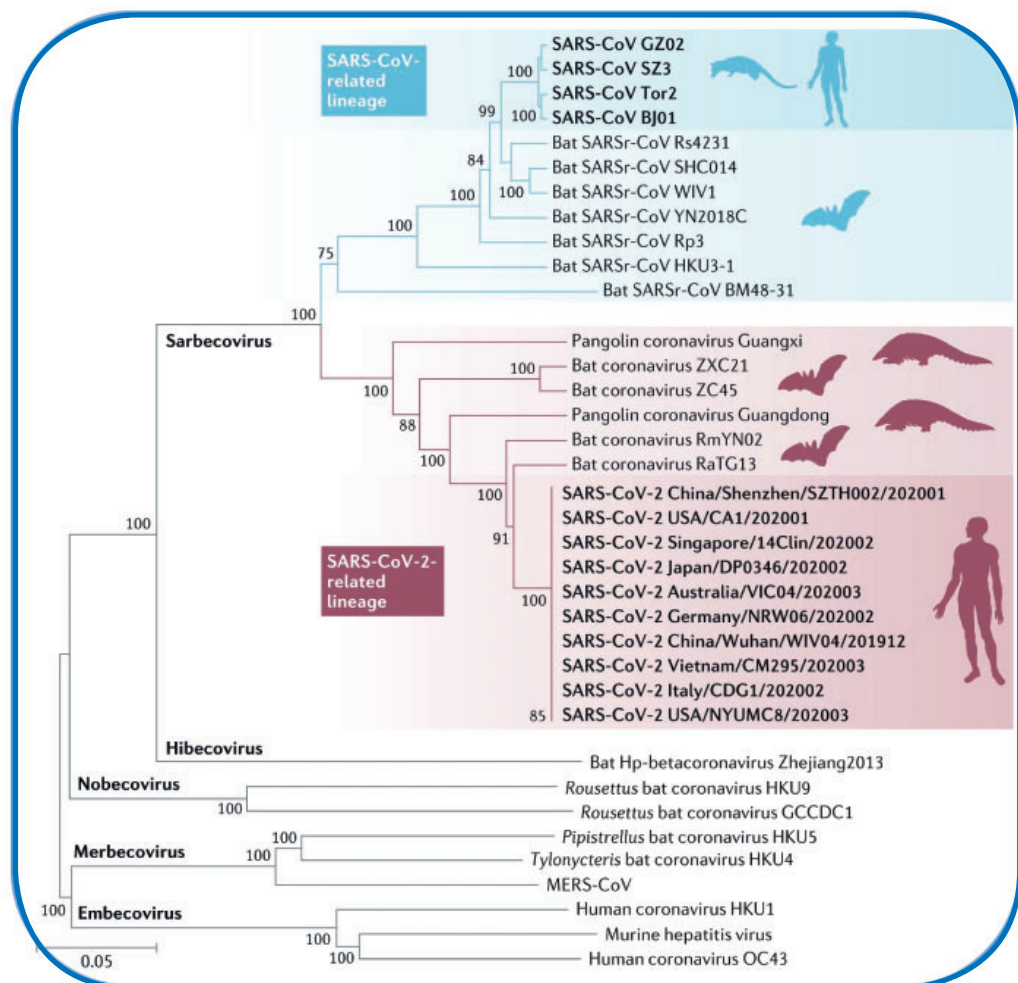
детельствуют, что люди болели оспой еще 3000 л. н., причём признаки заболевания обнаружены в мумифицированных останках возрастом 3570 л. Данные секвенирования древних геномов вируса натуральной оспы показали, что заболевание было распространено в Северной Европе в эпоху викингов (600–1050 гг. н.э.), т. е. на тысячу лет раньше, чем считали прежде. Филогенетически, древние геномы вируса составляют сестринскую кладу по отношению к современным вирусам натуральной оспы. Древняя и современная клады разошлись около 1700 л. н. и претерпели независимую эволюцию с инактивацией некоторых генов, вызванной специализацией к человеку. Однако, вирусные геномы были найдены только у 11 из 525 исследованных жителей эпохи викингов (т.е. у 2%), что может служить приблизительной оценкой масштабов заболевания. Вместе с тем, в останках человека более раннего времени вирусная ДНК не найдена, следовательно, конец эпохи викингов представляется наиболее вероятным временем панъевропейского распространения натуральной оспы (Mühlemann et al., 2020).

Эндосимбиотическая бактерия *Wolbachia* Hertig, 1936 подавляет способность комара *Aedes aegypti* (Linnaeus, 1762) к передаче арбовируса (арбовирусы способны вызывать у человека такие тяжёлые заболевания, как энцефалит, лихорадка Денге, жёлтая лихорадка), распространяясь через популяции диких комаров после их локального заселения. Однако недавние интродукции в г. Кэрнс (Австралия) продемонстрировали более медленное, чем ожидалось, пространственное распространение бактерии. Использование методов ландшафтной геномики и молекулярных анализов инфекции *Wolbachia* предоставило доказательства трёх процессов, которые, как предполагается, замедляют пространственное распространение бактерии через *Ae. aegypti*: барьеры (автомобильные магистрали) для расселения *Ae. aegypti*, более высокая частота распространения на большие расстояния (предполагают, что перемещение самок на большие расстояния может быть опосредовано транспортом) и потеря *Wolbachia* между поколениями. Такая модель исследования может быть применена и для других систем хозяин-симбионт, которые рассматриваются как кандидаты для биоконтроля риска заболеваний человека (Schmidt et al., 2018).

Понимание эволюции и видообразования патогенов и выявление детерминант генетической изменчивости имеет большое теоретическое и прикладное значение. Популяционно-генетический анализ крайне специфичного и лекарственно-устойчивого грибкового патогена млекопитающих *Pneumocystis jirovecii* Frenkel, 1976 не обнаружил доказательств географической структуризации популяции, позволяя предполагать недавнюю смену хозяина, что открывает новые перспективы его эволюции (Cissé et al., 2018). Полногеномные исследования ассоциаций *Streptomyces albidoflaus* (Rossi Doria 1891) Waksman and Henrici 1948 (Approved Lists 1980) (с неизвестными процессами видообразования, повсеместно распространённым в природе и представляющим собой источник антибиотиков) продемонстрировали стимулирование генетической изменчивости вида за счёт экологической адаптации, предполагая модель экологического видообразования с потоком генов у стрептомицетов (Li et al., 2019b).

В последнее время большое внимание было сосредоточено на исследованиях генома коронавирусов. Коронавирусная инфекция представляет собой опасное респираторное заболевание, вызываемое, в частности, коронавирусом SARS-CoV-2 (2019-nCoV) (класс IV по Балтимору – вирусы, содержащие одноцепочечную (+)РНК). За время её распространения зарегистрировано свыше 560 млн случаев заболевания по всему миру с более 6,3 млн летальных исходов, что делает пандемию COVID-19 одной из самых смертоносных в истории (Poorolajal, 2021). Как показал эволюционный анализ набора геномных данных человека, взаимодействие людей с коронавирусами, вероятно, началось более 20000 л.н. и ограничивалось популяцией предков современного населения Восточной Азии (Souilmi et al., 2021). Филогенетический анализ полногеномных последовательностей показал, что коронавирус тяжёлого острого респираторного синдрома 2 (SARS-CoV-2) объединяется с близкородственными вирусами летучих мышей и панголинов и вместе с SARS-CoV и коронавирусами, связанными с SARS летучих мышей (SARSr-CoV), образует сарбековирусы (рис. 5.6) (Hu et al., 2021). Несмотря на то, что геном SARS-CoV-2 похож на геномы SARS-CoV и MERS-CoV, препараты, ранее доказавшие свою эффективность, против нового вируса оказались совершенно неэффективными. Потенциальные противовирусные лекарства нацелены на различные этапы репликации коронавируса тяжёлого острого респираторного синдрома 2, от связывания с рецептором проникновения и слияния до репликации. Кроме того, иммуноглобулиновые и иммуномодулирующие препараты также являются потенциальными терапевтическими средствами (Hu et al., 2021). Исследования в области структурной геномики показывают, что даже минимальные изменения в геноме SARS-CoV-2 способны радикально трансформировать структуру белков-мишеней, что приводит к неэффективности доступных лекарств, и прежде всего это касается шиповидного белка (рис. 5.7) (Naqvi et al., 2020; Hu et al., 2021). В ходе эпидемиологических исследований были также получены убедительные доказательства расовых и этнических различий в пандемии SARS-CoV-2, которые потенциально опосредованы уникальными социальными детерминантами здоровья (Vahidy et al., 2020).

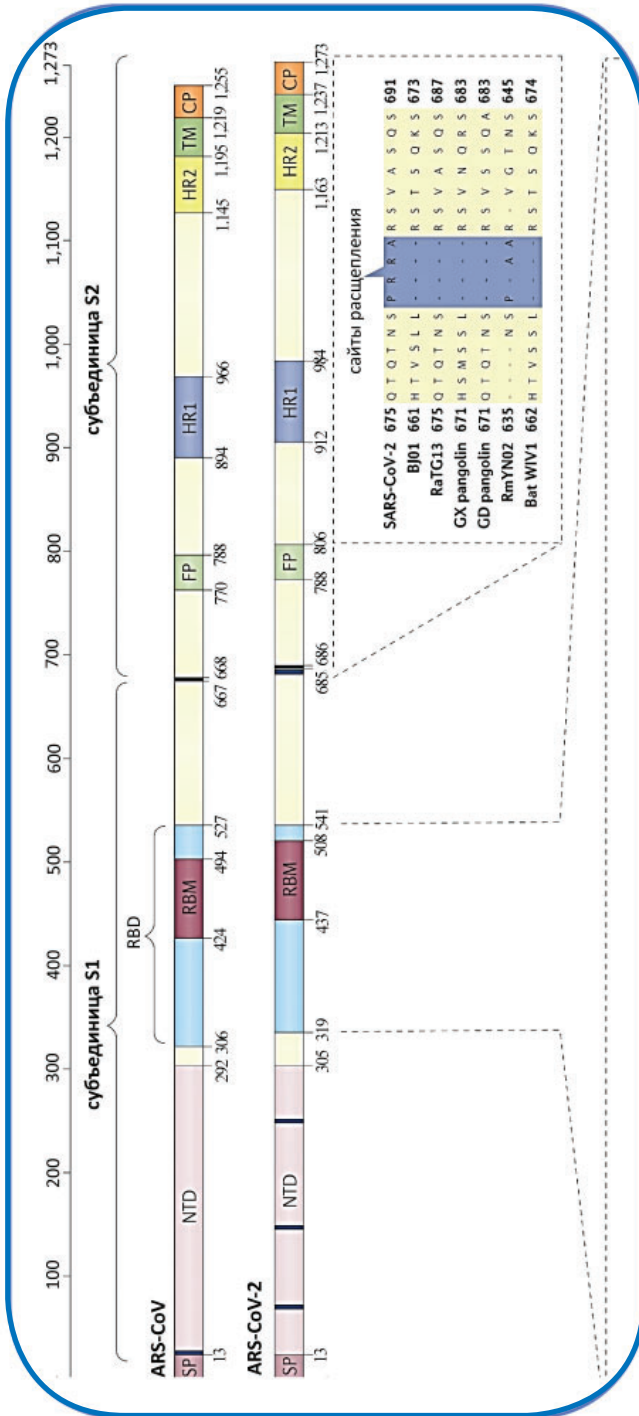
Паразитарные заболевания также широко распространены во всём мире. Они представляют серьёзную угрозу для здоровья человека и животных, приводящую к высокому социальному и экономическому ущербу (Thompson, Conlan, 2011). Паразитарными болезнями в мире поражено более 4,5 млрд. человек, а средняя заболеваемость населения Земли составляет 100 млн в год. По данным Всемирной Организации Здравоохранения 80% всех болезней у человека и животных связано с наличием в организме паразитов и продуктов их жизнедеятельности (Thompson, Conlan, 2011; Челомина, 2015). Геномные исследования представляют собой одно из новейших и многообещающих направлений современной паразитологии. Благодаря инновациям в области геномики, наши знания о биологии паразитов и патогенезе вызываемых ими заболеваний значительно расширились, а диагностические инструменты на основе данных секвенирования геномов все чаще применяются в клинической практике.



**Рисунок 5.6.** Филогенетическое древо полноразмерных геномных последовательностей SARS-CoV-2, SARSr-CoV и других бетакоронавирусов (из: Hu et al., 2021)

Малярия, вызываемая паразитическими протистами рода *Plasmodium* Marchiafava et Celli, 1985, только в 2020 г. привела к гибели 600 млн человек (в основном в африканских странах). Малярийная инфекция не идентифицируется в скелетных останках, поэтому о таких событиях можно судить только по анализу ДНК. Филогенетический анализ древней ДНК плазмодия (общий размер генома составляет около 25 Mb) предполагает два разных пути постконтактного проникновения *P. vivax* (Grassi & Feletti, 1890) и *P. falciparum* William H. Welch, 1897 в Америку, а также наличие связи между современной индийской и европейской линиями *P. falciparum* (Joseph, Lindo, 2023).

Болезнь Шагаса вызывает одноклеточный паразит *Trypanosoma cruzi* Chagas, 1909; даже при лечении специальными препаратами в 10% случаев наблюдается летальный



**Рисунок 5.7.** Ключевые различия в шиповидном белке SARS-CoV-2 и родственных коронавирусов (из: Hu et al., 2021).

Схематическая диаграмма спайкового белка (S) коронавируса тяжёлого острого респираторного синдрома (SARS-CoV) и SARS-CoV-2. Номера аминокислотных остатков каждой области соответствуют их положениям в S-белках SARS-CoV и SARS-CoV-2. Тёмно-синие блоки представляют вставки в S-белке. Вставки в аминокислотах 675–691 S-белка SARS-CoV-2 показаны в увеличенном виде в правом нижнем углу и выровнены с вставками других коронавирусов в том же регионе.

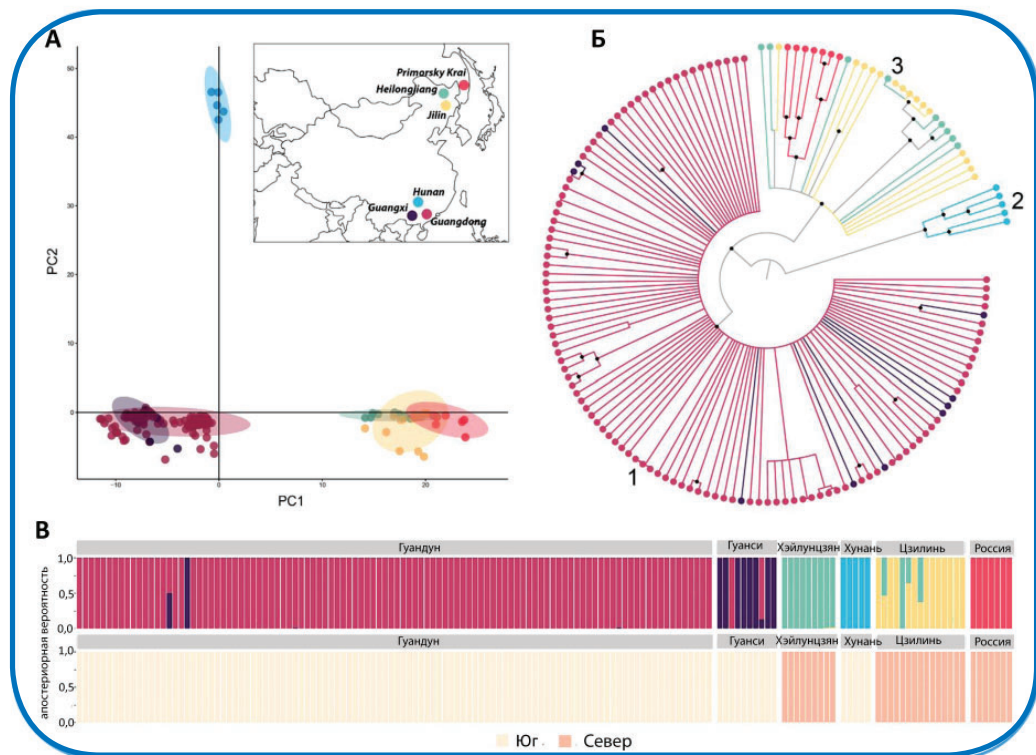
исход. Анализ ДНК южноамериканских мумий указал на наличие трипаносом среди древних популяций еще 9000 л. н. Филогенетические исследования показали, что коренное население Америки впервые столкнулось с *T. cruzi* как зоонозом от летучих мышей при заселении Анд и пустыни Атакама (Joseph, Lindo, 2023). Триатомовые клопы *Mepraia spinolai* (Poerter, 1934) являются одним из основных природных переносчиков болезни Шагаса на юге Южной Америки, распространённой в трёх средиземноморских климатических экорегионах Чили. В рамках ландшафтной геномики было проанализировано 2380 нейтральных однонуклеотидных полиморфизмов этого вида и обнаружена положительная связь его генетического разнообразия со средней зимней температурой и отрицательная связь со средней летней температурой и размером тела. Эти результаты указывают на факторы, которые могут повлиять на генетическое разнообразие *M. spinolai* и, следовательно, на эпидемиологическую ситуацию (San Juan et al., 2021).

Как заметил основатель отечественной гельминтологической науки академик Константин Иванович Скрябин, *«Всё живое на земном шаре попало в очень цепкую гельминтологическую паутину, которая плелась миллионы лет»*. Карл Циммер, известный американский писатель и популяризатор науки, отмечая огромное влияние паразитов на *«биологическое и социальное поведение целых видов, ...эволюцию флоры и фауны...»* подчеркнул, что всегда *«...паразит находит способ борьбы с иммунной системой... Погибнет в результате хозяин или останется жив, зависит только от того, что выгоднее паразиту»* (Циммер, 2017). Изучение геномов чрезвычайно важно для своевременной диагностики, профилактики и успешного лечения паразитарных заболеваний, хотя исследования популяций гельминтов осложнены из-за их больших и сложных геномов, огромных эффективных размеров популяции и слабых сигналов отбора генетических локусов (Челомина, 2016; Wit, Gillards, 2017).

Масштабный геномный анализ паразитических (цестоды, трематоды и нематоды) и непаразитических червей обнаружил появление, а также увеличение генных семейств в ключевых узлах филогенеза, связанных с паразитизмом (например, модулирующих иммунный ответ хозяина, обеспечивающих миграцию паразита через ткани хозяина или позволяющих ему питаться) (International Helminth Genomes Consortium, 2019). Полногеномный анализ показал, что паразитические черви (1) в целом обладают уменьшенным геномом по сравнению со свободноживущими филогенетическими родственниками, несмотря на то, что большинство генов у них общие; (2) имеют значительные генетические инновации (например, многокопийные семейства эффекторных генов и геномные домены, участвующие в их экспрессии и разнообразии) для паразитарной адаптации; (3) демонстрируют обширные клонально-специфические различия в белках основного метаболизма и семействах белков, используемых для разработки лекарственных средств; (4) обладают общими чертами между экто- и эндопаразитарными линиями и содержат новые потенциальные мишени для лекарств; (5) гены поверхностных антигенов могут располагаться в теломерных локусах, а вариации антигенов многократно возникали в различных системах (Hahn

et al., 2014; International Helminth Genomes Consortium, 2019; Chelomina, 2024b). С помощью сравнительной геномики также получены дополнительные доказательства адаптации представителей семейства печёночных трематод Opisthorchiidae к обитанию в желчных протоках и секреции этими трематодами белков, модулирующих пролиферацию клеток-хозяев. Сравнительный геномный анализ видов *Platyhelminthes* обнаружил связанные с эпидемиологической обстановкой свидетельства межвидовой гибридизации у шистосом, высокую вероятность существования криптических видов у китайской печёночной двуустки, а также открыл у трематод новый вид эндобактерии *Neorickettia* – *N. sennetsu* (Misao and Kobayashi 1956) (Dumler et al. 2001), тесно связанный с возбудителем лихорадки Сенэцу (сеннетсу) человека и Потомакской лошадиной лихорадки (Челомина, 2017; Oey et al., 2019a; Chelomina, 2024b). В целом, высокий внутривидовой генетический полиморфизм геномов трематод предполагает значительный эволюционный потенциал для быстрой адаптации к хозяевам, включая колонизацию новых видов, изменение климата, и реакцию на лекарства и вакцины (Cwiklinski et al., 2015; Pitaksakulrat et al., 2018; Oey et al., 2019b). У цестод (ленточные черви) сравнительная геномика выявила адаптивную эволюцию азиатского цепня при переходе на нового промежуточного хозяина. Высокие показатели дубликации генов и функциональной диверсификации могли частично стать причиной различий между *Taenia asiatica* Eom and Rim, 1993 и *T. saginata* Goeze, 1782. Согласно полученным данным, эти виды разошлись примерно 1,14 млн л.н., что совпадает с миграцией *Homo erectus* в Азию, т.е. видообразование *T. asiatica* предшествовало видообразованию *H. sapiens* (Wang et al., 2016b). Геномный анализ паразита рыб *Pomphorhynchus laevis* Müller, 1776 обнаружил значительную реорганизацию репертуара генов, повышенная пластичность которого за счёт интеграции чужеродной ДНК, вероятно, является одним из основных факторов, обеспечивших эволюционный успех скребней (Mauer et al., 2020).

Представления о возбудителях паразитарных заболеваний в пространстве и времени и возможность прогнозирования зоонозов даёт палеогеномика. Первое исследование популяционной геномики палеопаразита, круглого червя власоглава *Trichuris trichiura* Linnaeus, 1771, вызывающего трихоцефалёз, от которого страдают до 500000 человек во всём мире (относится к так называемым «забытым тропическим заболеваниям», т.е. которым правительства стран не уделяют достаточного внимания), подтвердило происхождение этого заболевания в Уганде и его последующее распространение с миграцией человека, а также выявило локальные области генетической дифференциации между географически различными популяциями. Неожиданным результатом оказалось близкое генетическое родство между образцами от угандийцев и бабуинов, возможно, представляющих собой зоонозный резервуар паразита. Для проведения этих исследований были использованы яйца власоглава, обнаруженные в копролитах человека на археологических раскопках в Европе и Северной Америке, датированных 7100 г.д.н.э. (Doyle et al., 2022). Следует отметить, что использование копролитов позволяет также прояснить модель



**Рисунок 5.8.** Структура популяции *Clonorchis sinensis* (из: Kinkar et al., 2023).

А – Анализ главных компонент (PCA) генетической дифференциации (карта мест отбора проб показана в левом верхнем углу); Б – Консенсусное филогенетическое дерево, полученное с помощью байесовского анализа. Ветви окрашены в зависимости от географического происхождения экземпляров (чёрные точки в узлах – абсолютные значения апостериорной вероятности); В – Гистограммы апостериорных вероятностей принадлежности, полученные методом дискриминантного анализа главных компонент (DAPC). Каждая особь представлена вертикальной линией, окрашенной в соответствии с генетическим кластером, к которому она была отнесена. Результат анализа с использованием мест отбора проб в качестве априорных групп показан сверху; нижний график представляет групповые назначения *de novo*.

поведения и рацион древнего человека (включая каннибализм), его миграции и взаимодействие с животными (Askari et al., 2022).

Первый широкомасштабный популяционный анализ китайской печёночной двуустки *Clonorchis sinensis* (Gobbold, 1875) (канцерогенная трематода, возбудитель клонорхоза, который поражает около 35 млн человек во всём мире) выявил различия между паразитами из северной и южной частей видового ареала (рис. 5.8), указывающие на существование биологически отличных внутривидовых вариантов. Такие варианты могут иметь различную эпидемиологию, патогенность и/или чувствительность к химиотерапии и, скорее всего, являются следствием глубокой адаптации паразита к промежуточным хозяевам и климатическим особенностям каждого региона (Kinkar et al., 2023). Кроме того, была обнаружена высокая гетерозиготность образцов *C. sinensis*, что предполагает наличие у паразита генетического или геномного

механизма, способствующего поддержанию гетерозиготности у отдельных особей в различных эндемичных зонах. Результаты также позволяют предположить, что печёночная двуустка развила способность легко адаптироваться к изменяющимся условиям окружающей среды и/или видам-хозяевам на протяжении своей эволюционной истории. Дальнейшая идентификация инвариантных генов из набора геномных данных может быть полезной в поиске новых мишеней для вмешательства в борьбе с этим паразитом, особенно учитывая высокую клиническую значимость клонорхоза. Это имеет большое значение, поскольку эффективной вакцины до сих пор не существует, а химиотерапия основана на единственном антигельминтном препарате – празиквантеле. Смешение различных биологических вариантов паразита, обладающих высоким адаптивным потенциалом, в результате расширения ареала или локальных интродукций может привести к непредсказуемым последствиям, радикально изменив как свойства самого паразита, так и эпидемиологическую ситуацию, что необходимо учитывать при разработке стратегий защиты населения от тяжёлых гельминтных инвазий (Kinkar et al., 2023).

В ходе геномных исследований трематод были обнаружены гены различных протеаз, киназ, фосфатаз, тегументальных белков и экскреторно-секреторных продуктов (ESP), белков и рецепторов, играющих важную роль в жизнедеятельности гельминтов. Некоторые из них были предложены в качестве потенциальных объектов для разработки вакцин, лекарств и диагностикомов (Yoo et al., 2011; Wang et al., 2014a; Челомина, 2015). Например, енолаза/фосфопируватгидратаза является многофункциональным ферментом гликолиза, играющим важную роль в патогенных заболеваниях и во взаимодействиях между хозяином и патогеном в качестве рецептора клеточной поверхности. Биохимическая и биологическая значимость енолазы предполагает её использование в качестве кандидата на вакцину для профилактики паразитарных заболеваний человека. Чтобы подтвердить биологическую важность енолазы для трематод, был проведён эксперимент по сайленсингу её гена у китайской двуустки *C. sinensis* с помощью РНК-интерференции. Эксперименты показали, что РНК-интерференция снижает экспрессию енолазы, и это имеет губительные последствия для трематоды (Wang et al., 2014a). Такой подход, безусловно, найдет применение в дальнейшем при идентификации функциональных генов этого паразита. Интересно, что нокадаун экспрессии генов катепсинов L и B (внутриклеточные протеазы, играющие ключевую роль в поддержании функций клеток центральной нервной системы), опосредованный РНК-интерференцией, предохраняет от вновь экцистированных личинок *Fasciola hepatica* (Linnaeus, 1758), а вакцина с катепсином L1 даёт частичную защиту против инфекции этой трематоды и последующего развития патологии (McGonigle et al., 2008; Zafra et al., 2013). Выяснение генной регуляции, основанной на РНК-интерференции, будет способствовать анализу биологических основ антигенной изменчивости и иммунного уклонения у паразитов (Liu et al., 2010), а также разработке антипаразитарных препаратов нового поколения (Bartel, 2004; Manzano-Román, Siles-Lukas, 2012; Britton et al., 2014).

Геномные и транскриптомные данные, а также методы биоинформатики облегчают поиск новых лекарственных препаратов, позволяя обнаружить паразитоспецифические метаболические пути и определить способы регуляции развития паразитов. В ходе геномных исследований пять генов, которых нет у позвоночных, были признаны в качестве предполагаемых лекарственных мишеней. Они включают (1) мембранные белки тетраспанины D76 и 18 (находятся на внешней оболочке трематод и функционируют как рецепторы для молекул хозяина; являются признанной мишенью вакцин против кровяной трематоды *Schistosoma mansoni* Sambon, 1907), (2) белок, взаимодействующий с фактором ARL6 (вовлечён в процессы гемопоэтического созревания, такие как транспорт белков, мембранный перенос и клеточная сигнализация), (3) содержащую N-концевой домен дофамин-бета-монооксигеназу (оксидаза со смешанной функцией, катализирующая превращение дофамина в норадреналин, участвует в нейрорегуляции), (4) протеолипид миелин (трансмембранный белок – вероятный структурный компонент миелина, способствующий стабильности и компактности слоистой структуры миелина), а также (5) START домен 3 (отвечает за транспортировку различных гидрофобных веществ между разными тканями, клетками и частями клеток эукариот) (Yoo et al., 2011).

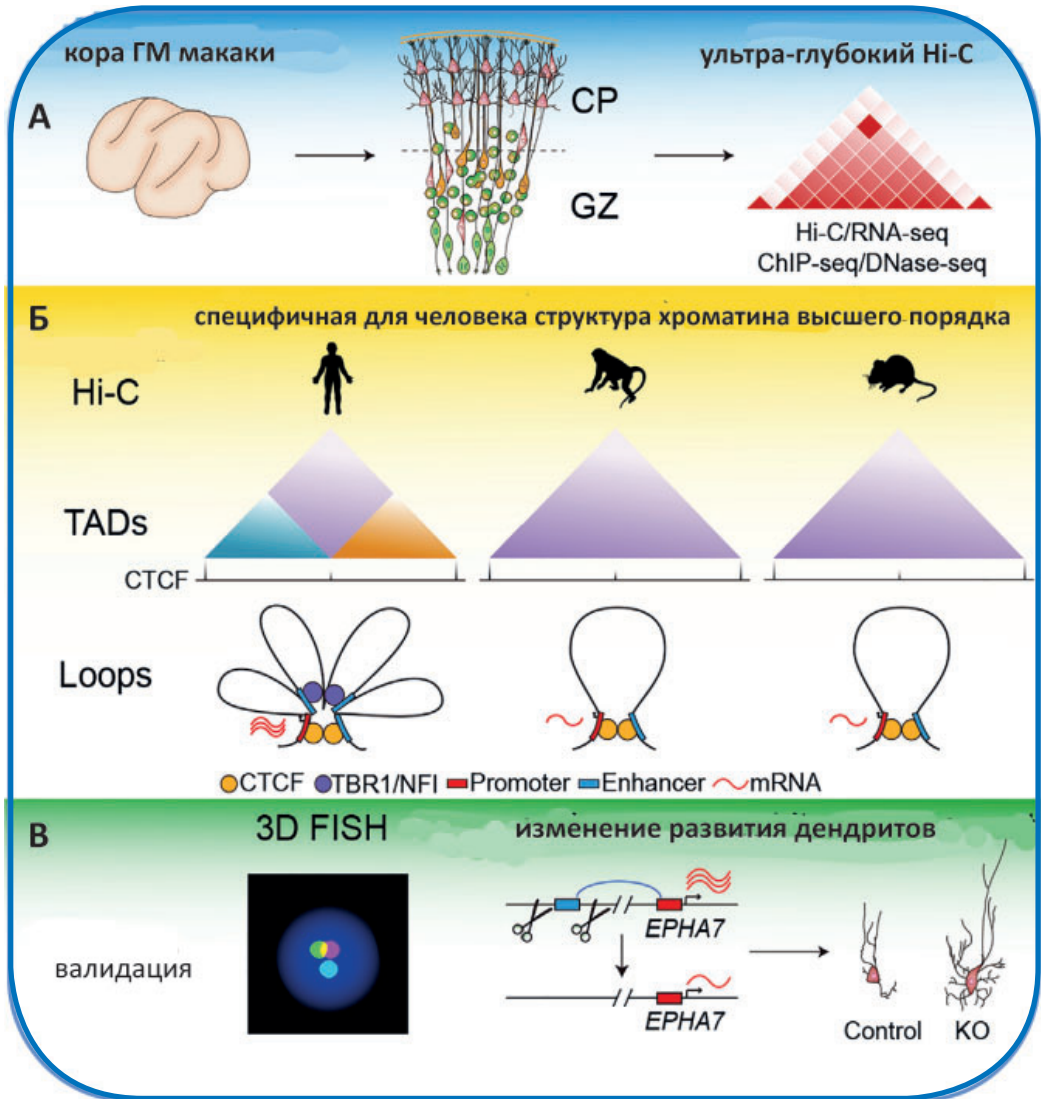
Считается доказанным, что как вирусные (например, вирусы гепатита В и С, папилломы и герпеса, вирус Эпштейна-Барра) и бактериальные (*Helicobacter pylori* и т.п.) инфекции, так и паразитарные инвазии (например, *Schistosoma haematobium*, Bilharz 1852, *Opisthorchis viverrini* (Poirier, 1886) Stiles & Hassal, 1896 и *C. sinensis*) приводят к онкологическим заболеваниям. Генотоксичность паразитарной инвазии (на примере лёгочной трематоды *Paragonimus heterotremus* Chen et Hsia, 1964) хорошо иллюстрируется с помощью анализа генома методом ДНК-комет, и список канцерогенных паразитов, видимо, будет увеличиваться (Chelomina et al., 2021). В медицинской практике антисмысловые мРНК могут быть использованы для блокирования малых интерферирующих РНК, которые действуют как онкогены, в то время как искусственные интерферирующие РНК могут быть получены (на основе дополнительных характеристик малых интерферирующих РНК для целевых мРНК) и использованы для блокирования экспрессии онкогенов (Челомина, 2015; Zhang et al., 2007; Han et al., 2016a; Yan et al., 2016).

Геномные подходы открывают исключительные возможности для изучения патологии, вызванной паразитами. Первый интегрированный полногеномный и эпигеномный анализы холангиокарциномы, вызванной инвазией *C. sinensis* (489 случаев из десяти стран), выявил новые драйверные гены заболевания, некодирующие промоторные мутации и разнообразные структурные варианты. Молекулярные ландшафты холангиокарциномы радикально отличаются по этиологии, подчёркивая, как многочисленные подтипы рака могут возникать в одном и том же органе в результате различных канцерогенных процессов (Jusakul et al., 2017).

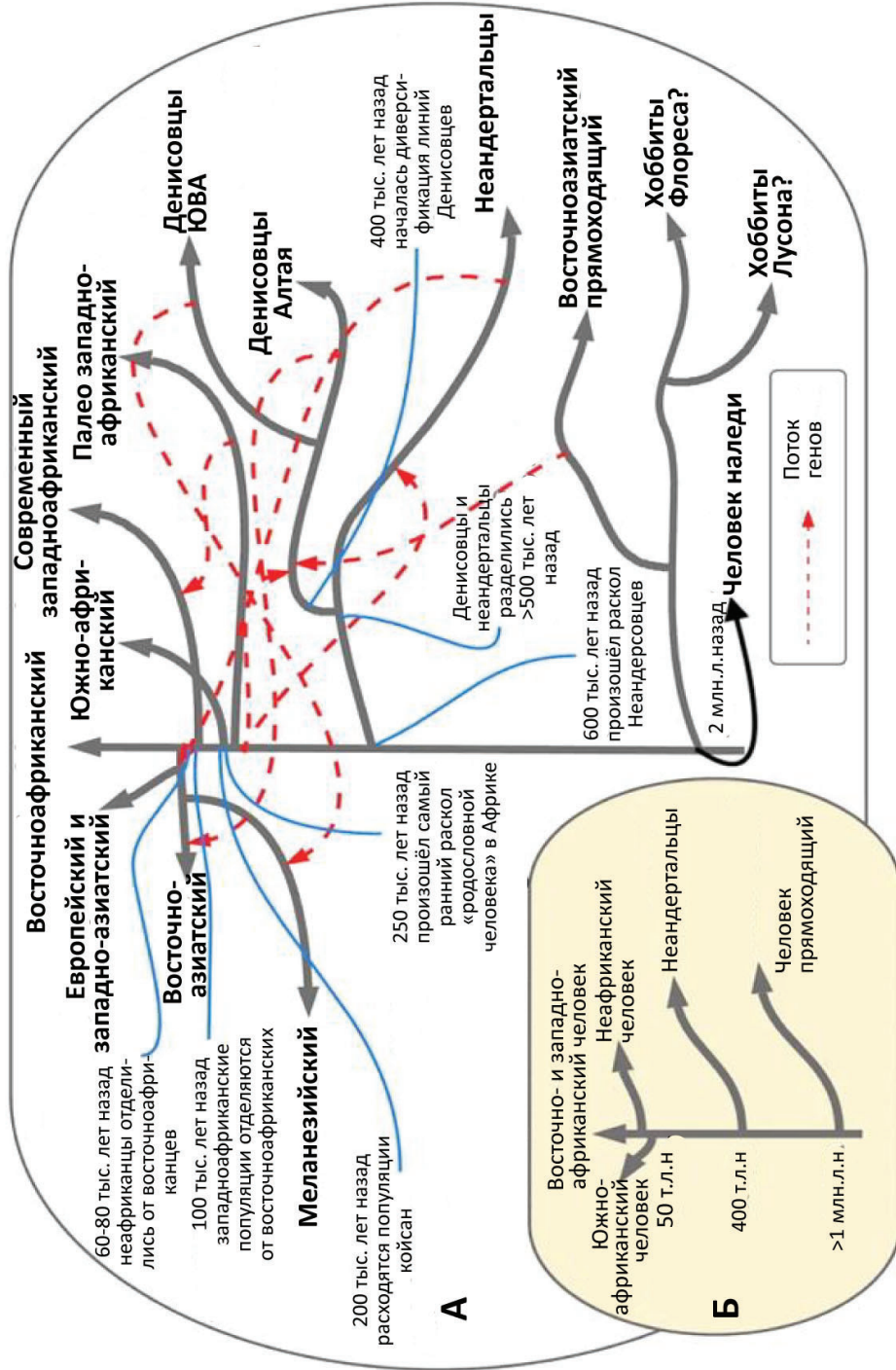
## 5.2. Антропология и археология

Высококачественные геномы архаичных людей позволяют исследовать недавнюю эволюцию человека и определять важные изменения, особенно связанные с мозгом, которые делают нас «полностью людьми» (Kuhlwilm, Voeckx, 2019). Удивительно, но согласно графу предковой рекомбинации геномов человека (279 геномов), неандертальца (два генома) и денисовца (один геном), только 1,5–7% генома современного человека уникальны для *Homo sapiens* Linnaeus, 1758 и связаны с работой головного мозга и нервной системы (Schaefer et al., 2021). ДНК неандертальца появилась в геноме человека около 74000 л.н., а всплеск развития специфичных для человека мутаций происходил дважды, 600000 и 200000 л.н., и мог быть вызван как природными явлениями, так и другими факторами (Schaefer et al., 2021). Недавние исследования показали, что сравнительный анализ трёхмерной организации генома важен для изучения регуляторных механизмов развития и эволюции мозга. Сопоставление трёхмерных геномов мозга человека, макаки и мыши позволило идентифицировать многие специфичные для человека изменения структуры хроматина, включая 499 топологически ассоциированных доменов (TAD) и 1266 петель хроматина. Петли были значительно обогащены взаимодействиями энхансер–энхансер, а регулируемые ими гены проявляли характерные для человека изменения экспрессии в субпластине (переходной зоне развивающегося мозга, критически важной для формирования и пластичности нейронных цепей). Кроме того, многие специфичные для человека изменения последовательностей находились в границах TAD и якорях петель, способных генерировать новые сайты связывания факторов транскрипции и структуры хроматина (Luo et al., 2021) (рис. 5.9).

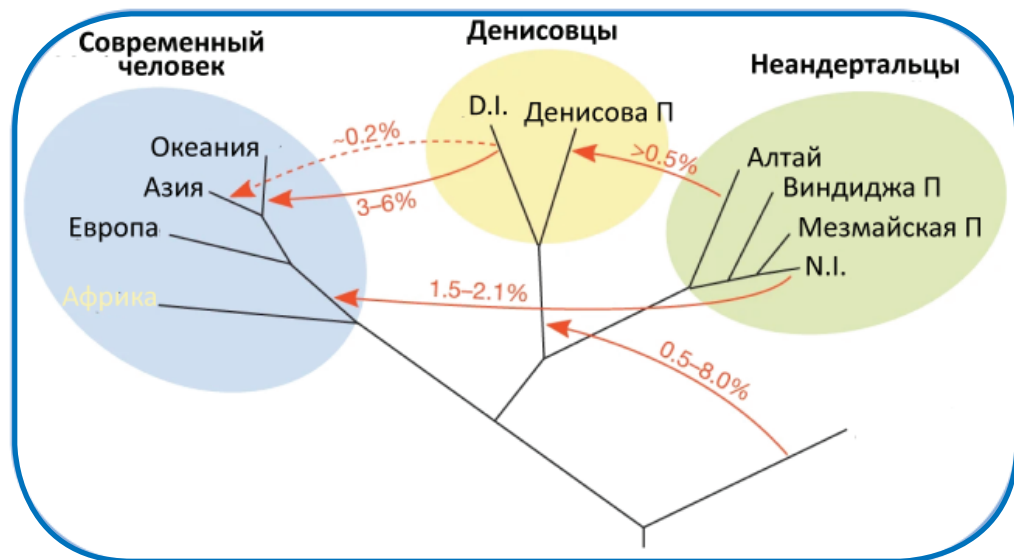
Геномные данные вымерших гоминид существенно дополняют наше представление о происхождении современного человека (рис. 5.10). За последние 200000–300000 л. между разными древними гоминидами происходило скрещивание, и следы этих событий сохранились в геномах современного человека (рис. 5.11). Установлено, что у ныне живущих людей 1–3% ДНК имеет неафриканское происхождение, при этом у современного населения Океании 2–4% ДНК унаследовано от денисовцев (Hubisz et al., 2020). Согласно геномному анализу, линии неандертальцев и денисовцев разошлись не позднее 390000 л.н. (Slon et al., 2018), а их общий предок отделился от предков современных людей 550000–765000 л.н. (Prüfer et al., 2014). События древней гибридизации были подтверждены результатами секвенирования генома из кости Денисовой пещеры на Алтае, показавшими, что находка принадлежит девочке, мать которой была неандерталкой, а отец – денисовцем (Slon et al., 2018). Интересно, что по имеющимся на сегодня данным, большая часть неандертальской ДНК в геноме современного человека получена в результате единственного смешения, которое произошло после выхода человека из Африки, тогда как с денисовцами современные люди смешивались в разных регионах несколько раз (Kerner et al., 2021b). Смешение с неандертальцами и денисовцами, вероятно, сыграло важную роль в истории биологической адаптации современного человека. От неандертальцев люди приобрели



**Рисунок 5.9.** Трёхмерные геномы мозга человека, макаки и мыши (из: Luo et al., 2021).  
 А – Hi-C карта головного мозга (ГМ) плода макаки в высоком разрешении; Б – специфичные для человека структуры хроматина и регуляторные изменения во время кортикогенеза; В – специфическая для человека петля, регулирующая ген *EPHA7*, влияет на развитие дендритов нейронов;  
 CP – корковая пластинка, состоящая преимущественно из постмитотических нейронов;  
 GZ – зародышевая зона, сформированная в основном из нейтральных предшественников.



**Рисунок 5.10.** Современное (2021 г., А) и более раннее (2000 г., Б) понимание происхождения человека. Неандертальцы – предки Денисовцев и Неандертальцев (из: <https://economistwritingeveryday.com/tag/paleogenomics/>)



**Рисунок 5.11.** Возможная модель событий потока генов в позднем плейстоцене (по: Prüfer et al., 2014).

D.I. – интроgressирующий денисовец, N.I. – интроgressирующий неандерталец, П – пещера

аллели, связанные с более светлой кожей, что способствовало выработке витамина Д в северных широтах с низким уровнем УФ. От денисовцев люди получили гаплотип, связанный с высотной толерантностью (например, в тибетских популяциях), а также гаплотипы с генами, играющими решающую роль в выработке тепла при воздействии холода (100% частота у гренландских инуитов – автохтонных народов Северной Америки) (Marciniak, Perry, 2017).

Анализ примерно 120 т.п.н. экзомной ДНК Y-хромосомы неандертальцев и современных людей выявил различия в кодировании белков, включая миссенс-мутации (приводящие к аминокислотным заменам) в генах, которые продуцируют специфичные для мужчин минорные антигены гистосовместимости. Считается, что некоторые из них вызывают материнский иммунный ответ во время беременности, и несовместимость одного или нескольких из этих генов сыграла определённую роль в репродуктивной изоляции. Время до появления самого последнего общего предка (TMRCA) Y-хромосомы неандертальца и современного человека оценивается примерно в 588000 л.н. (Mendez et al., 2016). Изучение популяции человека в масштабе всего генома показало, что гены, связанные с иммунной системой, могут быть основными мишенями положительного отбора; гены, связанные с репродукцией и фертильностью, вероятно, быстро эволюционируют; а гены, ассоциированные с персистенцией лактазы, эккринными железами и реакцией на гипоксию являются следствием недавней адаптации человека (Lachance, Tishkoff, 2013).

