

DOI: <https://doi.org/10.17816/ecogen609552>

Обзорная статья

## Обзор способов визуализации гетерозиготности в контексте природоохранных исследований

А.А. Томаровский<sup>1, 2</sup>, А.А. Тотиков<sup>1, 2</sup>, А.Р. Якупова<sup>3</sup>, А.С. Графодатский<sup>1</sup>, С.Ф. Кливер<sup>4</sup>

<sup>1</sup> Институт молекулярной и клеточной биологии Сибирского отделения РАН, Новосибирск, Россия;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия;

<sup>3</sup> Независимый исследователь, Мюнхен, Германия;

<sup>4</sup> Независимый исследователь, Копенгаген, Дания

### АННОТАЦИЯ

Оценка уровня гетерозиготности — одна из основных метрик в природоохранной биологии, поскольку она способствует корректной разработке программ по сохранению видов, находящихся под угрозой исчезновения. С развитием технологий полногеномного секвенирования появилась возможность более точно оценивать гетерозиготность не только на организменном уровне, но и на популяционно-видовом. Современные природоохранные исследования подразумевают обработку больших объемов полногеномных данных, что приводит к проблемам интерпретации и обуславливает необходимость изучения современных методов визуализации для наглядного и корректного представления результатов. В данном обзоре мы подробно рассматриваем основные типы визуализации оценок уровня гетерозиготности, полученных с использованием различных подходов. Мы подробно излагаем теорию, лежащую в основе каждого метода формирования изображения и обсуждаем их особенности на примере исследований немодельных видов с различным природоохранным статусом. Обзор позволяет получить представление об актуальных инструментах для оценки и последующей визуализации гетерозиготности, а также о текущих тенденциях в данной области.

**Ключевые слова:** природоохранная биология; геномика; популяционная генетика; генетическое разнообразие; визуализация данных.

### Как цитировать

Томаровский А.А., Тотиков А.А., Якупова А.Р., Графодатский А.С., Кливер С.Ф. Обзор способов визуализации гетерозиготности в контексте природоохранных исследований // Экологическая генетика. 2023. Т. 21. № 4. С. 383–400. DOI: <https://doi.org/10.17816/ecogen609552>

DOI: <https://doi.org/10.17816/ecogen609552>

Review Article

# Review of heterozygosity visualization approaches in the context of conservation research

Andrey A. Tomarovsky<sup>1, 2</sup>, Azamat A. Totikov<sup>1, 2</sup>, Aliya R. Yakupova<sup>3</sup>,  
Aleksandr S. Graphodatsky<sup>1</sup>, Sergey F. Kliver<sup>4</sup>

<sup>1</sup> Institute of Molecular and Cellular Biology, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia;

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia;

<sup>3</sup> Independent researcher, Munich, Germany;

<sup>4</sup> Independent researcher, Copenhagen, Denmark

## ABSTRACT

The assessment of heterozygosity level is one of the key metrics in conservation biology, as it contributes to the accurate design of conservation programs for endangered species. With the development of whole-genome sequencing technologies, it is now possible to more accurately estimate heterozygosity not only at the organismal level, but also at the population and species level. Contemporary conservation studies involve the processing of large volumes of whole-genome data, leading to problems of interpretation and necessitates the study of modern visualization methods for clear and correct presentation of results. In this review, we comprehensively examine the main types of visualization of heterozygosity assessments obtained using various approaches. We delve into the theory underlying each visualization method and discuss their characteristics using examples from studies of non-model species with different conservation statuses. The review provides insight into current tools for heterozygosity assessment and subsequent visualization, as well as current trends in this field.

**Keywords:** conservation biology; genomics; population genetics; genetic diversity; data visualization.

## To cite this article

Tomarovsky AA, Totikov AA, Yakupova AR, Graphodatsky AS, Kliver SF. Review of heterozygosity visualization approaches in the context of conservation research. *Ecological genetics*. 2023;21(4):383–400. <https://doi.org/10.17816/ecogen609552>

Received: 17.10.2023

Accepted: 22.11.2023

Published: 29.12.2023

## ВВЕДЕНИЕ

Природоохранная биология — это область науки, направленная на поддержание биологического разнообразия экосистем и защиту видов от вымирания [1]. По мере развития технологий полногеномного секвенирования и доступности высококачественных геномных данных, исследования в контексте природоохранной биологии стали проводиться на более высоком уровне [2]. Развитие подходов биоинформатики, в свою очередь, способствовало получению геномных сборок не только скаффолдного, но и хромосомного уровней, которые стали основой для изучения генетического разнообразия и адаптационного потенциала как для отдельной популяции, так и для вида в целом [3].