Доступность геномных данных для древних людей и архаичных гоминид даёт новое представление об адаптации человека к изменяющейся среде, сельскохозяйственному образу жизни и патогенам, а также способствует более глубокому пониманию эволюции человека и последствий его действий в контексте эволюционных процессов (Marciniak, Perry, 2017). Например, исследования древних геномов выявили очень постепенную эволюцию депигментации кожи у современных европейцев. Аллели, связанные с более светлой пигментацией кожи почти фиксированы, или наблюдаются с максимальной частотой среди европейцев во всем мире, причём варианты некоторых генов увеличились в частоте совсем недавно. Аналогична ситуация с физиологической адаптацией к высокогорной среде Гималайской дуги в Непале с низким содержанием кислорода (Marciniak, Perry, 2017). Классический пример – персистенция лактазы у человека, которая сформировалась во многих регионах мира, а географическое распространение адаптивного варианта и увеличение его частоты произошло за последние 5000 л. Большинство современных людей имеют увеличенное число копий гена амилазы (ответственного за переваривание крахмала) по сравнению с шимпанзе. При этом, первоначальная дупликация гена произошла после расхождения линии человека с линиями неандертальца и денисовца, и, возможно, это связано с земледелием (Marciniak, Perry, 2017).

К настоящему времени проанализированы многочисленные геномы древних людей, что в значительной мере прояснило вопросы истории популяций анатомически современного человека. Получены доказательства того, что доисторические миграции оказали глубокое и долгосрочное влияние на генетическое разнообразие евразийских популяций (Allentoft et al., 2024). Одна из важных и спорных проблем – заселение Европы. Палеогеномика показала, что заселение этой части Света является очень сложным процессом с множественными миграциями различных популяций, которые могли повлиять на современный генофонд европейцев. Согласно одной из моделей, по крайней мере три различных древних популяции внесли вклад в происхождение современных европейцев: коренные палеолитические западноевропейские охотники-собиратели; древние северные евразийцы, связанные с верхнепалеолитическими сибиряками; и ранние европейцы, относящиеся к восточным неолитическим земледельцам (Lan, Lindqvist, 2018). Полногеномный анализ древних европейцев указал на массовую миграцию на запад из понтийских степей около 4500 л. н. и обнаружил, что представители ямной культуры имеют общее происхождение с древними жителями Северной Евразии (Lan, Lindqvist, 2018). Более того, палеогеномные исследования выявили присутствие генетических типов, характерных для представителей ямной культуры, у населения медного и бронзового веков в Испании и Португалии (Olalde et al., 2019). Морская экспансия скандинавского населения в эпоху викингов (750–1050 г. н. э.) является масштабным событием мировой истории. Секвенирование геномов 442 человек из археологических раскопок в Европе и Гренландии выявило обширные миграционные потоки, предоставив доказательства значительного притока датского происхождения в Англию, приток шведов в Прибалтику, норвежцев – в Ирландию, Исландию и Гренландию, а в Скандинавию – народов

из других стран Европы (Margaryan et al., 2020). Палеогеномные данные указывают на преимущественно анатолийское происхождение континентальных неолитических земледельцев и иберийского – британских неолитических популяций. Примечательно, что согласно прогностическому анализу уровня пигментации кожи, мезолитический человек из Чеддера (пещера Гофа на юго-западе Англии) был темнокожим с голубыми или зелёными глазами, неолитический человек из пещеры Карсингтон–Пастчер (Дербишир, Англия) имел пигментацию кожи от средней до тёмной и карие глаза, а у западноевропейских мезолитических охотников-собирателей к 6000 г. до н. э. сосуществовали совершенно разные уровни пигментации (Brace et al., 2019).

Происхождение и развитие этрусской цивилизации, занимавшей в железном веке большую территорию центральной Италии, обсуждалось на протяжении веков. Эта культура известна выдающимися навыками в области металлургии, сложными культурными представлениями и своим вымершим неиндоевропейским языком. Существуют две гипотезы, первая из которых предполагает анатолийско-эгейское происхождение этрусков (что также указывали древнегреческие писатели), вторая – автохтонное развитие (поддержана археологами) (Posth et al., 2021). Геномные исследования 82 человек, живших от 800 г. до н. э. до 1000 г. н. э. в Этрурии и южной Италии, обнаружили компонент индоевропейского степного происхождения (т.е. люди, связанные с этрусской культурой, имели большую долю степных предков) и отсутствие недавней анатолийской примеси среди предполагаемых неиндоевропейских этрусков в железном веке (что не согласуется с первой гипотезой). На протяжении почти 800 л. после окончания бронзового века родственный этрускам генофонд в целом оставался однородным, несмотря на спорадическое присутствие выходцев из Ближнего Востока, северной Африки и центральной Европы. Однако местный генофонд резко изменился в период Римской империи, и примесь генов с восточным средиземноморским происхождением достигла 50%. Кроме того, были идентифицированы североευропейские компоненты, появившиеся в центральной Италии в раннем средневековье, которые сформировали генетический ландшафт современного итальянского населения (Posth et al., 2021).

Скифы в широком смысле являются кочевыми народами на территории от причерноморских степей до современного Китая (хотя археологи считают скифами только понтийское население, а остальных – кочевниками скифского типа). Они представляли собой множество конно-воинских кочевых культур, обитавших в евразийских степях, в том числе на территории современного Казахстана, в первом тысячелетии до н.э. Однако мало что известно об их происхождении и связях с другими культурами. Вместе с тем, предложены три гипотезы происхождения скифов. Гипотеза понтийско-каспийского происхождения базируется на предполагаемых иранских языках этих кочевых народов, гипотеза казахского степного происхождения поддерживается археологическими находками, в то время как множественное независимое происхождение от генетически разных групп опирается на общие культурные черты (Gnecchi-Ruscone et al., 2021). Полногеномные данные для 111 древних людей перво-

го тысячелетия до н.э. и н.э. из 39 археологических раскопок в степях центральной Азии не подтвердили их западное Понто-каспийское происхождение, они больше соответствовали гипотезе казахского степного происхождения. Тем не менее, идеи как единого, так и множественного происхождения поддержки не получили. Напротив, были обнаружены свидетельства формирования двух основных генофондов железного века, возникших на территориях вокруг Алтая и Урала (Gnecchi-Ruscione et al., 2021). Интересно, что неоднородность и географическая структуризация, наблюдаемая в железном веке, резко контрастирует с генетической однородностью среди современных казахов. Полагают, что казахский генофонд представляет смесь различных западно- и восточноевропейских источников со сложной демографической историей. Гомогенизация генофонда, вероятно, произошла в результате 400–летнего периода строгих экзогамных правил, появившихся с установлением Казахского ханства в XV веке (Gnecchi-Ruscione et al., 2021).

Аварцы (авары) – один из коренных народов Кавказа, считающих себя преемниками монгольской степной империи жужаней, разрушенной тюрками около 550 г. н.э. В шестом веке аварская элита обосновалась в Карпатском бассейне, о чем свидетельствуют археологические раскопки. Геномные исследования останков из элитных погребений с изысканным инвентарем из серебра и золота, расположенных в центральной части Аварской империи в Дунайско-Тисском междуречье, показали, что 90% жителей раннего и 70–80% позднего аварского периодов действительно имеют северо-восточное азиатское происхождение. Остальная родословная аварской элиты (10–20%) имеет западно-евразийское происхождение, вероятнее всего – из Северного Кавказа. Таким образом, аварцы за короткий промежуток времени мигрировали из Монголии на Кавказ, а затем – Карпаты, совершив самую быструю из известных дальних миграций человека. Отсутствие признаков инбридинга указывает на то, что либо мигрировала большая группа аварского населения со строгими брачными обычаями, либо миграция была непрерывной. Анализ захоронений аварского периода из соседних регионов выявил более высокое генетическое разнообразие, указав на генетическую гетерогенность местной элиты: от 40% до менее 5% её популяций имели предков из Северо-Восточной Азии, но большая часть была местного происхождения (Fregel, 2022).

Почти четверть населения мира с множеством разнообразных этнических групп и языковых слоёв проживает в Восточной Азии, куда входят Китай, Монголия, Корея и близлежащие острова. Однако генетическая история Восточной Азии, особенно Китая, только начинает изучаться (Yang et al., 2020b). Анализ полногеномных данных 166 жителей Восточной Азии, датированных 6000–1000 гг. до н.э., и 46 современников позволил сделать ряд интересных выводов и внести значительный вклад в понимание истории населения этой пока малоизученной области. Например, охотники-собираатели Монголии и бассейна реки Амур имеют общую родословную с современными носителями монгольского и тунгусского языков. Жители Тайваня в период примерно 13000–800 гг. н.э. имели 75% предков от линии (видимо происходящей от фермеров

Янцзы), распространённой у современных людей, говорящих на австронезийских, тай-кадайских и австроазиатских языках, и около 25% предков – от северной линии, связанной с фермерами Хуанхэ (но отличной от них). Предки скотоводов ямной культуры прибыли в западную Монголию после 3000 г. до н.э., однако были вытеснены представителями более ранних миграций. Более поздние потоки генов обусловлены мигрантами ямной и европейской родословной фермеров (после 2000 г. до н.э.), а также последующими мигрантами из Турана (Wang et al., 2021).

Древние геномы раскрывают родословную, связанную с ямной культурой, и потенциальный источник индоевропейских носителей на Тянь-Шане железного века. Недавние исследования геномов человека раннего бронзового века выявили массовую экспансию населения за счёт носителей ямной культуры из причерноморских и каспийских степей в Западную и Восточную Евразию, что, вероятно, сопровождалось распространением индоевропейских языков. Современное население из региона Синьцзян на северо-западе Китая имеет сложную популяционную историю, в которой прослеживаются генетические связи как с Восточной, так и с Западной Евразией. Анализ первых полногеномных данных о 10 древних жителях северо-восточного Синьцзяна (датируются примерно 2200 л.н.; стоянка железного века Ширенцигоу) показал, что они генетически смешаны с восточными и западными евразийцами. Большая часть восточно-евразийского происхождения связана с популяциями северо-восточной Азии, в то время как западно-евразийское происхождение лучше всего представлено (20–80%) ямно-подобными предками. Таким образом, данные предполагают западно-евразийское степное происхождение, по крайней мере, для части населения древнего Синьцзяна. Кроме того, подтверждается связанное с ямной культурой происхождение ныне вымерших тохарских языков в Таримской котловине на юге Синьцзяна (Ning et al., 2019).

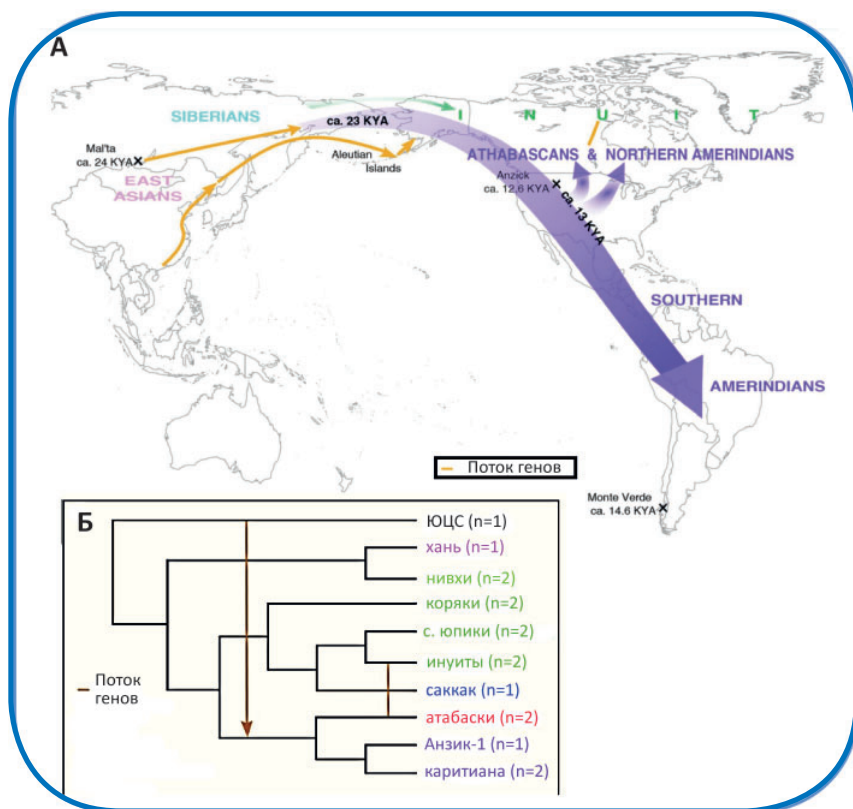
Высокогорные поперечные долины Гималайской дуги от Аруначал-Прадеша до Ладакха были одними из последних обитаемых мест, постоянно заселяемых доисторическими людьми. Современное население этих долин имеет культурное и языковое сходство с народами, проживающими сегодня на Тибетском плато, и считается потомком самых ранних обитателей Гималайской дуги, хотя это предположение неоднократно оспаривалось археологами. Секвенирование ядерного и митохондриального геномов восьми человек из Непала, датированных 3150–1250 гг. до н.э., показало, что для региона характерна многолетняя стабильность генетического состава населения вопреки выраженным изменениям в материальной культуре. Значительное генетическое сходство с современными высокогорными выходцами из Восточной Азии, а также общие высотные адаптивные аллели предполагают высокогорное восточноазиатское происхождение доисторических гималайских популяций (Jeong et al., 2016).

Современные японцы представляют собой уникальную популяцию, геномное разнообразие которой сформировалось в результате многочисленных миграций из Евразии. Анализ 12 геномов древних японцев показал, что население Японии происходит

от коренных охотников–собираателей–рыболовов периода Дзёмон (насаляли японские острова 20000–15000 л. н.), земледельцев периода Яёй (мигранты из северо-восточной Азии, прибывшие примерно 2300 л. н.) и более позднего притока восточноазиатского происхождения в период Кофун (250–538 г. н. э.). Эти три наследственных компонента остаются важными характеристиками современных популяций японцев (Cooke et al., 2021). С помощью широкомасштабного полногеномного секвенирования 9850 образцов из сети национальных центров биобанков, были проанализированы геномные характеристики, которые определяют генетический состав современного населения Японии с точки зрения популяционной генетики. Помимо обнаружения низкочастотных вредных или патогенных вариантов, было выявлено два эпизода сокращения популяции о. Хондо (Хонсю), в период Дзёмон (соответствует позднему неолиту) и в период Эдо (соответствует эпохе раннего Нового времени), в то время как популяция островов Рюкю испытала сокращение в период «отложения ракушек» (всемирное археологическое явление, чаще всего встречающееся на участках, прилегающих к морским берегам) позднего неолита. Кроме того, выявлены уникальные следы естественного отбора, многие из которых ассоциированы с изменениями в рационе японцев (например, на гены, связанные с метаболизмом алкоголя и липидов). Таким образом, геномное разнообразие японцев было сформировано событиями, тесно связанными с сельским хозяйством и производством продуктов питания (Kawai et al., 2023).

Вопрос о происхождении первых американцев остаётся во многом спорным, с противоречивыми интерпретациями, опирающимися на антропологические и генетические данные. Вместе с тем, существует общепринятое мнение, что предки современных коренных американцев проникли в Америку из Сибири через Берингов мост, и что это произошло позднее 14600 л.н. (рис. 5.12) (Raghavan et al., 2015). Черновые версии генома человека возрастом примерно 24000 л. с юга центральной Сибири подтвердили гипотезу сибирского происхождения современных коренных американцев. Было показано, что сибиряки имеют генетическое сходство как с западноевропейским населением, так и с современными коренными американцами, причём 14–38% предков коренного населения Америки могут происходить в результате потока генов из этой древней популяции (Raghavan et al., 2014). Дальнейшие палеогеномные исследования показали, что верхняя граница для первоначального расхождения предков коренных американцев с их предками из Восточной Азии составляет около 23000 л.н., затем следует короткий период изоляции (не более 8000 л.) и последующее проникновение и распространение по Америке. Диверсификация (вероятно, в пределах Америки), давшая начало северной и южной ветвям современных коренных американцев, произошла около 13000 л.н. (Raghavan et al., 2015).

Народы первых миграционных волн широко распространились и разделились на несколько популяций, а их потомки испытали различную степень изоляции и смешения. Например, анализ геномных профилей 51 представителя ашанинка из Амазонского Перу обнаружил неожиданно высокую степень геномной изменчивости, отражающую генетическую структурированность внутри этой группы. Согласно данным,



**Рисунок 5.12.** Происхождение и история населения коренных американцев (из: Raghavan et al. 2015).

А – Предки всех современных коренных американцев, включая американских индейцев и атабасков, произошли от единой волны миграции в Америку из Восточной Азии (фиолетовые стрелки) не позднее 23 т. л. н. отдельно от инуитов (зелёные стрелки); раскол между северной и южной ветвями коренных американцев произошел ~13 т. л. н.; показан поток генов из южной части центральной Сибири (Mal'ta) к предкам коренных американцев ~23 т. л. н. и примеси между инуитами, атабасками и северными индейцами (коричневые стрелки).

Б – Все коренные американцы образуют кладу, отдельную от инуитов, с потоком генов между некоторыми коренными американцами и американцами северной Арктики.

ЮЦС – народ южной части центральной Сибири, хань – народность Китая, с. юпики – сибирские юпики (юиты), саккак – палеоэскимос Гренландии, Анзик-1 – палеоиндеец, каритиана – народ в Бразилии.

предки ашанинка мигрировали с юга на север через Южную Америку, в результате чего произошло их значительное смешение с коренными популяциями Анд и Тихоокеанского побережья. Кроме того, было установлено существование генетической связи предков современных ашанинка с араваками и карибами в период раннего керамического века (Capodiferro et al., 2023). Геномные исследования также показали, что все древние народы Америки, за исключением более поздних арктических групп, таких как палеоинуиты и инуиты, имеют гораздо более тесные генетические связи с

современными коренными американцами (индейцами), чем с любыми другими популяциями в мире. Примечательно, что у одного древнего человека и у современных жителей из относительно небольшой области Бразилии, а также у современных коренных американцев, которые населяют регион тихоокеанского побережья Южной Америки, был обнаружен слабо выраженный Австралийский геномный сигнал (генетический вариант). Следовательно, есть вероятность, что и другие популяции могли внести свой вклад в формирование коренных народов Америки (Willerslev, Meltzer, 2021).

Популяции человека, занимающие обширные территории северной Евразии, от Фенноскандии до Чукотки, обладают ярко выраженным генетическим компонентом, который часто называют сибирским происхождением. Несмотря на множество современных самодийских популяций, таких как нганасаны, история этих народов по-прежнему остаётся загадочной, даже с учётом увеличивающегося числа прочитанных древних и современных геномов из Сибири (Gill et al., 2024). Масштабный геномный анализ привёл к ряду интересных выводов. Геномный ландшафт Внутренней Евразии сформировался в процессе сложной истории смешения между различными предковыми популяциями как восточного, так и западного происхождения. Генетический состав сибиряков среднего голоцена образовался в результате смешения трёх предков: древнего североевразийского (ANE), древнего палеосибирского (APS; предок, тесно связанный с предками коренных американцев) и древнего североазиатского (ANA). Во время последнего ледникового максимума предки ANE смешались с популяциями восточноевразийского происхождения и сформировали предковую популяцию коренных американцев. В начале среднего голоцена представители ANA-родословной стали появляться как на восточных, так и на западных берегах Байкала, вероятно, распространившись из соседних районов северо-восточного Китая. В среднем голоцене популяция APS присутствовала в районе Байкала и Якутии, и в каждом регионе локальная популяция APS сформировала генетическую основу для более поздних популяций. В этот период генетический профиль популяций Западного Байкала претерпел существенные изменения, со сдвигом от богатого ANA к богатому APS, что совпало с переходом от китойской к серово-глазковской культуре. Регион, простирающийся от Западного Байкала до Алтайских гор, мог служить рефугиумом для предков APS, и, вероятно, между Западным Байкалом и Алтаем существовала определённая связь, опосредованная потоком генов предков APS из районов Западного Байкала на Алтай. Среднеолитический человек из Якутии, относящийся к культуре Белькачи, оказался наиболее вероятным источником распространения сибирского происхождения в Фенноскандии и Гренландии; его генетический компонент обнаружен у северных сибиряков, северо-восточных европейцев, палеоэскимосов и древних атабасков (Gill et al., 2024).

Отдельный интерес для геномных исследований представляют предки одомашненных видов животных и растений. **Одомашнивание** (доместикация) диких животных и растений открыло новую эру в истории человечества, сыграв важную роль

в становлении цивилизации и в формировании человека современного типа. За последние 15000 л. в условиях изоляции от диких сородичей и адаптации к новой нише фенотипы и генотипы одомашненных видов сильно изменились. Николай Иванович Вавилов отмечал, что одомашнивание является формой экспериментальной эволюции, и для её успешного управления важно обладать историческим пониманием эволюционных процессов (Вавилов, 1987). Палеогеномные данные помогли реконструировать некоторые механизмы, с помощью которых происходила одомашнивание (Frantz et al., 2020).

Согласно археологическим свидетельствам, одомашнивание лошади началось примерно 5500 л.н. и оказало большое влияние на жизнь человека, коренным образом изменив его способность к быстрому передвижению на большие расстояния и ведение боевых операций. Имеется ряд работ, посвящённых изучению генетических процессов, лежащих в основе одомашнивания лошадей, в которых изучались древние геномы в различные временные отрезки их истории. Примечательно, что при этом удалось повысить верхний предел сохранения ДНК в костных останках млекопитающих с возраста 120000 л. (геном самого древнего белого медведя) до примерно 735000 л. (Lan, Lindqvist, 2019). Палеогеномные исследования показали, что современные одомашненные породы не происходят от самой ранней линии домашних лошадей Ботая, или Иберии и Анатолии, как предполагали археологические находки. Родиной современных лошадей могли быть западно-евразийские степи и особенно – нижнее Волго-Донье. Картирование изменений в популяциях на основе 273 древних геномов лошадей обнаружило, что современные домашние лошади вытеснили почти все местные популяции, поскольку они быстро распространялись по Евразии примерно с 2000 г. до н.э. Результаты отвергают широко признаваемую связь между верховой ездой и массовой экспансией скотоводов, носителей ямной культуры, в Европу около 3000 г. до н. э., что привело к распространению индоевропейских языков. Тем не менее, в Азии в начале второго тысячелетия до н. э. индоиранские языки, колесницы и лошади распространились вместе, сопутствуя синташтинской культуре (Librado et al., 2021). Кроме того, оказалось, что верховая езда включала в себя строгий отбор ключевых двигательных и поведенческих адаптаций в генах *GSDMC* и *ZFPPI1* (Librado et al., 2021), многочисленные аллели которых, связанные со скоростью бега, увеличились в частоте в последние 600–1100 л. (Fages et al., 2019), а вариант гена *DMRT3*, связанный с иноходью, появился в Англии примерно в 850 г., и затем распространился по континентальной Европе (Wutke et al., 2016). Полагают, что большинство пород современных лошадей произошло от персидской линии (Fages et al., 2019).

Собаки являются первыми и единственными полностью одомашненными крупными хищниками (волкоподобные хищники семейства псовых *Canidae*), и их одомашнивание стало важным событием в истории человечества. Сравнение митохондриальных геномов вымерших и современных популяций волков *Canis lupus* Linnaeus, 1758 показало, что современные волки произошли от популяции, существовавшей около 25000 л.н. Однако жили они не на юге Европы или Средней Азии, как ранее

предполагали зоологи, а на территории Берингии. Поселившись вначале в Евразии, они затем (около 15000 л.н.) колонизировали Северную Америку и часть Центральной Америки (Loog et al., 2020). Данные секвенирования генома плейстоценового волка с полуострова Таймыр показали, что он хорошо отличается как от предков собак, так и от современных волков. Радиоуглеродный метод определил его возраст в 35000 л., что позволило калибровать молекулярные часы псовых (Thalmann, Perri 2018). Время дивергенции волка и собаки по данным древних геномов составляет примерно 36900–41500 л., а одомашнивание собаки датируется 20000–40000 л. Причём получены убедительные доказательства генетической преемственности от палеолита к неолиту и, в некоторой степени, к настоящему времени (Botigué et al. 2017). Данные митохондриального генома показали, что линия предков самой древней собаки Северной Америки из Аляски возрастом 10150 л., отделилась от линии сибирских собак примерно 16700 л.н. Это совпадает с датировкой прибрежного маршрута северной части Тихого океана и поддерживает гипотезу о первой миграции людей и собак вдоль побережья, а не по центральному континентальному коридору (da Silva Coelho et al., 2021).

Древняя ДНК была также успешно экстрагирована из различных типов растительных тканей, сохранившихся в самых разных условиях (Schlumbaum et al., 2008). Первые генетические исследования древней ДНК растений внесли существенный вклад в понимание процессов одомашнивания, адаптации и распространения сельскохозяйственных культур (Palmer et al., 2009, 2012; Jaenicke-Despres et al., 2003). Древние геномы способствуют нашему пониманию механизмов и темпов эволюции растений и позволяют исследовать отбор при одомашнивании, а геномы гербарных растений помогают понять динамику инвазионной флоры (Gutaker, Burbano, 2017). Однако прогресс в восстановлении полных геномов древних растений значительно отстаёт от такового для животных и человека. Основная проблема связана с недостаточным количеством современных эталонных геномов, большим размером и сложной организацией генома растений, наличием в них большой доли повторяющихся элементов и различными уровнями пloidности (Lan, Lindqvist, 2019).

Палеогеномные исследования злаков раскрывают пути их расселения, механизмы адаптации и одомашнивания, что может служить основой для современной селекции. Эволюционная история кукурузы *Zea mays* L. ssp. *mays* довольно сложная. Археологические находки предполагают, что её одомашнивание произошло 10000–6250 л.н. на юге Мексики. Внешне современная кукуруза сильно отличается от своего древнего предка теосинте *Zea mays* subsp. *parviglumis* Iltis & Doebley, 1980, стебель которого напоминает кукурузный, но заканчивается он не початком, а небольшим колоском с зёрнами. Сравнительный анализ ядерной ДНК (экзоны 348 генов) из 32 археологических образцов кукурузы, охватывающих 6000 л. эволюции, с современными местными сортами обнаружил, что первоначальное, около 4000 л.н., продвижение кукурузы на юго-запад, вероятно, происходило по высокогорному маршруту, но примерно 2000 л.н. последовал поток генов из прибрежных низменных популя-

ций. Популяционный анализ геномных данных позволил дифференцировать отбор при одомашнивании на адаптацию к климатическим и культурным условиям юго-запада, выявляя локусы адаптации, относящиеся к засухоустойчивости и содержанию сахара (da Fonseca et al., 2015). Исследование палеогенома кукурузного початка возрастом 5310 л. из долины Теуакан в Мексике обнаружило, что этот древний образец представляет базальную линию всех современных разновидностей. Древняя кукуруза была частично одомашнена и содержала предковые аллельные варианты, отсутствующие в современных популяциях. Некоторые локусы одомашнивания уже утратили большую часть нуклеотидной изменчивости, свойственной теосинте, другие сохранили частичную изменчивость нуклеотидов, отсутствующую у современной кукурузы. Таким образом, этот образец является промежуточным вариантом между кукурузой и теосинте, а процесс одомашнивания происходил постепенно (Ramos-Madrigal et al., 2016; Vallebuena-Estrada et al., 2016).

Согласно археологическим данным, злаковый ячмень *Hordeum vulgare* L. (1753) был одомашнен примерно 10000 л. н. в «Плодородном полумесяце» и стал основной культурой неолитического земледелия. Реконструкция палеогеномов пяти 6000-летних зёрен ячменя, извлечённых из пещеры в Иудейской пустыне недалеко от Мёртвого моря, и их сравнение с последовательностями полного экзона современных образцов ячменя показали близкое родство древних растений с существующими местными сортами из Южного Леванта и Египта. Это согласуется с предполагаемым происхождением одомашненного ячменя в долине Верхнего Иордана. Местные сорта ячменя, выращиваемые в современном Израиле, не претерпевали значительной смены родословных за последние шесть тысячелетий, хотя существуют доказательства обмена генами между культивируемыми и симпатрическими дикими популяциями (Mascher et al., 2016).

Канарские острова были заселены в первом тысячелетии н.э., вероятно, выходцами из Северной Африки, которые развили фермерское хозяйство с ячменём в качестве основной культуры. В популяционном исследовании древней ДНК из двадцати одного археоботанического зерна ячменя с о. Гран-Канария, датированного периодом с 1050 по 1440 г. н.э. (с исключительно высокой сохранностью материала), было проведено генотипирование по 99 однонуклеотидным маркерам. Кроме того, был генотипирован 101 образец современных местных сортов Канарских островов и западного Средиземноморья. Археологический материал показал высокое генетическое сходство с существующими местными сортами Канар. Однако, образцы с Канарских островов сильно отличались как от иберийского, так и от материкового ячменя Северной Африки. На Канарах местные сорта самых восточных и западных островов генетически хорошо отличались между собой, что подтверждает наличие доиспанского выращивания ячменя на Лансароте (Hagenblad et al., 2017).

Прорастание семян финиковой пальмы *Phoenix dactylifera* L. (которая славится своими вкусными плодами) возрастом примерно 2000 л. облегчило сравнение её генома с геномами современных видов. Оказалось, что самые старые семена связаны с

современными сортами фиников из Западной Азии, тогда как более поздние семена обнаруживают большее генетическое родство с современными североафриканскими финиковыми пальмами. Древняя популяция *P. dactylifera* в Восточном Средиземноморье, по-видимому, содержит интрогрессированные сегменты генома критской пальмы *P. theophrasti* (Gros-Balthazard et al., 2021). Эти уникальные данные представляют ценный ресурс для разработки новых стратегий селекции финиковой пальмы, способных значительно улучшить вкусовые качества её плодов.

### 5.3. Экология: сохранение биоразнообразия и управление биоресурсами

В настоящее время большинство существующих экосистем находится под угрозой антропогенной деградации, а изменение климата считается одной из основных причин сокращения и/или утраты биоразнообразия в нынешнем столетии. Поэтому особенно актуальным становятся развитие и реализация стратегий восстановления, направленных на обеспечение устойчивости экосистем. Выбор наиболее подходящих источников семян для восстановления растительного покрова в сильно деградированных ландшафтах является одним из первых шагов в этом направлении. Решающее значение в выборе источника семян для лесовосстановления, улучшения управления лесами и прогнозирования генетического сдвига, связанного с изменением климата, имеет понимание адаптивной генетической изменчивости популяций в ответ на климатические изменения (Jia et al., 2020).

Стратегии поиска семян включают отбор локальных семян, высоко приспособленных к существующим местообитаниям, а также отбор генотипов, совместимых с будущими условиями (Rico, 2021). Считается, что подходы «прогнозирующего» восстановления, в которых используются семена с благоприятной адаптацией к будущему климату, являются перспективными для практического применения и способствующими долгосрочной устойчивости, тогда как местная адаптация может создать определённые трудности для усилий по восстановлению экосистем (Shryock et al., 2021). Генотипирование в рамках ландшафтной геномики 760 растений из 64 популяций подорожника *Plantago ovata* Forsk пустыни Мохаве (США) выявило 184 потенциально адаптивных локуса и доказательства того, что максимальная летняя температура может как ограничивать ареал *P. ovata*, так и стимулировать адаптивную дивергенцию в популяциях, подверженных воздействию более высоких температур. Согласно прогнозам, в пределах нынешнего ареала вида генотипы, адаптированные к теплу и засухе, к середине столетия испытают пятикратное расширение своей климатической ниши и смогут иметь ключевые адаптации для выживания в условиях нового климата. Именно генотипы, которые могут расширить пределы будущей среды обитания, являются приоритетными для сбора семян (авторы исследования рекомендовали восемь конкретных зон). Для их поиска подходят регионы, подверженные быстрому изменению климата, где местные генотипы больше не являются предпочтительными (Shryock et al., 2021).

Адаптацию, миграцию и вымирание считают основными реакциями видов на изменение климата, а генетическую изменчивость – «сырьём» для адаптации к новой среде (Chu et al., 2021). Принципы филогеографии и ландшафтной геномики были применены к эндемичному виду цветкового растения *Roscoea humeana* Balf. F. & W. Sm., произрастающего в горах хребта Хэндуан (Китай). Для анализа генетической структуры, разнообразия и изменений в эффективном размере популяции с течением времени были использованы в общей сложности 5902 однонуклеотидных полиморфизмов. Результаты показали, что современная генетическая структура и разнообразие *R. humeana* были сформированы под влиянием четвертичных колебаний климата и благодаря устойчивости микрорефугиумов в горах Хэндуан. Будущее выживание вида возможно в рефугиумах, находящихся в пределах области с количеством осадков в самой тёплой четверти года выше 560 мм, причём местная адаптация к засухоустойчивости может быть полезной для будущих условий (Chu et al., 2021).

Леса являются центральным компонентом ландшафтной матрицы, они обеспечивают среду обитания для множества видов и имеют огромное экономическое значение. Поэтому необходимо понимать, какие факторы влияют на их долгосрочную стабильность, и какова связь между геномными и ландшафтными изменениями. Согласно результатам нейтральной ландшафтной геномики, многие виды деревьев оказались устойчивыми к фрагментациям леса из-за высокого уровня потока пыльцы на большие расстояния. Очевидно, это обстоятельство также способствовало послеледниковой колонизации северных территорий. В то же время, ландшафтная геномика выявила географические барьеры, ограничивающие поток генов для ряда древесных видов (Balkenhol et al., 2017).

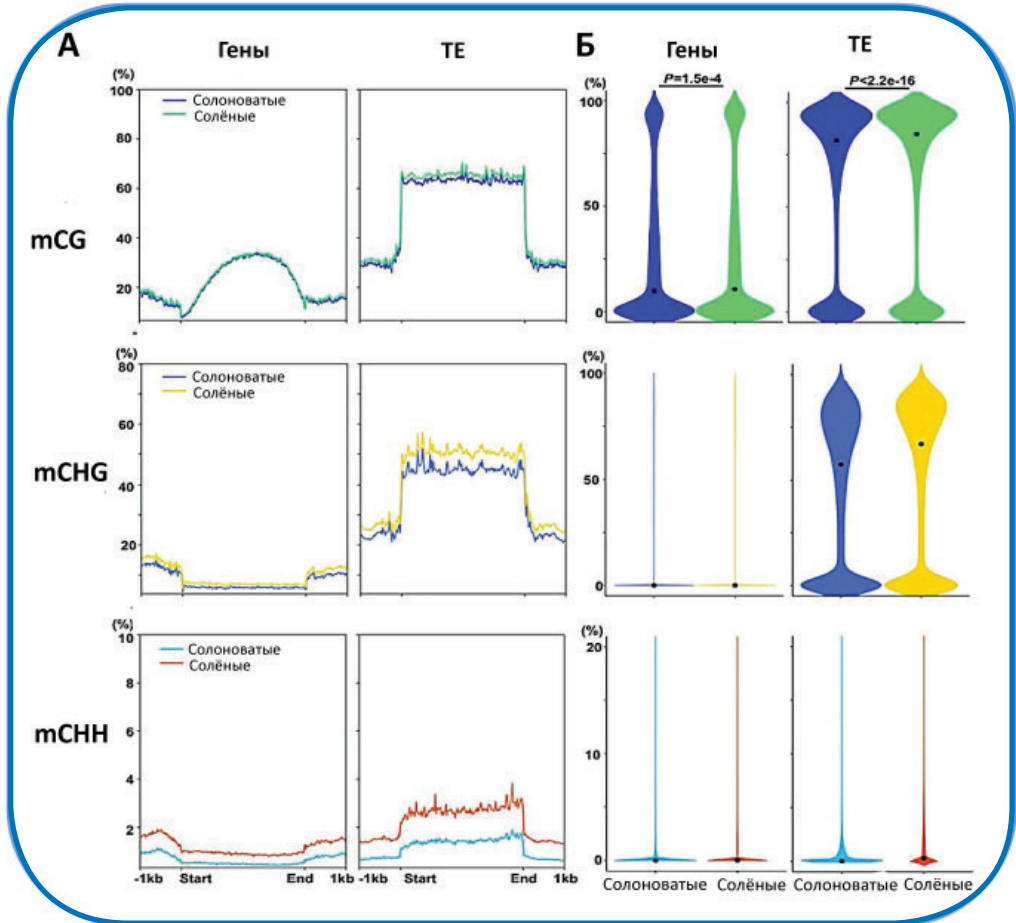
Исследования ландшафтной геномики долинного дуба *Quercus lobate* Née выявили гены, участвующие в адаптации к местному климату. Из примерно 10000 однонуклеотидных полиморфизмов, обнаруженных с помощью секвенирования генома, признаки естественного отбора были выявлены в более чем 600 SNP, а также была установлена нелинейная ассоциация SNP-кандидатов и переменных окружающей среды (Gugger et al., 2021). Изучение широко распространённого в Мексике высокогорного вида дуба *Quercus rugosa* Née (всего идентифицировано 5354 SNP для 103 особей на 17 участках Транс-Мексиканского вулканического пояса) позволило выявить 97 SNP, связанных с климатическими изменениями. При этом пространственные закономерности генетической вариабельности оказались наиболее сильно связаны с сезонностью осадков и географическим расстоянием. Кроме того, удалось определить регионы, где будущие популяции *Q. rugosa* могут оказаться в опасности из-за ожидаемой высокой скорости изменения климата (Martins et al., 2018). Эти данные являются хорошей иллюстрацией того, как геномный анализ с географической привязкой может стать полезным для стратегий сохранения и управления лесными ресурсами.

Подходы ландшафтной геномики были также использованы для изучения генетических основ, определяющих локальную адаптацию широко распространённой в

Китае лапины узкокрылой *Pterocarya stenoptera* DC. В общей сложности выявили 801 однонуклеотидный полиморфизм, который достоверно коррелировал с различными факторами окружающей среды. Были установлены гены-кандидаты, участвующие в адаптации к изменениям температуры, наличию воды и интенсивности света. Результаты исследования также показали, что естественный отбор, сопровождавший локальную адаптацию, способствовал успешному выживанию вида в гетерогенных условиях окружающей среды, но существенно не повлиял на его генетическую структуру. Напротив, солнечная радиация во время развития плодов, сезонность температур, годовая температура, осадки и влажность воздуха оказали значительное влияние на адаптивную дивергенцию популяций *P. stenoptera*. В целом, это исследование даёт необходимое представление о том, как виды растений адаптируются к различным экологическим нишам, а поскольку *P. stenoptera* является важным садовым растением, результаты исследования могут быть полезны для его интродукции, выращивания, защиты и контроля (Li et al., 2020b).

Хотя высотная адаптация давно привлекает внимание экологов и эволюционистов, генетические механизмы, лежащие в её основе, остаются недостаточно изученными. Некоторое представление о них дают геномные исследования сычуаньского тополя *Populus szechuanica* Schneid var. *tibetica* Schneid. Генотипирование и сравнение популяций тополя, распространённого на высоте 2000–3000 м горы Седжила на Цинхай-Тибетском плато (Китай), позволило чётко классифицировать две группы в соответствии с их высотным распределением. Кроме того, был обнаружен асимметричный поток генов от высокогорных к низинным популяциям, а также две «горячие точки» в геноме, одна из которых содержит четыре гена, связанных с высотной изменчивостью, а другая – десять генов, ассоциированных с реакцией на солнечную радиацию. Эти гены участвуют в устойчивости к абиотическим стрессам и в регуляции репродуктивных процессов (Zheng et al., 2020).

Мангровые заросли приспособлены к суровым условиям, таким как сильное ультрафиолетовое излучение, анаэробные почвы и колебания солёности в прибрежных зонах. Геномный (сборка *de novo*) и эпигеномный анализы бругиеры голокорневой *Bruguiera gymnorhiza* (L.) Savigny, одного из доминирующих видов мангровых зарослей, показал, что резкие морфологические изменения, ассоциированные со средой произрастания естественных популяций, связаны с изменением экспрессии тысяч генов. Более того, среда с высокой солёностью индуцировала гиперметилирование ДНК мобильных элементов в масштабах всего генома бругиеры (рис. 5.13), которое происходило одновременно с регуляцией транскрипции генов-модификаторов хроматина. Это свидетельствует о сильной эпигеномной регуляции мобильных элементов в геноме *B. gymnorhiza* в условиях высокой солёности. В целом, результаты исследования дают новое представление о регуляции эпигенома у растений мангровых зарослей и способствуют лучшему пониманию их адаптации к изменчивым и суровым прирдным условиям (Miryeganeh et al., 2022).



**Рисунок 5.13.** Метаграфики (А) и скрипичные графики (Б) уровней метилирования цитозина в контекстах CG, CHG и CHH для генов и мобильных элементов (TE) бругиеры голокорневой *Bruguiera gymnorhiza* в солоноватых и солёных условиях в естественной среде (из: Miryeganeh et al., 2022).

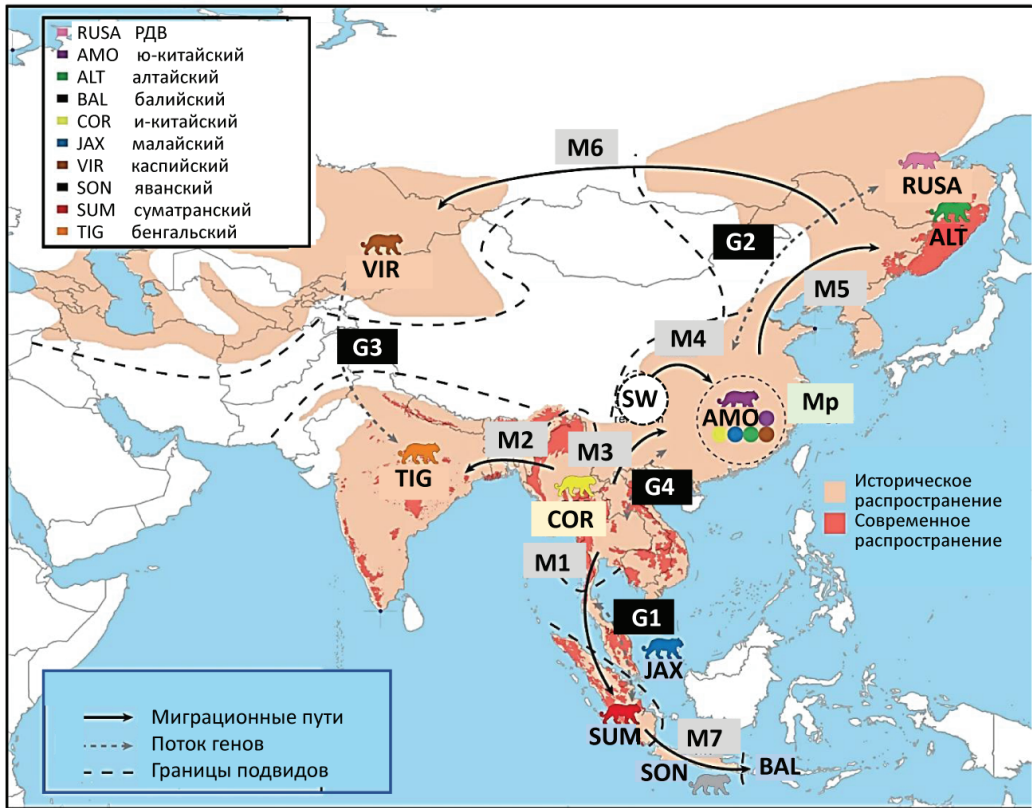
На дне современных и существовавших в прежние геологические эпохи озёр находятся осадочные образования, которые могут иметь высокий потенциал для исследований sedaDNA. В них содержится огромное количество информации, представляющей большую ценность для палеолимнологических реконструкций (Brasell et al., 2022). Исследование озёрных отложений на Шпицбергене возрастом более 8000 л. с помощью методов анализа sedaDNA позволило идентифицировать почти все макрофоссилии, включая шесть дополнительных (к уже известным) таксонов сосудистых растений, а также два таксона водорослей и 12 таксонов мохообразных. В целом флора продемонстрировала устойчивость к изменяющемуся климату, поэтому если будущее потепление будет ограничено  $2^{\circ}\text{C}$ , можно ожидать лишь незначительных

флористических изменений в этом регионе (хотя летопись голоцена не даёт аналогов большому потеплению) (Alsos et al., 2016). Исследования sedaDNA выявили прошлые изменения экосистемы и биоразнообразия на Тибетском нагорье (Jia et al., 2022). Анализ геномной ДНК человека и млекопитающих из верхнепалеолитических отложений пещеры Сацурблия (Грузия) возрастом 25000 л. обнаружил предковый компонент большинства людей послеледникового периода, геномы ранее неизвестной и, вероятно, вымершей кавказской линии волков, а также европейского бизона (Gelabert et al., 2021). Эти исследования дают важную информацию об историческом развитии флоры и фауны и являются ценным ресурсом для прогностических реконструкций.

Глобальное изменение окружающей среды коренным образом влияет на экологические системы, и широко распространено мнение, что урбанизированные ландшафты представляют собой деградировавшую окружающую среду, несовместимую с дикой природой. Однако некоторые виды смогли успешно адаптироваться в таких условиях. Например, ландшафтная геномика выявила доказательство местной адаптации лесных лягушек к урбанизации: 37 из 344 генотипированных SNP были оценены как находящиеся под прямым отбором, а 12 из 37 локусов имели умеренную поддержку ассоциации с типом окружающей среды. Эти данные свидетельствуют о способности городских ландшафтов управлять эволюционными процессами, причём аборигенные виды, сохраняющиеся в городской среде, могут иметь преимущества от местной адаптации к новым факторам отбора (Nomola et al., 2019).

Как отмечают авторы исследований, ни один другой вид не привлекает больше международных ресурсов и общественного внимания, чем тигр *Panthera tigris* Linnaeus, 1758 – харизматичный символ плейстоценовой мегафауны. Исторически этот хищник, который возник, эволюционировал и сохранился лишь в лесах Азии, населял обширные территории западнее Каспийского и Аральского морей, а также Северо-Восточной и Юго-Восточной Азии, включая Зондские острова (Sun et al., 2023a). В настоящее время насчитывается менее 4000 свободноживущих особей этого вида, покрывающих всего 7% их исторического ареала, для чёткого разграничения единиц управления и сохранения которых необходима ясность в отношении количества подвидов (предмет многолетних дискуссий). Полногеномное секвенирование 32 ваучерных образцов позволило выделить шесть статистически надёжных монофилетических клад, соответствующих существующим подвидам. Идентифицировано четырнадцать генов, связанных с адаптивной дивергенцией у современных тигров, а самый сильный сигнал потенциально положительного отбора обнаружен в гомозиготном районе размером 180 т. п. н., который отличал суматранского тигра от всех других подвидов. Этот район включает ген *ADH7*, гомозиготные мутанты которого имеют сниженную массу тела у млекопитающих, что может объяснить феномен островной карликовости, наблюдаемый у суматранского тигра. Ген *ADH7*, по-видимому, подвергся строгому отбору у этого подвида, возможно, в связи с его адаптацией к условиям тропических Зондских островов. Кроме того, полногеномный анализ ассоциаций выявил две аминокислотные замены в *SLC45A2* (транспортный белок, кодируется геном пигментации *Slc45a2*) и *CORIN* (трансмембранная протеаза), которые определяют





**Рисунок 5.15.** Предполагаемые пути распространения тигра, основанные на совместной интерпретации геномов современных и древних представителей вида (из: Sun et al., 2023a). Эволюционные пути реконструированы на основе полногеномной филогении, байесовского молекулярного датирования, моделирования миграции и потока генов, а также моделирования экологических ниш подходящих местообитаний. Указаны семь путей распространения (M1–M7) и четыре основных события потока генов после дивергенции (G1–G4) среди популяций тигров. Исторический рефугиум для тигров был постулирован на юго-западе Китая (SW), а сценарий генетического плавильного котла – в восточном Китае (Mp), что подчёркивает значение материкового Китая как трамплина во время экспансии тигров.

а один экземпляр возрастом 2000 л. оказался близким к современным амурским тиграм. Каспийский тигр, вероятно, отделился от предковой популяции Северо-Восточной Азии и испытал поток генов от южных бенгальских сородичей. Филогеномный анализ подтвердил подвидовой статус южно-китайского тигра, хотя и с митохондриальной парафилией, тем самым разрешив давние таксономические споры (Sun et al., 2022b, 2023a). Распределение митохондриальных гаплогрупп, подтверждённое биогеографическим моделированием, позволяет предположить, что Юго-Западный Китай был позднеплейстоценовым убежищем для реликтовой базальной линии тигров. Восстановление благоприятной среды обитания сделало Восточный Китай генетическим «плавильным котлом», способствующим слиянию расходящихся линий с юж-

но-китайскими тиграми и другими северными подвидами для дальнейшей эволюции (Sun et al., 2023a).

Для сохранения и управления биоразнообразием необходимо учитывать адаптивные способности видов, т. е. знать основные селективные силы, действующие на экосистемы и их возможное влияние на эволюционный потенциал популяций и видов (Nielsen et al., 2020a). Популяции находящегося под угрозой исчезновения каспийского полоза *Dolichophis caspius* (Gmelin, 1789) по всему ареалу становятся всё более разобщенными из-за изменения среды обитания. Анализ в рамках ландшафтной геномики, направленный на понимание динамики и адаптивного потенциала популяций, выявил 757 генотипов, коррелирующих с семью климатическими переменными. Причём оказалось, что колебания температуры от дня к ночи могут иметь более сильное влияние на распространение и адаптацию этого вида по сравнению с температурными колебаниями от лета к зиме. Несмотря на выраженную генетическую изоляцию, среди популяций широко распространены генотипы, связанные с определёнными переменными окружающей среды, предполагая наличие во всех популяциях *D. caspius* соответствующего адаптивного потенциала, необходимого для выживания (Mahtani-Williams et al., 2020). Исследование островных лис *Urocyon littoralis* (Baird, 1858) у побережья южной Калифорнии (США) показали, что общая геномная дифференциация популяций объясняется в основном генетическим дрейфом или местной адаптацией (предположительно 3,3–6,6% локусов подвержены отбору). Однако значимых ассоциаций этих локусов и данных окружающей среды обнаружено не было, хотя геномная дифференциация по выпадающим локусам соответствовала характеру морфологического сходства (Funk et al., 2016). Результаты геномного анализа, направленные на исследования процессов, управляющих популяционно-генетическими паттернами белогородного поползня *Sitta carolinensis* Latham, 1790 (семейство поползневых) на Небесных островах (изолированные горы, окружённые радикально отличающимися низменностями) Мадрейского архипелага (горная цепь, соединяющая северную оконечность Западной Сьерра-Мадре и южную часть плато Колорадо), выявили изоляцию окружающей средой и отсутствие изоляции расстоянием среди островных популяций вида. Интересно, что из 79 обнаруженных локусов, предположительно находящихся в процессе отбора, семь совпадали с кодирующими регионами у зебровой амадины (птица семейства вьюрковых ткачиков). Выбранные локусы имели выраженную связь с экстремальными климатическими условиями, т. е. максимальной температурой самого тёплого месяца и минимальным количеством осадков в самый засушливый месяц. Это первые результаты, предоставившие доказательства изоляции окружающей средой и отсутствие изоляции расстоянием у позвоночных островов Мадрейского архипелага, а также свидетельства потенциально адаптивной генетической изменчивости (Manthey, Moyle, 2015).

Гетерогенность ландшафта влияет на поток и миграцию генов, формируя в итоге генетическую структуру и изменчивость популяций, а понимание взаимодействия между ландшафтными переменными и генетической дифференциацией мо-

жет прояснить, как создаётся, поддерживается и распространяется биоразнообразие (Chan, Brown, 2019). Тропические Анды и межандийские долины образованы крайне неоднородными ландшафтами, где виды подвергаются сильному давлению отбора. На примере лягушек *Boana platanera* Escalona, Marca, Castellanos, Fouquet, Crawford, Rojas-Runjaic, Giaretta, Señaris & Castroviejo-Fisher, 2021, обитающих на банановых деревьях, была проанализирована взаимосвязь между генетической и фенотипической дифференциацией, связанной с окружающей средой, и потенциальными факторами изоляции окружающей средой вдоль градиента высоты. Картина дифференциации геномов оказалась существенно связанной с температурой, при которой *B. platanera* демонстрирует значительные различия голоса и размера тела с популяциями низменностей тёплой долины Магдалены и популяциями более прохладных и влажных тропических лесов на склонах Центральных Кордильер. Большая группа обнаруженных кандидатных локусов, связанных с колебаниями температуры, указывает на фенотипическую и геномную дифференциацию у *B. platanera*, распределённых по градиенту высот в тропических Андах Колумбии (Medina et al., 2021).

Ландшафтная геномика даёт возможность оценить долгосрочную функциональную взаимосвязь (имеющую важное значение для поддержания здоровых популяций диких животных), связывая переменные среды с пространственными моделями геномной изменчивости, возникающими в результате передвижений животных и их брачного поведения. Известно, что фрагментация и утрата среды обитания происходит всё чаще, существенно сокращая возможности передвижения высокоомобильных видов. Поэтому важным является понимание того, как естественная и антропогенная среда влияют на поток генов в таких ситуациях (LaCava et al., 2021). В качестве модели была охарактеризована генетическая структура и разнообразие мигрирующего оленя-мула *Odocoileus hemionus* (Rafinesque, 1817) на основе анализа 4051 SNP у 406 особей из различных местообитаний штата Вайоминг (США). Выделены три смешанные генетические группы *O. hemionus* и показано, что переменные среды, связанные с потоком генов, различались между группами. В целом данные указывают как на крупномасштабные, так и на региональные различия в функциональной связанности и на то, что антропогенные особенности ландшафта препятствуют такой взаимосвязи. Эта информация имеет несомненную ценность для планирования землепользования и природоохранных мероприятий (LaCava et al., 2021).

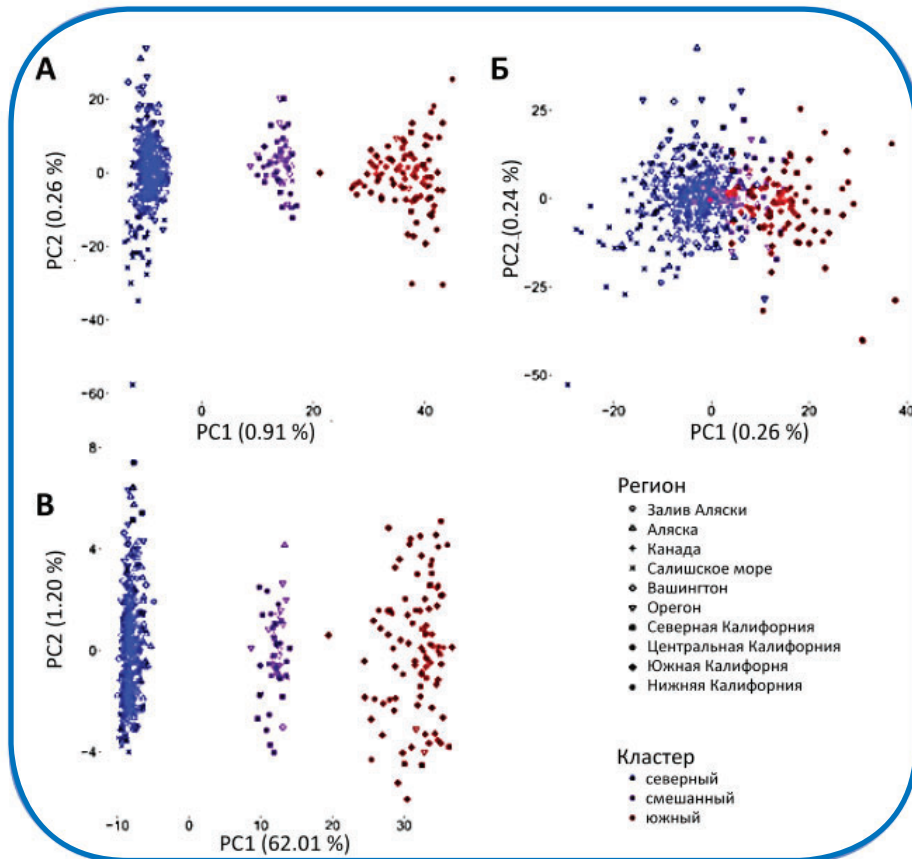
Водные экосистемы, особенно прибрежные, подвержены сильному антропогенному давлению и глобальным изменениям климата, что ставит под угрозу существование биологических сообществ. На формирование генетических (по нейтральным и адаптивным маркерам) структур морских популяций влияют такие факторы, как течения и градиенты температур, солёности и кислорода, особенно в прибрежных акваториях. В рамках морской ландшафтной геномики сравнивали предполагаемые экологические факторы отбора у трёх симпатрических видов южноафриканских морских беспозвоночных с контрастными экологией и историей жизни: морского ежа *Parechinus angulosus* (Leske, 1778), берегового краба *Cyclograpsus punctatus* H. Milne-Edwards, 1837 и морской улитки *Scutellastra granularis* (Linneus, 1758). Из трёх

видов только моллюск продемонстрировал значительную изоляцию расстоянием и изоляцию окружающей средой, обусловленную температурой поверхности моря, в то время как солёность морской поверхности и диапазон температуры воздуха сильнее коррелировали с геномной изменчивостью краба и ежа. Показательно, что на графике распределения частот аллелей выпадающие локусы вносили вклад в два кластера (Восточного и Западного побережья) только у *S. granularis* и *P. angulosus*. Эти данные предполагают, что для прогнозирования эволюционных закономерностей лучшими предикторами являются специфические для вида требования к среде обитания и реакции на экологические стрессы, а не сильные экологические градиенты в регионе. Несомненное достоинство исследования заключается в том, что оно опирается на мультимодельный подход (хотя результаты разных подходов имели отличия) и использует разные виды, что значительно повышает достоверность полученных результатов (Nielsen et al., 2020b).

Для управления промыслом и охраняемыми видами ключевое значение имеет определение точных границ популяций. Желтопёрый тунец *Thunnus albacares* (Bonnaterre, 1788) – один из важнейших объектов коммерческого рыболовства. Однако данные о структуре его популяции в локальном и глобальном масштабах были противоречивыми. Геномные исследования обнаружили 939 однонуклеотидных полиморфизмов, позволивших выявить генетическую дифференциацию между популяциями Атлантического, Индийского и Тихого океанов, в то время как анализ 33 SNP, предположительно находящихся в процессе отбора, идентифицировал отдельные популяции в Тихом океане и впервые – в Атлантическом океане. Эти данные необходимы для оценки запасов, предотвращения незаконной торговли и устойчивого управления популяциями тунца (Pecoraro et al., 2018).

Полосатый марлин *Kajikia audax* (Philippi, 1887) – другая крупная пелагическая рыба, широко распространённая в тропических и умеренных водах Тихого и Индийского океанов, имеющая промысловое и рекреационное значение. Полногеномный анализ вида, основанный на более чем 4000 однонуклеотидных полиморфизмов, выявил шесть генетически различных популяций, обнаружил подмножество из 59 SNP (предположительно подвергшихся естественному отбору и способных дифференцировать популяции), продемонстрировал стабильность частот аллелей на протяжении 3–5 поколений и подчеркнул важность юго-западной части Тихого океана для поддержания межконтинентального потока генов. Кроме того, эти исследования указали на значительные несоответствия между генетически различными популяциями полосатого марлина и границами запасов, признанными для управления рыболовством (Mamoozadeh et al., 2019).

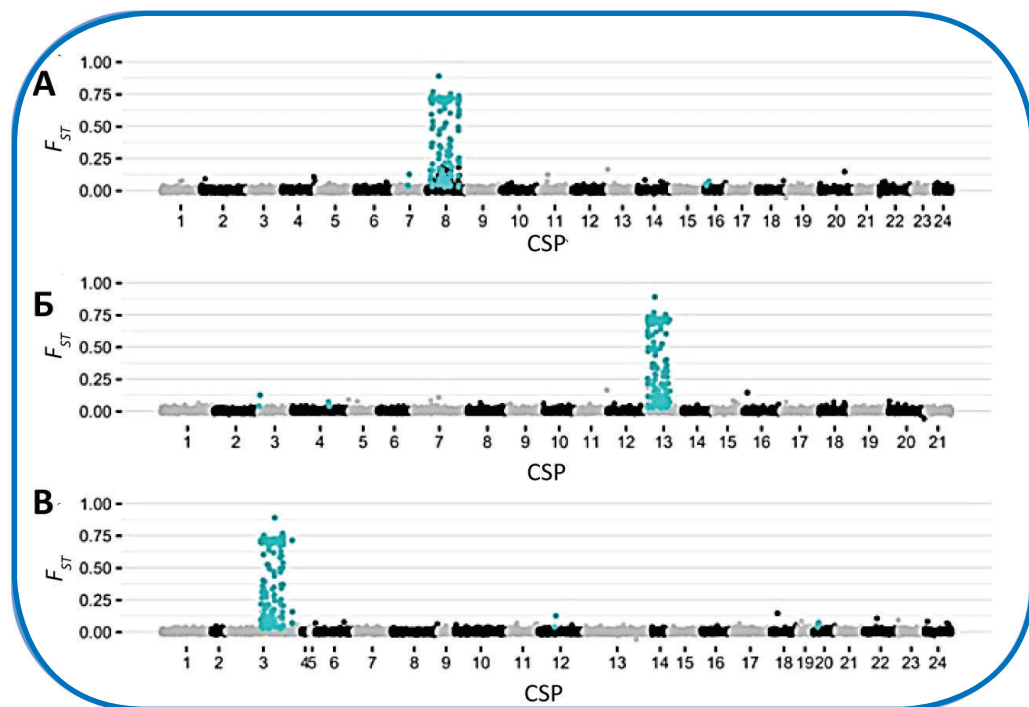
Геномный анализ генетической структуры популяции зубастого терпуга *Ophiodon elongatus* Girard, 1854, ценного промыслового вида, предоставил убедительные доказательства существования двух различных генетических кластеров, северного и южного (рис. 5.16). Генетический разрыв полностью обусловлен относительно небольшим количеством локусов, которые, вероятно, колокализуются на одной хромосоме.



**Рисунок 5.16.** Анализ главных компонент для маркеров RADseq зубастого терпуга *Ophiodon elongatus*, включая все локусы (А), исключая выпадающие локусы (Б), и включая только выпадающие локусы (В) с процентной вариацией для PC1 и PC2 (из: Longo et al., 2020).

Об этом свидетельствует сопоставление локусов RADseq (позволяет идентифицировать тысячи генетических маркеров, случайным образом распределённых по целевому геному) с черновой сборкой генома терпуга и геномов других видов костистых рыб на хромосомном уровне (рис. 5.17). Эти результаты дают новое представление о генетической структуре *O. elongatus* и определяют чёткие эволюционные единицы, которые можно использовать в рыболовстве (Longo et al., 2020).

Понимание того, как абиотические и биотические особенности окружающей среды влияют на перемещение паразитов, их хозяев и видов-переносчиков, имеет важное значение для борьбы с паразитарными болезнями. Пространственное распределение генетического разнообразия является результатом взаимодействий хозяин-переносчик-паразит. При этом, поток генов не только коррелирует с распространением эпидемиологически значимых признаков, таких как лекарственная устойчивость



**Рисунок 5.17.** Манхэттенские графики, сопоставляющие контиги RADseq со специфическими для локуса значениями  $F_{ST}$ , основанными на 42 участках отбора образцов для проекта сборки генома *Ophiodon elongatus*, встроенного в псевдомолекулы хромосомного масштаба (CSP), с использованием геномов (A) *Sebastes schlegelii*, корейский морской окунь; (Б) *Gasterosteus aculeatus*, трехиглая колюшка; и (В) *Larimichthys crocea*, большой жёлтый горбыль, в качестве ориентиров (из: Longo et al., 2020). Локусы, находящиеся под отбором, выделены бирюзовым цветом.

и вирулентность, но также регулирует местную адаптацию, появление новых фенотипов и их вторжение в незаражённые территории (Schwab et al., 2017).

Экзотические патогены кардинально изменили лиственные леса на востоке США, и многие виды оказались под угрозой исчезновения. Поэтому понимание последствий экзотических болезней для местных лесов имеет ключевое значение для эволюционной экологии и природоохранной биологии. Исследования популяционной генетики и ландшафтной геномики цветущего кизила *Cornus florida* L. выявили генетические признаки его адаптации к антракнозу – смертельно опасному заболеванию деревьев, вызываемому патогенным грибом *Discula destructiva* (Fr) Munk ex. H. Kern (1955). Некоторые из локусов (всего было идентифицировано 68 локусов, предположительно находящихся под положительным отбором) показали резкое изменение частот аллелей, хотя генетическое разнообразие между популяциями поражённых и экологически чистых территорий существенно не отличалось, а филогеографическая структура популяций была слабо выраженной (Pais et al., 2020). Эта информация

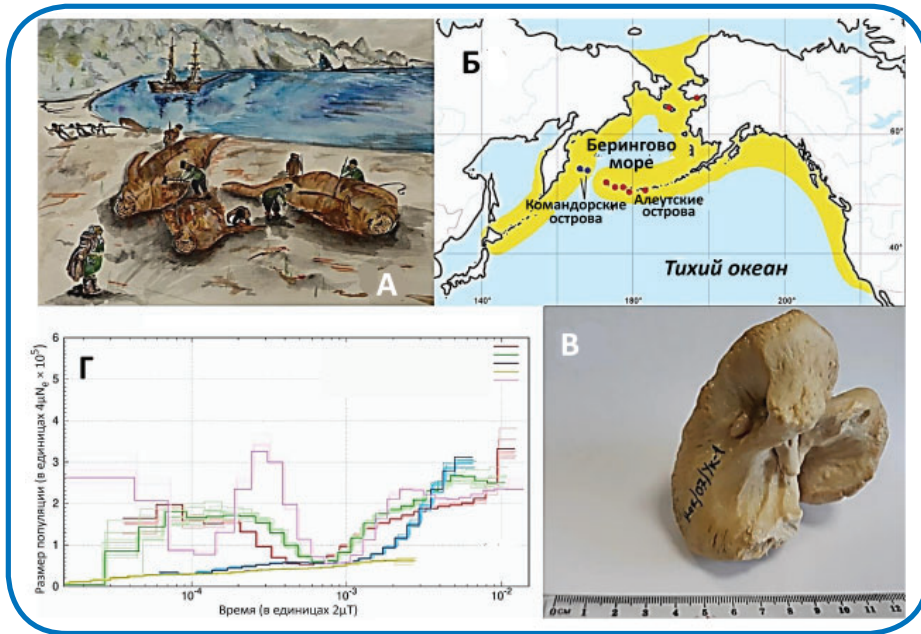
имеет большое значение для разработки планов сохранения видов и управления природными популяциями. В частности, она указывает на целесообразность внедрения деревьев с ценными генами-кандидатами из экологически чистых зон в поражённые участки, что также будет способствовать их сохранению *ex-situ* и размножению. Хотя снижение генетического разнообразия в поражённых антракозом районах ещё не заметно, демографические изменения с последующим отбором могут привести к сокращению генетического разнообразия, и в конечном счёте стимулировать эволюцию генов вирулентности у *D. destructiva* для преодоления имеющихся местных адаптаций в организме хозяина, поскольку у патогена более короткое время генерации (Pais et al., 2020).

Исследование паразитической бабочки голубянки алькон *Phengaris alcon* Denis & Schiffmüller, 1775 и двух её хозяев, муравья *Myrmica scabrinodis* Nylander, 1846 и горечавки болотной *Gentiana pneumonanthe* L., является одним из немногих примеров геномики ландшафтного сообщества. Самка бабочки откладывает яйца на бутоны горечавки, в которых они развиваются до гусениц, далее переходящих к муравьям, заставляя тех (при помощи химических сигналов) выращивать себя, как собственное потомство. Эксперименты показали, что несмотря на то, что некоторая генетическая изменчивость между популяциями видов-хозяев может быть объяснена абиотическими переменными, значительные изменения в генетической структуре бабочки голубянки могут быть связаны с генетической структурой хозяина, причём коэволюционный отбор, очевидно, очень важен для синхронизации генетической структуры этой системы паразит–хозяин (De Kort et al., 2014).

**5.3.1. Исчезнувшие виды.** Почти полные или частичные геномы были получены от вымерших видов млекопитающих, таких как шерстистый мамонт *Mammuthus primigenius* (Blumenbach, 1799), белый медведь *Ursus maritimus* Phipps, 1774, дикая лошадь *Equus ferus* Boddaert, 1785 и др. (Lan, Lindqvist, 2019) (таб. 5.2). Эти данные полезны для изучения генетических изменений, связанных с демографией и вымиранием, одомашниванием, межвидовой гибридизацией в прошлом, а также с сохранением видов. Шерстистый мамонт одним из первых привлёк внимание биологов в качестве объекта палеогеномики. Будучи одним из самых многочисленных видов мегафауны Северного полушария в плейстоцене и раннем голоцене, около 10000 л. н. вид вымер на материке, а примерно 3700 л. н. исчезли и островные популяции (Vartanyan et al., 2008). Сравнение высококачественных геномов шерстистого мамонта материковой популяции возрастом 45000 л. и последней выжившей островной популяции (о. Врангеля) обнаружило высокое сходство демографических траекторий, с древним бутылочным горлышком и резким сокращением эффективной численности островных популяций в начале голоцена. Кроме того, у островного образца обнаружили снижение полногеномной гетерозиготности и накопление вредных мутаций, что могло способствовать вымиранию последней популяции шерстистых мамонтов (Palkopoulou et al., 2015).

**Таблица 5.2.** Географическое распространение, таксономическая принадлежность и приблизительный возраст некоторых ключевых образцов животных, по которым были получены палеогеномные данные (по: Mitchell, Rawlence, 2021; van der Valk et al., 2021).

Вид	Географическое распространение	Возраст
Тасманийский тигр ( <i>Thylacinus cynocephalus</i> )	Австралия	~100 л.
Восточная равнинная горилла ( <i>Gorilla beringei graueri</i> )	Африка	~100 л.
Малый кустарниковый моа ( <i>Anomalopteryx didiformis</i> )	Новая Зеландия	>700 л.
Гигантский лемуру-коала ( <i>Megalapteryx edwardsi</i> )	Мадагаскар	~1500 л.
Шерстистый мамонт ( <i>Mammuthus primigenius</i> )	Беренгия	~4300 л.
Гигантская панда ( <i>Ailuropoda melanoleuca</i> )	Китай	~5000 л.
Гаур ( <i>Bos gaurus</i> )	Азия	~5200 л.
Пещерная гиена ( <i>Crocota crocuta spelaea</i> )	Евразия	Плейстоцен
Лютый волк ( <i>Canis dirus</i> )	Северная Америка	~11000 л.
Харингтонгипп ( <i>Haringtonhippus francisci</i> )	Северная Америка	~13000 л.
Шерстистый носорог ( <i>Coelodonta antiq-uitatis</i> )	Евразия	~18000 л.
Пещерный лев ( <i>Panthera spelaea</i> )	Евразия	~30000 л.
Серый волк ( <i>Canis lupus</i> )	Евразия	~35000 л.
Саблезубый кот ( <i>Homotherium latidens</i> )	Северная Америка	~47000 л.
Западный верблюд ( <i>Camelops hesternus</i> )	Северная Америка	>50000 л.
Полярный/белый медведь ( <i>Ursus mari-timus</i> )	Евразия	>110000 л.
Слон с прямыми бивнями/ Палеолоксодон ( <i>Palaeoloxodon antiquus</i> )	Евразия	>120000 л.
Пещерный медведь ( <i>Ursus spelaeus</i> )	Евразия	>270000 л.
Древняя лошадь ( <i>Equus sp.</i> )	Северная Америка	>560000 л.
Самый старый мамонт ( <i>Mammuthus</i> )	Восточная Сибирь	>1000000 л.



**Рисунок 5.18.** Секвенирование генома вымершей коровы Стеллера (из: Sharko et al., 2021).

А – Моряки Великой северной экспедиции Витуса Беринга (1741 г.) разделяют туши убитых стеллеровых коров *Hydrodamalis gigas*.

Б – Географическая карта региона Берингова моря, показывающая ареал стеллеровой коровы в позднем плейстоцене (жёлтый цвет), археологические памятники и места (красные кружки), где были описаны образцы стеллеровой коровы, а также распространение последней популяции *H. gigas* (синие кружки).

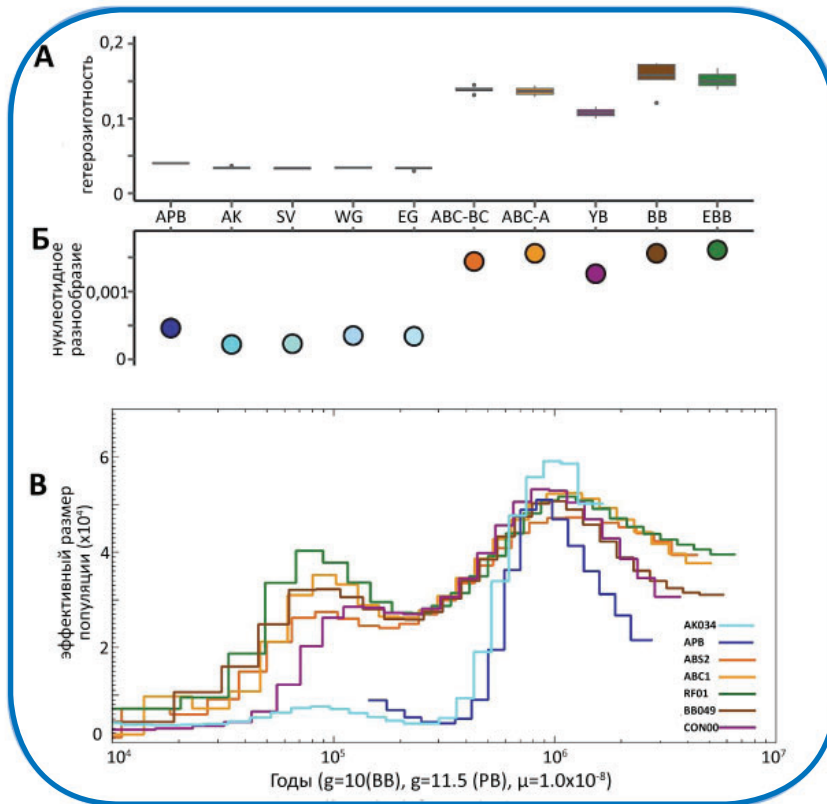
В – Каменистая кость *H. gigas*, которая использовалась для извлечения исторической ДНК.

Г – Истории размеров популяций стеллеровой морской коровы, дюгоной, шерстистых мамонтов и ленских лошадей.

Морская корова Стеллера (или капуста, Стеллерова корова) *Hydrodamalis gigas* (Zimmermann, 1780) – другой символ плейстоценовой мегафауны, исчезновение которого считается первым историческим вымиранием морского млекопитающего в результате деятельности человека (Le Duc et al., 2022). Геномные исследования *H. gigas* показали, что гетерозиготность последней популяции этого животного была низкой и сравнимой с последней инбредной популяцией шерстистого мамонта, населявшего о. Врангеля 4000 л.н., и примерно в два раза ниже, чем у ближайшего вида сирен – дюгонь *Dugong dugon* (Müller, 1776) (Sharko et al., 2021). Однако процесс вымирания *H. gigas* начался задолго до исчезновения и, следовательно, изменения окружающей среды также способствовали вымиранию (рис. 5.18) (Sharko et al., 2021; Le Duc et al., 2022). Демографический анализ показал, что *H. gigas* и *D. dugon* имели только одно бутылочное горлышко в своей эволюционной истории, в отличие от шерстистого мамонта, имевшего как минимум два таких события за последние сотни тысяч лет (Sharko et al., 2021). Популяции стеллеровой морской коровы сокращались по крайней мере в течение последних 500000 л. с возможной временной стабилизацией в периоды тёплого межледниковья (Le Duc et al., 2022).

Вымершая стеллерова корова была морским млекопитающим размером с кита, имевшим высокую морфологическую адаптацию к суровому прибрежному климату северной части Тихого океана и гемоглобин с фиксированной аминокислотной заменой, радикально влияющей на его функции (Signore et al., 2023). В процессе эволюции некоторые локусы, связанные с метаболическими, иммунными и гормональными системами, прошли положительный отбор (Sharko et al., 2021). Из 238 генов, которые у *H. gigas* эволюционируют значительно быстрее или существенно медленнее, 20 генов модулируют массу тела и энергетический обмен (Le Duc et al., 2022). Сравнительный геномный анализ выявил конвергентную эволюцию между стеллеровой коровой и китообразными, предполагая роль нескольких генов в адаптации к холодной водной среде, среди которых – инактивация генов липоксигеназы. У людей аллельные варианты потери функции в этих генах являются причиной аутосомно-рецессивного врождённого ихтиоза, характеризующегося толстым гиперкератотическим эпидермисом, повторяющим кожу стеллеровой коровы (похожей, согласно описанию Георга Стеллера, на кору дуба) (Le Duc et al., 2022).

Изучение генома белого медведя *U. maritimus*, основная среда обитания которого определяется протяжённостью арктического морского льда, представляет исключительный интерес в плане получения сведений об адаптации к экстремальным условиям, влиянии изменения климата и колебания ледникового покрова на эволюцию видов и биоразнообразия, а также прояснения эволюционной истории медведей. Белые и бурые медведи являются родственными видами с различными физиологическими и поведенческими адаптациями, которые развились за последние 500000 л. Однако геномный анализ показал, что некоторые существующие и вымершие популяции бурых медведей имеют относительно недавнее происхождение от белых медведей. Результаты ресеквенирования древнего генома с о. Шпицберген возрастом около 120000 л. и сравнение его с геномами современных видов, предполагают, что в прошлом белый медведь имел меньший поток генов от бурого медведя (Lan, Lindqvist, 2019). Анализ палеогенома белого медведя возрастом около 100000 л. обнаружил массивный поток генов от белого медведя к бурому, что не было зарегистрировано при изучении современных представителей вида. Исторический однонаправленный поток генов от белых медведей к бурым медведям, вероятно, произошёл из-за изменения климата, вызвавшего перекрытие ареалов двух видов (Wang et al., 2022a). Недавно проведённый геномный анализ выявил (рис. 5.19) древнее и резкое сокращение эффективной численности популяции белого медведя за последние 300000 л., что согласуется с его более низким генетическим разнообразием по сравнению с бурым медведем. Кроме того, получены свидетельства по крайней мере одного древнего, произошедшего, вероятно, более 150000 л. назад, события интрогрессии аллелей бурых медведей в геном предков белых медведей (Lan et al., 2022). Таким образом, исследования подчеркивают сложные пути ретикулярности, по которым может пойти эволюция в условиях резко изменяющегося климата.



**Рисунок 5.19.** Геномный анализ бурого и белого медведей (из: Lan et al., 2022).

А – Частоты аутомсомной гетерозиготности, Б – разнообразие нуклеотидов в популяциях полярных и бурых медведей, В – Оценки эффективного размера популяции с течением времени, показанные для одной репрезентативной особи каждой из популяций бурого медведя и белого медведя: АК034 – аляскинский белый медведь, АРВ – древний полярный медведь (Шпицберген), АВС2 – Чичаговский бурый медведь и АВС1 – Адмиралтейский бурый медведь (Эй-Би-Си/Малые Антильские острова), RF01 – Европейский бурый медведь, ВВ049 – Аляскинский бурый медведь, CON001 – Йеллоустонский бурый медведь, ВВ – бурый медведь, РВ – белый медведь;  $g$  – время генерации,  $\mu$  – частота мутаций.

Древние геномы могут предоставить информацию о приоритетах сохранения современных видов. Например, палеогеномные исследования показали, что прямой предок лошади Пржевальского (которая до сих пор считалась единственной настоящей дикой лошастью) когда-то был приручен. Это открытие подняло вопросы о природоохранном статусе вида (так как гибриды не являются приоритетными единицами сохранения) (Frantz et al., 2020). Голубая антилопа *Hippotragus leucophaeus* (Pallas, 1766) является единственным крупным африканским млекопитающим, которое вымерло в историческое время. Филогенетический анализ двух черновых геномов из исторического (приблизительно XIX в.) и ископаемого (возрастом около 9800–9300 л., самый старый палеогеном Африки) образцов подтвердил, что голу-

бая и чёрная антилопы являются сестринскими видами, и продемонстрировал поток древних генов от чалой антилопы к голубой. Геномное разнообразие голубой антилопы оказалось намного ниже, чем у чалых и чёрных антилоп, свидетельствуя о низком размере популяции вида. Однако его сохранение на протяжении всего голоцена предполагает, что основным фактором вымирания голубой антилопы было воздействие человека в колониальную эпоху (Hempel et al., 2022; Sheng et al., 2019). Ещё один пример связан с гигантской пандой *Ailuropoda melanoleuca* David, 1869. Исторически вид был широко распространён от северного Китая до юго-западной Азии, в то время как современный ареал ограничен фрагментированными горными массивами на восточной окраине Цинхай-Тибетского плато. Палеогеномный анализ выявил генетический вклад исчезнувшей гигантской панды в сохранившиеся популяции. Оказалось, что ядерный геном этого вида из провинции Юньнань возрастом около 5000 л. представляет собой генетически обособленную популяцию, которая отделилась до диверсификации современных популяций. Несмотря на то, что существующие популяции *A. melanoleuca* имеют достаточно высокий уровень генетического разнообразия, это лишь часть того, чем данный вид обладал до недавнего сокращения своего ареала (Sheng et al., 2019).

Иногда исследования древних геномов указывают на возможные проблемы в таксономии. Например, палеогеномы гиен продемонстрировали сложную эволюционную историю межконтинентального генного потока между африканскими пятнистыми и евразийскими пещерными гиенами (род *Crocuta* Kaup, 1828), позволяющую предположить, что предки рода гиен покинули Африку примерно в то же время, что и ранние представители рода *Homo* Linnaeus, 1758. Однако обнаруженное несоответствие между ядерной и митохондриальной филогениями и доказательства двустороннего потока генов могут усложнить имевшуюся таксономическую классификацию (Westbury et al., 2020). Палеогеномика прояснила эволюционную историю вымершего в голоцене рогатого крокодила. Вместе с тем, филогенетический анализ митохондриальных геномов 1300–1400-летних образцов свидетельствует в пользу родственных связей между родами *Voay* Brochu, 2007 и *Crocodylus* Laurenti, 1768 (рогатыми и настоящими крокодилами), что противоречит результатам исследования по систематике и морфологическим данным, помещающих *Voay* в подсемейство *Osteolaeminae* (карликовые крокодилы и их родственники) (Hekkala et al., 2021).

Птицы являются одним из самых богатых видами классом позвоночных, однако данные использования древней ДНК в изучении их эволюции практически отсутствуют и ограничиваются реконструкцией филогенетических отношений для уточнения таксономии и применением для сохранения и управления популяциями. При этом большая часть материала датируется поздним голоценом. В то же время, изучение древних геномов птиц чрезвычайно актуально для понимания механизмов, управляющих эволюцией и вымиранием, поскольку за последние несколько сотен лет у них произошло более четверти случаев исчезновения видов (Grealy et al. 2017).

Яркий пример – странствующий голубь *Ectopisces migratorius* (Linnaeus, 1766), который некогда был самой многочисленной птицей в мире. В 1800-х гг. числен-

ность его популяции оценивалась в 3–5 миллиарда особей, что должно было обеспечить защиту от вымирания, однако в начале XX века вид внезапно исчез. С помощью технологии высокопроизводительного секвенирования было восстановлено от 57% до 75% генома каждого странствующего голубя из четырех музейных экспонатов XIX века. Согласно геномным данным, эффективный размер популяции *E. migratorius* за последний миллион лет составлял около 1/10000 от предполагаемого числа особей 1800 г. и был не больше, чем у других широко распространённых видов птиц. Такой результат предполагает, что странствующий голубь не всегда был очень многочисленным и испытывал частые и резкие колебания численности в результате климатических, пищевых и других экологических изменений (что поддерживается анализом экологических ниш), увеличившими риски его вымирания. Вероятно, в конце 1800-х гг. тенденция к сокращению численности популяции произошла одновременно с воздействием человека, и сочетание этих двух факторов привело к стремительному исчезновению странствующего голубя. Это величайшее в истории вымирание, вызванное деятельностью человека, показывает, насколько уязвимыми могут быть даже многочисленные виды (Hung et al., 2014).

Крупные травоядные на протяжении миллионов лет влияли на наземные экосистемы, регулируя популяции мелких животных и разнообразие растений. Плиоценовая и плейстоценовая мегафауны крупных травоядных вызвала совместную эволюцию различных специализированных хищников, падальщиков и т. д., многие из которых, включая стервятника Нового Света *Coragyps occidentalis* (L. Miller, 1909), вымерли вместе с мегатравоядными в конце плейстоцена. *C. occidentalis* долгое время рассматривался предком или сестринским видом чёрного стервятника *C. atratus* (Linnaeus, 1758). Интересно, что согласно летописи окаменелостей *C. occidentalis* в основном обитал на больших высотах, в то время как *C. atratus* населял низменности. Секвенирование 14000 л. генома *C. occidentalis*, обнаруженного в перуанских Андах, показало, что этот вид произошёл от популяции *atratus* после того, как популяция современного чёрного стервятника колонизировала Высокие Анды 300000–400000 л. н. Стервятник Нового Света был значительно крупнее и имел несколько иные пропорции тела по сравнению с чёрным стервятником. Таким образом, вымерший вид *C. occidentalis* может быть примером прерывистой эволюции в результате сильного естественного отбора в экстремальных высокогорных условиях (Ericson et al., 2022).

#### 5.4. Сельское хозяйство и биотехнология

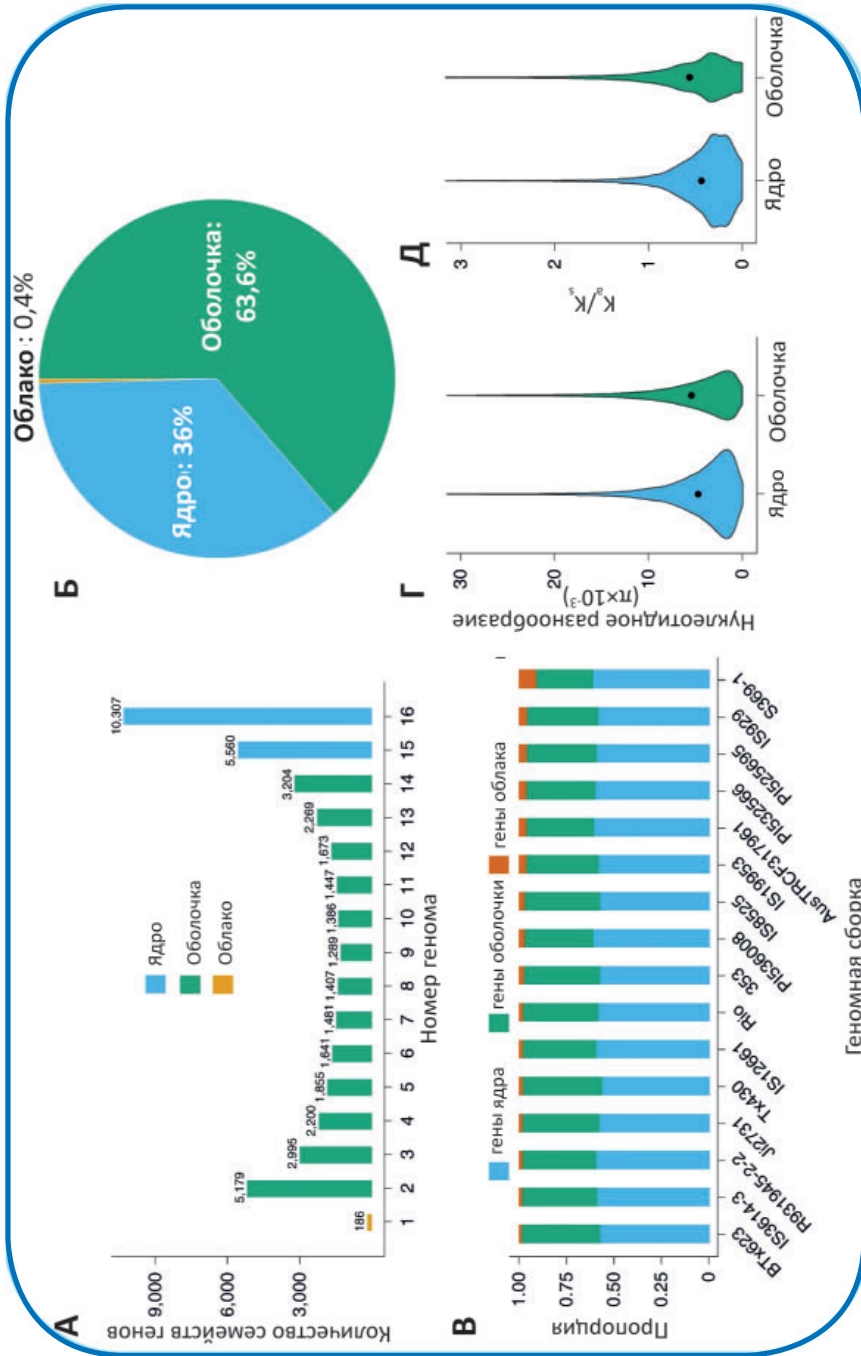
Геномика является надёжным инструментом для идентификации генов и областей генов, потенциально отобранных при одомашнивании, и улучшении пород животных и растений. Например, полногеномный анализ восковой тыквы *Benincasa hispida* (Thunb.) Cogn., приносящей гигантские плоды, выявил гены, которые, вероятно, ответственны за большой размер плода. Эта информация необходима для дальнейших исследований по селекции восковой тыквы (Xie et al., 2019). Пшеница является одной из основных зерновых культур и важным пищевым ресурсом. Се-

лекция злака осуществлялась в направлении повышения урожайности (например, по размеру зерна) и устойчивости к болезням и вредителям. Цвет чешуи – тоже важная характеристика сортов пшеницы, однако, быстрая идентификация генов-кандидатов с использованием стандартных подходов часто затруднена. Популяционная геномика и анализ гаплотипов у спельты и мягкой пшеницы выявили ген, регулирующий окраску чешуи (красная чешуя 1, *Rg-B1*), что чрезвычайно важно для работы селекционеров (Abrouk et al., 2021). Интересно, что у одомашненных животных по сравнению с дикими наблюдается повышенная плотность ряда мобильных генетических элементов и возникновение на их основе новых генов, например, гена синтицина (индуцирует формирование синцития), гена-регулятора синаптической пластичности *arc* и семейства генов *bex* (важны для развития и функционирования центральной нервной системы у плацентарных) (Глазко и др., 2022).