Под генетическим разнообразием понимают генетические различия в геномах между особями и популяциями одного вида, включая как нейтральные, так и существенные различия, влияющие на выживаемость и адаптацию к изменяющимся условиям среды [4]. Появление подходов полногеномного секвенирования позволило не только с большей достоверностью изучать демографическую историю и структуру популяций, но и способствовало более точной оценке генетического разнообразия [5]. Возможность и необходимость проведения многочисленных анализов для контроля биоразнообразия обусловлена прежде всего возможностью оценки текущего генетического состояния отдельных особей и популяций посредством оценки гетерозиготности [6]. Таким образом, исследователи могут разрабатывать программы по сохранению видов, находящихся под угрозой исчезновения, и впоследствии прилагать усилия по реинтродукции вымирающих видов, путем отбора особей для контролируемого скрещивания и разведения как в неволе, так и в дикой природе [7].

Поскольку изучение уровня генетического разнообразия и, в частности, гетерозиготности как одной из главных его метрик связано с большими объемами биологических данных, исследователи сталкиваются с определенными трудностями. Основная проблема заключается в сложности наглядного представления и восприятия больших объемов данных, позволяющих корректно изучить гетерозиготный диплоидный геном, не допустив при этом потери биологически значимой информации. Одним из способов преодоления этой проблемы является правильная визуализация полученных результатов. Внедрение эффективных способов представления изображения необходимо не только для представления результатов и облегчения интерпретации для формирования выводов, новых идей, гипотез и концепций, но и для выявления возможных проблем с качеством исходных данных. В современных исследованиях встречается множество различных способов графического отображения, однако подходы, лежащие в их основе, чаще всего остаются вне понимания читателей, поскольку их описание либо отсутствует,

либо приведено в общих чертах. Это может негативно отразиться на выборе наиболее оптимального способа визуализации гетерозиготности в будущих исследованиях. Таким образом, в связи с большими объемами полногеномных данных и сопутствующих им сложно организованных метаданных, важно подобрать корректный и лаконичный способ представления полученных результатов.

В данном обзоре мы сосредоточимся на достижениях в области графического представления результатов различных подходов по оценке гетерозиготности в контексте природоохранных исследований. Дополнительная цель данного обзора — дать читателю представление о текущих тенденциях в разработке инструментов для визуализации гетерозиготности.

## ГЕТЕРОЗИГОТНОСТЬ КАК ОДНА ИЗ КЛЮЧЕВЫХ МЕТРИК В ПРИРОДООХРАННОЙ БИОЛОГИИ

Стремительное развитие и последующее снижение стоимости высокопроизводительного секвенирования в последнее десятилетие сделало возможным проведение исследований в области природоохранной биологии в масштабах целого генома [8]. Например, использование данных полногеномного секвенирования с последующим объединением скаффолдов (скаффолдингом) до хромосомного уровня на основе данных Hi-C стало быстрым и недорогим способом генерации высококачественных *de novo* сборок хромосомного уровня у млекопитающих [9]. Интересно, что такой подход геномной сборки остается практически невозможным без визуализации карт контактов Hi-C и последующей ручной коррекции хромосомных взаимодействий в интерактивном режиме [10]. Доступность и качество данных полногеномного секвенирования позволили не только получать геномные сборки. В отличие от подходов низкого разрешения, таких как секвенирование отдельных ядерных маркеров методом Сэнгера или анализа вариабельных локусов коротких tandemных повторов (short tandem repeats, STR), которые остаются значимыми лишь при анализе большого количества образцов [11], полногеномное секвенирование позволило выявлять однонуклеотидные полиморфизмы (single nucleotide polymorphism, SNP), охватывая практически весь геном исследуемой особи.

Выявление и последующий анализ SNP позволяет не только оценить уровень генетической дифференциации между оставшимися популяциями, но и изучить генетическое разнообразие внутри рассматриваемых популяций [12]. Снижение уровня генетического разнообразия (генетическая эрозия) представляет собой серьезную угрозу для биоразнообразия, поскольку может способствовать накоплению вредоносных генетических вариантов в гомозиготном состоянии и снижению адаптационного потенциала особей, что в конечном итоге приводит к угрозе вымирания вида [13, 14].

В случае, если подобные явления становятся следствием скрещиваний близкородственных особей или, наоборот, скрещивания особей из отдаленных популяций, их можно характеризовать как аутбредную [15] или инбредную депрессию вида [16]. При организации мероприятий по реинтродукции исчезающих видов важно учитывать возможность возникновения обоих типов депрессии. Важно отметить, что вероятность инбредной депрессии как правило выше, чем аутбредной, в связи с чем последнюю часто недооценивают [15].

Несмотря на то что гетерозиготность играет ключевую роль в эволюции, предоставляя ресурс для поддержания генетического разнообразия внутри популяции, высокий уровень данного показателя нельзя считать однозначным признаком благосостояния изолированных на протяжении длительного времени популяций. В данном контексте низкая гетерозиготность, сформировавшаяся согласно положениям «генетической революции», может способствовать выживанию и, как следствие, проявляет положительные свойства [17]. Низкий уровень гетерозиготности в изолированных или полуизолированных популяциях, обусловленный постепенным отбором в пользу доминантных аллелей, в результате способствует снижению отрицательного воздействия инбридинга [18, 19]. Закрепление вредных мутаций в инбредных популяциях с низкой численностью в данном случае может быть нейтрализовано полигенными модификаторами, влияющими на экспрессию генов и, впоследствии, препятствующими проявлению нежелательных признаков [20]. Более того, согласно «теории эволюции со смещающимся равновесием», снижение гетерозиготности в субпопуляциях в конечном счете может способствовать увеличению изменчивости как в метапопуляции, так и между популяциями в случае потока генов между изначально изолированными группами [21]. Таким образом, уменьшение гетерозиготности в отдельных субпопуляциях может играть роль в поддержании генетического разнообразия на более высоких уровнях организации, способствуя эволюционному прогрессу в масштабах популяций и их взаимодействий. И хотя не существует объективного стандарта уровня гетерозиготности, поскольку значение данного показателя может отличаться у разных таксонов [22, 23], его корректная оценка способствует общему пониманию эволюционных процессов и имеет решающее значение в организации мероприятий по реинтродукции исчезающих видов.

Оценку гетерозиготности можно выполнить в различных масштабах генома. Прежде всего на уровне SNP в отдельно взятой позиции, вариант которой может быть в гомозиготном или в гетерозиготном состоянии. Классификация и функциональная аннотация подобных SNP способствует выявлению вредоносных мутаций в последовательностях генов и может дать представление об адаптивном потенциале особей [24]. Тем не менее в контексте природоохранной биологии под гетерозиготностью понимают в первую очередь долю гетерозиготных SNP

(в диапазоне от 0 до 1). При этом выделяют ожидаемую гетерозиготность (expected heterozygosity,  $H_e$ ) и наблюдаемую гетерозиготность (observed heterozygosity,  $H_o$ ) [25]. Ожидаемая гетерозиготность оценивается как разница единицы и суммы квадратов частот гомозиготных аллелей, тогда как наблюдаемая гетерозиготность определяется долей гетерозигот, усредненной по локусу или целому геному, и зависит как от величины генетической изменчивости в популяции, так и от уровня инбридинга [25, 26]. При этом значение наблюдаемой гетерозиготности для удобства масштабируется до среднего количества SNP на 1 тыс. п. н., в связи с чем данную оценку называют средней гетерозиготностью. Отметим, что для оценки инбридинга можно сравнить наблюдаемую гетерозиготность с ожидаемой гетерозиготностью. При наличии инбридинга ожидается, что последняя будет выше. Помимо этого, разработано множество альтернативных метрик генетического разнообразия. В частности, нуклеотидное разнообразие ( $\pi$ ), определяемое как среднее количество нуклеотидных различий на сайт, а также разнообразие гаплотипов ( $H$ ), отражающее вероятность того, что два случайных гаплотипа различны [27]. Отдельного внимания также заслуживает  $F$ -статистика Сьюэлла Райта ( $F_{st}$ ), которая является мерой генетической дифференциации популяций и применяется для оценки доли общей генетической вариации между ними [28].