Сорго является основной засухоустойчивой культурой для полумиллиарда человек в Африке и Азии, важным источником корма для животных во всём мире и сырьём для производства биотоплива. Пангеном, созданный по данным 16 полногеномных последовательностей культивируемых и дикорастущих растений сорго, продемонстрировал существенную изменчивость содержания генов, при которой 64% генных семейств отличались по наличию/отсутствию в разных геномах (рис. 5.20). Были определены варианты, подвергнутые селекции на качество во время одомашнивания сорго, и показано, что они имеют важные фенотипические проявления, которые могут способствовать увеличению урожая. В целом, сконструированный пангеном сорго представляет собой современную платформу для улучшения свойств этого растения и поиска новых полезных генов (Tao et al., 2021).

Как ценная сельскохозяйственная культура, земляника (*Fragaria* spp.) в последние годы стала модельной системой для различных фундаментальных и прикладных исследований. Анализ данных для эталонных геномов пяти разных видов *Fragaria* L., а также результатов повторного секвенирования генома для 128 дополнительных образцов, идентифицировал множественные независимые одонуклеотидные мутации гена *MYB10* (играет главную роль в регуляции метаболизма флавоноидов/фенилпропаноидов), связанные с белыми плодами и общими для разных видов. Кроме того, были получены полностью разрешённые и надёжно подтверждённые филогении и время дивергенции для большинства диплоидных видов земляники, а также выявлен новый диплоидный вид, построен пангеном *Fragaria* и изучена эволюционная динамика генных семейств (рис. 5.21). Эти эталонные геномы и наборы данных представляют собой мощный ресурс для дальнейших исследований земляники, в том числе в сельскохозяйственных целях (Qiao et al., 2021).

Изучение индивидуальных геномов представителей разных популяций и пород тибетских кур предоставило данные генетических основ искусственного отбора при сельскохозяйственном производстве, а также естественного отбора для успешного проживания в условиях высокогорья. При этом несколько генов были идентифицированы как кандидаты на хозяйственные признаки кур и другие фенотипические



**Рисунок 5.20.** Пангеном сорго (из: Тао et al., 2021).

А – количество представленных семейств генов на геном; Б – содержание генов ядра, оболочки и облака (встречающихся во всех, в двух и более, или единичных геномах соответственно); В – содержание генов ядра, оболочки и облака на геном; Г – скрипичные графики, показывающие разнообразие нуклеотидов генов ядра и генов оболочки, Д – значения Ka/Ks генов ядра и генов оболочки; чёрные точки указывают средние значения; К1 – гены ядра, К2 – гены оболочки, К3 – гены облака.

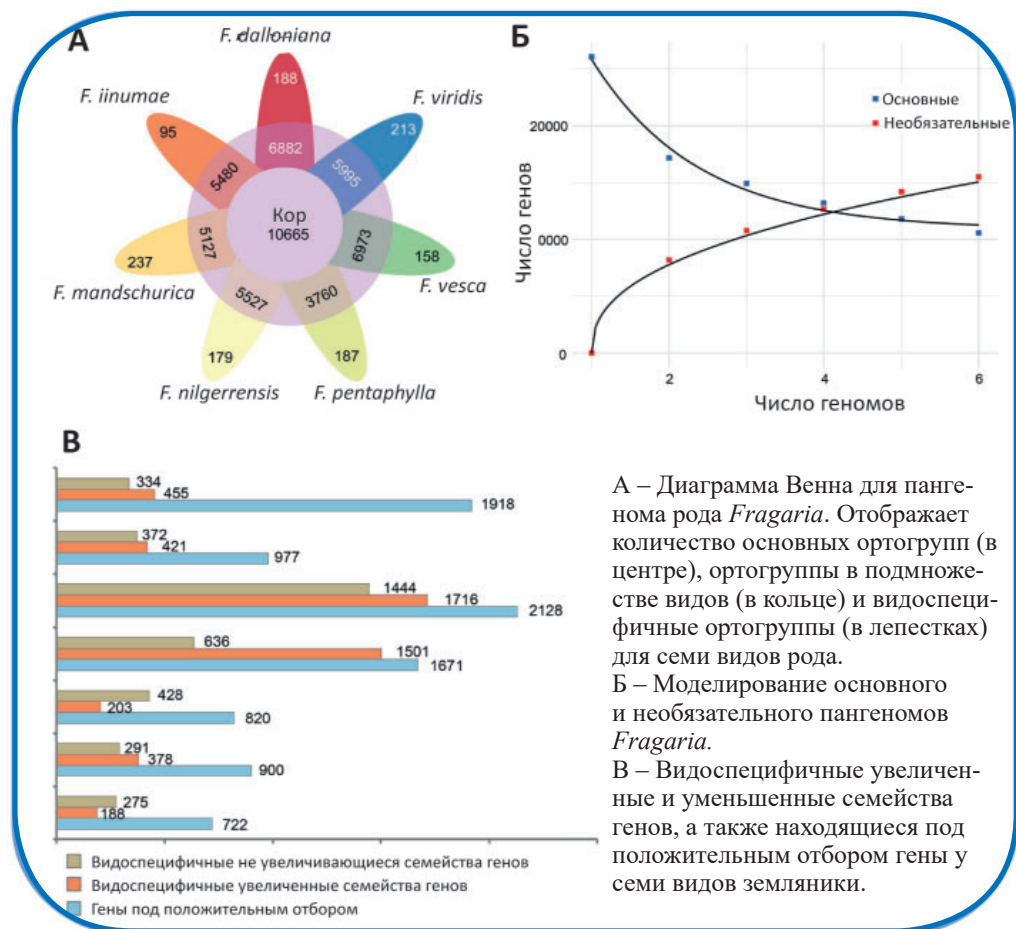
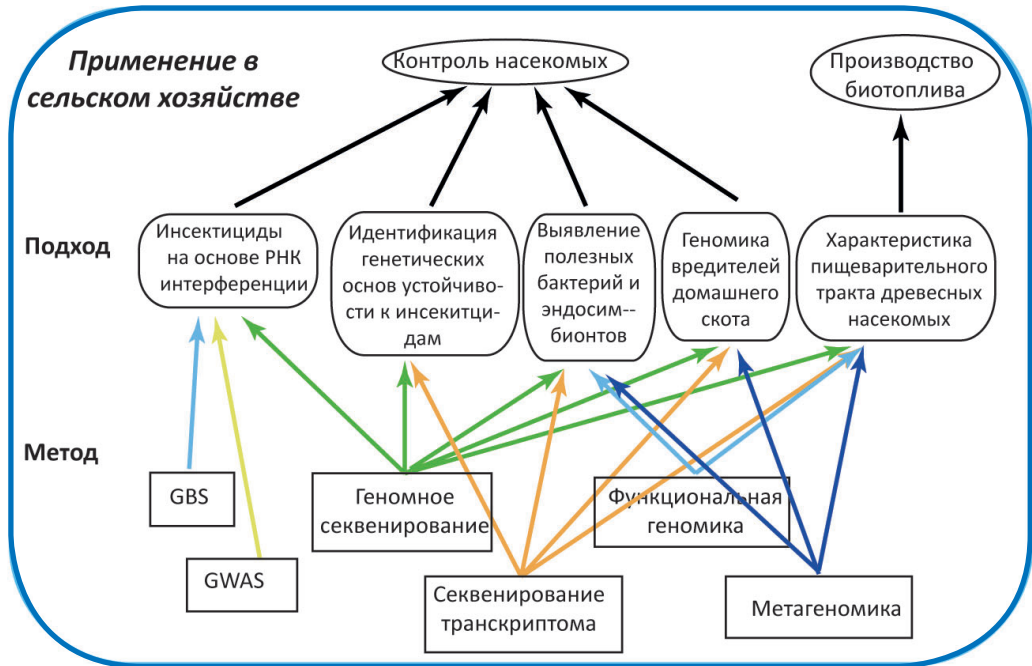


Рисунок 5.21. Эволюция семи диплоидных геномов *Fragaria* (из: Qiao et al., 2021).

признаки (Li et al., 2019a). В другой работе популяционный анализ геномов липицианской лошади выявил гены (*SPRY4*, *NDFIP1*, *IMPDH2* и *HSP90AB1*), участвующие в развитии меланомы лошадей; эти данные представляют большой практический интерес для ветеринарии (Grilz-Seger et al., 2018). В сельскохозяйственной энтомологии геномика используется для разработки новых мер борьбы с вредителями сельскохозяйственных видов, выявления антигенов-кандидатов для вакцинации скота и управления устойчивостью к пестицидам. Кроме того, метагеномные методы могут характеризовать микробные сообщества термитов и жуков-древоточцев и с помощью транскриптомных подходов выявить молекулярные основы их способности переваривать древесину – потенциальный механизм производства биотоплива (рис. 5.22) (Poelchau et al., 2016).

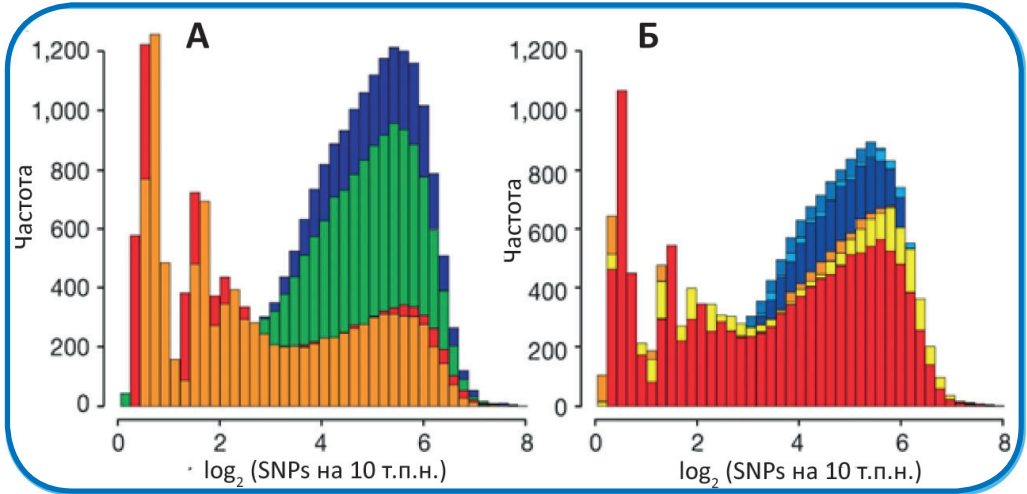


**Рисунок 5.22.** Применения экологической геномики насекомых в сельском хозяйстве (по: Poelchau et al., 2016).

Перечислены различные методы и подходы, используемые для решения проблем борьбы с насекомыми и производства биотоплива, а также их взаимосвязи друг с другом. GBS – генотипирование с помощью секвенирования.

Исследование микробиома является одним из основных подходов к разработке эффективных мер профилактики метаболических нарушений у животных на молочных фермах. Анализ биоразнообразия и метаболических функций микробиома рубца у молочных коров показал, что физиологические периоды являются важным фактором, определяющим относительное изобилие неатрибутируемых бактерий из кандидатных семейств *vadinBE97* и *WCHV1-41*, функции которых практически не изучены. По сравнению с сухостойными, новотельными и раздойными животными, наиболее выраженные изменения метаболического потенциала микробиоты, а именно – угнетение различных типов обмена веществ (энергетического, белкового, углеводного, липидного, включая синтез жирных кислот) в химусе рубца, наблюдаются у коров в период стабилизации и спада лактации (Лаптев и др., 2021).

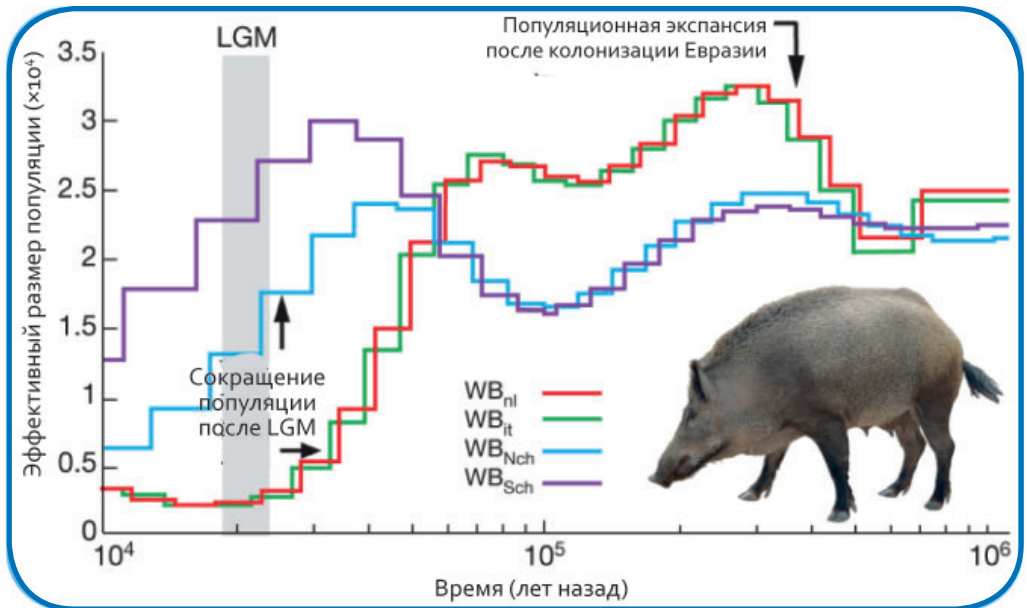
Дикие свиньи появились в Юго-Восточной Азии и впоследствии распространились по всему евразийскому континенту. На протяжении 10000 л., от одомашнивания до современных методов разведения, люди сформировали геномы домашних свиней с различными уровнями гетерозиготности, разными демографическими историями (рис. 5.23, 5.24) и т. д., и которые, по-видимому, даже в период одомашнивания продолжали достаточно свободно скрещиваться с дикими сородичами. Поэтому гене-



**Рисунок 5.23.** Распределение гетерозиготности дикой *Sus scrofa* и домашней *S. s. domestica* свиньи (из: Groenen et al., 2012).

А – дикая свинья: синий – южнокитайская, зелёный – северокитайская, оранжевый – итальянская, красный – голландская.

Б – породы: голубой – китайские породы, красный и жёлтый – европейские породы.



**Рисунок 5.24.** Демографическая история дикого кабана (из: Groenen et al., 2012).

Последний ледниковый максимум (LGM) выделен серым цветом; WBnl – кабан, Нидерланды; WBIt – дикий кабан, Италия; WB\_Nch – дикий кабан, северный Китай; WB\_Sch – дикий кабан, южный Китай.

тическое разнообразие современных и древних популяций свиней наиболее полно соответствует модели, включающей постоянный поток генов от диких кабанов к домашним породам. Сравнения геномов диких и домашних популяций показали глубокий филогенетический раскол между европейскими и азиатскими дикими кабанями примерно 1 млн л.н. и быструю эволюцию генов, связанных с иммунным ответом и обонянием. Свиньи имеют самый большой арсенал функциональных генов обонятельных рецепторов, что отражает важность обоняния для этого всеядного животного (Groenen et al., 2012). Сравнение полных геномов европейских диких кабанов и домашних свиней выявило три подвергавшихся отбору гена (*NR6A1*, *PLAG1* и *LCORL*), которые связаны с увеличением длины тела и числа позвонков (наиболее характерные признаки домашней свиньи) (Rubin et al., 2012). Интересно, что согласно данным по древним геномам, аллельный вариант гена *MC1R* (кодирует рецептор меланокортина 1), детерминирующий чёрный цвет шерсти и соответственно потерю маскировочного окраса, подвергся положительному отбору во время одомашнивания свиней очень рано, поскольку его частота составляла ~50% уже около 8000 л.н., тогда как в дикой природе он почти отсутствует (~2%) (Frantz et al., 2020). Исследование полных геномов свиней беркширской породы обнаружило ряд находящихся под направленным отбором генов, связанных с регуляцией липидного метаболизма, распределением внутримышечного жира и типом мышечных волокон (Jeong et al., 2015), а сравнение геномов ландрасов и свиней йоркширской породы показал, что у них отбору подвергались различные гены и биохимические пути (Wang et al., 2014b). Последовательность генома свиньи обеспечивает важный ресурс для дальнейшего улучшения этого значимого для человека вида домашнего скота, а идентификация многих предполагаемых болезнетворных вариантов расширяет потенциал свиньи как биомедицинской модели (Groenen et al., 2012).

**Редактирование геномов** – мощный инновационный инструмент получения растений и животных с заданными признаками, позволяющий за относительно короткий срок повысить качество и эффективность производства сельскохозяйственной продукции. Введение гена мыши *UCP1* (адипонектина – гормона, который синтезируется и секретируется белой жировой тканью) в эндогенный локус *UCP1* свиньи (из семейства *Artiodactyl Suidae*) привело к улучшению способности поддерживать температуру тела в период холодового воздействия. Адаптация свиньи к холоду снижает неонатальную смертность поросят и уменьшает экономические затраты на обогрев помещений для содержания животных (Zheng et al., 2017). Селекция на мраморность мяса крупного рогатого скота привела к накоплению рецессивных мутаций, которые вызывают генетические заболевания. У гомозиготных телят, мутантных по гену изолейцил-тРНК-синтетазы (*IARS*), наблюдается задержка внутриутробного развития, в результате чего они или погибают внутриутробно, или рождаются нежизнеспособными. С помощью системы CRISPR-Cas9 мутантный ген фермента *IARS* был успешно восстановлен у эмбрионов телят, которые затем развивались в здоровое потомство (Ikeda et al., 2017). В результате редактирования генома животных с помощью системы Cas9-gRNA (нуклеаза Cas9 + направляющая РНК) было продемонстрировано

увеличение подшёрстка кашемировых коз (являющееся востребованным качеством) за счёт частичной потери функции гена длины шерсти *FGF5* (Wang et al., 2016c).

Развитие технологий редактирования генома растений открывает широкие возможности для селекции. Эффективный, точный и целенаправленный мутагенез посредством редактирования генома заложил основу для многих стратегий селекции следующего поколения, которые произведут революцию в сельском хозяйстве будущего (рис. 5.25). Например, одним из перспективных направлений является одомашнивание *de novo*. Все основные сельскохозяйственные культуры произошли от диких предков, которые были одомашнены на протяжении многих тысячелетий. Однако несмотря на то, что одомашнивание значительно повышает урожайность сельскохозяйственных культур, а также обеспечивает идеальную архитектуру растения и лёгкий сбор урожая, со временем оно приводит к генетическому горлышку бутылки, и как следствие – к уменьшению генетического разнообразия и потере стрессоустойчивости. Из-за полигенности многих полезных признаков (таких, как устойчивость к абиогенным стрессам) прямое и обратное скрещивание с дикими сородичами малоэффективно. Альтернативной стратегией, причём значительно ускоряющей процесс, является одомашнивание диких видов *de novo* путём редактирования генома. Так, с использованием CRISPR–Cas9 технологии созданы мутантные линии риса с улучшенными сельскохозяйственными признаками, а также было успешно выполнено мультиплексное редактирование генов (требуется введения нескольких РНК-гидов, каждый из которых направляет белок Cas9 к своей мишени) для частичного одомашнивания дикого томата при сохранении его стрессоустойчивости. Следует отметить, что использование одомашнивания *de novo* для преобразования не представляющих хозяйственную ценность культур (сорго, просо, лебеда) в новые суперкультуры может быть эффективным способом обеспечения глобального снабжения продовольствием (Gao, 2021). Известно также, что количественные признаки имеют большое агротехническое значение. Они полигенны и контролируются локусами количественных признаков (QTL), которые чрезвычайно трудно изучать и манипулировать ими, так как их наследование не подчиняется менделевским законам. Технологию редактирования QTL с помощью CRISPR–Cas9 системы или высокопроизводительного экрана редактирования (улучшенный высокопроизводительный конвейер редактирования генома) QTL-кандидатов можно использовать для введения нескольких желаемых количественных аллелей непосредственно в элитные сорта сельскохозяйственных культур (Gao, 2021).

Применение РНК-интерференции также является эффективной технологией для улучшения агрономически важных растений с целью повышения их урожайности и устойчивости к стрессовым факторам окружающей среды. Как и CRISPR–Cas9, эта технология может оказаться важной с точки зрения глобальной продовольственной безопасности и развития устойчивого сельского хозяйства. У растений биотический стресс вызывается живыми организмами – вирусами, бактериями, грибами, насекомыми, паукообразными, нематодами и сорняками, что приводит примерно к 40% по-

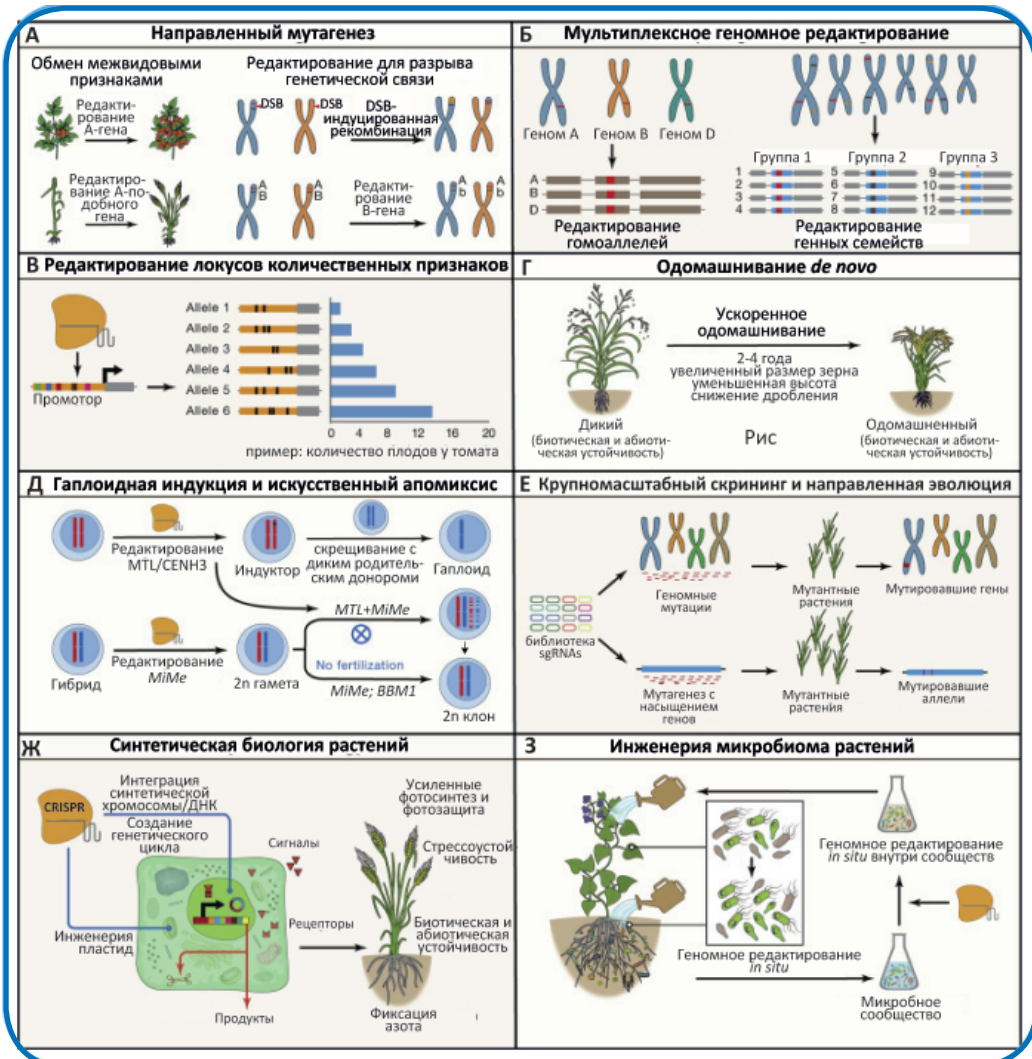


Рисунок 5.25. Стратегии улучшения урожая на основе редактирования генома (по: Gao, 2021).

А – Обмен межвидовыми признаками и разрыв генетической связи с помощью мутагенеза посредством редактирования генома; Б – Мультиплексное геномное редактирование гомеоаллелей и семейств генов; В – Редактирование локусов количественных признаков для получения новых аллелей и признаков; Г – Схематическая диаграмма ускоренного одомашнивания дикого риса посредством редактирования генома; Д – Гаплоидная индукция и искусственный апомиксис через геномную модификацию эндогенных генов; Е – Крупномасштабный скрининг и направленная эволюция для обнаружения признаков с помощью CRISPR; Ж – CRISPR-опосредованная синтетическая биология растений, в которой поведение растительных клеток изменяется для улучшения роста растений и производства продуктов; З – Изменение микробиома растений для улучшения роста сельскохозяйственных культур и их устойчивости к патогенам.

теперь общего урожая шести основных продовольственных и товарных культур (Rajput et al., 2021). РНК-интерференция успешно используется для придания сельскохозяйственным растениям устойчивости к биотическим, а также к абиотическим стрессам (таким как солёность, изменение температуры, влажность и т. д.) (Rajput et al., 2021). Кроме этого, РНК-интерференция применяется для получения плодов с определёнными качествами, например, без косточек и с увеличенным сроком годности, для развития растений с мужской стерильностью, для изменения цвета лепестков, улучшения питания растений и для производства вторичных метаболитов. Например, линии чая «Камелия» (из листьев *Camellia sinensis* (L.) Kuntze) с низким содержанием кофеина были разработаны путём подавления гена синтазы кофеина с использованием РНК-интерференции; с помощью опосредованного РНК-интерференцией подавления гена инозитол-пентакисфат 2-киназы были получены зёрна пшеницы *Triticum* L. 1753 с высоким содержанием Zn и Fe и со сниженным уровнем непитательной фитиновой кислоты; сверхэкспрессия гена *F3'5'H* (*PCFH*) и подавление гена *CmF3'H* у хризантемы приводили к повышенному содержанию цианидина и более ярко-красной окраске цветка (Rajput et al., 2021).

Средиземноморская плодовая муха *Ceratitis capitata* (Wiedemann, 1824) является **насекомым-вредителем**, обладающим уникальной способностью вторгаться и адаптироваться к экологическим нишам в тропических и субтропических регионах мира. Геномный анализ представил важную информацию о биологии этого одного из самых опасных и широко распространённых сельскохозяйственных вредителей, что должно значительно улучшить средства контроля численности и инвазивного потенциала популяций *C. capitata*. Кроме того, эти данные будут способствовать дальнейшим сравнительным функциональным и структурным исследованиям геномов насекомых, обеспечивая более глубокое понимание эволюции отдельных семейств генов (Papanicolaou et al., 2016).

**Загрязнение окружающей среды** – глобальная проблема, требующая скорейшего решения и разработки новых современных биотехнологий. Особое значение приобретает эффективная очистка сточных вод. Биологическая очистка сточных вод представляет собой процесс, при котором микробный метаболизм сложных сообществ превращает загрязняющие вещества в малотоксичные или нетоксичные продукты. Чтобы сделать его эффективным, требуется глубокое понимание разнообразия и сложности микробных сообществ. Метагеномика может идентифицировать состояние микробов в их естественной среде более эффективно, чем традиционные молекулярные и микробиологические методы. Метагеномный анализ применяется для оценки микробных сообществ при биологической очистке сточных вод, как, например, биологическое удаление фосфора и азота бактериями, снижение содержания тяжёлых металлов микробными сообществами с акцентом на вклад микробного и метаболического разнообразия, для изучения генов устойчивости к антибиотикам. В дальнейшем для разработки наиболее эффективных технологий биологической очистки сточных вод потребуются такие области метагеномных исследований, как мультиомика (Chen et al., 2022a).

В связи с выбросами парниковых газов особый интерес представляет изучение *микроорганизмов, участвующих в анаэробном окислении метана*. Метагеномный анализ подземных вод, отложений и болотистой почвы, где образуется и окисляется метан, выявил необычайно большие и разнообразные последовательности ДНК, которые за способность интегрировать гены были названы «боргами» (в честь цивилизации существ из фильма «Звёздный путь»). Представляя собой новый и эволюционно отчётливо возникший тип внехромосомного элемента архей, большинство генов и повторяющихся последовательностей боргов являются новыми, но многие гены, по-видимому, ассимилированы метан-окисляющими археями *Methanoperedens*. Из-за способности генов боргов расширять окислительно-восстановительные и дыхательные способности, реагировать на изменения окружающей среды и, вероятно, окислять метан, они могут играть заметную роль в контроле выброса парниковых газов. Помимо высокой научной значимости, полученные данные являются ценным ресурсом для биотехнологического использования (Al-Shayeb et al., 2022).

## ЗАКЛЮЧЕНИЕ

Придёт время, когда наука опередит фантазию.

Жюль Верн

Нравственность должна быть полярной звездой науки.

Станислас де Буффлер

В монографии представлена лишь малая часть огромного фактического материала, накопленного за более чем тридцатилетний период геномных исследований. В ней собраны и систематизированы ключевые идеи о геномике, её значении в науке и повседневной жизни, чтобы дать представление об этой области исследований всем, кто интересуется современной биологией.

Возможность получать данные о последовательностях ДНК в масштабе генома произвела революцию в биологии с созданием ряда новых областей исследования. Геномика обеспечила надёжные ключи прогнозирования развития фауны и флоры в условиях изменения климата и повышения антропогенной нагрузки; создала мощный ресурс для расшифровки механизмов эволюции эукариотического генома и реконструкции глобального древа жизни; сформировала рациональную систему управления видами и высокоэффективную геномную медицину; заложила основы для дальнейшего развития одного из перспективных направлений – синтетической геномики. Геномные исследования ясно показали, что иерархическая классификация биологических видов отражает иерархию организации их генетического материала. Благодаря палеогеномным исследованиям мы многое узнали об исчезнувших видах и истории человечества и достигли больших успехов в реконструкции генома универсального общего предка, его метаболизма и экологического контекста. Использование полных геномов со всеми кодирующими и некодирующими последовательностями помогло обнаружить новые свидетельства динамичности генома, выявить и понять многие ранее скрытые геномные эффекты и т. д.

Успехи в секвенировании нуклеиновых кислот позволили геномике стать ведущей дисциплиной с точки зрения как роста и доступности данных, так и объёма необработанной информации, что позволяет извлекать из этих крупномасштабных хранилищ новые практические идеи (Navarro et al., 2019). Сочетание интеллектуального анализа (в естественных науках тесно связанного с математическим моделированием) и биофизического моделирования, несомненно, повысит интерпретируемость данных и обеспечит дальнейший прогресс геномики. Уже сейчас используются быстрые алгоритмы для выравнивания последовательностей, скрытые Марковские модели для аннотирования геномов. Большой вклад в геномику внесла наука о сетях и графах. Наблюдается приток решений для глубокого обучения, импортированных из машинного обучения, и ожидается, что многие архитектуры нейронных сетей могут быть перенесены в биологические исследования (Navarro et al., 2019). На основе ана-

лиза вторичных структур ДНК с использованием нейросети AlphaFold3 выдвинута новая гипотеза, согласно которой современный генетический код мог сформироваться благодаря тинкерам – самоорганизующимся молекулярным комплексам. Тинкеры появились благодаря соответствию между нуклеотидами низкой сложности, флиппонами (особые участки ДНК, способные образовывать альтернативные вторичные структуры), и простыми пептидами, используя металлы для катализа их начальной репликации. *«В целом, предлагаемая схема не требует мира ДНК, РНК или белков для объяснения происхождения жизни»* (Herbert, 2025), что кардинально меняет наш взгляд на эту одну из самых интересных и давних проблем в истории человечества.

Исследования показали, что большие языковые модели (LLM), являющиеся передовыми разработками искусственного интеллекта, могут понимать структуру языка ДНК, позволяя подойти к разным слоям генетического кода и решать многие задачи в области биологии генома. Языковые модели – *«не только ключ к предсказанию генотипа-фенотипа и информация о том, что в ДНК делает нас людьми, но и сведения о предрасположенности к болезням и реакциях на лечение»*, т.е. они могут значительно ускорить развитие персонализированной медицины (Sanabria et al., 2024). Как геном кодирует форму организма, и какова природа этого «геномного кода» на протяжении многих лет описывали с помощью разных сравнений (аналогий/метафор). Основываясь на последних разработках в области машинного обучения и нейронауки, недавно была предложена новая аналогия: геном кодирует генеративную модель организма. Это означает, что геном не кодирует ни форму организма, ни процессы развития напрямую, но включает в себя сжатое пространство скрытых переменных – последовательностей ДНК. Скрытые переменные вместе создают энергетический ландшафт, контролирующий процессы развития для создания новых особей определённого типа. Аналогия генеративной модели объясняет сложную генетическую архитектуру большинства признаков и эволюционизируемость процессов развития (Mitchell, Cheney, 2024). Для человека всегда особое значение имело старение; философы веками обсуждали его причины, а алхимики искали эликсир молодости. Недавно была разработана новая языковая модель (GPT-4b micro), направленная на оптимизацию конкретных факторов, влияющих на продолжительность жизни, и которая предлагает способы реинжиниринга белковых молекул для повышения их функции. По предварительным данным модель способна изменить факторы Яманаки (белки, которые могут превращать обычные клетки в стволовые), сделав их в десятки раз более эффективными. Эта технология открывает совершенно новые горизонты в медицине, поскольку может не только продлить жизнь, но и существенно улучшить её качество.

Разнообразии высокотехнологичных геномных подходов и сложных аналитических методов требует определенной компетенции для корректного проведения экспериментов и интерпретации данных. Использование передовых технологий также обязывает исследователей придерживаться нравственных принципов, моральных, этических и правовых норм. Поэтому были созданы различные международные до-

кументы, такие как «Всеобщая декларация о геноме человека и правах человека» (ЮНЕСКО, 1997), «Всеобщая декларация о биоэтике и правах человека» (ЮНЕСКО, 2005), «Декларация о клонировании человека» (ООН, 2005) и другие. Главным документом для геномной медицины (включающей генную терапию и технологию CRISPR–Cas9) является «Конвенция о биомедицине и правах человека» 2005 г. В статье 13 данной Конвенции заявлено, что «Вмешательство в геном человека, направленное на его модификацию, может быть осуществлено только в профилактических, терапевтических или диагностических целях и только при условии, что подобное вмешательство не направлено на изменение генома наследников данного человека». В России основные принципы этико-правового регулирования в данной сфере отражены в Федеральном законе «О государственном регулировании в области геноинженерной деятельности» (1996, N 86-ФЗ; действующая редакция, 2016).

В завершение – высказывания двух выдающихся умов человечества:

«Наука не является и никогда не будет являться законченной книгой» – Альберт Эйнштейн;

«Наука должна быть самым высшим воплощением Отечества, ибо из всех народов, первым будет всегда тот, который опередит другие в области мысли и умственной деятельности» – Луи Пастер.

## ЛИТЕРАТУРА

- Байрамов А.В., Ермакова Г.В., Кучерявый А.В., Зарайский А.Г. 2021.** Геномные дубликации как основа эволюционного успеха позвоночных. *Онтогенез* 52(3):170–194.
- Белозерский А.Н., Чигирев С.Д. 1936.** О нуклеиновом комплексе ростков семян фасоли. *Биохимия* 1(1): 134–146.
- Боркин Л.Я., Литвинчук С.Н. 2013.** Гибридизация, видообразование и систематика животных. Труды Зоологического института РАН 317(S2): 83.
- Вавилов Н.И. 1987.** Теоретические основы селекции. М.: Наука. – 512 с.
- Глазко В.И., Глазко Т.Т. 2010.** Ландшафтная геномика. *Известия Тимирязевской сельскохозяйственной академии* (5): 130–143.
- Глазко В.И., Косовский Г.Ю., Глазко Т.Т. 2022.** Геномные источники разнообразия как драйверы доместикиции (обзор). *Сельскохозяйственная биология* 57(5):832–851.
- Голубовский М.Д. 2000.** Век генетики: эволюция идей и понятий. СПб.: Борей Арт, 262 стр.
- Гольцов А.Ю., Мукосей И.С., Кочеткова Т.О. и др. 2019.** Детекция хромосомных перестроек в коротком плече 4-й и 12-й хромосом как пример полногеномного подхода при проведении неинвазивного ДНК-скрининга. *Вестник Российского государственного медицинского университета* (3):16–19.
- Комар А.А. 2019.** Распределение синонимических кодонов в мРНК определяет путь котрансляционного сворачивания белка в клетке. *Молекулярная биология* 53(6):883–898.
- Комар А.А. 2021.** Код внутри кода: как кодоны осуществляют тонкую настройку сворачивания белка в клетке. *Биохимия* 86(8):1176–1194.
- Корочкин Л.И. 2006.** Клонирование. Фрязино: Век 2. 64 с.
- Лаптев Г.Ю., Йылдырым Е.А., Дуняшев Т.П. и др. 2021.** Биоразнообразие и метаболические функции микробиома рубца у молочных коров в разные физиологические периоды. *Сельскохозяйственная биология* 56(4):619–640.
- Разин С.В., Ульянов С.В., Гаврилов А.А. 2019.** 3D геномика. *Молекулярная биология* 53(6):802–812.
- Соколова Е.Е., Власов П.К., Егорова Т.В. и др. 2020.** Влияние A/G-состава 3'-контекстов стоп-кодонов на терминацию трансляции у эукариот. *Молекулярная биология* 54(5):837–848.
- Тетушкин Е.Я. 2013.** Ландшафтная и экологическая геномика. *Генетика* 49(10):1133–1148.
- Тюлько Ж.С., Якименко В.В. 2020.** Анализ использования кодонов в кодирующих последовательностях подсемейства Coronavirinae. <https://covid19.neicon.ru/publication/10370>
- Хесин Р.Б. 1984.** Непостоянство генома. М.: Наука. 472 с.
- Циммер К. 2017.** Паразит – царь природы: Тайный мир самых опасных существ на Земле. Альпина Паблишер.
- Челомина Г.Н. 2006.** Древняя ДНК. *Генетика* 42(3):293–309.
- Челомина Г.Н. 2011.** Малые интерферирующие РНК паразитов: современный статус и перспективы использования. *Вестник ДВО РАН* (4):88–96.
- Челомина Г.Н. 2015.** Клонорхоз: эпидемиология и генетика. Владивосток: Дальнаука.
- Челомина Г.Н. 2016.** Гельминтотерапия, или парадоксы паразито-хозяйственных отношений. *Вестник ДВО РАН* (185):83–89.

- Челомина Г.Н. 2017.** Геномика и транскриптомика китайской печеночной двуустки *Clonorchis sinensis* (Opisthorchiidae, Trematoda). *Молекулярная биология* 51(2):215–226.
- Челомина Г.Н. 2021.** Ландшафтная геномика. *Биота и среда природных территорий* 9(4): 122–134.
- Челомина Г.Н. 2022.** Палеогеномика. *Биота и среда природных территорий* 10(4):64–87.
- Челомина Г.Н. 2023.** Филогеномика. *Биота и среда природных территорий* 11(2):83–99.
- Челомина Г.Н. 2024.** Геномика гибридов. *Биота и среда природных территорий* 12(1):56–77.
- Шаптуренко М.Н., Хотылева Л.В. 2016.** Гетерозис: современные тенденции в изучении молекулярных механизмов. *Вавиловский журнал генетики и селекции* 20(5):683–694.
- Abdulla A.Z., Salari H., Tortora M.M. et al. 2023.** 4D epigenomics: deciphering the coupling between genome folding and epigenomic regulation with biophysical modeling. *Curr Opin Gen Dev* 79: 102033.
- Abecasis G.R., Altshuler D., Auton A. et al. 2010.** A map of human genome variation from population-scale sequencing. *Nature* 467(7319):1061–73.
- Abecasis G.R., Auton A., Brooks L.D. et al. 2012.** An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.
- Abrouk M., Athiyannan N., Müller T. et al. 2021.** Population genomics and haplotype analysis in spelt and bread wheat identifies a gene regulating glume color. *Commun Biol* 4:375.
- Adams K.L., Wendel J.F. 2005.** Polyploidy and genome evolution in plants. *Curr Opin Plant Biol* 8(2):135–141.
- Akgül B., Stadler P.F., Hawkins L.J. et al. 2022.** 44 Current challenges in miRNomics. *Methods Mol Biol* 2257:423–438.
- Akhunov E.D., Sehgal S., Liang H. et al. 2013.** Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. *Plant Physiology* 161:252–265.
- Aksit M.A., Ling H., Pace R.G. et al. 2022.** Pleiotropic modifiers of age-related diabetes and neonatal intestinal obstruction in cystic fibrosis. *Am J Hum Genet* 109(10):1894–1908.
- Alexander D.H., Novembre J., Lange K. 2009.** Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664
- All of Us Research Program Investigators. 2019.** The “All of Us” research program. *N Engl J Med* 381(7):668–676.
- Allentoft M.E., Sikora M., Refoyo-Martínez A. et al. 2024.** Population genomics of post-glacial western Eurasia. *Nature* 625(7994):301–311.
- Al-Shayeb B., Schoelmerich M.C., West-Roberts J. et al. 2022.** Borgs are giant genetic elements with potential to expand metabolic capacity. *Nature* 610:731–736
- Alsos I.G., Sjögren P., Edwards M.E. et al. 2016.** Sedimentary ancient DNA from lake Skartjørna, Svalbard: assessing the resilience of arctic flora to holocene climate change. *Holocene* 26:627–642.
- Amores A., Force A., Yan Y.L. et al. 1998.** Zebrafish hox clusters and vertebrate genome evolution. *Science* 282(5394):1711–1714.
- An J.Y., Lin K., Zhu L. et al. 2018.** Genome-wide *de novo* risk score implicates promoter variation in autism spectrum disorder. *Science* 362:1270.
- Anderson E.C., Thompson E.A. 2002.** A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160:1217–1229.