Значительно большее число SNP, получение которых стало возможно благодаря использованию данных полногеномного секвенирования, способствовало повышению точности вышеперечисленных показателей и открыло новые возможности для оценки гетерозиготности [29, 30]. Это позволило проводить оценку средней гетерозиготности для отдельных геномных локусов путем вычисления оценок в окнах — перекрывающихся или неперекрывающихся геномных участках фиксированного размера [31]. Полученные для отдельных геномных участков оценки для удобства также масштабируются до SNP / 1 тыс. п. н. Результатом такого подхода стал переход от сравнительного анализа гетерозиготности целого генома к анализу ее распределения, что позволило не только провести оценку доверительного интервала для среднего значения, но и предоставило информацию о геномных локусах с повышенным и пониженным вследствие отбора генетическим разнообразием [24]. В случае анализа некоторого количества особей (выбор числа которых зависит от размера исходной популяции) полученную информацию можно использовать для установления степени родства и оценки генетической структуры популяции, что будет способствовать принятию своевременных природоохранных мер, направленных на снижение вероятности генетической эрозии [32]. Хотя для оценки гетерозиготности на уровне популяции обычно требуется значительное количество особей, использование подходов полногеномного секвенирования позволило проводить анализ на малом числе особей без потери информации благодаря качеству и покрытию исходных данных [33].

Изучение гетерозиготности можно условно разделить на анализ участков с повышенным и пониженным уровнем гетерозиготности. Непрерывные области генома с пониженным уровнем гетерозиготности (повышенный уровень гомозиготности) называют протяженными участками гомозиготности (runs of homozygosity, ROH) [34]. Такие участки возникают из-за инбридинга, причина которого состоит в наследовании идентичных гаплотипов от каждой родительской особи [34]. Как правило, свидетельством недавнего инбридинга являются длинные участки ROH, тогда как признак древнего инбридинга — чередование коротких фрагментов ROH. Это обусловлено тем, что по мере увеличения числа поколений увеличивается и вероятность возникновения событий рекомбинации [35]. Как следствие, ROH в большей степени распространены в небольших изолированных популяциях, в которых отсутствует поток генов, в то время как более крупные популяции, напротив, имеют короткие ROH и меньшее их количество. Известно, что инбридинг увеличивает частоту рецессивных заболеваний, а также приводит к увеличению частоты и протяженности ROH в популяции, что в свою очередь способствует проявлению вредных эффектов генетических рецессивных аллелей при их выходе в гомозиготное состояние у инбредных особей [36, 37]. В конечном итоге этот процесс приводит к снижению репродуктивного успеха, известного как инбредная депрессия [38]. Выявление участков ROH также играет роль в изучении генетической структуры и демографической истории популяций. Наконец, анализ ROH позволяет устанавливать специфические признаки отбора, которые проявляются в фиксации определенного аллеля под высоким давлением отбора в популяции [39]. Такая фиксация приводит к увеличению гомозиготности в геномных областях, окружающих эти аллели [40]. Таким образом, определение ROH и их последующая визуализация являются неотъемлемой частью оценки генетического разнообразия и анализа структуры популяции в природоохранных исследованиях.

Участки с повышенным уровнем гетерозиготности (runs of heterozygosity, ROHet или heterozygosity-rich regions, HRR) описаны в литературе не так хорошо, как ROH. В отличие от последних, идентификация ROHet и оценка коэффициента гетерозиготности позволяют получить представление не только о уровне генетического разнообразия и эволюционной истории популяции, но и обнаружить конкретные области генома, где поддержание высокого генетического разнообразия может быть особенно полезным для выживания [41]. Поиск и локализация участков ROHet зачастую происходит наряду с выявлением ROH, что может быть также использовано для изучения особей из недавно смешанных близкородственных популяций, геномы которых представляют набор чередующихся сегментов повышенной гомозиготности и гетерозиготности [42]. Определение координат и последующий анализ происхождения таких участков позволяет изучить

генетические основы заболеваний и морфологических особенностей исследуемой группы особей [43]. Таким образом, выявление регионов ROH и ROHet с последующей визуализацией позволяет не только примерно оценить эффективный размер численности, но и предоставляет возможности для изучения структуры и анализа происхождения популяции.

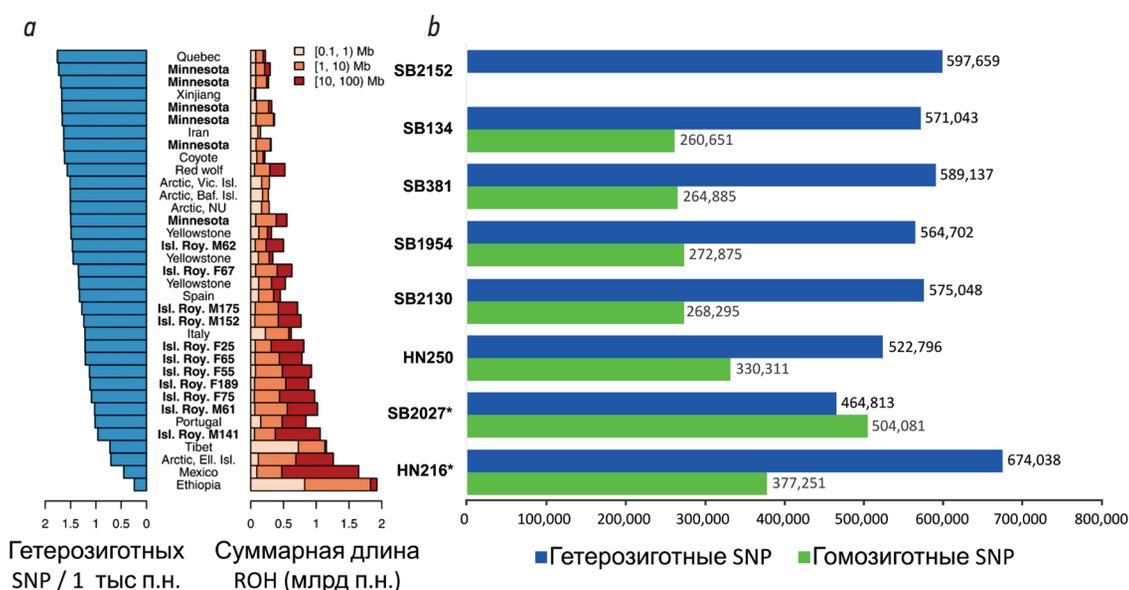
## ВЫБОР СПОСОБА ВИЗУАЛИЗАЦИИ ГЕТЕРОЗИГОТНОСТИ

Изучение генетического разнообразия может происходить на уровне одной особи, в пределах нескольких особей из одной или нескольких популяций и путем оценки и сравнения гетерозиготности особей различных популяций. Поскольку геномные данные, как правило, на некотором этапе анализа превращаются в сложно организованный набор табличных данных, выбор метода графического представления оценок гетерозиготности играет большую роль. Грамотное формирование изображения помогает формулировать и проверять гипотезы в процессе исследования, а также упрощает интерпретацию полученных результатов и ускоряет их понимание аудиторией при представлении. В данной главе описаны различные способы визуализации оценок гетерозиготности на примере исследований немодельных видов с различным природоохранным статусом.

### Линейчатые диаграммы и диаграммы Венна

Оценка наблюдаемой и ожидаемой гетерозиготности на уровне всего генома или доступной последовательности (например, гена или STR) — один из первых подходов, используемых исследователями для анализа генетического разнообразия. Учитывая, что полученные оценки представляют собой набор, состоящий из одного числа для каждой исследуемой особи, они не подразумевают какого-то специфического формата для хранения и часто формируются в универсальных программах для статистической обработки и визуализации данных, таких как Microsoft Excel. Отметим, что масштабированная до SNP / 1 тыс. п. н. оценка средней гетерозиготности зачастую недостаточно точно отражает действительное генетическое состояние особи, что может приводить к ошибкам интерпретации. Для сравнительного анализа полученных оценок подойдут классические подходы визуального представления информации, используемые для сравнительного анализа, такие как линейчатые диаграммы и диаграммы Венна.

Линейчатые диаграммы — один из ранних способов визуализации данных [46]. Столбчатые линейчатые диаграммы (bar chart) активно используются в современных исследованиях вне зависимости от типа используемых данных. Как правило, данный подход применяется в случае, когда необходимо провести сравнение оценок гетерозиготности множества особей из различных популяций



**Рис. 1.** Примеры визуализации гетерозиготности при помощи линейчатых диаграмм: *a* — средний уровень гетерозиготности (слева) и суммарная длина ROH (справа) в геномах волков из популяции острова Айл-Ройал; *b* — количество гетерозиготных и гомозиготных вариантов в двух подвидах черной антилопы. SB2027\* и HN216\* — особи южного подвида, остальные — замбийского. Оригинальные изображения из [44, 45]