- Arkhipova I.R., Burns K.H., Lesage P. 2024. Controlling and controlled elements: highlights of the year in mobile DNA research. *Mobile DNA* 15(1): 1–4.
- Arnold M.L. 1992. Natural hybridization as an evolutionary process. *Annu Rev Ecol Syst* 23(1):237–261.
- Arumugam M., Raes J., Pelletier E. et al. 2011. Enterotypes of the human gut microbiome. *Nature* 473(7346): 174–180.
- Askari Z., Ruehli F., Bouwman A. et al. 2022. Genomic palaeoparasitology traced the occurrence of *Taenia asiatica* in ancient Iran (Sassanid Empire, 2th cent. CE–6th cent. CE). *Sci Rep* 12(1):1–9.
- Austin M., Mellow M., Tierney W.M. 2014. Fecal microbiota transplantation in the treatment of *Clostridium difficile* infections. *Am J Med* 127(6):479–483.
- Avery O., MacLeod C., McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *Exp Med* 79:137–158.
- Aw T., Schlauch K., Keeling C.I. et al. 2010. Functional genomics of mountain pine beetle (*Dendroctonus ponderosae*) midguts and fat bodies. *Genomics* 11:215.
- Awasthi M.K., Ravindran B., Sarsaiya S. et al. 2020. Metagenomics for taxonomy profiling: tools and approaches. *Bioengineered* 11(1):356–374.
- Awasthi S., Berglund A., Abraham-Miranda J. et al. 2021. Comparative genomics reveals distinct immune-oncologic pathways in African American men with prostate cancer. *Clin Cancer Res* 27:320–329.
- Bah S.Y., Morang’a C.M., Kengne-Ouafu J.A. et al. 2018. Highlights on the application of genomics and bioinformatics in the fight against infectious diseases: challenges and opportunities in Africa. *Front Genet* 9:575.
- Bahado-Singh R. O., Vishweswaraiah S., Aydas B. et al. 2021. Artificial intelligence and leukocyte epigenomics: Evaluation and prediction of late-onset Alzheimer’s disease. *PloS One* 16(3):e0248375.
- Baker D., Sali A. 2001. Protein structure prediction and structural genomics. *Science* 294:93–96.
- Baker E.N. 2007. Structural genomics as an approach towards understanding the biology of tuberculosis. *J Struct Funct Genomics* 8:57–65.
- Balabanova L., Nedashkovskaya O., Otstavnykh N. et al. 2023. Computational insight into intraspecies distinctions in *Pseudoalteromonas distincta*: Carotenoid-like synthesis traits and genomic heterogeneity. *Int J Mol Sci* 24(4): 41–58.
- Balkenhol N., Dudanic R. ., Krutovsky K.V. et al. 2017. Landscape genomics: understanding relationships between environmental heterogeneity and genomic characteristics of populations. In: Rajora O.P. (ed.), *Population Genomics: Concepts, Approaches and Applications*. Springer. pp. 261–322.
- Balkenhol N., Fortin M.-J. 2016. Basics of study design: sampling landscape heterogeneity and genetic variation for landscape genetic studies. In: Balkenhol N., Cushman S., Storfer A., Waits L. (eds.) *Landscape Genetics: Concepts, Methods, Applications*. Wiley: West Sussex. pp. 58–75.
- Bao W., Kapitonov V.V., Jurka J. 2010. Ginger DNA transposons in eukaryotes and their evolutionary relationships with long terminal repeat retrotransposons. *Mobile DNA* 1:313.
- Barski A., Cuddapah S., Cui K, Roh T.Y., Schones D.E. et al. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823–837.
- Bartel D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297.

- Bashir T., Chandra Mishra R., Hasan M.M. et al. 2018.** Effect of hybridization on somatic mutations and genomic rearrangements in plants. *Int J Mol Sci* 19(12):37–58.
- Barton N.H., Hewitt G.M. 1985.** Analysis of hybrid zones. *Annu Rev Ecol Evol Syst* 16:113–148.
- Berkel C., Cacan E. 2019.** Single-cell epigenomics in cancer research. *Biomed J Sci Techn Res* 21:15966–15973.
- Bernaola-Galván P., Carpena P., Gómez-Martín C., Oliver J.L. 2023.** Compositional structure of the genome: a review. *Biology* 12(6):849.
- Berns K., Hijmans E.M., Mullenders J. et al. 2004.** A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 428:431–437.
- Berto S., Liu Y., Konopka G. 2020.** Genomics at cellular resolution: insight into cognitive disorders and their evolution. *Human Mol Gene* 19(R1):19.
- Bhattacharya S., Roche R., Shuvo M.H., Bhattacharya D. 2021.** Recent advances in protein homology detection propelled by inter-residue interaction map threading. *Front Mol Biosci* 8:643–752.
- Bicknell B., Liebert A., Borody T. et al. 2023.** Neurodegenerative and neurodevelopmental diseases and the gut-brain axis: The potential of therapeutic targeting of the microbiome. *Int J Mol Sci* 24(11): 9577.
- Bikard D., Euler C. W., Jiang W. et al. 2014.** Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nature Biotechnology* 32(11):11461150.
- Bilitewski U. 2009.** DNA Microarrays: an introduction to the technology. *Methods Mol Biol* 509:114.
- Blanc G., Wolfe K.H. 2004.** Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *The Plant Cell* 16(7):1679–1691.
- Blanckaert A., Payseur B.A. 2021.** Finding hybrid incompatibilities using genome sequences from hybrid populations. *Mol Biol Evol* 38(10):46164627.
- Blischak P.D., Chifman J., Wolfe A.D., Kubatko L.S. 2018.** HyDe: a python package for genome-scale hybridization detection. *Systematic Biology* 67:821–829.
- Blommaert J. 2020.** Genome size evolution: towards new model systems for old questions. *Proceedings of the Royal Society B* 287(1933):20201441.
- Blommaert J., Riss S., Hecox-Lea B. et al. 2019.** Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. *BMC Genomics* 20(1):112.
- Bock D.G., Cai Z., Elphinstone C. et al. 2023.** Genomics of plant speciation. *Plant Commun* 4:100599.
- Boeke J.D., Church G., Hessel A. et al. 2016.** The genome project–write. *Science* 353:126127.
- Bolte C.E., Phannareth T., Fitzpatrick M.C. et al. 2024.** Genomic insights into hybrid zone formation: The role of climate, landscape, and demography in the emergence of a novel hybrid lineage. *Mol Ecol* 33(14):e17430.
- Bond D.R., Uddipto K., Enjeti A., Lee H.J. 2020.** Single-cell epigenomics in cancer: charting a course to clinical impact. *Epigenomics* 12(13):11391151.
- Bond J.E., Garrison N.L., Hamilton C.A. et al. 2014.** Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution. *Curr Biol* 24(15):17651771.
- Bos K.I., Harkins K.M., Herbig A. et al. 2014.** Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514(7523):494–497.

- Botigué L.R., Song S., Scheu A. et al. 2017.** Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat Commun* 8(1):1–11.
- Bouwman B.A., Crosetto N., Bienko M. 2023.** A GC-centered view of 3D genome organization. *Curr Opin Genet Dev* 78:102020.
- Bowers J.E., Chapman B.A., Rong J., Paterson A.H. 2003.** Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422(6930): 433–438.
- Brace S., Dickmann Y., Booth T.J. et al. 2019.** Ancient genomes indicate population replacement in Early Neolithic Britain. *Nat Ecol Evol* 3(5):765–771.
- Bradshaw A.J., Backman T.A., Ramírez-Cruz V. et al. 2022.** DNA authentication and chemical analysis of Psilocybe mushrooms reveal widespread misdeterminations in Fungaria and inconsistencies in metabolites. *Appl Environ Microbiol* 88(24): e01498-22.
- Brasell K.A., Pochon X., Howarth J. et al. 2022.** Shifts in DNA yield and biological community composition in stored sediment: implications for paleogenomic studies. *Metabarcoding and Metagenomics* 6:e78128.
- Brenner S.E. 2001.** A tour of structural genomics. *Nat Rev Genet* 2(10):801–809.
- Brenner S.E., Levitt M. 2000.** Expectations from structural genomics. *Protein Sci* 9(1):197-200.
- Brent M.R. 2008.** Steady progress and recent breakthroughs in the accuracy of automated genome annotation. *Nat Rev Genet* 9:6273.
- Brindley P.J., Laha T., McManus D.P., Loukas A. 2003.** Mobile genetic elements colonizing the genome of metazoan parasites. *Trends Parasitol* 19:79–87.
- Britton C., Winter A.D., Gillan V., Devaney E. 2014.** microRNAs of parasitic helminths—Identification, characterization and potential as drug targets. *International Journal for Parasitology: Drugs and Drug Resistance* 4(2):85–94.
- Brockhurst M.A., Harrison E., Hall J.P. et al. 2019.** The ecology and evolution of pangenomes. *Curr Biol* 29(20):R1094R1103.
- Brown P.D., Walsh E.J. 2019.** Genome size and lifestyle in gnesiotrochan rotifers. *Hydrobiologia* 844(1):105–115.
- Brown S.D.M., Lad H.V. 2019.** The dark genome and pleiotropy: challenges for precision medicine. *Mamm Genome* 30:212–216.
- Brownstein C.D., Yang L., Friedman M., Near T.J. 2022.** Phylogenomics of the ancient and species-depauperate gars tracks 150 million years of continental fragmentation in the Northern Hemisphere. *Syst Biol* 72(1) 213–227.
- Brunson K., Reich D. 2019.** The promise of paleogenomics beyond our own species. *Trends Genet* 35(5):319–329.
- Brynildsrud O.B., Pepperell C.S., Suffys P. et al. 2018.** Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. *Sci Adv* 4(10):eaat5869.
- Buggs R.J., Elliott N.M., Zhang L. et al. 2010.** Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytol* 186(1): 175–183.
- Burki F. 2014.** The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb Perspect Biol* 6(5):a016147.
- Burki F., Shalchian-Tabrizi K., Pawlowski J. 2008.** Phylogenomics reveals a new ‘megagroup’ including most photosynthetic eukaryotes. *Biology Letters* 4(4):366–369.

- Caciro-Dias G., Brelsford A., Meneses-Ribeiro M. et al. 2023.** Hybridization in late stages of speciation: Strong but incomplete genome-wide reproductive isolation and ‘large Z-effect’ in a moving hybrid zone. *Mol Ecol* 32:4362–4380.
- Cahan S.H., Nguyen A.D., Zhou Y. 2022.** Population genomics supports multiple hybrid zone origins of socially hybridogenetic lineages of *Pogonomyrmex harvester* ants. *Evolution* 76(5):10161032.
- Cai C., Tihelka E., Giacomelli M. et al. 2022.** Integrated phylogenomics and fossil data illuminate the evolution of beetles. *R Soc Open Sci* 9(3):211771.
- Callaway E. 2023.** ‘Truly gobsmacked’: ancient-human genome count passes 10,000. *Nature* 617:20.
- Camargo A. P., Roux S., Schulz F. et al. 2024.** Identification of mobile genetic elements with geNomad. *Nat Biotechnol* 42:1303–1312.
- Cao B., Wang X., Yin W. et al. 2024.** The human microbiota is a beneficial reservoir for SARS-CoV-2 mutations. *mBio* 15(5):e03187-23.
- Capodiferro M.R., Osorio A.M.C., Migliore N.R. et al. 2023.** The multifaceted genomic history of Ashaninka from Amazonian Peru. *Curr Biol* 33(8):1573–1581.
- Carter J.K., Kimball R.T., Funk E.R. et al. 2023.** Estimating phylogenies from genomes: A beginners review of commonly used genomic data in vertebrate phylogenomics. *J Hered* 114(1): 1–13.
- Cello J., Paul A.V., Wimmer E. 2002.** Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* 297:1016-1018.
- Chan K.O., Brown R.M. 2019.** Elucidating the drivers of genetic differentiation in Malaysian torrent frogs (Anura: Ranidae: *Amolops*): A landscape genomics approach. *Zool J Linn Soc* 190:65–78.
- Chandonia J.-M., Brenner S.E. 2006.** The impact of structural genomics: expectations and outcomes. *Science* 311:347-351.
- Charlesworth B. 2010.** Molecular population genomics: a short history. *Genet Res* 92:397–411.
- Chaudhary B., Flagel L., Stupar R.M. et al. 2009.** Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics* 182: 503–517.
- Chelomina G.N. 2024a.** Genomics. Handbook of Molecular Biotechnology. P. 421–440.
- Chelomina G.N. 2024b.** Genomic and transcriptomic analyses of *Clonorchis sinensis* infection. In: Yi-Wei Tang, Musa Y. Hindiyeh, Dongyou Liu, Andrew Sails, Paul Spearman, Jing-Ren Zhang (Eds) *Molecular Medical Microbiology*, 3<sup>rd</sup> ed. 32213234.
- Chelomina G.N., Kukla S.P., Chelomin V.P., Doanh P.N. 2021.** Genotoxicity of *Paragonimus heterotremus* infection in a rat model of simultaneous pulmonary and hepatic Paragonimiasis. *Biomedicines* 9(9):1180.
- Chen A., Chen D., Chen Y. 2018.** Advances of DNase-seq for mapping active gene regulatory elements across the genome in animals. *Gene* 667:8394.
- Chen E., Huang X., Tian Z. et al. 2019.** The genomics of *Oryza* species provides insight into rice domestication and heterosis. *Annu Rev Plant Biol* 70:639665.
- Chen G., Bai R., Zhang Y. et al. 2022a.** Application of metagenomics to biological wastewater treatment. *Science of The Total Environment* 807:150737.
- Chen H., Guo M., Dong S. et al. 2023.** A chromosome-scale genome assembly of *Artemisia argyi* reveals unbiased subgenome evolution and key contributions of gene duplication to volatile terpenoid diversity. *Plant Commun* 4(3):100516.
- Chen L., Zhu Y., Ren X. et al. 2022b.** Heterosis and differential DNA methylation in soybean hybrids and their parental lines. *Plants* 11(9):1136.

- Chen M.Y., Teng W.K., Zhao L. et al. 2021a.** Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *The ISME Journal* 15(1):211–227.
- Chen N., Nedoluzhko A. 2023.** Ancient DNA: the past for the future. *BMC Genom* 24:309.
- Chen X., Rechavi O. 2022.** Plant and animal small RNA communications between cells and organisms. *Nat Rev Mol Cell Biol* 23(3):185–203.
- Chen R., Huangfu L., Lu Y. et al. 2021b.** Adaptive innovation of green plants by horizontal gene transfer. *Biotechnol Adv* 46:107671.
- Chen S.Y., Schenkel F.S., Melo A.L. 2022c.** Identifying pleiotropic variants and candidate genes for fertility and reproduction traits in Holstein cattle via association studies based on imputed whole-genome sequence genotypes. *BMC Genom* 23(1):122.
- Chen Z.J., Ni Z. 2006.** Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* 28: 240–252.
- Chénais B. 2022.** Transposable elements and human diseases: mechanisms and implication in the response to environmental pollutants. *Int J Mol Sci* 23(5): 2551.
- Chénais B., Caruso A., Hiard S., Casse N. 2012.** The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* 509(1): 7–15.
- Cheng S., Xian W., Fu Y. et al. 2019.** Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* 179(5):1057–1067.
- Chiang V.S. C., DeRosa H., Park J.H., Hunter R.G. 2022.** The role of transposable elements in sexual development. *Front Behav Neurosci* 16:923732.
- Chu X., Gugger P.F., Li L. et al. 2021.** Responses of an endemic species (*Roscoea humeana*) in the Hengduan Mountains to climate change. *Divers Distrib* 27(11): 2231–2244.
- Chua P.Y.S., Crampton-Platt A., Lammers Y., Alsos I.G., Boessenkool S., Bohmann K. 2021.** Metagenomics: A viable tool for reconstructing herbivore diet. *Mol Ecol Resour* 21(7):2249–2263.
- Cissé O.H., Ma L., Wei Huang D. et al. 2018.** Comparative population genomics analysis of the mammalian fungal pathogen *Pneumocystis*. *mBio* 9: e0038118.
- Clamp M., Fry B., Kamal M. et al. 2007.** Distinguishing protein-coding and noncoding genes in the human genome. *PNAS* 104(49):19428–19433.
- Cloutier A., Sackton T. B., Grayson P. et al. 2019.** Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. *Syst Biol* 68(6):937955.
- Coghlan A., Tyagi R., Cotton J.A. et al. 2019.** Comparative genomics of the major parasitic worms. *Nat Genet* 51:163–174.
- Collens A.B., Katz L.A. 2021.** Opinion: genetic conflict with mobile elements drives eukaryotic genome evolution, and perhaps also eukaryogenesis. *J Heredity* 112(1):140–144.
- Cooke N.P., Mattiangeli V., Cassidy L.M. et al. 2021.** Ancient genomics reveals tripartite origins of Japanese populations. *Sci Adv* 7(38):eabh2419.
- Cooper M.D., Alder M.N. 2006.** The evolution of adaptive immune systems. *Cell* 124:815–822.
- Coradini A.L.V., Hull C.B., Ehrenreich I.M. 2020.** Building genomes to understand biology. *Nat Commun* 11:61–77.
- Corradi N., Pombert J.F., Farinelli L. et al. 2010.** The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat Commun* 1(1):77.

- Cowart D.A., Murphy K.R., Cheng C.H.H. 2018.** Metagenomic sequencing of environmental DNA reveals marine faunal assemblages from the West Antarctic Peninsula. *Marine Genomics* 37:148–160.
- Cui R., Schumer M., Kruesi K. et al. 2013.** Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution* 67(8):2166–2179.
- Cwiklinski K., Dalton J.P., Dufresne P.J. et al. 2015.** The *Fasciola hepatica* genome: gene duplication and polymorphism reveal adaptation to the host environment and the capacity for rapid evolution. *Genome Biol* 16:71–82.
- Da Fonseca R.R., Smith B.D., Wales N. et al. 2015.** The origin and evolution of maize in the Southwestern United States. *Nat Plants* 1(1):1–5.
- da Silva Coelho F.A., Gill S., Tomlin C.M. et al. 2021.** An early dog from southeast Alaska supports a coastal route for the first dog migration into the Americas. *Proc R Soc Lond B* 288(1945):20203103.
- Damas J., Corbo M., Kim J. et al. 2022.** Evolution of the ancestral mammalian karyotype and syntenic regions. *PNAS* 119(40):e2209139119.
- Davidson B.L., Boudreau R.L. 2007.** RNA interference: a tool for querying nervous system function and an emerging therapy. *Neuron* 53(6):781–788.
- Davidson C., Wordsworth B.P., Cohen C.J. et al. 2023.** Chromosome conformation capture approaches to investigate 3D genome architecture in ankylosing spondylitis. *Front Genet* 14:1129207.
- De Koning A. P., Gu W., Castoe T.A. et al. 2011.** Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7:e1002384.
- De Kort H., Vandepitte K., Bruun H.H. et al. 2014.** Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species, *Alnus glutinosa*. *Mol Ecol* 23:4709–4721.
- De Moya R.S., Yoshizawa K., Walden K.K. et al. 2021.** Phylogenomics of parasitic and nonparasitic lice (Insecta: Psocodea): combining sequence data and exploring compositional bias solutions in next generation data sets. *Syst Biol* 70(4):719–738.
- De Wit E., De Laat W. 2012.** A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 26(1):11–24.
- Derelle E., Ferraz C., Rombauts S. et al. 2006.** Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *PNAS* 103 (31):11647–16752.
- Devaney E., Winter A.D., Britton C. 2010.** MicroRNAs: a role in drug resistance in parasitic nematodes. *Trends Parasitol* 26:428–433.
- Dewey F.E., Grove M.E., Pan C. et al. 2014.** Clinical Interpretation and Implications of Whole-Genome Sequencing. *JAMA* 311(10):1035–1045.
- Dittrich C., Hoelzl F., Smith S. et al. 2024.** Genome assembly of the dyeing poison frog provides insights into the dynamics of transposable element and genome-size evolution. *Genome Biol Evol* 16(6):evae109.
- Diz A.P., Skibinski D.O. 2024.** Patterns of admixture and introgression in a mosaic *Mytilus galloprovincialis* and *Mytilus edulis* hybrid zone in SW England. *Mol Ecol* 33(3):e17233.
- Dogan M.V., Knight S., Dogan T.K. et al. 2021.** External validation of integrated genetic-epigenetic biomarkers for predicting incident coronary heart disease. *Epigenomics* 13(14):10951112.
- Dong W., Liu Y., Li E. et al. 2022.** Phylogenomics and biogeography of *Catalpa* (Bignoniaceae) reveal incomplete lineage sorting and three dispersal events. *Mol Phylogenet Evol* 166:107330.

- dos Reis M., Inoue J., Hasegawa M. et al. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc Lond B* 279(1742):34913500.
- Doyle S.R., S e M.J., Nejsum P. et al. 2022. Population genomics of ancient and modern *Trichuris trichiura*. *Nat Commun* 13(1):1–12.
- Drukewitz S.H., von Reumont B.M. 2019. The significance of comparative genomics in modern evolutionary venomomics. *Front Ecol Evol* 7:163.
- Dukka B.K.C. 2017. Recent advances in sequence-based protein structure prediction. *Brief Bioinform* 18(6):10211032.
- Duruz S., Sevane N., Selmoni O. et al. 2019. Rapid identification and interpretation of gene-environment associations using the new R.SamBada landscape genomics pipeline. *Mol Ecol Res* 19:1355–1365.
- Edger P., Poorten T., VanBuren R., Hardigan M., Colle M., McKain M. et al. 2019. Origin and evolution of the octoploid strawberry genome. *Nature Genetics* 51(3): 541–547.
- Eid J., Fehr A., Gray J. et al. 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 323(5910):133–138.
- Eisen J.A. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* 8(3):163–167.
- Eisen J.A., Kaiser D., Myers R.M. 1997. Gastrogenomic delights: a movable feast. *Nat Med* 3(10):1076.
- Eisenstein M. 2012. Oxford Nanopore announcement sets sequencing sector abuzz. *Nat Biotechnol* 30:295–296.
- Ejigu G.F., Jung J. 2020. Review on the computational genome annotation of sequences obtained by next-generation sequencing. *Biology* 9(9):295.
- Elbarbary R.A., Lucas B.A., Maquat L.E. 2016. Retrotransposons as regulators of gene expression. *Science* 351(6274):aac7247.
- Elgvin T., Trier C., T rresen O. et al. 2017. The genomic mosaicism of hybrid speciation. *Sci Adv* 3(6):e1602996,
- Ellis L.L., Huang W., Quinn A.M. et al. 2014. Intrapopulation genome size variation in *D. melanogaster* reflects life history variation and plasticity. *PLoS Genetics* 10(7):e1004522.
- Emery N.J., Clayton N.S. 2004. The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 306(5703):1903–1907.
- Eren A.M., Kiefl E., Shaiber A. et al. 2021. Community-led, integrated, reproducible multi-omics with anvio. *Nat Microbiol* 6(1):3–6.
- Ericson P.G., Irestedt M., Zuccon D. et al. 2022. A 14,000-year-old genome sheds light on the evolution and extinction of a Pleistocene vulture. *Commun Biol* 5(1):1–9.
- Esau C.C., Monia B.P. 2007. Therapeutic potential for microRNAs. *Adv Drug Deliv Rev* 59(23):101114.
- Fadji A.E., Babalola O.O. 2020. Metagenomics methods for the study of plant-associated microbial communities: a review. *J Microbiol Methods* 170:105860.
- Fadrosh D.W., Ma B., Gajer P. et al. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2:6.
- Fages A., Hangh j K., Khan N. et al. 2019. Tracking five millennia of horse management with extensive ancient genome time series. *Cell* 177(6):1419–1435.

- Falcon F., Tanaka E.M., Rodriguez-Terrones D. 2023.** Transposon waves at the water-to-land transition. *Curr Opin Genet Dev* 81:102059.
- Farhat S., Modica M. V., Puillandre N. 2023.** Whole genome duplication and gene evolution in the hyperdiverse venomous gastropods. *Mol Biol Evol* 40(8), msad171.
- Fedoroff N.V. 2012.** Transposable elements, epigenetics, and genome evolution. *Science* 338(6108): 758–767.
- Feldman M., Harbeck M., Keller M. et al. 2016.** A high-coverage *Yersinia pestis* genome from a sixth-century Justinianic Plague victim. *Mol Biol Evol* 33(11):2911–2923.
- Fernández P., Amice R., Bruy D. et al. 2024.** A 160 Gbp fork fern genome shatters size record for eukaryotes. *iScience* 27(6):109889
- Fernandez A.M., Dodd R.S. 2018.** Using CRISPR as a gene editing tool for validating adaptive gene function in tree landscape genomics. *Front Ecol Evol* 6:76.
- Fiers W., Contreras R., Duerinck F. et al. 1976.** Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 260:500–507.
- Fischer E.C., Hashimoto K., Zhang Y. et al. 2020.** New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nat Chemical Biol* 16(5):570–576.
- Fitzgerald J., Morris D.W., Donohoe G. 2020.** Cognitive genomics: recent advances and current challenges. *Curr Psychiatry Rep* 22:2.
- Fleischmann R.D., Adams M.D., White O. et al. 1995.** Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science* 269: 496–512.
- Flutre T., Permal E., Quesneville H. 2012.** Transposable element annotation in completely sequenced eukaryote genomes. In: Grandbastien M.A., Casacuberta J.M. (eds.). *Plant Transposable Elements. Topics in Current Genetics* 24:17–40. Springer, Heidelberg, Germany.
- Flyamer I.M., Gassler J., Imakaev M. et al. 2017.** Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 544:110.
- Flynt A.S., Lai E.C. 2008.** Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet* 9(11):831–842.
- Fong J.J., Sung Y.H., Ding L. 2020.** Comparative analysis of the fecal microbiota of wild and captive beal's eyed turtle (*Sacalia bealei*) by 16S rRNA gene sequencing. *Front Microbiol* 11:570890.
- Fox J.M., Erill I. 2010.** Relative codon adaptation: a generic codon bias index for prediction of gene expression. *DNA Res* 17:185–196.
- Francis J.M., Zhang C.Z., Maire C.L. et al. 2014.** EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discovery* 4(8):956–971.
- Frantz L.A.F., Bradley D.G., Larson G., Orlando L. 2020.** Animal domestication in the era of ancient genomics. *Nat Rev Genet* 21(8):449–460.
- Fregel R. 2022.** The origin of the Carpathian Avar elites revealed. *Cell Genomics* 2(6):100143.
- Fremin B.J., Bhatt A.S. 2020.** A combined RNA-Seq and comparative genomics approach identifies 1,085 candidate structured RNAs expressed in human microbiomes. *bioRxiv* 2020-03.
- Fremin B.J., Bhatt A.S. 2021.** Comparative genomics identifies thousands of candidate structured RNAs in human microbiomes. *Genome Biol* 22(1):116.
- Fu X., Shen Y. 2024.** Synthetic genomics: repurposing biological systems for applications in engineering biology. *ACS Synth Biol* 13(5):1394–1399.

- Funk W.C., Lovich R.E., Hohenlohe P.A. et al. 2016.** Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). *Mol Ecol* 25(10):2176–2194.
- Gaczorek T.S., Chechetkin M., Dudek K. et al. 2023.** Widespread introgression of MHC genes in Iberian *Podarcis* lizards. *Mol Ecol* 32(14):4003–4017.
- Gamez S., Srivastav S., Akbari O.S., Lau N. C. 2020.** Diverse defenses: A perspective comparing Dipteran piwi-piRNA pathways. *Cells* 9(10):2180.
- Gantuz M., Morales A., Bertoldi M.V. et al. 2022.** Hybridization and polyploidization effects on LTR-retrotransposon activation in potato genome. *J Plant Res* 135(1):81–92.
- Gao C. 2021.** Genome engineering for crop improvement and future agriculture. *Cell* 184(6):1621–1635.
- Garel M.E.C., Philippe V. 2022.** Estimating of additive, dominance, and epistatic genetic variance in eucalypt hybrid population. *Silvae Genet* 71(1):39–46.
- Garlapati D., Charankumar B., Ramu K. et al. 2019.** A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. *Rev Environ Sci Biotechnol* 18:389–411.
- Gasperini M., Findlay G.M., McKenna A. et al. 2017.** CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *Am J Hum Genet* 101:192–205.
- Gelabert P., Sawyer S., Bergström A. et al. 2021.** Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. *Current Biol* 31(16):3564–3574.
- Georgiev G.P. 1984.** Mobile genetic elements in animal cells and their biological significance. *Eur J Biochem* 145:203–220.
- Giakoumis M., Pinilla-Buitrago G.E., Musher L.J. et al. 2023.** Evidence of introgression, ecological divergence and adaptation in *Asterias* sea stars. *Mol Ecol* 32(20): 5541–5557.
- Giani A.M., Gallo G.R., Gianfranceschi L., Formenti G. 2020.** Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J* 18:919.
- Gibson D.G., Benders G.A., Andrews-Pfannkoch C. et al. 2008.** Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319(5867):1215–1220.
- Gibson D.G., Glass J.I., Lartigue C. et al. 2010.** Creation of a bacterial cell controlled by a chemically synthesized gene. *Science* 329:52–56.
- Gilbert C., Schaack S., Pace J.K. et al. 2010.** A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Science* 464:1347–1350.
- Gill H., Lee J., Jeong C. 2024.** Reconstructing the genetic relationship between ancient and present-day Siberian populations. *Genome Biol Evol* 16(4):evae063.
- Ginsburg G.S., Willard H.F. 2009.** Genomic and personalized medicine: foundations and applications. *Transl Res* 154(6):277–287.
- Gladyshev E.A., Meselson M., Arkhipova I.R. 2008.** Massive horizontal gene transfer in bdelloid rotifers. *Science* 320:1210–1213.
- Gnecchi-Ruscione G.A., Khussainova E., Kahbatkyzy N. et al. 2021.** Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. *Sci Adv* 7(13):eabe4414.
- Goffeau A., Barrell B.G., Bussey H. et al. 1996.** Life with 6000 genes. *Science* 274:563–567.
- Goh H.H., Ng C.L., Loke K.K. 2018.** Functional Genomics. *Adv Exp Med Biol* 1102:11–30.
- Goldman A.D., Becerra A. 2024.** A new view of the last universal common ancestor. *J Mol Evol* 92(5):659–661.

- Golicz A.A., Batley J., Edwards D. 2016. Towards plant pangenomics. *Plant Biotechnol J* 14(4):10991105.
- Golicz A.A., Bayer P.E., Bhalla P.L. et al. 2020. Pangenomics comes of age: from bacteria to plant and animal applications. *Trends Genet* 36(2):132–145.
- Gomez M.S., Brown M.J., Pironon S. et al. 2024. Genome size is positively correlated with extinction risk in herbaceous angiosperms. *New Phytol* 243(6):2470–2485.
- Gompert Z., Mandeville E. G., Buerkle C. A. 2017. Analysis of population genomic data from hybrid zones. *Annu Rev Ecol Evol Syst* 48: 207–229.
- Goodwin S., McPherson J.D., McCombie W.R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17(6):333–351.
- Goto Y., Akahori R., Yanagi I., Takeda K.I. 2020. Solid-state nanopores towards single-molecule DNA sequencing. *Journal of Human Genetics* 65(1):69–77.
- Gozashti L., Roy S.W., Thornlow B. et al. 2022. Transposable elements drive intron gain in diverse eukaryotes. *PNASi* 119(48): e2209766119.
- Graur D. 2022. Slaying (yet again) the brain-eating zombie called the “Isochore Theory”: A segmentation algorithm used to “confirm” the existence of isochores creates “Isochores” where none exist. *Int J Mol Sci* 23(12):6558.
- Grealy A., Phillips M., Miller G. et al. 2017. Eggshell palaeogenomics: Palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. *Mol Phylogenet Evol* 109:151–163.
- Grilz-Seger G., Mesarič M., Cotman M. et al. 2018. Runs of homozygosity and population history of three horse breeds with small population size. *J Equine Vet Sci* 71:27–34.
- Groenen M.A., Archibald A.L., Uenishi H. et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491(7424):393–398.
- Gros-Balthazard M., Flowers J.M., Hazzouri K.H. et al. 2021. The genomes of ancient date palms germinated from 2,000 y old seeds. *PNAS* 118(19):e2025337118.
- Grummer J.A., Behergary L.B., Bernatchez L. et al. 2019. Aquatic landscape genomics and environmental effects on genetic variation. *Trends Ecol Evol* 34(7):641–654.
- Guerra-Assunção J.A., Enright A.J. 2012. Large-scale analysis of microRNA evolution. *BMC Genomics* 13:218.
- Guerreiro M.P.G. 2014. Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in *Drosophila*. *Mobile Genetic Elements* 4(4):e88992.
- Gugger P.F., Fitz-Gibbon S.T., Albarrán-Lara A. et al. 2021. Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales. *Mol Ecol* 30:406–423.
- Guigó R., Flicek P., Abril J.F. et al. 2006. EGASP: The human ENCODE Genome Annotation Assessment Project. *Genome Biol* 7:S2.131.
- Gupta A., Swaminathan G., Martin-Garcia J., Navas-Martin S. 2012. MicroRNA, hepatitis C virus, and HCV/HIV-1 co-infection: new insights in pathogenesis and therapy. *Viruses* 4:24852513.
- Gupta N., Verma V.K. 2019. Next-generation sequencing and its application: empowering in public health beyond reality. *Microbial Technology for the Welfare of Society* 17:313–341.
- Gustaffson A. 1946. The effect of heterozygosity on variability and vigour. *Hereditas* 32:263–286.
- Gutaker R.M., Burbano H.A. 2017. Reinforcing plant evolutionary genomics using ancient DNA. *Curr Opin* 36:38–45.

- Hackl T., Laurenceau R., Ankenbrand M.J. et al. 2023.** Novel integrative elements and genomic plasticity in ocean ecosystems. *Cell* 186(1):47–62.
- Hagenblad J., Morales J., Leino M.W., Rodríguez–Rodríguez A.C. 2017.** Farmer fidelity in the Canary Islands revealed by ancient DNA from prehistoric seeds. *J Archaeol Sci* 78:78–87.
- Haghani A., Li C.Z., Robeck T.R. et al. 2023.** DNA methylation networks underlying mammalian traits. *Science* 381(6658):eabq5693.
- Hahn C., Fromm B., Bachmann L. 2014.** Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparasitic neodermata. *Genome Biol Evol* 6:1105–1117.
- Han S., Tang Q., Lu X. et al. 2016a.** Dysregulation of hepatic microRNA expression profiles with *Clonorchis sinensis* infection. *BMC Infect Dis* 16:72431.
- Han T.S., Zheng Q.J., Lv J.X. et al. 2024.** Intraspecific genome size variation in *Rorippa indica* reveals a tropical adaptation by genomic enlargement. *bioRxiv* 2024.10.13.618028.
- Han Y., Wu Z., Wu T. et al. 2016b.** Tumor-suppressive function of long noncoding RNA MALAT1 in glioma cells by downregulation of MMP2 and inactivation of ERK/MAPK signaling. *Cell Death Dis* 7(3):e2123.
- Hardison R.C. 2003.** Comparative Genomics. *PLoS Biol* 1(2):156160.
- Harfe B.D. 2005.** MicroRNAs in vertebrate development. *Curr Opin Genet Dev* 15:410–415.
- Harris B.J., Harrison C.J., Hetherington A.M., Williams T.A. 2020.** Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. *Current Biol* 30(11):2001–2012.
- Harrison R.G. (Ed.). 1993.** Hybrid Zones and the Evolutionary Process. Oxford University Press, USA.
- Hart T., Chandrashekhar M., Aregger M. et al. 2015.** High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163(6):15151526.
- Hauth A., Galupa R., Servant N. et al. 2022.** Deciphering high-resolution 3D chromatin organization via capture Hi-C. *J Vis Exp* (188):e64166.
- Head S.R., Komori H.K., LaMere S.A. et al. 2014.** Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* 56:61.
- Hehmeyer J., Spitz F., Marlow H. 2023.** Shifting landscapes: the role of 3D genomic organizations in gene regulatory strategies. *Curr Opin Genet Dev* 81:102064.
- Hekkala E., Gatesy J., Narechania A. et al. 2021.** Paleogenomics illuminates the evolutionary history of the extinct Holocene “horned” crocodile of Madagascar, *Voay robustus*. *Commun Biol* 4(1):505.
- Hempel E., Bibi F., Faith J.T. et al. 2022.** Blue turns to gray: paleogenomic insights into the evolutionary history and extinction of the blue antelope (*Hippotragus leucophaeus*). *Mol Biol Evol* 39(12):msac241.
- Hendricks S., Anderson E.C., Antao T. et al. 2018.** Recent advances in conservation and population genomics data analysis. *Evol Appl* 11(8):1197–1211.
- Herbert A. 2025.** Flipons and the origin of the genetic code. *Biol Lett* 21(1):20240635.
- Hertel J., Lindemeyer M., Missal K. et al. 2006.** The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7(1):115.
- Hickey G., Monlong J., Ebler J. et al. 2024.** Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nat Biotechnol* 42:663–673.

- Hidalgo O., Pellicer J., Christenhusz M.J. et al. 2017.** Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*. *Bot J Linn Soc* 183(4):509514.
- Higuchi R., Bowman B., Freiberger M. et al. 1984.** DNA sequences from the Quagga, an extinct member of the horse family. *Nature* 312:282–284.
- Hime P.M., Lemmon A.R., Lemmon E.C.M. et al. 2021.** Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Syst Biol* 70(1):49–66.
- Hirsch P., Tagirdzhanov A., Kushnareva A. et al. 2024.** ABC-HuMi: The Atlas of biosynthetic gene clusters in the human microbiome. *Nucleic Acids Res* 52(D1):D579–D585.
- Hjelmen C.E. 2024.** Genome size and chromosome number are critical metrics for accurate genome assembly assessment in Eukaryota. *Genetics* 227(4): iyae099.
- Hjort L., Rushiti F., Wang S-J. et al. 2021.** Intergenerational effects of maternal post-traumatic stress disorder on offspring epigenetic patterns and cortisol levels. *Epigenomic* 13 (12):967–980.
- Ho A.T., Hurst L.D. 2022.** Stop codon usage as a window into genome evolution: mutation, selection, biased gene conversion and the TAG paradox. *Genome Biol Evol* 14(8):evac115.
- Hoencamp C., Dudchenko O., Elbatsh A.M. et al. 2021.** 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* 372(6545):984–989.
- Hohenlohe P.A., Funk W.C., Rajora O.P. 2020.** Population genomics for wildlife conservation and management. *Mol Ecol* 30(1):62–82.
- Holley R.W., Apgar J., Everett G.A. et al. 1965.** Structure of a ribonucleic acid. *Science* 147:1462–1465.
- Homola J.J., Loftin C.S., Cammen K.M. et al. 2019.** Replicated landscape genomics identifies evidence of local adaptation to urbanization in wood frogs. *J Hered* 110:707–719.
- Hong S., Lim Y.P., Kwon S-Y. et al. 2021.** Genome-wide comparative analysis of flowering-time genes; insights on the gene family expansion and evolutionary perspective. *Front Plant Sci* 12:702243.
- Hook P.W., Timp W. 2023.** Beyond assembly: the increasing flexibility of single-molecule sequencing technology. *Nat Rev Genet* 24(9) 627–641.
- Hooper D.M., Griffith S.C., Price T.D. 2019.** Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. *Mol Ecol* 28(6):1246–1262.
- Hotaling S., Desvignes T., Sproul J.S. et al. 2023.** Pathways to polar adaptation in fishes revealed by long-read sequencing. *Mol Ecol* 32(6):1381–1397.
- Hotaling S., Kelley J.L., Frandsen P. 2021.** Toward a genome sequence for every animal: Where are we now? *PNAS* 118(52):e2109019118.
- Hu B., Guo H., Zhou P., Shi, Z.L. 2021.** Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 19(3):141–154.
- Hu P., Zhang W., Xin H., Deng G. 2016.** Single cell isolation and analysis. *Front Cell Dev Biol* 4: 116.
- Hu X., Mckenzie C.A., Smith C. et al. 2023.** The remarkable complexity of the brain microbiome in health and disease. *bioRxiv* 2023.02.06.527297.
- Huang Q.X., Cheng X.Y., Mao Z.C. et al. 2010.** MicroRNA discovery and analysis of Pinewood nematode *Bursaphelenchus xylophilus* by deep sequencing. *PLoS ONE* 5:e13271.
- Hubisz M.J., Williams A.L., Siepel A. 2020.** Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination Graph. *PLoS Genet* 16(8):e1008895.

- Hugenholtz P., Goebel B.M., Pace N.R. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180:4765–4774.
- Hung C.M., Shaner P.J.L., Zink R.M. et al. 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *PNAS* 111(29):10636–10641.
- Huo J., Wu L., Zang Y. 2021. Eight-gene prognostic signature associated with hypoxia and ferroptosis for gastric cancer with general applicability. *Epigenomics* 13(11):875–890.
- Ikeda M., Matsuyama S., Akagi S. et al. 2017. Correction of a disease mutation using CRISPR/Cas9-assisted genome editing in Japanese black cattle. *Sci Rep* 7(1):17827.
- International Helminth Genomes Consortium. 2019. Comparative genomics of the major parasitic worms. *Nat Genet* 51(1):163–174.
- Ito J., Suqimoto R., Nakaoka H. et al. 2017. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet* 13(7):e1006883.
- Isemann M., Stoddart M.J., Schmelzeisen R. et al. 2023. Basic principles of RNA interference: Nucleic acid types and in vitro intracellular delivery methods. *Micromachines* 14(7):13–21.
- Jaenicke-Despres V., Buckler E.S., Smith B.D. et al. 2003. Early allelic selection in maize as revealed by ancient DNA. *Science* 302(5648):1206–1208.
- Jenuwein T., Allis C.D. 2001. Translating the histone code. *Science* 293(5532):1074–1080.
- Jeong C., Ozga A.T., Witonsky D.B. et al. 2016. Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *PNAS* 113:7485–7490.
- Jeong H., Song K. D., Seo M. et al. 2015. Exploring evidence of positive selection reveals genetic basis of meat quality traits in Berkshire pigs through whole genome sequencing. *BMC Genetics* 16(1):19.
- Jia K., Zhao W., Maier P.A. et al. 2020. Landscape genomics predicts climate change-related genetic offset for the widespread *Platyclus orientalis* (Cupressaceae). *Evol Appl* 13:665–676.
- Jia W., Anslan S., Chen F. et al. 2022. Sedimentary ancient DNA reveals past ecosystem and biodiversity changes on the Tibetan Plateau: Overview and prospects. *Quat Sci Rev* 293:107703.
- Jiang X., Song Q., Ye W., Chen Z.J. 2021. Concerted genomic and epigenomic changes accompany stabilization of *Arabidopsis* allopolyploids. *Nat Ecol Evol* 5(10):1382–1393.
- Johnston D.A. 2006. Genomes and genomics of parasitic flatworms. In: Maule A.G., Marks N.J. (Eds.) *Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology and Physiology*. CAB International, Wallingford Press, pp. 3780.
- Johnston M.J., Nikolic A., Ninkovic N. et al. 2019. High-resolution structural genomics reveals new therapeutic vulnerabilities in glioblastoma. *Genome Res* 29:1211–1222.
- Joos R., Boucher K., Lavelle A. et al. 2025. Examining the healthy human microbiome concept. *Nat Rev Microbiol* 23:192–205.
- Joseph S.K., Lindo J. 2023. The evolutionary history of infectious disease in the ancient Americas and the pathogenic consequences of European contact. *Am J Biol Anthropol.* 182(4):532–541.
- Jothi R., Cuddapah S., Barski A. et al. 2008. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucl Acids Res* 36:5221–5231.
- June V., Xu D., Papoulas O. et al. 2023. Protein nonadditive expression and solubility contribute to heterosis in *Arabidopsis* hybrids and allotetraploids. *Front Plant Sci* 14:1252564.
- Jusakul A., Cutcutache I., Yong C.H. et al. 2017. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. *Cancer Discov* 7:1116–1135.
- Kachroo A.H., Laurent J.M., Yellman C.M. et al. 2015. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* 348(6237):921–925.