**Fig. 1.** Examples of visualization of heterozygosity using line diagrams: *a* — mean heterozygosity level (left) and total ROH length (right) in the genomes of wolves from the Isle Royale population; *b* — number of heterozygous and homozygous variants in two subspecies of black antelope. SB2027\* and HN216\* are individuals from the southern subspecies, the others from the Zambian subspecies. Original images from [44, 45]

одного или нескольких близкородственных видов. Каждый столбец диаграммы в таком случае соответствует особи из отдельной популяции, отражая выбранную метрику гетерозиготности, рассчитанную для всего генома или рассматриваемого региона. Сравнительный анализ генетического разнообразия с использованием столбчатых диаграмм был выполнен в работе по изучению инбридинга у особей серых волков из популяций Миннесоты и острова Айл-Ройал [44]. Используя разнонаправленные горизонтальные диаграммы, исследователям удалось показать как среднюю гетерозиготность в геномах рассматриваемых особей, так и суммарные длины ROH в разном диапазоне отсечек по длине (рис. 1, *a*). Таким образом, удалось установить, что волки с острова Айл-Ройал содержат значительно меньше гетерозигот и значительно больше гомозигот, чем волки из Миннесоты, а также установлена повышенная гомозиготность вредоносных SNP. В результате в данном исследовании подтвердили инбредную депрессию популяции серых волков острова Айл-Ройал.

При помощи столбчатых диаграмм часто проводят прямое сравнение общего количества гетерозиготных и гомозиготных SNP, как это было сделано для восьми особей черных антилоп двух разных подвигов (рис. 1, *b*) [45]. Отметим, что для удобства визуализации нескольких параметров, линейчатые диаграммы различных показателей часто выделяют разным цветом. Среди особенностей столбчатых диаграмм можно выделить возможность отображения погрешностей в виде линейных отрезков,

в пределах которых могут находиться значения гетерозиготности. Это, к примеру, было реализовано в исследовании, посвященном оленю Давида. В этой работе визуализацию количества генетических вариантов выполнили при помощи круговых линейчатых диаграмм (pie chart), благодаря чему удалось подчеркнуть значительную разницу в процентном соотношении участков генома с низкой и высокой гетерозиготностью [47]. Разновидности линейчатых диаграмм также используются в других исследованиях, например, в работах по изучению иберийской рыси, морской выдры и гигантской выдры, льва и снежного барса [48–50].

Отдельного внимания заслуживает способ отображения как уникальных, так и общих SNP для нескольких особей при помощи диаграммы Венна, которая представляет собой схему возможных пересечений между несколькими подмножествами (как правило, от 2 до 3) [52]. Таким образом, удается не только визуально сравнить количественное соотношение гетерозиготных и/или гомозиготных вариантов между особями, но и показать количество общих вариантов в пределах выборки. Данный способ был использован в работе, посвященной изучению популяционной истории белого и бурого медведя (рис. 2), благодаря чему исследователи не только сравнили суммарное количество SNP для каждого вида, но и выявили значительно меньшее количество общих вариантов для белого и черного медведя, чем между бурым и черным, а также белым и бурым медведем [51]. Диаграммы Венна были также использованы в работе по изучению

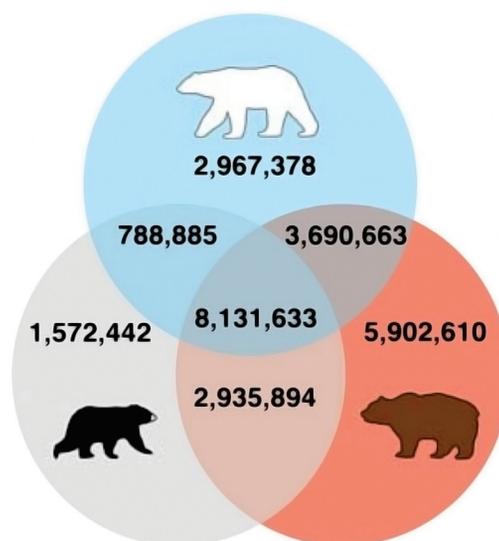
тасманийского дьявола, где удалось показать количество общих и уникальных гетерозиготных SNP двух исследуемых особей из разных регионов Тасмании (Австралия) [53]. Важно отметить, что, хотя диаграммы Венна довольно часто используются для сравнения гетерозиготности нескольких особей, они становятся неприменимы, если рассматривается более 6–10 особей. В таком случае визуализацию множеств SNP приходится осуществлять в форме эллипсов, а не окружностей и, как следствие, с добавлением новых образцов читабельность такой диаграммы заметно снижается.

### Диаграммы распределения

Диаграммы распределения используются как для сравнительного анализа набора оценок гетерозиготности для множества особей, так и для оценок одной особи, рассчитанных в окнах фиксированного размера. В первом случае данные часто сохраняются в форматах CSV или TSV, каждый столбец в которых соответствует отдельной группе особей (популяция или вид). Во втором — неформальным стандартом хранения геномных участков и их аннотации (в данном случае, оценок средней гетерозиготности) является формат BED (browser extensible data). Отметим, что уже на этом этапе анализа исследователи неизбежно вынуждены прибегнуть к обработке полученных данных при помощи различных инструментов и библиотек для языков программирования, что обусловлено объемом полученных данных. Для визуализации распределений существует несколько видов диаграмм: гистограммы (histogram), диаграммы размаха (box plot), скрипичные диаграммы (violin plot), а также графики кумулятивной суммы (cumulative sum chart).

Визуализацию распределения оценок гетерозиготности часто можно встретить в виде гистограммы, которая на первый взгляд напоминает столбчатую сравнительную диаграмму. Однако, в отличие от последней, гистограмма (или график плотности) позволяет получить представление о форме распределения значений гетерозиготности или гомозиготности, что открывает возможности для выявления закономерностей и потенциальных выбросов. Гистограммы обычно строятся путем нанесения значений гетерозиготности на ось *x* и частоты встречаемости (плотности) данной оценки на ось *y*. Полученный график не только отображает максимальный и минимальный уровни гетерозиготности, но и подчеркивает форму распределения этой характеристики. Гистограммы были успешно использованы в работе, посвященной байкальской нерпе, где с их помощью удалось сопоставить гетерозиготность особей этого вида с другими представителями ластоногих (рис. 3, *a*). Подобная визуализация гетерозиготности уже стала стандартом и используется во множестве природоохранных исследований [55–59].

Диаграммы размаха выгодно отличаются от гистограмм за счет более компактного вида и отображения большего количества полезной информации [60].