- Kachroo A.H., Vandelloo M., Greco B.M., Abdullah M. 2022.** Humanized yeast to model human biology, disease and evolution. *Dis Model Mech* 15(6):dmm049309.
- Kachroo P., Eraso J.M., Beres S.B. et al. 2019.** Integrated analysis of population genomics, transcriptomics and virulence provides novel insights into *Streptococcus pyogenes* pathogenesis. *Nat Genet* 51:548–559.
- Kadakkuzha B.M., Puthanveetil S.V. 2013.** Genomics and proteomics in solving brain complexity. *Mol Biosyst* 9(7):1807–1821.
- Kalaentzis K., Arntzen J. W., Avcı A. et al. 2023.** Hybrid zone analysis confirms cryptic species of banded newt and does not support competitive displacement since secondary contact. *Ecology and Evolution* 13(9):e10442
- Kambayashi C., Kakehashi R., Sato Y. et al. 2022.** Geography-dependent horizontal gene transfer from vertebrate predators to their prey. *Mol Biol Evol* 39(4):msac052.
- Kaneko-Ishino T., Ishino F. 2012.** The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. *Front Microbiol* 3:211.
- Kantidze O.L., Gurova K.V., Studitsky V.M., Razin S.V. 2020.** The 3D genome as a target for anticancer therapy. *Trends Mol Med* 26(2):141–149.
- Kaul S., Koo H.L., Jenkins J. et al. 2000.** Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6814):796–815.
- Kawai Y., Watanabe Y., Omae Y. et al. 2023.** Exploring the genetic diversity of the Japanese population: insights from a large-scale whole genome sequencing analysis. *PLoS Genet* 19(12):e1010625.
- Ke R., Marco M., Hauling T. et al. 2016.** Fourth generation of next-generation sequencing technologies: promise and consequences. *Hum Mutat* 37:1363–1367.
- Ke F., Vasseur L. 2024.** Contribution of GC content to differentiated region formation along a speciation continuum. *bioRxiv* 2024.09.02.610918.
- Keough K.C., Whalen S., Inoue F. et al. 2023.** Three-dimensional genome rewiring in loci with human accelerated regions. *Science* 380(6643):eabm1696.
- Kerner G., Laval G., Patin E. et al. 2021a.** Human ancient DNA analyses reveal the high burden of tuberculosis in Europeans over the last 2,000 years. *Am J Hum Genet* 108(3):517–524.
- Kerner G., Patin E., Quintana–Murci L. 2021b.** New insights into human immunity from ancient genomics. *Curr Opin Immunol* 72:116–125.
- Khandia R., Gurjar P., Kamal M.A., Greig N.H. 2024.** Relative synonymous codon usage and codon pair analysis of depression associated genes. *Sci Rep* 14(1): 3502.
- Kim S., Cho Y.S., Kim H.M. et al. 2016.** Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. *Genome Biol* 17(1):112.
- Kimoto M., Hirao I. 2022.** Genetic code engineering by natural and unnatural base pair systems for the site-specific incorporation of non-standard amino acids into proteins. *Front Mol Biosci* 9:851646.
- Kimoto M., Tan H.P., Tan Y.S. et al. 2023.** Success probability of high-affinity DNA aptamer generation by genetic alphabet expansion. *Phil Trans R Soc B* 378(1871):20220031.
- Kinkar L., Korhonen P.K., Saarma U. et al. 2023.** Genome-wide exploration reveals distinctive northern and southern variants of *Clonorchis sinensis* in the Far East. *Mol Ecol Res* 23 (4):833–843.
- Klar A. J. 1998.** Propagating epigenetic states through meiosis: where Mendel's gene is more than a DNA moiety. *Trends Genet* 14(8):299–301.

- Klein S.J., O'Neill R.J. 2018. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res* 26:523.
- Knief U., Bossu C.M., Saino N. et al. 2019. Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat Ecol Evol* 3(4):570–576.
- Koelsche C., Schrimpf D., Stichel D. et al. 2021. Sarcoma classification by DNA methylation profiling. *Nat Commun* 12(1):498.
- Koga A., Iida A., Hori H. et al. 2006. Vertebrate DNA transposon as a natural mutator: the medaka fish Tol2 element contributes to genetic variation without recognizable traces. *Mol Biol Evol* 23(7):1414–1419
- Kong S., Kubatko L. S.2021. Comparative performance of popular methods for hybrid detection using genomic data. *Syst Biol* 70(5): 891–907.
- Kong S., Li R., Tian Y. et al. 2023. Single-cell omics: A new direction for functional genetic research in human diseases and animal models. *Frontiers Genetics* 13:1100016.
- Konig H., Frank D., Heil R., Coenen C. 2013. Synthetic genomics and synthetic biology applications between hopes and concerns. *Curr Genomics* 14(1):11–24.
- Koonin E.V. 2009. Evolution of genome architecture. *Int J Biochem Cell Biol* 41(2):298–306.
- Koren O., Spor A., Felin J. et al. 2011. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *PNAS* 108(supplement\_1):4592–4598.
- Korhonen P.K., Young N.D., Gasser R.B. 2016. Making sense of genomes of parasitic worms: Tackling bioinformatic challenges. *Biotechnol Adv* 34 (5):663–686.
- Kotari I., Kosiol C., Borges R. 2024. The patterns of codon usage between chordates and arthropods are different but co-evolving with mutational biases. *Mol Biol Evol* 41(5):msae080.
- Kovina A.P., Petrova N.V., Gushchanskaya E.S. et al. 2017. Evolution of the genome 3D organization: comparison of fused and segregated globin gene clusters. *Mol Biol Evol* 34(6):1492–1504.
- Krasheninnikova K., Diekhans M., Armstrong J. et al. 2020. halSynteny: a fast, easy-to-use conserved synteny block construction method for multiple whole-genome alignments. *GigaScience* 9:1–5.
- Krug P.J., Caplins S.A., Algosó K. et al. 2022. Phylogenomic resolution of the root of Panpulmonata, a hyperdiverse radiation of gastropods: new insight into the evolution of air breathing. *Proc R Soc B*, 289(1972):20211855.
- Kuhlwilm M., Boeckx C. 2019. A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. *Sci Rep* 9:8463.
- Kumar R.P., Senthilkumar R., Singh V., Mishra R.K. 2010. Repeat performance: how do genome packaging and regulation depend on simple sequence repeats? *Bioessays* 32(2):165–174.
- Kumar V., Muratani M., Rayan N.A. et al. 2013. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat Biotechnol* 31:615–622.
- Kurata N., Moore G., Nagamura Y. et al. 1994. Conversation of genome structure between rice and wheat. *BioTechnol* 12:276–278.
- Kushwaha B., Nagpure N.S., Srivastava S. et al. 2023. Genome size estimation and its associations with body length, chromosome number and evolution in teleost fishes. *Gene* 864: 147294.
- Kyriacou R.G., Mulhair P.O., Holland P.W. 2024. GC content across insect genomes: phylogenetic patterns, causes and consequences. *J Mol Evol* 92(2):138–152.
- LaCava M.E.F., Gagne R.B., Gustafson K.D. et al. 2021. Functional connectivity in a continuously distributed, migratory species as revealed by landscape genomics. *Ecography* 47: 987–999.

- Lachance J., Tishkoff S. A. 2013.** Population genomics of human adaptation. *Ann Rev Ecol Evol Systematics* 44:123–143.
- Lack J.B., Weider L. J., Jeyasingh P.D. 2018.** Whole genome amplification and sequencing of a *Daphnia* resting egg. *Mol Ecol Res* 18(1):118–127.
- Lähnemann D., Köster J., Fischer U. et al. 2021.** Accurate and scalable variant calling from single cell DNA sequencing data with ProSolo. *Nat Commun* 12(1):6744.
- Lähnemann D., Köster J., Szczurek E. et al. 2020.** Eleven grand challenges in single-cell data science. *Genome Biol* 21:135.
- Lamolle G., Fontenla S., Rijo G. et al. 2019.** Compositional analysis of flatworm genomes shows strong codon usage biases across all classes. *Front Genet* 10:771.
- Lamolle G., Iriarte A., Musto H. 2022.** Codon usage in the flatworm *Schistosoma mansoni* is shaped by the mutational bias towards A+ T and translational selection, which increases GC-ending codons in highly expressed genes. *Mol Biochem Parasitol* 247:111445.
- Lan T., Leppälä K., Tomlin C. et al. 2022.** Insights into bear evolution from a Pleistocene polar bear genome. *PNAS* 119(24):e2200016119.
- Lan T., Lindqvist C. 2018.** Paleogenomics: genome-scale analysis of ancient DNA and population and evolutionary genomic inferences. In: Om P.R. (ed.) *Population Genomics: Concepts, Approaches and Applications*, 138. Springer International Publishing, Cham.
- Lan T., Lindqvist C. 2019.** Technical advances and challenges in genome-scale analysis of ancient DNA. In: Lindqvist C., Rajora O.P. (eds), *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, 329. Springer International Publishing, Cham.
- Lander E.S., Linton L.M., Birren M. et al. 2001.** Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Larsen C.S., Crespo F. 2022.** Paleosyndemics: A bioarchaeological and biosocial approach to study infectious diseases in the past. *Centaurus* 64(1):181–196.
- Le Duc D., Velluva A., Cassatt–Johnstone M. et al. 2022.** Genomic basis for skin phenotype and cold adaptation in the extinct Steller’s sea cow. *Sci Adv* 8(5):eabl6496.
- Lee H., Gurtowski J., Yoo S. et al. 2016.** Third-generation sequencing and the future of genomics. *bioRxiv* 048603.
- Li A., Hao C., Wang Z. et al. 2022a.** Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. *Mol Plant* 15(3):504–519.
- Li C., Kasinski A.L. 2020.** *In vivo* cancer-based functional genomics. *Trends Cancer* 6:1002–1017.
- Li D., Li Y., Li M. et al. 2019a.** Population genomics identifies patterns of genetic diversity and selection in chicken. *BMC Genom* 20:263.
- Li H., Durbin R. 2024.** Genome assembly in the telomere-to-telomere era. *Nat Rev Genet* 25(9): 658–670.
- Li H., Yang Y., Hong W. et al. 2020a.** Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduc Target Ther* 5(1):1.
- Li K., Yan F., Duan Z. et al. 2022b.** Widespread of horizontal gene transfer events in eukaryotes. *bioRxiv* 2022.07.26.501571.
- Li L.F., Cushman S.A., He Y.X. et al. 2020b.** Landscape genomics reveals genetic evidence of the local adaptation in a widespread woody species, the Chinese wingnut (*Pterocarya stenoptera* C. DC). *J Syst Evol* 60(2):386–397.

- Li Y., Pinto-Tomás A.A., Rong X. et al. 2019b. Population genomics insights into adaptive evolution and ecological differentiation in streptomycetes. *Appl Environ Microbiol* 85: e02555–18.
- Li Y., Shi W., Wasserman W.W. 2018. Genome-wide prediction of cis-regulatory regions using supervised deep learning methods. *BMC Bioinform* 19:202.
- Li Y., Zhang X-X., Mao R-L. et al. 2017. Ten years of landscape genomics: challenges and opportunities. *Front Plant Sci* 8:21–36.
- Liao W.W., Asri M., Ebler J. et al. 2023. A draft human pangenome reference. *Nature* 617(7960):312–324.
- Librado P., Khan N., Fages A. et al. 2021. The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* 598(7882):634–640.
- Link C.D. 2021. Is there a brain microbiome? *Neuroscience Insights* 16: 26331055211018709.
- Litvinchuk S.N., Matiushova M.I., Bezman-Moseyko O.S. et al. 2023. Distribution, population systems, and peculiarities of hybrid gametogenesis in water frogs (*Pelophylax esculentus* complex) in the Dniester River valley (Moldova). *Russian Journal of Herpetology* 30(2):88–100.
- Liu D., Hunt M., Tsai I.J. 2018. Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics* 19:1–13.
- Liu H.J., Yan J. 2019. Crop genome-wide association study: a harvest of biological relevance. *Plant J* 97(1):8–18.
- Liu L., Li Y., Li S. et al. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 25:1364
- Liu Q., Tuo W., Gao H., Zhu X.Q. 2010. MicroRNAs of parasites: current status and future perspectives. *Parasitol Res* 107:501–507.
- Liu S., Wu I., Yu Y-P. et al. 2021. Targeted transcriptome analysis using synthetic long read sequencing uncovers isoform reprogramming in the progression of colon cancer. *Commun Biol* 4:506.
- Liu W., Zhang Y., He H. et al. 2022. From hybrid genomes to heterotic trait output: Challenges and opportunities. *Curr Opin Plant Biol* 66:102193.
- Liu Y., Weyrich L. S., Llamas B. 2020. More arrows in the ancient DNA quiver: use of paleoepigenomes and paleomicrobiomes to investigate animal adaptation to environment. *Mol Biol Evol* 37:307–319.
- Liu Y., Yang Q., Zhao F. 2021. Synonymous but not silent: the codon usage code for gene expression and protein folding. *Annu Rev Biochem* 90(1): 375–401.
- Liu Y.C., Sun X., Driscoll C. et al. 2018. Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. *Curr Biol* 28(23): P3840–49
- Longo G.C., Lam L., Basnett B. et al. 2020. Strong population differentiation in lingcod (*Ophiodon elongatus*) is driven by a small proportion of the genome. *Evol Appl* 13(10):2536–2554.
- Loog L., Thalmann O., Sinding M.-H.S. et al. 2020. Ancient DNA suggests modern wolves trace their origin to a late Pleistocene expansion from Beringia. *Mol Ecol* 29:1596–1610.
- Lopes F., Oliveira L. R., Beux Y. et al. 2023. Genomic evidence for homoploid hybrid speciation in a marine mammal apex predator. *Sci Adv* 9(18):eadf6601.
- Loyfer N., Magenheimer J., Peretz A. et al. 2023. A DNA methylation atlas of normal human cell types. *Nature* 613(7943):355–364.
- Lu J., Huang P., Sun J., Liu J. 2023a. DupScan: predicting and visualizing vertebrate genome duplication database. *Nucl Acids Res* 51(D1):D906–D912.

- Lu M., Xueying Q., Hexiang P. et al. 2021.** Genome-wide associations between alcohol consumption and blood DNA methylation: evidence from twin study. *Epigenomics* 13(12):939–951.
- Lu Y., Li M., Gao Z. et al. 2025.** Advances in whole genome sequencing: methods, tools, and applications in population genomics. *Int J Mol Sci* 26(1):372.
- Lu Y., Rice E., Du K. et al. 2023b.** High resolution genomes of multiple *Xiphophorus* species provide new insights into microevolution, hybrid incompatibility, and epistasis. *Genome Res* 33(4):557–571.
- Luikart G., Kardos M., Hand B.K. et al. 2018.** Population genomics: advancing understanding of nature. In: Rajora O.P. (ed.), *Population Genomics: Concepts, Approaches and Applications*, p. 3–79. Springer International Publishing.
- Luo S.J., Liu Y.C., Xu X. 2019b.** Tigers of the world: Genomics and conservation. *Annu Rev Anim Biosci* 7:521–548.
- Luo X., Liu, Y., Dang D. et al. 2021.** 3D Genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis. *Cell* 184(3):723–740.
- Ma Z.S. 2024.** Metagenome comparison (MC): A new framework for detecting unique/enriched OMUs (operational metagenomic units) derived from whole-genome sequencing reads. *Comput Biol Med* 180:108852.
- Mahtani-Williams S., Fulton W., Desvars-Larrive A. et al. 2020.** Landscape genomics of a widely distributed snake, *Dolichophis caspius* (Gmelin, 1789) across Eastern Europe and Western Asia. *Genes* 11(10):1218.
- Magoulopoulou A., Salas S. M., Tiklová K. et al. 2023.** Padlock probe-based targeted *in situ* sequencing: Overview of methods and applications. *Annu Rev Genomics Hum Genet* 24:133–150.
- Maldonado L.L., Stegmayer G., Milone D.H. et al. 2018.** Whole genome analysis of codon usage in *Echinococcus*. *Mol Biochem Parasitol* 225:54–66.
- Mamoozadeh N.R., Graves J.E., McDowell J.R. 2020.** Genomewide SNPs resolve spatiotemporal patterns of connectivity within striped marlin (*Kajikia audax*), a broadly distributed and highly migratory pelagic species. *Evol Appl* 13(4):677–698
- Manguso R.T., Pope H.W., Zimmer M.D. et al. 2017.** *In vivo* CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* 547:413–418.
- Manthey J. D., Moyle R. G. 2015.** Isolation by environment in white-breasted nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: A landscape genomics approach. *Mol Ecol* 24:3628–3638.
- Manzano-Román R., Siles-Lucas M. 2012.** MicroRNAs in parasitic diseases: potential for diagnosis and targeting. *Mol Biochem Parasitol* 186(2):81–86.
- Marciniak S., Perry G.H. 2017.** Harnessing ancient genomes to study the history of human adaptation. *Nat Rev Genet* 18:659–674.
- Margaryan A., Lawson D. J., Sikora M. et al. 2020.** Population genomics of the Viking world. *Nature* 585(7825):390–396.
- Margulies M., Egholm M., Altman W.E. et al. 2005.** Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380.
- Marquardt J., Blanckaert A., Nichols R.A. et al. 2022.** Cyto-nuclear incompatibilities across a bluebell hybrid zone in northern Spain do not prevent genome-wide introgression of neutral markers. *Evolutionary Journal of the Linnean Society* 1:zkac00.
- Marri P.R., Paniscus M., Weyand N.J. et al. 2010.** Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. *PLoS One* 5(7):e11835.

- Marschall T. 2018.** Computational pan-genomics: status, promises and challenges. *Brief Bioinform* 19: 118–135.
- Marsden R.L., Lewis T.A., Orengo C.A. 2007.** Towards a comprehensive structural coverage of completed genomes: a structural genomics viewpoint. *BMC Bioinform* 8:86.
- Martellacci L., Quaranta G., Patini R. et al. 2019.** Literature review of metagenomics and culturomics of the peri-implant microbiome: current evidence and future perspectives. *Materials* 12:3010.
- Martin S. H., Davey J. W., Jiggins C. D. 2015.** Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol Biol Evol* 32(1):244–257.
- Martinez J., Klasson L., Welch J.J., Jiggins F.M. 2021.** Life and death of selfish genes: comparative genomics reveals the dynamic evolution of cytoplasmic incompatibility. *Mol Biol Evol* 38(1):215.
- Martins K., Gugger P. F., Llanderal-Mendoza J. et al. 2018.** Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evol Appl* 11:1842–1858.
- Martinson E.O., Martinson V.G., Edwards R., Werren J.H. 2016.** Laterally transferred gene recruited as a venom in parasitoid wasps. *Mol Biol Evol* 33(4):1042–1052.
- Mascher M., Schuenemann V.J., Davidovich U. et al. 2016.** Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat Genet* 48(9):1089–1093.
- Mashanov V.S., Zueva O.R., Garcia-Arrarás J.E. 2012.** Retrotransposons in animal regeneration. *Mobile Genet Elements* 2:244–246.
- разин**
- Mat-Sharani S., Quay D.H., Chyan L.N., Firdaus-Raih M. 2019.** Structural Genomics. In: Ranganathan Sh., Gribskov M., Nakai K., Schönbach C. (eds.). *Encyclopedia of Bioinformatics and Computational Biology*, 3:722–728. Elsevier.
- Matsumura H., Ito A., Saitoh H. et al. 2005.** SuperSAGE. *Cell Microbiol* 7(1):11–18.
- Mauer K., Hellmann S. L., Groth M. et al. 2020.** The genome, transcriptome, and proteome of the fish parasite *Pomphorhynchus laevis* (Acanthocephala). *PLoS One* 15(6):e0232973.
- Mauro V.P., Chappell S.A. 2014.** A critical analysis of codon optimization in human therapeutics. *Trends Mol Med* 20:604–613.
- Maxam AM, Gilbert W. 1977.** A new method for sequencing DNA. *PNAS* 74(2):560–564.
- McClintock B. 1984.** The significance of responses of the genome to challenge. *Science* 226: 792–801
- McCoy R.C., Taylor R.W., Blauwkamp T.A. et al. 2014.** Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. *PLoS ONE* 9(9):e106689.
- McGonigle L.L. 2008.** A novel target for chemotherapeutic intervention in the liver fluke, *Fasciola hepatica* (Doctoral dissertation, Queen’s University of Belfast).
- McGowen M.R., Tsagkogeorga G., Álvarez-Carretero S. et al. 2020.** Phylogenomic resolution of the cetacean tree of life using target sequence capture. *Syst Biol* 69(3):479–501.
- Medina R., Wogan G.O.U., Bi K. et al. 2021.** Phenotypic and genomic diversification with isolation by environment along elevational gradients in a neotropical treefrog. *Mol Ecol* 30(16):4062–4076.
- Meesil W., Muangpat P., Sitthisak S. et al. 2023.** Genome mining reveals novel biosynthetic gene clusters in entomopathogenic bacteria. *Sci Rep* 13(1):20764.

- Melouanea A., Ghanemia A., Yoshioka M. et al. 2019.** Functional genomics applications and therapeutic implications in sarcopenia. *Mutat Res Rev Mutat Res* 781:175–185.
- Mendel G. 1866.** Experiments on Plant Hybridization. Versuche über Pflanzen-Hybriden. *Verh. Naturforsch. Ver Brünn* 4:347.
- Mendez F.L., Poznik G.D., Castellano S. et al. 2016.** The Divergence of neanderthal and modern human Y chromosomes. *Am J Hum Genet* 98: 728–734.
- Mendum T.A., Schuenemann V.J., Roffey S. et al. 2014.** *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. *BMC Genomics* 15(1):1–8.
- Mérot C., Llaurens V., Normandeau E. et al. 2020.** Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nat Commun* 11:670.
- Messing J., Crea R., Seeburg P.H. 1981.** A system for shotgun DNA sequencing. *Nucl Acids Res* 9:309–321.
- Metzler D., Knief U., Peñalba J.V. et al. 2021.** Assortative mating and epistatic mating-trait architecture induce complex movement of the crow hybrid zone. *Evolution* 75(12): 31543174.
- Meyer A., Schloissnig S., Franchini P. et al. 2021.** Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* 590(7845):284–289.
- Meyer M., Kircher M., Gansauge M.T. et al. 2012.** A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222–226.
- Milián-García Y., Amato G., Gatesy J. et al. 2020.** Phylogenomics reveals novel relationships among Neotropical crocodiles (*Crocodylus* spp.). *Mol Phylogenet Evol* 152:106924.
- Miller W., Drautz D.I., Ratan A. et al. 2008.** Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387–390.
- Min J.W., Haegeman G., Ysebaert M. et al. 1972.** Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* 237:82–88.
- Miryeganeh M., Marlétaz F., Gavriouchkina D. et al. 2022.** *De novo* genome assembly and in natura epigenomics reveal salinity-induced DNA methylation in the mangrove tree *Bruguiera gymnorhiza*. *New Phytologist* 233(5):2094–2110.
- Misevic G. 2021.** Single-cell omics analyses with single molecular detection: challenges and perspectives. *J Biomed Res* 35(4):264.
- Mitchell K.J., Cheney N. 2025.** The genomic code: The genome instantiates a generative model of the organism. *Trends in Genetics* (in Press) doi.org/10.1016/j.tig.2025.01.008.
- Mitchell K.J., Rawlence N.J. 2021.** Examining natural history through the lens of palaeogenomics. *Trends Ecol Evol* 36:258–267.
- Moguel B., Bobes R.J., Carrero J.C. et al. 2015.** Transfection of platyhelminthes. *BioMed Res Int* 2015:ID 206161.
- Mohr J.L. 1952.** Protozoa as indicators of pollution. *The Scientific Monthly* 74(1):79.
- Montague M.J., Li G., Gandolfi B. et al. 2014.** Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *PNAS* 111(48):17230–17235.
- Montero C.I., Shea Y.R., Jones P.A. et al. 2008.** Evaluation of pyrosequencing technology for the identification of clinically relevant non-dematiaceous yeasts and related species. *Eur J Clin Microbiol Infect Dis* 27:821–830.
- Moody E.R., Álvarez-Carretero S., Mahendrarajah T.A. et al. 2024.** The nature of the last universal common ancestor and its impact on the early Earth system. *Nat Ecol Evol* 8(9):1654–1666.

- Moran B.M., Payne C., Langdon Q. et al. 2021.** The genomic consequences of hybridization. *eLife* 10:e69016.
- Moran Y., Genikhovich G., Gordon D. et al. 2012.** Neurotoxin localization to ectodermal gland cells uncovers an alternative mechanism of venom delivery in sea anemones. *Proc Royal Soc B: Biol Sci* 279(1732):1351–1358.
- Moreno-Ayala R., Junker J.P. 2021.** Single-cell genomics to study developmental cell fate decisions in zebrafish. *Brief Funct Genomics* 20(6):420–426.
- Moreno-Hagelsieb G., Treviño V., Pérez-Rueda E. et al. 2001.** Transcription unit conservation in the three domains of life: a perspective from *Escherichia coli*. *Trends Genet* 17(4):175–177.
- Morrell P.L., Buckler E.S., Ross-Ibarra J. 2012.** Crop genomics: advances and applications. *Nat Rev Genet* 13(2):85–96.
- Moss M.J., Chamness L.M., Clark P.L. 2023.** The effects of codon usage on protein structure and folding. *Annu Rev Biophys* 53(1):87–108.
- Mousa W.K., Chehadeh F., Husband S. 2022.** Recent advances in understanding the structure and function of the human microbiome. *Front Microbiol* 13:825338.
- Muhammad I.I., Kong S.L., Abdullah S.N.A. et al. 2020.** RNA-seq and ChIP-seq as complementary approaches for comprehension of plant transcriptional regulatory mechanism. *Int J Mol Sci* 21:167.
- Mühlemann B., Vinner L., Margaryan A. et al. 2020.** Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science* 369(6502):eaaw8977.
- Mullis K., Scharf S., Saiki R. et al. 1986.** Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb Sympos Quant Biol* 51(1):263–273.
- Murat F., Peer Y.V.D., Salse J. 2012.** Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. *Genome Biology and Evolution* 4(9):917–928.
- Myers E.A., Pyron R.A. 2024.** The first complete assembly for a lungless urodelan with a »miniaturized« genome, the Northern Dusky Salamander (Plethodontidae: *Desmognathus fuscus*). *bioRxiv* 2024.04.30.591895.
- Nagabhushana A., Mishra R. K. 2016.** Finding clues to the riddle of sex determination in zebrafish. *J Biosci* 41:145–155.
- Naqi A.A.T., Fatima K., Mohammad T. et al. 2020.** Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim Biophys Acta Mol Basis Dis* 1866(10):165878.
- Nam N.N., Do H.D.K., Loan T.K.T., Lee N.Y. 2023.** Metagenomics: An effective approach for exploring microbial diversity and functions. *Foods* 12(11):2140.
- Natola L., Seneviratne S. S., Irwin D. 2022.** Population genomics of an emergent tri-species hybrid zone. *Mol Ecol* 31(20):5356–5367.
- Navarro F.C., Mohsen H., Yan C. et al. 2019.** Genomics and data science: an application within an umbrella. *Genome Biol* 20(1):109.
- Nedoluzhko A., Sharko F., Tsygankova S. et al. 2022.** Intergeneric hybridization of two stickleback species leads to introgression of membrane-associated genes and invasive TE expansion. *Front Genet* 13:863547.
- Nguyen N.L. 2022.** Ancient environmental genomics: An Introduction. *Authorea*. DOI:10.22541/au.166012157.74707420/v2
- Nielsen E.S., Begger M., Henriques R. et al. 2020a.** A comparison of genetic and genomic approaches to represent evolutionary potential in conservation planning. *Biol Conserv* 251:108770.

- Nielsen E.S., Henriques R., Beger M. et al. 2020b.** Multi-model seascape genomics identifies distinct environmental drivers of selection among sympatric marine species. *BMC Evol Biol* 20:121.
- Niemann H., Petersen B. 2016.** The production of multi-transgenic pigs: update and perspectives for xenotransplantation. *Transgenic Res* 25:361–374.
- Niiranen L., Leciej D., Edlund H. et al. 2022.** Epigenomic modifications in modern and ancient genomes. *Genes* 13(2):178.
- Nikolakakis Z.L., Schield D.R., Westfall A K. et al. 2022.** Evidence that genomic incompatibilities and other multilocus processes impact hybrid fitness in a rattlesnake hybrid zone. *Evolution* 76(11):2513–2530.
- Ning C., Wang C. C., Gao S. et al. 2019.** Ancient genomes reveal Yamnaya-related ancestry and a potential source of Indo-European speakers in Iron Age Tianshan. *Curr Biol* 29(15): 2526–2532.
- Nirenberg M. 2004.** Historical review: Deciphering the genetic code – a personal account. *Trends Biochem Sci* 29(1):46–54.
- Nishihara H. 2019.** Retrotransposons spread potential cis-regulatory elements during mammary gland evolution. *NUCLEIC ACIDS RES* 47(22):11551–11562.
- Niu B., Paulson J.N., Zheng X., Kolter R. 2017.** Simplified and representative bacterial community of maize roots. *PNAS* 114:2450–2459.
- Niu W., Lu Z.J., Zhong M. et al. 2011.** Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*. *Genome Res* 21:245–254.
- Nowell R. W., Wilson C. G., Almeida P. et al. 2021.** Evolutionary dynamics of transposable elements in bdelloid rotifers. *eLife* 10:e63194.
- Nowoshilow S., Schloissnig S., Fei J.F. et al. 2018.** The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554(7690):50–55.
- Nurk S., Koren S., Rhie A. et al. 2022.** The complete sequence of a human genome. *Science* 376(6588):44–53.
- Nyrén P. 2015.** The history of pyrosequencing. *Methods Mol Biol* 1315:3–15.
- Obbard D.J., Gordon K.H., Buck A.H. et al. 2009.** The evolution of RNAi as a defence against viruses and transposable elements. *Phil Trans R Soc B Biol Science*. 364:99–115.
- O’Brien S.J., Stanyon R. 1999.** Ancestral primate viewed. *Nature* 402(6760):365–366.
- Oey H., Zakrzewski M., Gravermann K. et al. 2019a.** Whole-genome sequence of the bovine blood fluke *Schistosoma bovis* supports interspecific hybridization with *S. haematobium*. *PLoS Pathog* 15(1):e1007513.
- Oey H., Zakrzewski M., Narain K. et al. 2019b.** Whole-genome sequence of the oriental lung fluke *Paragonimus westermani*. *Gigascience* 8:1–8.
- Olalde I., Mallick S., Patterson N. et al. 2019.** The genomic history of the Iberian Peninsula over the past 8000 years. *Science* 363(6432):1230–1234.
- Orlando L., Gilbert M. T., Willerslev E. 2015.** Reconstructing ancient genomes and epigenomes. *Nat Rev Genet* 16(7):395–408.
- Ostrov N., Nyerges A., Chiappino-Pepe A. et al. 2020.** Synthetic genomes with altered genetic codes. *Curr Opin Syst Biol* 24:32–40.
- Pääbo S. 1985.** Preservation of DNA in Ancient Egyptian Mummies. *J Archeol Sci* 12:411–417.
- Pachiadaki M. G., Sintés E., Bergauer K. et al. 2017.** Major role of nitrite-oxidizing bacteria in dark ocean carbon fixation. *Science* 358(6366):1046–1051.

- Pais A.L., Whetten R.W., Xiang Q.Y. 2020.** Population structure, landscape genomics, and genetic signatures of adaptation to exotic disease pressure in *Cornus florida* L. Insights from GWAS and GBS data. *J Syst Evol* 58:546–570.
- Palkopoulou E., Mallick S., Skoglund P. et al. 2015.** Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr Biol* 25(10):1395–1400.
- Palmer S.A., Moore J.D., Clapham A.J. et al. 2009.** Archaeogenetic evidence of ancient Nubian barley evolution from six to two-row indicates local adaptation. *PLoS One* 4(7):e6301.
- Palmer S.A., Clapham A.J., Rose P. et al. 2012.** Archaeogenomic evidence of punctuated genome evolution in *Gossypium*. *Mol Biol Evol* 29(8):2031–2038.
- Pandey M., Singhal B. 2021.** Metagenomics: adding new dimensions in bioeconomy. *Biomass Conv Bioref* 13:7461–7480.
- Park J.C., Bae S. 2024.** Current status of genome editing technologies: special issue of BMB Reports in 2024. *BMB Rep* 57(1):1.
- Parvathy S.T., Udayasuriyan V., Bhadana V. 2022.** Codon usage bias. *Mol Biol Rep* 49(1): 539–565.
- Patterson N., Moorjani P., Luo Y. et al. 2012.** Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Pavlopoulos G.A., Oulas A., Iacucci E. et al. 2013.** Unraveling genomic variation from next generation sequencing data. *BioData Mining* 6:13.
- Pease J. B., Hahn M. W. 2015.** Detection and polarization of introgression in a five-taxon phylogeny. *Systematic biology*, 64(4), 651–662.
- Pecoraro C., Babbucci M., Franch R. et al. 2018.** The population genomics of yellowfin tuna (*Thunnus albacares*) at global geographic scale challenges current stock delineation. *Sci Rep* 8:13890.
- Pecoraro C., Zudair I., Galimbert G. et al. 2020.** When size matters: The gonads of larger female yellowfin tuna (*Thunnus albacares*) have different fatty acid profiles compared to smaller individuals. *Fish Res* 232:105726.
- Pei L., Li G., Lindsey K. et al. 2021.** Plant 3D genomics: the exploration and application of chromatin organization. *New Phytol* 230:1772–1786.
- Pellicer J., Fay M.F., Leitch I.J. 2010.** The largest eukaryotic genome of them all? *Bot J Linn Soc* 164(1):10–15.
- Peltzer A., Jäger G., Herbig A. et al. 2016.** EAGER: efficient ancient genome reconstruction. *Genome Biol* 17:60.
- Pervez M. T., Abbas S. H., Moustafa M. F. et al. 2022.** A comprehensive review of performance of next-generation sequencing platforms. *BioMed Res Int* 2022:3457806.
- Peterson K.J., Beavan A., Chabot P.J. et al. 2022.** MicroRNAs as indicators into the causes and consequences of whole-genome duplication events. *Mol Biol Evol* 39(1):msab344.
- Pevsner J. 2009.** *Bioinformatics and Functional Genomics* (2nd ed.). 1160 pp. Hoboken, NY: Wiley-Blackwell.
- Pezzino S., Sofia M., Greco L.P. et al. 2023.** Microbiome dysbiosis: a pathological mechanism at the intersection of obesity and glaucoma. *Int J Mol Sci* 24(2):1166.
- Phillips P.C. 2008.** Epistasis the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9:855–867.
- Pickrell J., Pritchard J. 2012.** Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* 8:e1002967.

- Pinakhina D., Chekunova E. 2020.** The conceptual landscape of ancient DNA research: Current state and future prospects. *Int J Conserv Sci* 11(3):703–714.
- Pincheira-Donoso D., Harvey L.P., Johnson J.V. et al. 2023.** Genome size does not influence extinction risk in the world's amphibians. *Functional Ecology* 37(1):190–200.
- Pinho C., Hey J. 2010.** Divergence with gene flow: models and data. *Annu Rev Ecol Evol Syst* 41:215–230.
- Pitaksakulrat O., Webster B.L., Webster J.P. et al. 2018.** Phylogenetic relationships within the *Opisthorchis viverrini* species complex with specific analysis of *O. viverrini* sensu lato from Sakon Nakhon, Thailand by mitochondrial and nuclear DNA sequencing. *Infect Genet Evol* 62:86–94.
- Plomin R., Spinath F.M. 2004.** Intelligence: genetics, genes, and genomics. *J Pers Soc Psychol* 86(1):112–129.
- Plotkin J.B., Kudla G. 2011.** Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 12:32–42.
- Poelchau M.F., Coates B.S., Childers C.P. et al. 2016.** Agricultural applications of insect ecological genomics. *Curr Opin Insect Sci* 13:61–69.
- Poelstra J.W., Vijay N., Bossu C.M. et al. 2014.** The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344(6190):1410–1414.
- Pont C., Wagner S., Kremer A. et al. 2019.** Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biol* 20(1):1–17.
- Poorolajal J. 2021.** The global pandemics are getting more frequent and severe. *J Res Health Sci* 21(1): e00502
- Pop M. 2009.** Genome assembly reborn: recent computational challenges. *Brief Bioinform* 10(4):354–366.
- Portik D.M., Feng X., Benoit G. et al. 2024.** Highly accurate metagenome-assembled genomes from human gut microbiota using long-read assembly, binning, and consolidation methods. *bioRxiv* 2024.05.10.593587.
- Posth C., Zaro V., Spyrou M. A. 2021.** The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. *Sci Adv* 7(39):eabi7673.
- Prado-Martinez J., Sudmant P.H., Kidd J.M. et al. 2013.** Great ape genetic diversity and population history. *Nature* 499(7459):471–475.
- Prasad A., Chirom O., Prasad M. 2021.** Insect herbivores benefit from horizontal gene transfer. *Trends Plant Sci* 26(11):1096–1097.
- Pritchard J.K., Stephens, P. Donnelly. 2000.** Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Prüfer K., Racimo F., Patterson N. et al. 2014.** The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481): 43–49.
- Qiao Q., Edger P.P., Xue L. et al. 2021.** Evolutionary history and pan-genome dynamics of strawberry (*Fragaria* spp.). *PNAS* 118(45):e2105431118.
- Qin J., Li R., Raes J. et al. 2010.** A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65.
- Qu Y., Shang X., Zeng Z. et al. 2023.** Whole-genome duplication reshaped adaptive evolution in a relict plant species, *Cyclocarya paliurus*. *Genomics Proteomics Bioinformatics* 21(3):455–469
- Raghavan M., Skoglund P., Graf K. E. 2014.** Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.

- Raghavan M., Steinrücken M., Harris K. et al. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250):aab3884.
- Rai A., Yamazaki M., Saito K. 2019. A new era in plant functional genomics. *Curr Opin Syst Biol* 15:58–67.
- Rainha J., Rodrigues J.L., Rodrigues L.R. 2021. CRISPR-Cas9: A powerful tool to efficiently engineer *Saccharomyces cerevisiae*. *Life* 11(1):13
- Rajput M., Choudhary K., Kumar M. et al. 2021. RNA interference and CRISPR/Cas gene editing for crop improvement: Paradigm shift towards sustainable agriculture. *Plants* 10(9):1914.
- Ramirez A. H., Sulieman L., Schlueter D. J. et al. 2022. The All of Us Research Program: data quality, utility, and diversity. *Patterns* 3(8):100570.
- Ramos–Madrigal J., Smith B.D., Moreno–Mayar J.V. et al. 2016. Genome sequence of a 5,310–year–old maize cob provides insights into the early stages of maize domestication. *Curr Biol* 26(23):3195–3201.
- Rand A.C., Jain M., Eizenga J.M. et al. 2017. Mapping DNA methylation with high throughput nanopore sSequencing. *Nat Methods* 14(4):411–413.
- Rasmussen S., Allentoft M.E., Nielsen K. et al. 2015. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* 163(3):571–582.
- Razin S.V., Ulianov S.V. 2020. Divide and rule: phase separation in eukaryotic genome functioning. *Cells* 9:2480.
- Redmond A.K., Casey D., Gundappa M.K. et al. 2023. Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. *Nat Commun* 14(1):2879.
- Rellstab C., Gugerli F., Eckert A.J. et al. 2015. A practical guide to environmental association analysis in landscape genomics. *Mol Ecol* 24:4348–4370.
- Rey-Iglesia A., Gopalakrishnan S., Carøe C. et al. 2019. MobiSeq: De novo SNP discovery in model and non-model species through sequencing the flanking region of transposable elements. *Mol Ecol Resour* 19(2):512–525.
- Rice E.S., Green R.E. 2019. New approaches for genome assembly and scaffolding. *Annu Rev Anim Biosci* 7(1):17–40.
- Richard G.F. 2020. Eukaryotic pangenomes. In: Tettelin H, Medini D (Eds) *The Pangenome: Diversity, Dynamics and Evolution of Genomes* [Internet].pp. 253–291 Cham (CH): Springer.
- Rico Y. 2021. A comprehensive landscape genomics approach for seed sourcing strategies in landscapes under varying degrees of habitat disturbance. *Mol Ecol Resour* 21(1):14–17.
- Rhie A., Nurk S., Cechova M. et al. 2023. The complete sequence of a human Y chromosome. *Nature* 621(7978):344–354.
- Rohl C., Strauss C., Misura K., Baker D. 2004. Protein structure prediction using Rosetta. *Methods Enzymol* 383:66–93.
- Romesberg F.E. 2023. Discovery, implications and initial use of semi-synthetic organisms with an expanded genetic alphabet/code. *Philos Trans R Soc B* 378(1871):20220030.
- Ronaghi M. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11:3–11.
- Rosser N., Edelman N. B., Queste L.M. et al. 2022. Complex basis of hybrid female sterility and Haldane’s rule in *Heliconius* butterflies: Z-linkage and epistasis. *Mol Ecol* 31(3): 959–977.
- Roth F.P., Lipshitz H.D., Andrews B.J. 2009. Q&A: epistasis. *J Biol* 8:35.
- Roth S.C. 2019. What is genomic medicine? *JMLA* 107(3):442.

- Rowley M.J., Corces V.G. 2018.** Organizational principles of 3D genome architecture. *Nat Rev Genet* 19(12):789–800.
- Rubin C.J., Megens H.J., Barrio A.M. et al. 2012.** Strong signatures of selection in the domestic pig genome. *PNAS* 109(48):19529–19536.
- Runemark A., Trier C., Eroukhmanoff F. et al. 2018.** Variation and constraints in hybrid genome formation. *Nat Ecol Evol* 2(3):549–556.
- Runemark A., Vallejo-Marin M., Meier J.I. 2019.** Eukaryote hybrid genomes. *PLoS Genet* 15(11):e1008404.
- Rusk N. 2011.** Torrents of sequence. *Nat Methods* 8(1):44.
- Sáez M.T., Hofreiter M., Straube N. 2024.** Shark genome size evolution and its relationship with cellular, life-history, ecological, and diversity traits. *Sci Rep* 14(1):8909.
- San Juan E., Araya-Donoso R., Véliz D. et al. 2021.** Genetic diversity in a restricted-dispersal kissing bug: The centre–periphery hypothesis halfway. *Mol Ecol* 30(19):4660–4672.
- Sanabria M., Hirsch J., Joubert P.M., Poetsch A.R. 2024.** DNA language model GROVER learns sequence context in the human genome. *Nat Mach Intell* 6:911–923.
- Sandler R. 2014.** The ethics of reviving long extinct species. *Conserv Biol* 28(2):354–360.
- Sanger F., Nicklen S., Coulson A.R. 1977.** DNA sequencing with chain-terminating inhibitors. *PNAS* 74(12):5463–5467.
- Sanger F., Thompson E.O.P. 1953.** The amino-acid sequence in the glycol chain of insulin. II. The investigation of peptides from enzymic hydrolysates. *Biochem J* 53:366–374.
- Satam H., Joshi K., Mangrolia U. et al. 2023.** Next-generation sequencing technology: current trends and advancements. *Biology* 12(7):997.
- Saxena K.B., Sultana R., Gangashetty P.I. et al. 2024.** Population breeding in Pigeonpea: a review. *Legume Res* 47(9):1437–1445.
- Schaack S., Gilbert C., Feschotte C. 2010.** Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol* 25:537–546.
- Schaefer N.K., Shapiro B., Green R.E. 2021.** An ancestral recombination graph of human, Neanderthal, and Denisovan genomes. *Sci Adv* 7(29):eabc0776.
- Schartl M., Woltering J. M., Irisarri I. et al. 2024.** The genomes of all lungfish inform on genome expansion and tetrapod evolution. *Nature* 634(8032):96–103.
- Schiml S., Puchta H. 2016.** Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. *Plant Meth* 12(1):8.
- Schindler D. 2020.** Genetic engineering and synthetic genomics in yeast to understand life and boost biotechnology. *Bioengineering* 7:137.
- Schindler D., Dai J., Cai Y. 2018.** Synthetic genomics: a new venture to dissect genome fundamentals and engineer new functions. *Curr Opin Chem Biol* 46:5662.
- Schlumbaum A., Tensen M., Jaenicke–Després V. 2008.** Ancient plant DNA in archaeobotany. *Veg Hist Archaeobot* 17(2):233–244.
- Schmidt T.L., Filipović I., Hoffmann A. A., Rašić G. 2018.** Fine-scale landscape genomics helps explain the slow spatial spread of *Wolbachia* through the *Aedes aegypti* population in Cairns, Australia. *Heredity* 120:386–395.
- Schubert M., Ermini L., Sarkissian C.D. et al. 2014.** Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc* 9(5):105682.

- Schuenemann V.J., Singh P., Mendum T.A. et al. 2013. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 341(6142):179–183.
- Schwabl P., Llewellyn S., Landguth E.L. et al. 2017. Prediction and prevention of parasitic diseases using a landscape genomics framework. *Trends Parasitol* 33:264–275.
- Schwörer C., Leunda M., Alvarez N. et al. 2022. The untapped potential of macrofossils in ancient plant DNA research. *New Phytol* 235(2):391.
- Szczesnak A., Segata N., Qin X. et al. 2011. The genome of th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe* 10(3):260–272.
- Seib K.L., Zhao X., Rappuoli R. 2012. Developing vaccines in the era of genomics: a decade of reverse vaccinology. *Clin Microbiol Infect* 18:109–116.
- Seoighe C., Gehring C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends in Genetics* 20(10):461–464.
- Session A. M., Rokhsar D. S. 2023. Transposon signatures of allopolyploid genome evolution. *Nat Commun* 14(1):3180.
- Setubal J.C. 2021. Metagenome-assembled genomes: concepts, analogies, and challenges. *Biophys Rev* 13:905–909.
- Shalem O., Sanjana N.E., Hartenian E. et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343(6166):84–87.
- Shapiro B., Hofreiter M. 2014. A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science* 343(6169):1236573.
- Shapiro J.A. 2023. Engines of innovation: biological origins of genome evolution. *Biol J Linn Soc* 139(4):441–456.
- Sharko F.S., Boulygina E.S., Tsygankova S.V. et al. 2021. Steller's sea cow genome suggests this species began going extinct before the arrival of Paleolithic humans. *Nat Commun* 12(1):2215.
- Sharma P.P. 2023. The impact of whole genome duplication on the evolution of the Arachnids. *Integr Comp Biol* 63(3):825842.
- Shendure J., Findlay G.M., Snyder M.W. 2019. Genomic medicine—progress, pitfalls, and promise. *Cell* 177(1):45–57.
- Sheng G.L., Basler N., Ji X.P. et al. 2019. Paleogenome reveals genetic contribution of extinct giant panda to extant populations. *Current Biol* 29(10):1695–1700.
- Shergalis A., Bankhead III A., Luesakul U. et al. 2018. Current challenges and opportunities in treating glioblastoma. *Pharmacol Rev* 70(3):412.
- Shi Y.M., Hirschmann M., Shi Y.N. et al. 2022. Global analysis of biosynthetic gene clusters reveals conserved and unique natural products in entomopathogenic nematode-symbiotic bacteria. *Nat Chem* 14(6):701–712.
- Shin H., Park J.E., Park H.R. et al. 2022. Admixture of divergent genomes facilitates hybridization across species in the family Brassicaceae. *New Phytol* 235(2):743–758.
- Shryock D.F., Washburn L., DeFalco L.A. et al. 2021. Harnessing landscape genomics to identify future-climate resilient genotypes in a desert annual. *Mol Ecol* 30:698–717.
- Signore A.V., Fago A., Weber R. et al. 2023. Evolution of an extreme hemoglobin phenotype contributed to the sub-Arctic specialization of extinct Steller's sea cows. *eLife* 12:e85414.
- Simakov O., Bredeson J., Berkoff K. et al. 2022. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci Adv* 8(5):eabi5884.

- Simonti C.N., Lachance J. 2021.** Ancient DNA reveals that few GWAS loci have been strongly selected during recent human history. *bioRxiv* 2021.04.13.439742.
- Singh A., Ramakrishna G., Kaila T. et al. 2022.** Next-generation sequencing technologies: approaches and applications for crop improvement. In: Wani S.H., Kumar A. (eds) *Genomics of Cereal Crops. Springer Protocols Handbooks*. pp. 31–94. Humana, New York.
- Singh R., Ming R., Yu Q. 2016.** Comparative analysis of GC content variations in plant genomes. *Trop Plant Biol* 9:136–149.
- Slatko B.E., Gardner A.F., Ausubel F.M. 2018.** Overview of next generation sequencing technologies. *Curr Protoc Mol Biol* 122(1):e59.
- Slon V., Mafessoni F., Vernot B. et al. 2018.** The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* 561:113–117.
- Slyusarev G.S., Starunov V.V., Bondarenko A.S. et al. 2020.** Extreme genome and nervous system streamlining in the invertebrate parasite *Intoshia variabili*. *Curr Biol* 30(7):1292–1298.
- Smith C.F., Nikolakis Z.L., Perry B.W. et al. 2023.** The best of both worlds? Rattlesnake hybrid zones generate complex combinations of divergent venom phenotypes that retain high toxicity. *Biochimie* 213:176–189.
- Smith L.M., Sanders J.Z., Kaiser R.J. et al. 1986.** Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674–679.
- Smoller J.W., Andreassen O.A., Edenberg H.J. et al. 2019.** Psychiatric Genetics and the Structure of Psychopathology. *Mol Psychiatry* 24(3):409–420.
- Sohn J.I., Nam J. W. 2018.** The present and future of de novo whole-genome assembly. *Brief Bioinform* 19(1):23–40.
- Song H., Béthoux O., Shin S. et al. 2020.** Phylogenomic analysis sheds light on the evolutionary pathways towards acoustic communication in Orthoptera. *Nat Commun* 11(1):4939.
- Souilmi Y., Lauterbur M. E., Tobler R. et al. 2021.** An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. *Curr Biol* 31(16): 3504–3514.
- Sproul J.S., Hotaling S., Heckenhauer J. et al. 2023.** Analyses of 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges. *Genome Res* 33(10): 1708–1717.
- Spyrou M.A., Bos K.I., Herbig A. et al. 2019.** Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat Rev Genet* 20(6):323–340.
- Staden R. 1979.** A strategy of DNA sequencing employing computer programs. *Nucl Acids Res* 6:2601–2610.
- Steindorff A.S., Carver A., Calhoun S. et al. 2021.** Comparative genomics of pyrophilous fungi reveals a link between fire events and developmental genes. *Environ Microbiol* 23(1):99–109.
- Stelzer C.P., Blommaert J., Waldvogel A.M. et al. 2021a.** Comparative analysis reveals within-population genome size variation in a rotifer is driven by large genomic elements with highly abundant satellite DNA repeat elements. *BMC Biol* 19:117.
- Stelzer C.P., Pichler M., Hatheuer A. 2021b.** Linking genome size variation to population phenotypic variation within the rotifer, *Brachionus asplanchnoidis*. *Commun Biol* 4(1):596.
- Stern D.B., Anderson N.W., Diaz J.A., Lee C.E. 2022.** Genome-wide signatures of synergistic epistasis during parallel adaptation in a Baltic Sea copepod. *Nat Commun* 13(1):4024.
- Stiller J., Zhang G. 2019.** Comparative phylogenomics, a stepping stone for bird biodiversity studies. *Diversity* 11(7):115.

- Storfer A., Antolin M.F., Manel S. et al. 2016.** Genomic approaches in landscape genetics. In: Balkenhol N., Cushman S., Storfer A., Waits L.(eds.) *Landscape Genetics: Concepts, Methods, Applications*. West Sussex. Wiley, pp. 249–164.
- Suda K., Hayashi S. R., Tamura K. et al. 2022.** Activation of DNA transposons and evolution of piRNA genes through interspecific hybridization in *Xenopus* Frogs. *Front Genet* 13:766424.
- Sun L., Ma X., Zhang B. et al. 2022a.** From polymerase engineering to semi-synthetic life: artificial expansion of the central dogma. *RSC Chem Biol* 3(10):1173–1197.
- Sun X., Liu Y. C., Tiunov M.P. et al. 2022b.** Ancient DNA reveals China as a historical genetic melting pot in tiger evolution. *bioRxiv* 2022.09.14.507899.
- Sun X., Liu Y. C., Tiunov M. P. et al. 2023a.** Ancient DNA reveals genetic admixture in China during tiger evolution. *Nat Ecol Evol* 7(11):1914–1929.
- Sun Z., Peng J., Lv Q. et al. 2023b.** Dissecting the genetic basis of heterosis in elite super-hybrid rice. *Plant Physiol* 192(1):307–325.
- Swain M.T., Tsai L.J., Assefa S.A. et al. 2012.** A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protocols* 7(7):1260–1284.
- Tao Y., Luo H., Xu J. et al. 2021.** Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat Plant* 7(6):766–773.
- Tao Y., Zhao X., Mace E. et al. 2019.** Exploring and exploiting pan-genomics for crop improvement. *Mol Plant* 12(2):156–169.
- Terwilliger T.C., Stuart D., Yokoyama S. 2009.** Lessons from structural genomics. *Annu Rev Biophys* 38:371–383.
- Tettelin H., Masignani V., Cieslewicz M.J. et al. 2005.** Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *PNAS* 102(39):13950–13955.
- Thalmann O., Perri A.R. 2018.** Paleogenomic inferences of dog domestication. In: Ch. Lindqvist, O.P. Rajora (eds.) *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, pp. 273–306. Springer Cham.
- Thao T.T.N., Labroussaa F., Ebert N. et al. 2020.** Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* 582:561565.
- The *C. elegans* Sequencing Consortium. 1998.** Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282(5396):2012–2018.
- The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. 2020.** *Nature* 578(7793):82–93.
- Thomas M.C., Macias F., Alonso C., López M.C. 2010.** The biology and evolution of transposable elements in parasites. *Trends Parasitol* 26:350–362.
- Thompson R.C.A., Conlan J.V. 2011.** Emerging issues and parasite zoonoses in the SE Asian and Australasian region. *Vet Parasitol* 181:69–73.
- Timofeeff-Ressovsky N.W., Zimmer K.G., Delbrück M. 1935.** Über die Natur der Genmutation und der Genstruktur, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. *Mathematische-Physikalische Klasse, Fachgruppe VI, Biologie* 13:189–245.
- Torri A., Jaeger J., Pradeu T., Saleh M. C. 2022.** The origin of RNA interference: Adaptive or neutral evolution? *PloS Biol* 20(6):e3001715.
- Touchman J. 2010.** Comparative genomics. *Nat Educ Knowl* 3(10):13.
- Tranchant-Dubreuil C., Rouard M., Sabot F. 2018.** Plant pangenome: impacts on phenotypes and evolution. *Annu Plant Rev online* 2(2):453–478.

- Tsai I.J., Zaroweiecki M., Holroyd N. et al. 2013.** The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496:57–63.
- Tusso S., Suo F., Liang Y. et al. 2022.** Reactivation of transposable elements following hybridization in fission yeast. *Genome Res* 32(2):324–336.
- Tyagi S., Kabade P.G., Gnanapragasam N. et al. 2023.** Codon usage provide insights into the adaptation of rice genes under stress condition. *Int J Mol Sci* 24(2):1098.
- Ulianov S.V., Razin S.V. 2021.** The two waves in single-cell 3D genomics. *Semin Cell Dev Biol* 121:143–152.
- Undheim E.A., Jenner R.A. 2021.** Phylogenetic analyses suggest centipede venom arsenals were repeatedly stocked by horizontal gene transfer. *Nat Commun* 12(1):818.
- Vahidy F.S., Nicolas J.C., Meeks J.R. et al. 2020.** Racial and ethnic disparities in SARS-CoV-2 pandemic: analysis of a COVID-19 observational registry for a diverse US metropolitan population. *BMJ Open* 10(8):e039849.
- Vakhrusheva O. A., Mnatsakanova E. A., Galimov Y. R. et al. 2020.** Genomic signatures of recombination in a natural population of the bdelloid rotifer *Adineta vaga*. *Nat Commun* 11(1):6421.
- Valiskova B., Gregorova S., Lustyk D. et al. 2022.** Genic and chromosomal components of Prdm9-driven hybrid male sterility in mice (*Mus musculus*). *Genetics* 222(1):iyac116.
- Vallebuena–Estrada M., Rodríguez–Arévalo I., Rougon–Cardoso A. et al. 2016.** The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *PNAS* 113(49):14151–14156.
- Van Damme K., Cornetti L., Fields P.D., Ebert D. 2022.** Whole-genome phylogenetic reconstruction as a powerful tool to reveal homoplasy and ancient rapid radiation in waterflea evolution. *Syst Biol* 71(4):777–787.
- Van Der Valk T., Pečnerová P., Díez-del-Molino D. et al. 2021.** Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* 591(7849):265–269.
- Van El C.G., Cornel M.C., Borry P. et al. 2013.** Whole-genome sequencing in health care. *Eur J Hum Genet* 21(6):580–584.
- Varadi M., Velankar S. 2023.** The impact of AlphaFold Protein Structure Database on the fields of life sciences. *Proteomics* 23(17):2200128.
- Vartanyan S.L., Arslanov K.A., Karhu J.A. et al. 2008.** Collection of radiocarbon dates on the mammoths (*Mammuthus primigenius*) and other genera of Wrangel Island, northeast Siberia, Russia. *Quat Res* 70(1):51–59.
- Velandia-Huerto C.A., Yazbeck A.M., Schor J., Stadler P.F. 2022.** Evolution and phylogeny of microRNAs—protocols, pitfalls, and problems. *Methods Mol Biol* 2257:211–233.
- Venancio T.M., Wilson R.A., Verjovski-Almeida S., DeMarco R. 2010.** Burst of transposition from non-long terminal repeat retrotransposon families of the RTE clade in *Schistosoma mansoni*. *Int J Parasitol* 40:743–749.
- Venter J.C. 2003.** A part of the human genome sequence. *Science* 299(5610):1183–1184.
- Venter J.C., Adams M.D., Myers E.W. et al. 2001.** The sequence of the human genome. *Science* 291:1304–1351.
- Vernot B., Stergachis A.B., Maurano M.T. et al. 2012.** Personal and population genomics of human regulatory variation. *Genome Res* 22:1689–1697.
- Vicario S., Mason C.E., White K.P., Powell J.R. 2008.** Developmental stage and level of codon usage bias in *Drosophila*. *Mol Biol Evol* 25:2269–2277.

- Vollger M.R., Dishuck P.C., Harvey W.T. et al. 2023. Increased mutation and gene conversion within human segmental duplications. *Nature* 617(7960): 325–334.
- Vohnoutová M., Sedláková A., Symonová R. 2023. Abandoning the isochore theory can help explain genome compositional organization in fish. *Int J Mol Sci* 24(17):13167.
- Waddington C.H. 1942. The epigenotype. *Endeavour* 1:18–20.
- Wagne H.H., Forti M.J. 2016. Basics of spatial data analysis: linking landscape and genetic data for landscape genetic studies. In: Balkenhol N., Cushman S., Storfer A., Waits L. (eds). *Landscape Genetics: Concepts, Methods, Applications*. West Sussex, Wiley. pp. 77–98.
- Wagner D.N., Curry R.L., Chen N. et al. 2020. Genomic regions underlying metabolic and neuronal signaling pathways are temporally consistent in a moving avian hybrid zone. *Evolution* 74(7):1498–1513.
- Wagner D.M., Klunk J., Harbeck M. et al. 2014. *Yersinia pestis* and the plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect Dis* 14(4):319–326.
- Waldvogel A.-M., Feldmeyer B., Rolshausen G. et al. 2020. Evolutionary genomics can improve prediction of species' responses to climate change. *Evol Lett* 4:4–18.
- Wallen Z. D., Demirkan A., Twa G. et al. 2022. Metagenomics of Parkinson's disease implicates the gut microbiome in multiple disease mechanisms. *Nat Commun* 13(1):6958.
- Walsh A.M., Kortschak R.D., Gardner M.G. et al. 2013. Widespread horizontal transfer of retrotransposons. *PNAS* 110(3):1012–1016.
- Walsh J., Billerman S. M., Butcher B. G. et al. 2023. A complex genomic architecture underlies reproductive isolation in a North American oriole hybrid zone. *Commun Biol* 6(1):154.
- Wang B., Hou M., Shi J. et al. 2023a. *De novo* genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. *Nat Genetics* 55(2):312–323.
- Wang B., Wu B., Liu X. et al. 2024. Whole genome sequencing reveals autooctoploidy in the Chinese sturgeon and its evolutionary trajectories. *Genom Proteom Bioinform* 22(1):qzad002.
- Wang C.C., Yeh H.Y., Popo A.N. 2021. Genomic insights into the formation of human populations in East Asia. *Nature* 591(7850):413–419.
- Wang H., Sun S., Ge W. et al. 2020. Horizontal gene transfer of Fhb7 from fungus underlies *Fusarium* head blight resistance in wheat. *Science* 368(6493):eaba5435.
- Wang J., Li X., Wang L., Li J. et al. 2016a. A novel long intergenic noncoding RNA indispensable for cleavage of mouse two-cell embryos. *EMBO Rep* 17:1452–1470.
- Wang J., Tian L., Lee H.S. et al. 2006. Genome-wide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172(1):507–517.
- Wang J., Yuan L., Tang J. et al. 2023b. Transposable element and host silencing activity in gigantic genomes. *Fron Cell Dev Biol* 11:1124374.
- Wang L, Jiang S, Chen C. et al. 2018a. Synthetic genomics: from DNA synthesis to genome design. *Angew Chem Int Ed* 57:1748–1756.
- Wang M.S., Murray G.G., Mann D. et al. 2022a. A polar bear paleogenome reveals extensive ancient gene flow from polar bears into brown bears. *Nat Ecol Evol* 6(7):936–944.
- Wang N., Li C., Kuang L. et al. 2022b. Pan-mitogenomics reveals the genetic basis of cytonuclear conflicts in citrus hybridization, domestication, and diversification. *PNAS* 119(43):e2206076119.
- Wang S., Wang S., Luo Y. et al. 2016b. Comparative genomics reveals adaptive evolution of Asian tapeworm in switching to a new intermediate host. *Nat Commun* 7(1):12845.
- Wang T., Guan C., Guo J. et al. 2018b. Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362(6420).

- Wang X., Cai B., Zhou J. et al. 2016c. Disruption of FGF5 in Cashmere goats using CRISPR/Cas9 results in more secondary hair follicles and longer fibers. *PLoS One* 11(10):e0164640.
- Wang X., Chen W., Tian Y. et al. 2014a. Surface display of *Clonorchis sinensis* enolase on *Bacillus subtilis* spores potentializes an oral vaccine candidate. *Vaccine* 32(12):13381345.
- Wang Y., Shen Y., Gu Y. et al. 2018c. Genome writing: current progress and related applications. *Genom Proteom Bioinform* 16:10-16.
- Wang Y., Tang Z., Sun Y. et al. 2014b. Analysis of genome-wide copy number variations in Chinese indigenous and western pig breeds by 60 K SNP genotyping arrays. *PloS One* 9(9): e106780.
- Wang Z., Kang M., Li J. et al. 2022c. Genomic evidence for homoploid hybrid speciation between ancestors of two different genera. *Nat Commun* 13(1):1987.
- Wang Z., Liao B.-Y., Zhang J. 2010. Genomics patterns of pleiotropy and the evolution of complexity. *PNAS* 107(42):18034–18039.
- Waterston R.H., Lindblad-Toh K., Birney E. et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520–562.
- Watson J.D., Crick F.H. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171(4356):737–738.
- Watson M., Stott K. 2019. Disordered domains in chromatin-binding proteins. *Essays in Biochemistry* 63(1):147–156.
- Watters K.E., Kirkpatrick J., Palmer M. J., Koblenz G.D. 2021. The CRISPR revolution and its potential impact on global health security. *Pathog Glob Health* 115(2):80–92.
- Weekes D., Krishna S.S., Constantina B. et al. 2010. TOPSAN: a collaborative annotation environment for structural genomics. *BMC Bioinform* 11: 426.
- Wei X., Qiu J., Yong K. et al. 2021. A quantitative genomics map of rice provides genetic insights and guides breeding. *Nat Genet* 53(2):243–253.
- Weiner S.A., Toth A.L. 2012. Epigenetics in social insects: a new direction for understanding the evolution of castes. *Genet Res Int* 2012:609810.
- Weisrock D.W., Hime P.M., Nunziata S.O. et al. 2021. Surmounting the large-genome “problem” for genomic data generation in salamanders. In: Hohenlohe P.A., Rajora O.P. (eds) *Population Genomics: Wildlife*. pp. 115–142. Springer, Cham.
- Wells J.N., Feschotte C. 2020. A field guide to eukaryotic transposable elements. *Annu Rev Genet* 54(1):539–561.
- Westbury M. V., Hartmann S., Barlow A. et al. 2020. Hyena paleogenomes reveal a complex evolutionary history of cross–continental gene flow between spotted and cave hyena. *Sci Adv* 6(11):eaay0456.
- Whirl-Carrillo M., McDonagh E.M., Hebert J.M. et al. 2012. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 92(4):414–417.
- Whitfill T., Oh J. 2019. Recoding the metagenome: microbiome engineering *in situ*. *Curr Opin Microbiol* 50:28–34.
- Wijayawardena B.K., DeWoody J.A., Minchella D.J. 2015. The genomic proliferation of transposable elements in colonizing populations: *Schistosoma mansoni* in the new world. *Genetica* 143:287–298.
- Wijayawardena B.K., Minchella D.J., DeWoody J.A. 2013. Hosts, parasites, and horizontal gene transfer. *Trends Parasitol* 29:329–338.

- Wilkinson G.S., Adams D.M., Haghani A. et al. 2021.** DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nat Commun* 12(1):1615.
- Willerslev E., Meltzer D.J. 2021.** Peopling of the Americas as inferred from ancient genomics. *Nature* 594(7863):356-364.
- Wit J., Gilleard J.S. 2017.** Resequencing helminth genomes for population and genetic studies. *Trends Parasitol* 33(5):388-399.
- Wolfe J. M., Fournier G. P. 2018.** Horizontal gene transfer constrains the timing of methanogen evolution. *Nat Ecol Evol* 2(5):897-903.
- Wooley J.C., Godzik A., Friedberg I.A. 2010.** Primer on Metagenomics. *PLoS Comput Biol* 6(2): e1000667.
- Wu R., Kaiser A.D. 1968.** Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J Mol Biol* 35:52337.
- Wu Y., Yan Y., Zhao Y. et al. 2021.** Genomic bases underlying the adaptive radiation of core landbirds. *BMC Ecol Evol* 21:114.
- Wu Z., Byrd D.A., Wan Y. et al. 2022.** The oral microbiome and breast cancer and nonmalignant breast disease, and its relationship with the fecal microbiome in the Ghana Breast Health Study. *Int J Cancer* 151(8):1248-1260.
- Wu X., Xu M., Yang J. R., Lu J. 2024.** Genome-wide impact of codon usage bias on translation optimization in *Drosophila melanogaster*. *Nat Commun* 15:8329.
- Wutke S., Andersson L., Benecke N. et al. 2016.** The origin of ambling horses. *Current Biol* 26(15):R697-R699.
- Xiao J., Zhang Z., Wu J., Yu J. 2015.** A Brief Review of Software Tools for Pangenomics. *Genom Proteom Bioinform* 13:73-76
- Xie D., Xu Y., Wang J. et al. 2019.** The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. *Nat Commun* 10:51-58.
- Xiong T., Li X., Yago M., Mallet J. 2022.** Admixture of evolutionary rates across a butterfly hybrid zone. *eLife* 11:e78135.
- Xu D., Zhang Y. 2012.** *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* 80:1715-1735.
- Xu M. J., Liu Q., Nisbet A.J. et al. 2010.** Identification and characterization of microRNAs in *Clonorchis sinensis* of human health significance. *BMC Genom* 11:521.
- Xue Z., Yuan J., Chen, F. et al. 2022.** Genome-wide association meta-analysis of 88,250 individuals highlights pleiotropic mechanisms of five ocular diseases in UK Biobank. *EBioMedicine* 82:104-161.
- Yadav S.P. 2007.** The wholeness in suffix -omics, -omes, and the word Om. *J Biomol Techniques* 18:277.
- Yan C., Shen L.P., Ma R. et al. 2016a.** Characterization and identification of differentially expressed microRNAs during the process of the peribiliary fibrosis induced by *Clonorchis sinensis*. *Infect Genet Evol* 43:321-328.
- Yan M., Man S., Sun B. et al. 2023.** Gut liver brain axis in diseases: the implications for therapeutic interventions. *Signal Transduct Target Ther* 8(1):443.
- Yan M., Wang C., He B. et al. 2016b.** Aurora-A kinase: a potent oncogene and target for cancer therapy. *Med Res Rev* 36(6):1036-1079.
- Yang H., Li Q. 2022.** The DNA methylation level is associated with the superior growth of the hybrid crosses in the Pacific oyster *Crassostrea gigas*. *Aquaculture* 547:737421.

- Yang J., Anishchenko I., Park H. et al. 2020a.** Improved protein structure prediction using predicted interresidue orientations. *PNAS* 117(3):1496–1503.
- Yang M. A., Fan X., Sun B. et al. 2020b.** Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* 369(6501):282–288.
- Yang M., Wang X., Ren D. et al. 2017.** Genomic architecture of biomass heterosis in *Arabidopsis*. *PNAS* 114(30):8101–8106.
- Yang Y., Gao J., Wang J. et al. 2018.** Sixty-five years of the long march in protein secondary structure prediction: the final stretch? *Brief Bioinform* 19(3):482–494.
- Yeoman C.J., Chia N., Yildirim S. et al. 2011.** Towards an evolutionary model of animal-associated microbiomes. *Entropy* 13(3):570–594.
- Yoneji T., Fujita H., Mukai T., Su’etsugu M. 2021.** Grand scale genome manipulation via chromosome swapping in *Escherichia coli* programmed by three one megabase chromosomes. *Nucl Acids Res* 49(15):8407–8418.
- Yoo M-J., Szadkowski.E., Wendel J. F. 2013.** Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110:171–180.
- Yoo W.G., Kim D.W., Ju J.W. et al. 2011.** Developmental transcriptomic features of the carcinogenic liver fluke, *Clonorchis sinensis*. *PLoS Negl Trop Dis* 5(6):e1208.
- Young A.D., Gillung J.P. 2020.** Phylogenomics – principles, opportunities and pitfalls of big-data phylogenetics. *Syst Entomol* 45(2):225–247.
- Yu J., Blomb J., Glaeser S.P. et al. 2017.** A review of bioinformatics platforms for comparative genomics. Recent developments of the EDGAR 2.0 platform and its utility for taxonomic and phylogenetic studies. *J Biotechnol* 261:2–9.
- Yu T., Ning J., Chen M. et al. 2023a.** Potential involvement of DNA methylation in hybrid sterility in hermaphroditic *Argopecten* scallops. *Mar Biotechnol (NY)* 25(5):701–717.
- Yu X., Qin M., Qu M. et al. 2023b.** Genomic analyses reveal dead-end hybridization between two deeply divergent kiwifruit species rather than homoploid hybrid speciation. *Plant J* 115(6):1528–1543.
- Zafra R., Pérez-Écija R.A., Buffoni L. et al. 2013.** Early and late peritoneal and hepatic changes in goats immunized with recombinant cathepsin L1 and infected with *Fasciola hepatica*. *J Comp Pathol* 148(4):373–384.
- Zarubin M., Murugova T., Ryzhykau Y. et al. 2024.** Structural study of the intrinsically disordered tardigrade damage suppressor protein (Dsup) and its complex with DNA. *Sci Rep* 14(1): 22910.
- Zaytsev K., Bogatyreva N.S., Fedorov A.N. 2024.** Exploring individual codon influence on protein expression: a predictive approach. *bioRxiv* 2024.06.30.601029.
- Zeng X., Lin J. 2017.** Factors influencing horizontal gene transfer in the intestine. *Anim Health Res Rev* 18(2):153–159.
- Zewde M., Kiyotani K., Park J.H. et al. 2018.** The era of immunogenomics/immunopharmacogenomics. *J Hum Genet* 63(8):865–875.
- Zhang B.L., Chen W., Wang Z. et al. 2023a.** Comparative genomics reveals the hybrid origin of a macaque group. *Sci Adv* 9(22):eadd3580.
- Zhang B., Pan X., Cobb G.P. Anderson T.A. 2007.** microRNAs as oncogenes and tumor suppressor. *Devel Biol* 302:1–12.
- Zhang C., Zhang T., Luebert F. et al. 2020.** Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Mol Biol Evol* 37(11):3188–3210.

- 
- Zhang L., Chaturvedi S., Nice C.C. et al. 2023b.** Population genomic evidence of selection on structural variants in a natural hybrid zone. *Mol Ecol* 32(6): 1497–1514.
- Zhang L., Chen F., Zeng Z. et al. 2021.** Advances in metagenomics and its application in environmental microorganisms. *Front Microbiol* 12:766364.
- Zhang M., Qiu X. 2023.** Genetic basis of genome size variation of wheat. *Funct Integr Genomics* 23(3):285.
- Zhang Y., de Lima C. B., Labrecque R., Sirard M. A. 2023c.** Whole-genome DNA methylation analysis of the sperm in relation to bull fertility. *Reproduction* 165(5):557–568.
- Zhang Y., Li G. 2020.** Advances in technologies for 3D genomics research. *Sci China Life Sci* 63:811–824.
- Zheng C., Tan L., Sang M. et al. 2020.** Genetic adaptation of Tibetan poplar (*Populus szechuanica* var. *tibetica*) to high altitudes on the Qinghai–Tibetan Plateau. *Ecol Evol* 10(20):10974–10985.
- Zheng Q., Lin J., Huang J. et al. 2017.** Reconstitution of UCP1 using CRISPR/Cas9 in the white adipose tissue of pigs decreases fat deposition and improves thermogenic capacity. *PNAS* 114(45):9474–9482.
- Zhou C., Jin J., Peng C. et al. 2019.** Comparative genomics sheds light on the predatory lifestyle of accipitrids and owls. *Sci Rep* 9(1):2249.
- Zhou H., Hu Y., Ebrahimi A. et al. 2021.** Whole genome-based insights into the phylogeny and evolution of the Juglandaceae. *BMC Ecol Evol* 21(1):116.
- Zhou T. 2019.** Small non-coding RNAs as epigenetic regulators. In: Ferguson B.S. (ed). *Nutritional Epigenomics*, pp. 37–47. Academic Press.
- Zhu Q., Mai U., Pfeiffer W. et al. 2019.** Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat Commun* 10(1):5477.
- Zou T., Kuang W., Yin T. et al. 2022.** Uncovering the enigmatic evolution of bears in greater depth: The hybrid origin of the Asiatic black bear. *PNAS* 119(31):e2120307119.

## СПИСОК ТЕРМИНОВ

- Анализ примесей (Admixture analysis).** Метод определения географического происхождения исследуемого объекта на основе анализа его генетической родословной; известен также как биогеографический анализ происхождения. Генетическая примесь происходит при скрещивании особей из ранее изолированных популяций.
- Аннотация генома.** Процесс добавления биологической информации к известным нуклеотидным последовательностям ДНК; предполагает поиск или предсказание генов, а также регуляторных элементов и повторов. Выделяют *структурную* (идентификация геномных элементов: рамок считывания, кодирующих участков, повторов, мотивов и др.) и *функциональную* (определение биохимических или биологических функций белков, уровней и механизмов регуляции экспрессии генов и др.) аннотации.
- База данных.** Совокупность данных, организованных для удобного поиска и обработки хранимой информации с помощью компьютерных систем. Крупнейшими базами данных с информацией о нуклеотидных последовательностях ДНК являются GenBank, DDBJ, ENA SRA, Ensembl, dbSNP.
- Базовый инструмент поиска локального выравнивания (BLAST).** Алгоритм, разработанный для быстрого приблизительного (эвристического) поиска в базах данных белков или последовательностей ДНК. В настоящее время под названием BLAST объединён ряд программ, специализированных под различные типы данных.
- Биннинг.** Разделение контигов на группы, соответствующие различным организмам. Для биннинга разработаны различные алгоритмы, например, на основе композиционного состава ДНК, частоты использования кодонов или анализа глубины прочтений. Считается, что использование секвенирования HiFi (высокоточные длинные прочтения), улучшенных методов сборки метагенома и дополнительных стратегий биннинга обеспечат быструю и высокоэффективную каталогизацию микробных геномов в сложных микробиомах.
- Вариация числа копий (CNV).** Вид генетического полиморфизма, к которому относят различия индивидуальных геномов по числу копий хромосомных сегментов размером от одной тысячи до нескольких млн пар оснований. CNV возникают в результате несбалансированных хромосомных перестроек, таких как делеции и дупликации.
- Взаимодействия генов.** Выделяют разные типы взаимодействия аллельных и неаллельных генов: *полное доминирование* (доминантные аллели полностью подавляют проявление рецессивных аллелей; фенотип гетерозигот не отличается от фенотипа гомозигот по доминанте), *неполное доминирование* (более слабый рецессивный признак не может быть полностью подавлен доминантным; у гибридов наблюдается промежуточный характер наследования) и *кододоминирование* (аллели в полной мере проявляют своё действие); *плейотропия* (множественное действие гена; способность одного гена влиять на

несколько фенотипических признаков), **комплементарность** (одновременное действие нескольких доминантных генов обеспечивает новое фенотипическое проявление признака); **комплементация** (дополняющее друг друга действие двух форм/аллелей одного гена или разных генов одного хромосомного набора; иногда отождествляется с термином комплементарность), **эпистаз** (проявление одного гена находится под влиянием другого неаллельного ему гена/генов), **полимерия** (взаимодействие неаллельных множественных генов, одинаково влияющих на развитие одного и того же признака; такие гены называются **аддитивными**/суммирующими).

**Выравнивание последовательностей.** Компоновка последовательностей, которая позволяет выявлять области их сходства, связанные со структурным, функциональным или эволюционным родством. Существующие алгоритмы можно разделить на две категории: медленные, но обеспечивающие высокую точность, и быстрые, однако не гарантирующие оптимального решения; основаны на эвристических или вероятностных методах. **Множественное выравнивание** обычно используется для поиска консервативных регионов в группах последовательностей, которые гипотетически являются эволюционно связанными.

**Захват конформации хромосом (Hi-C).** Высокопроизводительная геномная технология, сочетающая захват конформации хроматина и секвенирование следующего поколения для полногеномного картирования конформации хромосом с высоким разрешением. Использует подход «все против всех», который позволяет охарактеризовать укладку хроматина на уровне всего генома (в отличие от методов 3C и 4C, которые оценивают частоты взаимодействия двух интересующих областей, или одной области с остальной частью генома соответственно).

**Ген.** Наследственный фактор, который несёт информацию об определённом признаке или функции организма, и который является структурной и функциональной единицей наследственности. Формы одного и того же гена, расположенные в одинаковых участках (локусах) гомологичных хромосом называются аллелями; локус гомологичных хромосом, содержащий одинаковые аллели – гомозиготный, а разные аллели – гетерозиготный. **Рецессивный** ген определяет развитие признака только в гомозиготном состоянии, а **доминантный** ген проявляется как в гомозиготном, так и в гетерозиготном состоянии. Определение принадлежности аллелей к одной из хромосом называется **фазированием**. Гены, кодирующие информацию о первичной структуре белка/РНК – **структурные**; контролирующие и направляющие деятельность структурных генов – **функциональные**. **Гомеологичные** гены – гены-ортологи в субгеномах, которые произошли от общего предкового гена. **Гомологичные** гены – копии одного и того же гена, расположенные на гомологичных хромосомах; имеют сходную последовательность нуклеотидов, общее происхождение и контролируют один и тот же признак. Гены **ортологи** (ортологичные гены) – произошедшие от общего предка гомологичные гены из разных организмов, разошедшиеся в процессе видообразования и обычно сохраняющие одну и ту же функцию в процессе эволюции. Гены **паралоги**

(паралогичные гены) – гомологичные гены в геноме организма, которые образовались путем дубликации предкового гена, а затем дивергировали в результате мутаций и отбора или при дрейфе генов; в пределах одного вида могут эволюционировать с приобретением новых функций.

**Генетический код.** Система записи наследственной информации в молекулах нуклеиновых кислот, основанная на определенном чередовании последовательностей нуклеотидов в ДНК или РНК, образующих кодоны для соответствующих аминокислот в белках.

**Генные сети.** Графические изображения группы координированно функционирующих генов, взаимодействующих друг с другом как через свои первичные продукты (РНК и белки), так и через разнообразные метаболиты и другие вторичные продукты функционирования генных сетей. Координированное функционирование генов в составе генных сетей обеспечивает формирование фенотипических характеристик организмов (молекулярных, биохимических, клеточных, физиологических, морфологических, поведенческих и т. д.). **Гены-концентраторы (*hub genes*)** в генной сети взаимодействуют со многими другими генами и считаются важными для регуляции сети в целом; обычно играют ключевую роль в биологических процессах и регуляции генов; являются главными в развитии заболеваний.

**Геном.** Весь генетический материал клетки или организма, необходимый для построения и поддержания его жизнедеятельности. Большая часть ДНК в государственном проекте «Геном человека» получена от единственного анонимного донора из Буффало (кодовое название RP11). **Геномика** – раздел молекулярной генетики, посвящённый изучению генома.

**Геномный браузер.** Графический интерфейс, предназначенный для визуализации информации из биологической базы данных, связанной с геномными данными. Позволяет визуализировать и просматривать целые геномы с аннотированными данными, перемещаться по геному, просматривать многочисленные функции, анализировать и исследовать взаимосвязи между различными геномными элементами. Например, геномный браузер UCSC предлагает широкий спектр инструментов, в том числе для визуализации генетической изменчивости, экспрессия генов и эпигенетических модификаций; Ensembl предоставляет инструменты для сравнительной геномики, позволяя сравнивать структуры генов, выравнивать геномы и выявлять синтению; IGV часто используется для визуализации и аннотации геномных данных, охватывающих геномные вариации, уровни экспрессии генов и структуру хроматина.

**Гетерозис.** Увеличение жизнеспособности гетерозиготных организмов по сравнению с гомозиготными родительскими организмами у растений и животных. Выделяют **истинный** (ВРН) и **потенциальный** (МРН) гетерозис. ВРН или гетерозис лучшего родителя – превосходство гибрида F1 по какому-либо признаку над лучшим из родителей. МРН или среднеродительский гетерозис – превосходство гибрида F1 над среднеродительскими показателями признака.

**Гистоновый код.** Набор посттрансляционных модификаций определённых аминокислотных остатков гистонов на N-терминальном конце

(ацетилирование, метилирование, фосфорилирование, убиквитинирование, рибозилирование), которые вовлекаются в сборку, поддержание и изменение структуры хроматина. Служит одним из основных эпигенетических механизмов, контролирующих включение или выключение генов и передачу программы этого контроля по наследству от клетки к клетке.

**Горизонтальный перенос генов (HGT).** Процесс передачи генетического материала от одного организма к другому, который не является его потомком.

**Дезоксирибонуклеиновая кислота (ДНК).** Макромолекула, обеспечивающая хранение, передачу из поколения в поколение и реализацию генетической программы у многих организмов. *ДНК кодирующая* – цепь ДНК, содержащая ту же последовательность, что и цепь мРНК, которая используется в качестве матрицы для трансляции (смысловая РНК); другая цепь ДНК представляет собой цепь-матрицу. *Денатурация ДНК* – обратимое разделение двухцепочечной молекулы на две отдельные цепи.

**ДНК линкерная.** Синтетический двухцепочечный фрагмент ДНК, содержащий сайт узнавания эндонуклеазой рестрикции.

**ДНК-адаптеры.** Небольшие фрагменты ДНК известной последовательности. Используются как участки, с которых начинается амплификация или секвенирование ДНК-библиотеки.

**ДНК-библиотека (геномная библиотека).** Набор ДНК-фрагментов всего генома любого организма; фрагменты фланкированы идентичными ДНК-адаптерами и содержат различные геномные вставки.

**ДНК-микрочип.** Набор из многих тысяч различных иммобилизованных на субстрате последовательностей нуклеотидов; применяют для одновременного определения экспрессии тысяч генов.

**Имунопреципитация хроматина с последующим секвенированием (ChIP-seq).** Метод, используемый для анализа ДНК-белковых взаимодействий и поиска потенциальных сайтов связывания белка; основан на иммунопреципитации хроматина (ChIP) и высокоэффективном секвенировании ДНК.

**Интрон.** Область гена, разделяющая экзоны и не несущая информации об аминокислотной последовательности белкового продукта. Транскрибируется, но удаляется из первичного транскрипта до начала трансляции; являются важнейшей частью регуляции экспрессии генов.

**Искусственная бактериальная хромосома (BAC).** Синтетическая молекула ДНК, используемая для клонирования длинных последовательностей ДНК и секвенирования геномов.

**Искусственный интеллект (ИИ).** Интеллект, демонстрируемый машинами, включая компьютерные системы, которые имитируют человеческий интеллект. Классифицируют на гипотетический ИИ (ASI), который способен не только воспроизводить, но и превосходить максимальные способности человека; общий ИИ (AGI), сильный, но ниже, чем ASI; и слабый/узкий (ANI), предназначенный для выполнения строго определённых приложений. *Машинное обучение* – направление ИИ, сосредоточенное на создании систем, которые обучаются и развиваются на основе получаемых ими данных,

- специализированных под конкретные задачи. Используются два основных типа алгоритмов машинного обучения: контролируемое обучение и самостоятельное обучение. **Глубокое обучение** – совокупность методов машинного обучения, основанных на обучении представлениям, а не на специализированных алгоритмах под конкретные задачи. Включает искусственные нейронные сети (вычислительные системы, основанные на принципах биологических нейронных сетей) и глубокие нейронные сети (искусственные нейронные сети, которые имеют большое количество слоёв).
- Кодон.** Последовательность из трех нуклеотидов (триплет) в цепи ДНК или РНК, которая кодирует определённую аминокислоту или терминацию синтеза полипептидной цепочки. **Стоп-кодон (терминирующий кодон)** обеспечивает терминацию трансляции (UAG, UAA, UGA); раньше такие кодоны называли нонсенс-кодонами. **Предпочтение кодонов** – проявление неравных частот встречаемости синонимичных кодонов в кодирующих областях генома.
- Коллинеарность.** В генетике означает соответствие между последовательностью триплетов ДНК и последовательностью аминокислот в полипептидной цепи. В сравнительной геномике – совпадение порядка расположения блоков последовательностей.
- Конвейер (pipeline, workflow).** Способ решения конкретных задач по обработке данных с помощью объединения программ более общего назначения в цепочку таким образом, чтобы информация, выдаваемая одной программой, попадала на вход следующей.
- Контиг (contig).** Набор перекрывающихся последовательностей фрагментов ДНК (ридов/прочтений), полученных из одного биологического источника (организма, ткани, клетки) в результате секвенирования.
- Конформация.** Пространственная структура комплексного соединения, необходимая для обеспечения его биологической функции.
- Ландшафтная геномика.** Область изучения нейтральных и адаптивных вариаций в масштабах всего генома экологически разнообразных популяций в гетерогенных ландшафтах для решения новых или ранее неразрешимых задач; часто использует подход секвенирования с сокращённым представлением (RRS).
- Локусы количественных признаков (QTLs).** Участки ДНК, либо содержащие гены, либо сцепленные с генами, которые отвечают за тот или иной количественный признак (относятся к полигенным эффектам).
- Локусы-выбросы или выпадающие локусы (outlier loci).** Локусы, которые имеют значительные отличия частоты аллелей относительно предположения о нейтральности, означающей отсутствие направленного отбора.
- Маркерная экспрессирующаяся последовательность (EST).** Секвенированный сайт из экспрессируемого гена, который «метит» участок несеквенированной кДНК рядом с ним; используют для картирования генов.
- Метагеномика.** Область исследования генетического материала, полученного непосредственно из образцов окружающей среды или клинических проб

методом секвенирования. Часто изучается набор генов всех находящихся в образце микроорганизмов, который называется метагеномом. Метагеномный анализ позволяет определить видовое разнообразие исследуемого образца (включая новые виды) без необходимости выделения и культивирования микроорганизмов. Также называют геномикой окружающей среды, экогеномикой, геномикой сообществ или микробиомикой. Согласно методам секвенирования, выделяют целевую метагеномику (на основе ампликонов), предназначенную для видовой идентификации, и метагеномику целого генома, предоставляющую информацию о видовом и функциональном составе сообщества.

**Метилирование ДНК.** Модификация молекулы ДНК путём присоединения метильной группы ( $-CH_3$ ) к цитозину в составе CpG-динуклеотида. Один из способов регуляции работы генома без изменения нуклеотидной последовательности ДНК.

**Метод дробовика (shotgun).** Получение случайной массивированной выборки клонированных фрагментов ДНК организма, на основе которых может быть составлена его геномная библиотека.

**Мобильные генетические элементы (МГЭ/ТЕ).** Последовательности ДНК, способные перемещаться внутри генома и выполняющие разнообразные функции. Существует несколько классов ТЕ, отличающихся по строению и способу перемещения. Присутствуют в геномах большинства организмов, составляя существенную часть геномной ДНК многих изученных видов, и имеют неслучайное распределение. Являются источником мутаций и перестройки генома, модифицируют регуляторную сеть, создают гены и РНК, влияют на зародышевую линию и соматические клетки, обеспечивая пластичность и адаптируемость генома, а также выполняют функции механизма обмена генетическим материалом между организмами. Совокупность всех мобильных генетических элементов в геноме называют *мобиломом*.

**Мутации.** Наследуемые структурные и функциональные изменения последовательности ДНК; могут возникать вследствие замены, утраты или вставки оснований в пределах гена.

**Нанопоровое секвенирование.** Высокоэффективные методы секвенирования ДНК/РНК третьего поколения, основанные на использовании нанопор (биологических или из синтетического материала), которые чувствительны к нуклеиновым кислотам.

**Некодирующая РНК.** РНК, которая не транслируется в белок, но выполняет другие функции. К ним относят транспортные и рибосомные РНК, малые ядерные и ядрышковые РНК, интерферирующие РНК и длинные некодирующие РНК.

**Однонуклеотидные варианты (SNV).** Однонуклеотидный полиморфизм, встречающийся в исследуемой популяции с частотой менее 1%. Как правило, такие варианты недостаточно хорошо охарактеризованы, например, обнаружены только у одного индивида; часто возникают сложности разделения SNV и ошибок секвенирования.

**Однонуклеотидный полиморфизм (SNP).** Отличия в один нуклеотид (А, Т, G или С) в геноме (или в другой сравниваемой последовательности) представителей одного вида или между гомологичными участками гомологичных хромосом; встречается в исследуемой популяции с частотой более 1%.

**Палеогеномика.** Область исследования молекул ДНК, сохранившихся с течением времени; основана на реконструкции и анализе геномной информации из музейных, археологических или палеонтологических образцов. Древняя ДНК может быть использована для тестирования связей между событиями окружающей среды и эволюционными изменениями в популяциях, уточнения эволюционных отношений между видами и калировки молекулярных часов, выявления криптических взаимоотношений между исчезнувшими и современными популяциями и т.д.

**Пангеном (супрагеном).** Совокупность всех генов рассматриваемой группы организмов (как правило, монофилетической). Пангеном можно разделить на коровый пангеном/пангеном ядра, пангеном оболочки и облачный пангеном/пангеном облака, содержащий гены, присутствующие у всех особей (универсальные гены), в некоторых (адаптивные гены) или одной (уникальные гены) линии соответственно. Пангеном вида – это объединённый референсный геном, содержащий разнообразные видоспецифичные генетические варианты и учитывающий различия между отдельными представителями вида.

**Пиросеквенирование.** Метод секвенирования путём синтеза, основанный на обнаружении пирофосфатов, освобождающихся при присоединении к растущей цепи ДНК нуклеотидов (хемилюминесценция). Геном Джеймса Д. Уотсона был первым индивидуальным геномом человека, секвенированным с использованием платформы Roche 454 NGS и завершённым за два месяца.

**Покрытие нуклеотида (глубина секвенирования).** Количество прочтений нуклеотида в процессе секвенирования; позволяет рассчитать среднее покрытие генома.

**Полимеразная цепная реакция (ПЦР).** Метод увеличения (до  $10^{12}$  раз) концентрации (амплификация) исследуемой последовательности ДНК *in vitro*; основан на многократном избирательном копировании определённого участка ДНК при помощи ферментов. **Эмульсионная ПЦР (ЭПЦР)** проводится в эмульсии, где в каждой капле масла амплифицируется единичная молекула ДНК (фрагмент ДНК). В **мостиковой ПЦР** фрагмент одноцепочечной ДНК для амплификации расположен в виде «мостика»: один конец пришит к поверхности, другой держится за счёт комплементарных взаимодействий со вторым иммобилизованным олигонуклеотидом. **ПЦР в нанополлимерах** использует репликацию по принципу катящегося кольца для амплификации небольших фрагментов геномной ДНК в нанополлимеры ДНК.

**Полимеразы.** Ферменты, основной биологической функцией которых является синтез цепей ДНК или РНК при транскрипции и репликации ДНК. Тип нового нуклеотида определяется по принципу комплементарности с шаблоном, с которого ведётся считывание.

- Полногеномный анализ ассоциаций (GWAS).** Поиск связи генотипа с различными фенотипическими признаками (включая наследственные заболевания) в масштабе всего генома.
- Полупроводниковое секвенирование.** Метод секвенирования путём синтеза, работающий, как и пиросеквенирование, с помощью регистрации присоединения каждого следующего нуклеотида. Не требует модификации нуклеотидных оснований, а также хемилюминесцентных или флуоресцентных меток: при присоединении нуклеотида ДНК-полимеразой к растущей цепочке ДНК высвобождается ион водорода, изменяющий pH раствора.
- Популяционная геномика.** Крупномасштабное сравнение последовательностей ДНК популяций с использованием различных генетических маркеров без необходимости полногеномных данных; изучение масштабов и причин изменчивости генома популяций, включая совместные эффекты отбора и сцепления. Играет важную роль в молекулярной и эволюционной биологии, экологии, в сохранении биоразнообразия и здоровья человека.
- Прочтения (риды, reads).** Отдельная последовательность нуклеотидов, полученная в результате секвенирования.
- Референсный/эталонный геном (референсная/эталонная сборка генома).** База данных, содержащая информацию о последовательности нуклеиновых кислот абстрактного представителя биологического вида. Не является результатом секвенирования какого-либо конкретного индивида; записывается в гаплоидном виде отдельно для каждой хромосомы.
- Рибонуклеиновая кислота (РНК).** Макромолекула, образующаяся у многих организмов в результате считывания с ДНК (транскрипции); необходима в процессе синтеза белков, а также многочисленных регуляторных процессов в клетке. Ответственна за хранение, передачу из поколения в поколение и реализацию генетической программы у некоторых вирусов.
- РНК-интерференция.** Подавление экспрессии гена на стадии транскрипции, трансляции, деаденилирования или деградации мРНК при помощи малых молекул РНК (siRNA, microRNA и piRNA). Консервативный биологический ответ на двухцепочечную РНК, который опосредует устойчивость как к эндогенным паразитарным, так и к экзогенным патогенным нуклеиновым кислотам и регулирует экспрессию генов, кодирующих белки, а также участвует в процессах развития и дифференцировки. РНК-интерференция обнаружена у многих видов животных, растений, грибов; предполагают, что она существовала у последнего общего предка эукариот.
- РНК-сплайсинг.** Вырезание определённых нуклеотидных последовательностей из молекул РНК и соединение последовательностей, сохраняющихся в «зрелой» молекуле, в ходе процессинга (созревания) РНК.
- Сборка генома.** Объединение большого числа прочтений в одну (хромосома) или несколько (контиги, скаффолды) длинных последовательностей, в результате чего должна быть восстановлена исходная последовательность генома. Выделяют *сборку de novo*, при которой последовательность собирается только на основе имеющихся коротких прочтений, и *сборку с использованием*

*референса*, при которой прочтения выравниваются на уже собранный геном того же или близкого вида. Сборка генома *de novo* сложна, поэтому используется большое число различных алгоритмов, многие из которых основаны на решении задачи поиска Эйлера пути в графе де Брёйна. В зависимости от качества, сборки геномов подразделяют на стандартные черновые, черновые высокого качества, улучшенные черновые геномы высокого качества, ориентированные на аннотации улучшенные сборки, почти законченные и финишные сборки.

**Секвенирование. Секвенирование полимеров** – определение последовательности нуклеотидов в генах и геномах, РНК, или аминокислотных оснований в белке. **Секвенирование *de novo*** – секвенирование нового для науки (ранее не прочитанного) генома; хорошим для секвенирования *de novo* считается 50–100 кратное покрытие нуклеотидов. **Ресеквенирование** – повторное определение последовательности ДНК организма с уже расшифрованным геномом. **Квантовое секвенирование РНК** одновременно выявляет модификации m6A и 5mC; метод применим для комплексного анализа сайтов метилирования в эпитранскриптом, что открывает новую эру в биологии РНК.

**Секвенирование ДНК отдельных клеток (single-cell DNA sequencing).**

Секвенирование генома отдельных (единичных) клеток. Пробоподготовка включает изоляцию и обогащение единичных клеток, приготовление их суспензий, а также полногеномную амплификацию для создания ДНК-библиотек.

**Секвенирование парных концов.** Метод, позволяющий секвенировать оба конца молекулы ДНК; сохраняет информацию о принадлежности прочтений к одной паре. Участок ДНК между парными прочтениями, который не секвенируется, называется «вставкой», размер которой может варьироваться от нескольких десятков до нескольких сотен оснований в зависимости от метода приготовления библиотеки.

**Секвенирование по Сенгеру.** Метод секвенирования первого поколения; известен как «метод терминаторов», «дидезоксиметод» или «метод обрыва цепи». Основан на использовании дидезоксинуклеотидов для терминации цепи в ходе ПЦР. В результате реакции получается набор копий одного фрагмента, отличающихся длиной на один нуклеотид, и крайний нуклеотид на каждом из фрагментов имеет флуоресцентную (первоначально – радиоактивную) метку; детекция флуоресценции осуществляется в ходе капиллярного электрофореза.

**Секвенирование с сокращенным представлением (RRS).** Обычно применяются три подхода: секвенирование ДНК, связанной с сайтами рестрикции; целевой захват последовательности, позволяющий секвенировать любую область генома, например, экзоны; и технология гибридизации на микрочипах, которая позволяет обнаруживать и типировать вариации ДНК в нескольких сотнях геномных локусов параллельно без предварительной информации о последовательностях. Охват генома очень низкий (до 5%).

**Секвенирование следующего (второго) поколения (NGS).** Технология секвенирования (может быть основана на разных методах), позволяющая одновременно «прочитать» сразу большое число фрагментов ДНК; включает этапы подго-

товки библиотеки и амплификации. За один рабочий цикл прибора происходит получение от сотен мегабаз до гигабаз нуклеотидных последовательностей. Выделяют три основных типа NGS: полногеномное (WGS), экзомное (WES) и таргетное (TS) секвенирование.

**Секвенирование третьего поколения (NNGS/TGS).** Подходы секвенирования, основанные на более простой пробоподготовке, не требующие амплификации и имеющие сокращённое время цикла. Одиночные прочтения значительно длиннее, что обеспечивает более непрерывную реконструкцию генома. Длина прочтений может достигать сотен тысяч и даже миллионов пар нуклеотидов, но их качество по сравнению с NGS пока ниже.

**Секвенирование четвертого поколения.** Методы секвенирования, позволяющие сохранять пространственные координаты последовательностей РНК и ДНК с субклеточным разрешением; интегрировали технологию нанопорового секвенирования в реальном времени без амплификации и повторных циклов и традиционные способы анализа изображений.

**Синтения.** В классической генетике под синтением понимается расположение каких-либо локусов на одной и той же хромосоме, наблюдаемое в разных наборах хромосом и у разных видов; если также совпадает и порядок локусов в хромосоме, это называется коллинеарностью. В сравнительной геномике синтения – это сохранённый порядок генов на хромосомах родственных видов, который является результатом происхождения от общего предка. **Консервативная** (или общая) синтения обозначает смежные выравненные области, которые сохраняют порядок и ориентацию выравнивания, допуская при этом микроперестройки внутри синтенных областей. Выделяют также **макросинтению** (коллинеарность порядка генов в масштабе всей хромосомы), **микросинтению** (коллинеарность порядка небольшого числа генов на участке хромосомы) и **мезосинтению** (содержание генов в хромосоме сохраняется, но коллинеарность отсутствует). Эволюционные перестройки генома (инсерции, делеции, транслокации) могут привести как к приросту синтении, так и к её потере.

**Синтенный блок.** Крупномасштабные гомологичные участки между разными геномами. Формально блок синтении строится следующим образом: находятся гомологичные участки между парой (или большим количеством) геномов (например, участки с хорошим выравниванием или просто гены). Эти участки называются якорями. Далее якоря группируются в синтенные блоки по определённым правилам и каждому блоку присваивается свой номер для комбинаторной интерпретации задачи поиска геномных перестроек.

**Синтетическая геномика.** Область исследований, направленных на создание химически синтезированных полных геномов или их больших частей, что позволяет вносить множественные инженерные изменения в генетическом материале организмов, ведущие к более глубокому пониманию биологических систем.

**Система CRISPR.** Короткие последовательности ДНК бактерий и архей, продуцирующие небольшие некодирующие РНК для обнаружения и устранения вирусных геномов за счёт спаривания и направленного расщепления нуклеазой.

В настоящее время разработаны методы высокоизбирательного активирования и ингибирования генов, основанные на этой системе.

**Системная биология.** Междисциплинарное научное направление, образовавшееся на стыке биологии и программирования, и ориентированное на изучение сложных взаимодействий в живых системах.

**Скаффолд (scaffold).** Серия контигов, расположенных в правильном порядке, но необязательно соединённых в одну непрерывную последовательность.

**Сравнительная или эволюционная геномика.** Область исследований, посвящённая сравнению содержания и организации геномов у разных видов по данным полных последовательностей генома. Оценивает биологическое сходство и различие между организмами и определяет их эволюционные взаимоотношения; помогает раскрыть механизмы эволюции, приближая к разгадке происхождения эукариотического генома. Обычно включает анализ семейств генов, эволюционного развития, дупликации всего генома, селективного давления и т.д.

**Структурная геномика.** Область исследований содержания и организации геномной информации по данным полных последовательностей генома. Основная цель – идентифицировать все гены генома и определить трёхмерные формы всех важных биологических макромолекул, сосредоточив внимание на белках. Классифицирует гены и некодирующие элементы и устанавливает их локализацию в геноме, создавая генетические и физические карты высокого разрешения.

**Таргетное (целевое) секвенирование (TS).** Тест, позволяющий сосредоточиться на конкретных областях генома или отдельных генах. Использует два разных подхода: широкий целевой захват с помощью метода гибридизации и проектирование панели ампликонов на основе ПЦР. TS является важным рутинным методом в клинических исследованиях, благодаря высокой достоверности, точности и относительно низкой стоимости.

**Трёхмерная геномика (3D геномика).** Область изучения механизмов формирования и поддержания структурных и функциональных доменов генома, установлении взаимодействий промотор–энхансер и способах их влияния на пространственную организацию генома.

**Факторы транскрипции.** ДНК-связывающие белки, обеспечивающие прочтение и интерпретацию генетической информации (самостоятельно или в комплексе с другими белками). Способствуют инициации программы повышения или понижения уровня транскрипции гена.

**Фенотип.** Внешние проявления активности одного или более генов в организме или клетке.

**Филогеномика.** В глобальном масштабе – наука о реконструкции эволюционной истории жизни на Земле. Включает две основные цели: установить филогенетические отношения между таксонами и получить представления о механизмах молекулярной эволюции, а также определить предполагаемые функции последовательностей ДНК (или белков), опираясь на данные филогенетических сравнений.

- Финишированный геном.** Полная последовательность ДНК всех хромосом без разрывов, с высоким качеством и низким процентом ошибок.
- Функциональная геномика.** Область исследований, посвящённая описанию биологической функции генов и их продуктов (белков, РНК) и, в конечном счёте, всех компонентов генома с целью понять взаимосвязь между геномом организма и его фенотипом. Фокусируется на таких динамических аспектах, как транскрипция генов, трансляция, регуляция экспрессии генов и белок-белковые взаимодействия; опирается на данные полных последовательностей генома.
- Хромосомные перестройки.** Тип мутаций, которые изменяют структуру хромосом: *дупликация* (повторение/удвоение участка хромосомы), *инверсия* (поворот участка хромосомы на 180°, т.е. изменение порядка генов на обратный), *индели* (вставка или потеря одного, или нескольких нуклеотидов, вплоть до крупных фрагментов хромосом, содержащих миллионы нуклеотидов), *транслокация* (перенос участка хромосомы на негомологичную хромосому).
- Эгоистичная ДНК (избыточная/мусорная ДНК).** Большая часть эукариотического генома с неизвестными функциями (или не выполняющая определённых функций), лишённая активных генов; предположительно реплицирует только себя и использует клетку хозяина и её синтезирующие способности исключительно для этой цели. Включает сателлитную ДНК, спейсерную ДНК и другие типы умеренно и высокоповторяющихся, а также большую часть уникальных нуклеотидных последовательностей. Различия в размерах геномов у разных таксономических групп обусловлены в значительной мере их содержанием, однако единой концепции о её возникновении и роли в эволюционных процессах пока не существует.
- Экзомное секвенирование (WES).** Стратегия секвенирования всех кодирующих белок генов в геноме (т.е. экзома), предполагающая выбор только тех участков ДНК, которые кодируют белки (экзонов – сегментов ДНК, которые представлены в зрелой мРНК эукариот) и их последующее секвенирование.
- Энхансер.** Цис-регуляторный сегмент ДНК, содержащий сайты связывания факторов транскрипции. Может располагаться на разном расстоянии от промотора; обеспечивает ускорение транскрипции.
- Эпигеном.** Совокупность химических модификаций в ДНК и гистоновых белках, передающихся потомству через трансгенерационное эпигенетическое наследование. Эпигеном участвует в регуляции экспрессии генов, развитии, дифференцировке тканей и подавлении мобильных элементов; может динамически изменяться под воздействием условий окружающей среды
- Эпигеномика.** Один из подразделов геномики, посвящённый изучению изменений всей совокупности эпигенетических модификаций генетического материала клетки (эпигенома), которые влияют на экспрессию и регуляцию генов, участвуют во многих важных клеточных процессах (например, дифференцировка, развитие, репарация ДНК и онкогенез).

## ПРИЛОЖЕНИЕ

### О ЖИЗНИ ЗАМЕЧАТЕЛЬНЫХ ЛЮДЕЙ, КТО ОТКРЫВАЛ НАМ МИРОЗДАНЬЕ

#### Похвала уму

*Безумие и разум равноценны,  
Как равноценны в Мире свет и тьма.  
В них два пути, пока мы в Мире пленны,  
Пока замкнуты наши терема.*

*И потому мне кажется желанной  
Различность и причудливость умов.  
Ум английский, и светлый, и туманный,  
Как Море вокруг несчетных островов.*

*Бесстыдный ум француза, ум немецкий,  
Строительный, тяжелый и тупой,  
Ум русский, иступленно-молодецкий,  
Ум скандинавский, вещей и слепой.*

*Испанский ум, как будто весь багряный,  
Торячий, как роскошный цвет гвоздик,  
Ум итальянский, сладкий, как обманы,  
Утонченный, как у Мадонны лик.*

*Как меч, как властный голос, ум латинский,  
Ум эллинский, язык полубогов.  
Индийский ум, кошмарно-исполинский,  
Свод радуги, богатство всех тонов.*

*Я вижу, волны мира многопенны,  
Я здесь стою на звонком берегу,  
И кто б ты ни был, Дух, пред кем мы пленны,  
Привет мой всем, и брату, и врагу.*

К. Бальмонт

## ДЖЕЙМС УОТСОН



Недавно американскому биологу Джеймсу Дьюи Уотсону исполнилось 97 лет (р. 1928 г., Чикаго, США). Дж. Уотсон – лауреат Нобелевской премии по физиологии и медицине 1962 г., присуждённой за открытие двойной спирали ДНК (совместно с Фрэнсисом Криком и Морисом Уилкинсом). Впервые об одном из самых главных открытий XX века учёные объявили на научной конференции в Бельгии в 1953 г.

С 1968 г. Дж. Уотсон возглавил Лабораторию Колд Спринг Харбор в Лонг-Айленде (Нью-Йорк), которая со временем стала одним из ведущих центров молекулярной биологии. В 1990 г. совместно с Национальной организацией здравоохранения Дж. Уотсон участвовал в запуске масштабного научного проекта «Геном человека»; в 1990–1992 гг. был его главой. Вышел из проекта в знак протеста против попыток запатентовать геномные данные. Награждён премией Альберта Ласкера, Президентской медалью Свободы и медалью Джона Карти, медалью двойной спирали Лаборатории Колд Спрингс Харбор и многими другими..

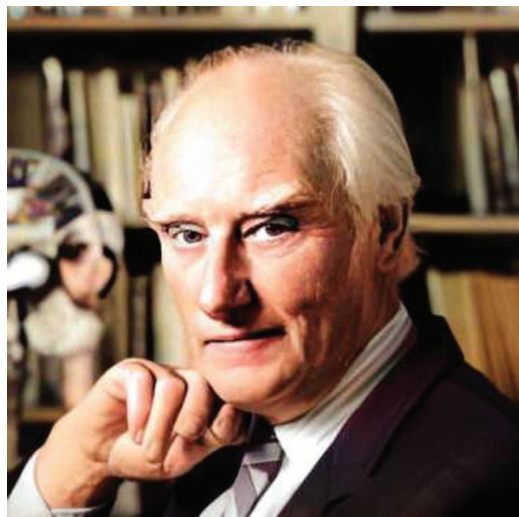
В 2002 г. ему был пожалован титул рыцаря Британской империи, а в 2008 г. он стал почётным доктором Московского государственного университета имени М. В. Ломоносова (МГУ). В Американском музее естественной истории (Нью-Йорк) Дж. Уотсону установлен прижизненный монумент. Он первый человек, чей геном полностью расшифрован.

В 2002 г. ему был пожалован титул рыцаря Британской империи, а в 2008 г. он стал почётным доктором Московского государственного университета имени М. В. Ломоносова (МГУ). В Американском музее естественной истории (Нью-Йорк) Дж. Уотсону установлен прижизненный монумент. Он первый человек, чей геном полностью расшифрован.

Историю своих исследований ДНК Дж. Уотсон описал в книге «Двойная спираль» (1969), которая до сих пор считается одной из лучших в жанре научно-популярной литературы. Он также выпустил автобиографию «Избегайте заnudства. Уроки жизни, прожитой в науке» (2010).



<https://indicator.ru/label/dzhejms-uotson>; <https://m.readly.ru/author/17036/books/>



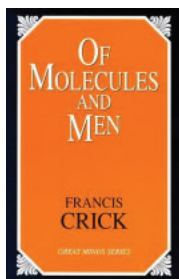
## ФРЭНСИС КРИК

Фрэнсис Гарри Комптон Крик (8 июня 1916, Нортгемптон, Англия – 28 июля 2004, Сан-Диего, Калифорния, США) – британский молекулярный биолог, биофизик и нейробиолог. Лауреат Нобелевской премии по физиологии и медицине 1962 г. совместно с Дж. Уотсоном и М. Уилкинсом «за открытия, касающиеся молекулярной структуры нуклеиновых кислот и их значения для передачи информации в живых системах». Спиралевидную структуру ДНК изобразила жена Ф. Крика, художница Одайл Спид. Ф. Крик сформулировал центральную догму молекулярной биологии: генетическая информация передаётся в клетке в одном направлении от ДНК к

РНК, а затем к белку. Дальнейшие исследования были связаны с теоретической нейробиологией.

Ф. Крик – почётный член Лондонского королевского общества, Королевского общества Эдинбурга, Королевской ирландской академии, Американской ассоциации содействия развитию наук, Американской академии наук и искусств и американской Национальной академии наук. Многочисленные награды включают премию Альберта Ласкера, Международную премию Гайрднера, Королевскую медаль, медаль Копли Королевского общества. В военные годы занимался созданием мин. Биологию начал изучать после войны, когда прочитал известную книгу Эрвина Шрёдингера «Что такое жизнь? Физические аспекты живой клетки», в которой автор задаётся вопросом «Как можно пространственно–временные события, происходящие в живом организме, объяснить с позиции физики и химии?».

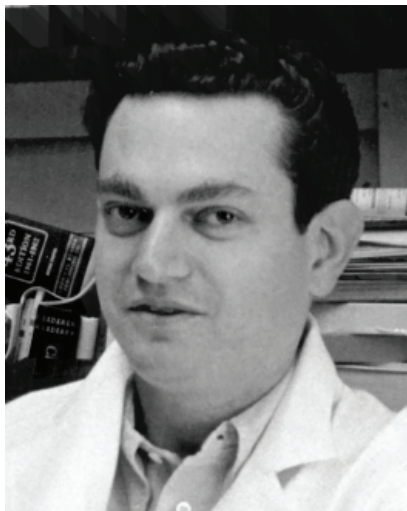
Крик обычно говорил быстро и громко, у него был заразительный смех и хорошее чувство юмора. Один коллега описал его как *«интеллектуальную электростанцию с лукавой улыбкой в состоянии мозгового штурма.... Он находил микроскопические недостатки в логике. В комнате, где было много учёных, Фрэнсис постоянно добивался положения чемпиона»*.



[https://www.cellbiol.ru/book/biografii/frensis\\_krik](https://www.cellbiol.ru/book/biografii/frensis_krik)

<https://ru.wikipedia.org/wiki/>; <https://ru.ruwiki.ru/wiki/>

## МАРШАЛЛ НИРЕНБЕРГ



Маршалл Уоррен Ниренберг (10 апреля 1927 – 15 января 2010, Нью-Йорк) – американский биохимик и генетик, лауреат Нобелевской премии по физиологии и медицине в 1968 г. (совместно с Робертом Холли и Харом Гобиндом Кораной) «за расшифровку генетического кода и его роли в синтезе белков». Член Национальной академии наук США и член Американской академии искусств и наук. Среди наград – Национальная научная медаль США, медаль Пристли, медаль Франклина, премия Альберта Ласкера и др. В своей Нобелевской лекции М. Ниренберг сказал: *«Передача информации от нуклеиновой кислоты к белковому синтезу происходит последовательно, в соответствии с систематическим кодом по относительно простым правилам. Каждая единица нуклеиновой*

*кислоты определяет вид отобранных молекул, их положение относительно предыдущей отобранной молекулы и время события относительно предыдущего. Таким образом, нуклеиновая кислота функционирует одновременно как матрица для других молекул и как биологические часы».*

Защитник социального сознания и высокой ответственности учёных, он отмечал, что рост научных знаний может привести к *«власти человека сформировать его собственную биологическую судьбу. Такая власть может использоваться мудро или неблагоразумно для пользы или вреда человечеству».* В 1992 г. М. Ниренберг подписал "Предупреждение человечеству», а в 1997 году, вместе с 148 выдающимися химиками и двумя лауреатами Нобелевской премии, убедил американский Сенат ратифицировать Соглашение о химическом оружии; выражал свою озабоченность о распространении ядерного оружия. В 2002 г. присоединился к сорока американским нобелевским лауреатам в поддержку терапевтического клонирования.

*«Я не могу сказать, что всегда хотел быть ученым. Но мне всегда была интересна природа, и мне улыбнулась удача при её изучении. Я думаю, интерес к науке – лишь эволюция неустрашимого интереса к тайнам и загадкам природы»* – М. Ниренберг.

The image shows a handwritten table of the genetic code. The columns represent codons (triplets of bases) and the rows represent anticodons (triplets of bases). The table is filled with numbers and letters, representing the assignment of amino acids to specific codons. Some cells are circled in red, and there are handwritten notes and symbols throughout the table.

<https://ru.wikipedia.org/wiki/>

<https://ru.wiki.ru/wiki/>

<https://indicator.ru/biology/nobelevskie-laureaty-marshall-uorren-nirenberg.htm>



## ГЕОРГИЙ АНТОНОВИЧ ГАМОВ



Георгий Антонович Гамов (1904 г., Одесса – 1968 г., Боулдер, США) – советский и американский физик-теоретик, астрофизик и популяризатор науки. В 1933 г. покинул СССР, в 1940 г. получил гражданство США. Член-корреспондент АН СССР (с 1932 по 1938 г., восстановлен посмертно в 1990 г.), Член Национальной академии наук США (1953).

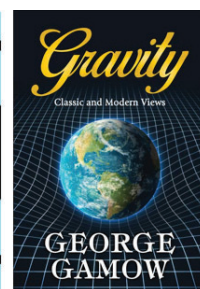
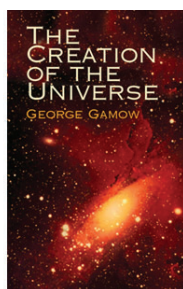
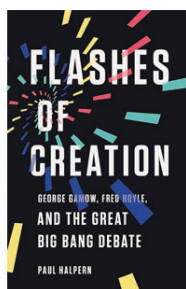
Является автором первой количественной теории альфа-распада, одним из основоположников теории «горячей Вселенной» и одним из пионеров применения ядерной физики к вопросам эволюции звёзд. Он первый чётко сформулировал проблему генетического кода. Широкую известность Гамову принесли его научно-популярные произведения, в которых живым и доступным языком рассказано о современных научных представлениях.

Гамов был не только крупным учёным, но и неординарным человеком, что неоднократно отмечали его друзья и современники. Историки науки считают, что в XX столетии был лишь один человек, три открытия которого заслуживали Нобелевских премий. Это Георгий Гамов. Три его гениальные теории: теория альфа распада, теория горячей Вселенной, идея генетического кода – не только определили науку XX века, но и сформировали повестку дня следующего столетия. Станислав Улам, один из самых талантливых математиков XX века, писал, что Гамов был наделён *«чрезвычайно высокой способностью видеть глубинную связь между разными идеями в науке и даже в искусстве»*. Во многих публикациях о Гамове отмечается, что он *«трижды Нобелевский не лауреат»*.

На мемориальной доске в университете имени Джорджа Вашингтона (Вашингтон, округ Колумбия) отмечено, что профессор физики Георгий Гамов работал в этом университете в 1934–1956 годах. Перечислены наиболее выдающиеся достижения его как учёного, отмечен значительный вклад в популяризацию науки:

[http://www.hrono.ru/biograf/bio\\_g/gamov\\_ga.html](http://www.hrono.ru/biograf/bio_g/gamov_ga.html);

<https://ihst.ru/projects/emigrants/gamov.htm>



## СВАНТЕ ПААБО



Сванте Паабо (р. 1955 г., Стокгольм) – шведский биолог, Лауреат Нобелевской премии по физиологии и медицине 2022 г. Член Шведской королевской академии наук, Леопольдины, иностранный член Национальной академии наук США, Французской академии наук и Лондонского королевского общества.

Сын биохимика Суне Бергстрёма, лауреата Нобелевской премии по физиологии и медицине 1982 г. Изучал русский язык в Школе переводчиков Шведских вооружённых сил в Упсале, в Упсальском университете – египтологию, коптский язык и историю науки.

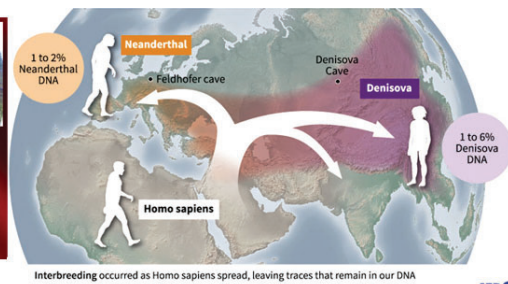
Работал в университетах Упсалы, Цюриха, Калифорнии в Беркли (в лаборатории Аллана Уилсона), Мюнхена; с 1997 г. – директор департамента генетики в институте эволюционной ан-

тропологии в Лейпциге.

С. Паабо стал одним из основателей новой научной дисциплины – палеогенетики. В 1985 г. впервые в истории извлёк генетический материал из человеческих мумий; в 2010 г. группа Паабо, изучившая ДНК из фрагмента кости, найденной в Денисовой пещере на Алтае, пришла к заключению о существовании в древности ранее неизвестного вида гоминид – денисовского человека, а в 2016 г. ими был сделан вывод о не менее трёх эпизодах скрещивания между неандертальцами и различными группами человека разумного.

Благодаря С. Паабо научное сообщество получило представление о том, что гены вымерших «родственников» влияют на физиологию современных людей. Сообщение об открытии так называемого «гена языка» FOXP2 вызвало широкий общественный резонанс и научную дискуссию.

<https://trends.rbc.ru/trends/social/633be393a794791f480d4d9>



<https://scfh.ru/papers/v-poiskakh-utrachennykh-genomov-ot-neandertaltsa-k/>

## ФРЕДЕРИК СЕНГЕР

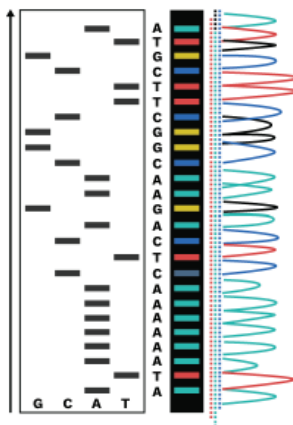


Фредерик Сенгер (1918 г., Рендкомб, Англия – 2013 г., Кембриджшир, Англия) – английский биохимик, один из пяти человек, получивших две Нобелевские премии (наряду с Марией Кюри, Лайнусом Полингом, Джоном Бардином и Барри Шарплессом) и первый учёный в истории, получивший две Нобелевские премии по химии (в 1958 г. «За установление структур белков, особенно инсулина» и в 1980 г. «За фундаментальные исследования биохимических свойств нуклеиновых кислот, в особенности рекомбинантных ДНК» – совместно с Уолтером Гилбертом и Полом Бергом).

Среди многочисленных наград также Королевская медаль, медаль Г. Копли Лондонского королевского общества, премия Альберта Ласкера.

Учёный как минимум в третьем поколении. «Человек, который научил нас читать». Ф. Сенгера называли «обезоруживающе скромным»: когда ему решили дать рыцарское звание, он отказался – не хотел, чтобы его называли сэром.

Ф. Сенгер ушел из науки в 65 лет, посвятив себя семье и парусному спорту. В 1992 г. фондом Wellcome Trust и Британским советом по медицинским исследованиям был основан Исследовательский центр Ф. Сенгера (ныне Институт Сенгера по изучению генома), названный в честь ученого. В 2007 г. Британское биохимическое общество получило грант от фонда Wellcome Trust на каталогизирование и хранение 35 лабораторных журналов, в которые Ф. Сенгер записывал свои исследования с 1944 по 1983 г.



<https://dic.academic.ru/dic.nsf/ruwiki/19721>

[https://elementy.ru/nauchno-populyarnaya\\_biblioteka/435176/Frederik\\_Senger](https://elementy.ru/nauchno-populyarnaya_biblioteka/435176/Frederik_Senger)

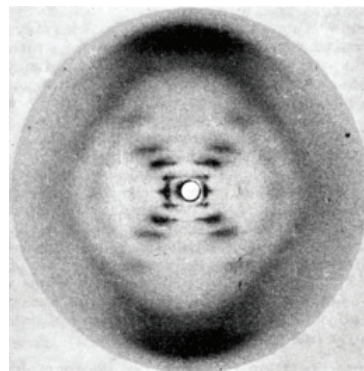
<https://ru.wikipedia.org/wiki>



Розалинд Франклин (1920 – 1958 гг., Лондон, Великобритания) – английский биофизик и учёная-рентгенограф. Известна в основном своей работой над получением рентгенограмм структуры ДНК. Сделанные ею снимки отличались особой чёткостью и подготовили почву для выводов о структуре ДНК, сделанных Дж. Уотсоном и Ф. Криком.

Противоречия, окружавшие фигуру Франклин при жизни, выяснились только после её смерти. Ф. Крик признает: *«Боюсь, что мы всегда относились к ней, скажем, снисходительно»*. Коллега из Кембриджа П. Кавендиш писал в письме: *«Предполагается, что всю работу проводил Уилкинс, а мисс Франклин, очевидно, дура»*. Заведующий лабораторией Дж. Рендалл многозначительно сказал ей *«прекратить работу по исследованию нуклеиновой кислоты»*, когда она ушла из лаборатории Королевского колледжа.

В 2002 г. вышла книга Б. Мэддокс «Розалинд Франклин: забытая леди ДНК», в которой были освещены ранее неизвестные подробности работы Р. Франклин и её причастности к открытию структуры ДНК. Полученные Франклин данные, по словам Ф. Крика, были фактически использованы ими при выдвижении в 1953 г. гипотезы Уотсона-Крика о структуре ДНК. Неопубликованные черновики её работ показывают, что она действительно выявила В-форму спирали ДНК. Однако в серии из трёх статей о ДНК в журнале «Nature» её работу опубликовали последней, ей предшествовала статья Дж. Уотсона и Ф. Крика, в которой лишь отчасти признавались доказательства Франклин в поддержку их гипотезы. О возможности того, что Р. Франклин сыграла основную роль в формировании гипотезы, не было известно до тех пор, пока Дж. Уотсон не представил свету свою автобиографическую повесть «Двойная спираль» в 1968 г., что позже вдохновило нескольких человек на изучение истории открытия ДНК.



Признание работ пришло значительно позднее, начиная с 1982 г. Например, общество Йота-Сигма-Пи назначила Р. Франклин почётным членом; Birkbeck University of London School of Crystallography открыл лабораторию им. Розалинд Франклин; Национальная портретная галерея добавила портрет Р. Франклин к портретам Ф. Крика, Дж. Уотсона и М. Уилкинсона; The U.S. National Cancer Institute установил премию Р. Франклин для женщин в науке; в честь Р. Франклин назван проект по онлайн-обучению биоинформатике, программированию и молекулярной биологии «Rosalind»; Марсоход миссии Экзомарс-2020 получил название «Розалинд Франклин», тогда же директор ESA заявил: *«это имя напоминает нам, что потребность к исследованию заключена в наших генах, а наука – в нашей ДНК»*

<https://ru.wikipedia.org/wiki>

## ГРЕГОР ИОГАНН МЕНДЕЛЬ



Грегор Иоганн Мендель (р. 1822 г., Хейнциндорф, Австрийская империя – 1884 г., Брюнн, Австро-Венгрия) – чешско-австрийский биолог-генетик, монах-августинец, аббат. Основоположник учения о наследственности. Открытие им закономерностей наследования моногенных признаков (известны сейчас как законы Менделя) стало первым шагом на пути к современной генетике.

Мендель активно участвовал в общественной жизни региона. В 1876 году он стал одним из директоров Моравского ипотечного банка, также был депутатом Моравского ландтага.

В 2022 г. в честь 200-летия Менделя, учёные эксгумировали его тело и секвенировали геном. В его ДНК были выявлены генетические варианты, связанные с диабетом, про-

блемами с сердцем и болезнями почек.

На окраине Брно в Старобрненском монастыре августинцев установлена мемориальная доска и памятник возле палисадника, посвященные великому биологу. В музее Менделя имеются его рукописи, документы и рисунки. Также есть старинный микроскоп и другие инструменты, которые учёный использовал в работе.

В 1970 г. Международный астрономический союз присвоил имя Менделя кратеру на обратной стороне Луны. Имя Менделя носит чешская научная полярная станция на острове Джеймса Росса у берегов Антарктиды..



<https://ru.wikipedia.org/wiki/mendel>  
<https://www.darwinmuseum.ru/news/200-letie-gregora-mendelya>  
<https://www.npr.org/sections/health-shots/2022/12/30/1142202365/gregor-mendel-genetics-dna-analyzed>



## БАРБАРА МАК-КЛИНТОК

Барбара Мак-Клинток (1902 г., Хартфорд, Коннектикут, США – 1992 г., Хантингтон, Нью-Йорк, США) – американская учёная-цитогенетик, лауреат Нобелевской премии по физиологии и медицине («За открытие мобильных генетических элементов»). Разработала метод визуализации хромосом клеток кукурузы, сделала множество фундаментальных открытий в цитогенетике, включая рекомбинацию наследственной информации в результате кроссинговера во время мейоза, составила первую генетическую карту кукурузы,

показала роль теломер и центромер, открыла транспозоны. Интересно, что «при рождении ей было дано имя Элеонора, но его сменили в четырёхмесячном возрасте, поскольку родители посчитали, что имя Барбара больше подходит ребёнку по характеру».

«Чтобы войти в историю науки, нужно обладать не только талантом и работоспособностью, но зачастую еще и долголетием, и железной уверенностью в себе. А если ученый – женщина, то этих качеств требуется вдвое больше. Барбара МакКлинток их имела и смогла не только открыть «прыгающие» гены, но и отстоять свои идеи и получить Нобелевскую премию в 81 год». «Это сейчас, почти 80 лет спустя, мы знаем, что в нашей ДНК гораздо больше нерабочих элементов, чем нормальных, кодирующих белок, генов, а транспозоны только в геноме человека занимают почти 45%. Сейчас мы также знаем, что работой любого гена можно управлять. Тогда обе эти идеи казались кошунством».

«Её первый доклад, посвященный транспозонам, был встречен каменным молчанием. Кто-то просто предпочёл не реагировать на смелые заявления, а кто-то открыто объявил Барбару сумасшедшей». «К 1970-м годам стало ясно, что Барбара, по сути, определила современную генетику – и на неё полился поток общественного признания. Президент Соединенных Штатов лично вручил ей в 1971 году национальную научную медаль. О ней снимали телепередачи, писали книги, приглашали читать лекции в ведущих университетах страны. Всё это её мало трогало – в первую очередь Барбару интересовали её кукурузное поле и эксперимент». Спустя еще 10 лет, в 1983 году, 81-летняя Барбара МакКлинток, узнав, что получила самую престижную премию в науке, улыбнулась и сказала: *«Спасибо. Это очень мило. Спасибо за премию. Вы знаете, она мне не очень-то и нужна. Раньше вы не верили, сейчас вы верите – хорошо. Оставьте меня в покое, мне надо работать».*

[https://www.peoples.ru/medicine/physiologist/barbara\\_mcclintock/](https://www.peoples.ru/medicine/physiologist/barbara_mcclintock/)

<https://knife.media/mcclintock/>



## КРЕЙГ ВЕНТЕР



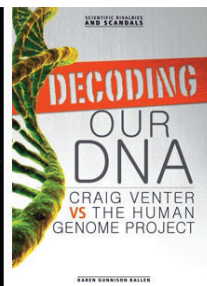
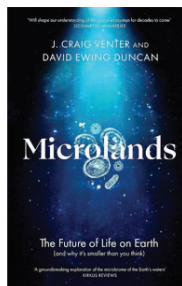
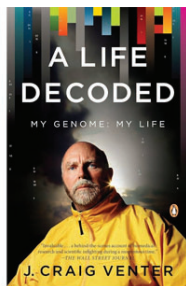
Джон Крейг Вентер (р. 1946 г., Солт-Лейк-Сити, США) – основатель Celera Genomics, Института геномных исследований и Института Крейга Вентера, а также Human Longevity Inc. Доктор философии, член Национальной академии наук США и Американского философского общества.

«Американский биолог и бизнесмен-организатор науки. Крейг Вентер прославился своими успехами в расшифровке генома чело-

века и предложивший эту услугу всем желающим за деньги. В научном мире сльвёт авантюристом, тем не менее этот его прорыв вызвал благосклонные и даже восторженные отзывы от генетиков всего мира. Сейчас он признанный лидер самых радикальных и перспективных направлений современной генетики и первопроходец в области синтеза искусственной жизни». Его ДНК вошла в число пяти, отобранных «Celera» для использования в частном геномном проекте.

*Аннотация к книге "Расшифрованная жизнь. Мой геном, моя жизнь":*

«Крейг Вентер – один из ведущих ученых нашего времени, внёсший огромный вклад в развитие геномики. В феврале 2001 года Вентер опубликовал полностью секвенированный геном человека. Его замечательные мемуары – честный, откровенный рассказ о своей жизни, в которой было и небогатое детство, и война во Вьетнаме, и общение с выдающимися учеными, научившими его любить науку и честно служить ей. "Расшифрованная жизнь" – ещё и рассказ о том, как сегодня делаются открытия и как нелегко приходится тем, кто пытается отстаивать новое».



[https://elementy.ru/bookclub/author/5273063/kreyg\\_venter](https://elementy.ru/bookclub/author/5273063/kreyg_venter)

<https://www.labirint.ru/books/586508/>

Научное издание

*Галина Николаевна Челомина*

**ГЕНОМИКА**

Оригинал-макет Е.В.Сундуковой

Подписано к печати 15.07.2025 г.  
Формат 70×100/16. Усл. п. л. 24,05. Уч.-изд. л. 22,97.  
Тираж 100 экз. Заказ 7

---

Отпечатано:  
ИП Мироманова И.В.  
690106, г. Владивосток, ул. Нерчинская, 42-102