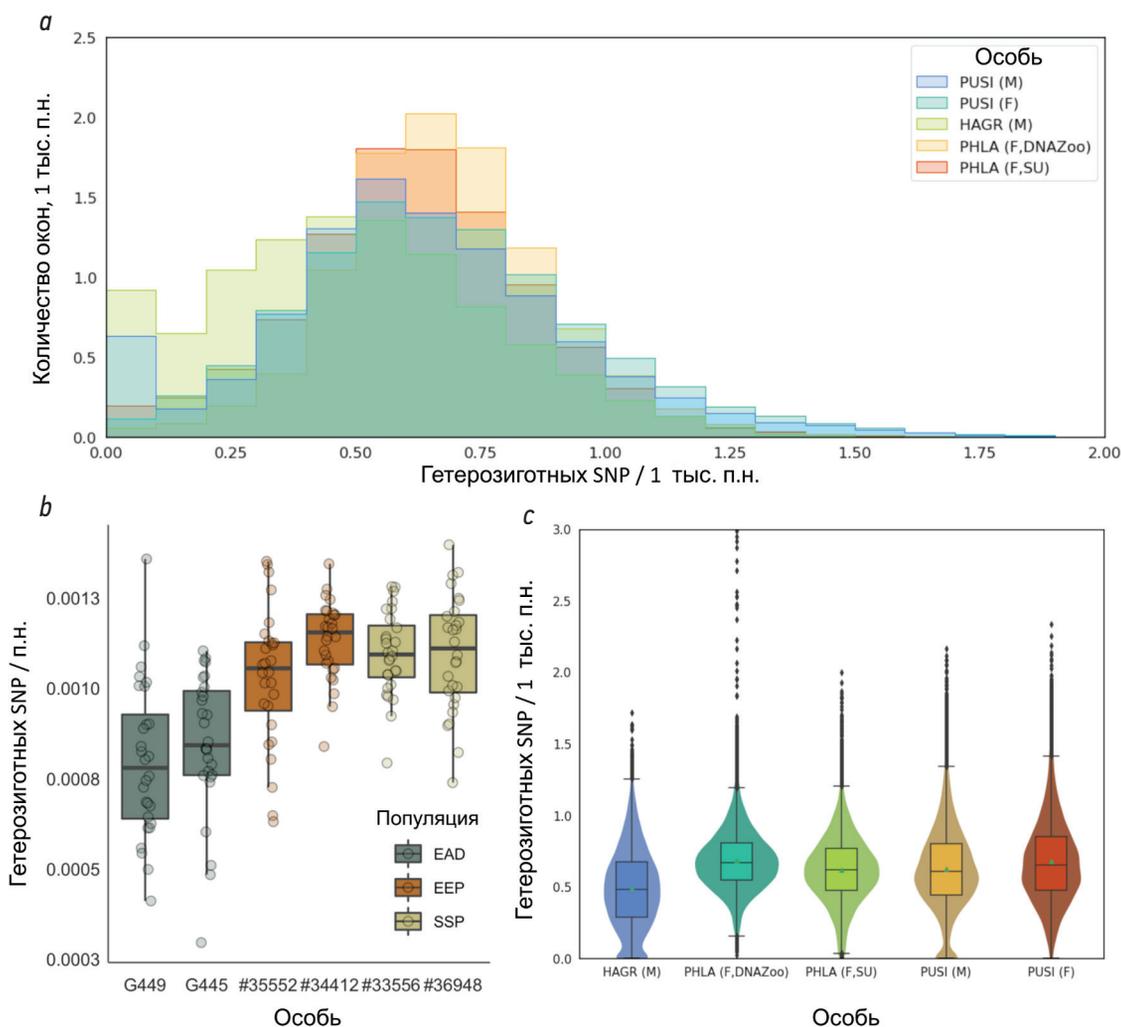


**Рис. 2.** Пример визуализации уникальных и общих SNP для белого, бурого и черного медведя при помощи диаграммы Венна. Оригинальное изображение из [51]

**Fig. 2.** Example of visualization of unique and shared SNPs for polar, brown and black bear using a Venn diagram. Original image from [51]

Этот подход позволяет показать среднее значение (или медиану), нижний и верхний квартили, минимальное и максимальное значение и выбросы. Помимо этого, подход позволяет эффективно проводить визуализацию сразу для ряда особей на одном графике, что удобно для сравнительного анализа. На графике с диаграммами размаха одна ось представляет собой шкалу со значениями гетерозиготности, а вторая — с идентификаторами особей. Так, сравнительный анализ особей разных видов млекопитающих был выполнен в работе, посвященной анализу новых перспектив и возможностей при получении геномных сборок хромосомного уровня [61]. Помимо анализа распределения оценок гетерозиготности одной особи, диаграммы размаха можно использовать для визуализации распределений оценок нескольких особей из одной популяции. Так, в работе по изучению особей сахарийского орикса были показаны различия в полногеномной гетерозиготности между популяциями этого вида (рис. 3, *b*) [54]. Дополнительно при помощи диаграмм размаха в этой же работе была показана взаимосвязь между оценкой гетерозиготности и размером популяции.

Разновидностью диаграмм размаха являются скрипичные диаграммы, которые более эффективно справляются с визуализацией асимметричных и мультимодальных распределений [62]. Скрипичные диаграммы были использованы в работе, посвященной чернолапому хорьку, где с их помощью провели сравнение распределений гетерозиготности близких видов из подсемейства *Mustelinae* [56]. Важно сказать, что диаграммы размаха и скрипичные диаграммы часто используют одновременно, как это было сделано в публикации по изучению байкальской нерпы (рис. 3, *c*) [55].



**Рис. 3.** Примеры визуализации гетерозиготности при помощи диаграмм распределений: *a* — гистограммы распределения гетерозиготности по геному для пяти выборок ластоногих. Гетерозиготные SNP подсчитывались в окнах размером 1 млн п. н. и масштабировались в SNP / 1 тыс. п. н.; *b* — оценки гетерозиготности в диаграмме размаха для множества особей сахарийского орикса из различных популяций; *c* — скрипичные диаграммы распределения гетерозиготности для двух особей байкальской нерпы, двух особей пятнистой нерпы и одной особи серой нерпы. Оригинальные изображения из [54, 55]

**Fig. 3.** Examples of heterozygosity visualization using distribution diagrams: *a* — histograms of heterozygosity distribution across the genome for five samples of pinnipeds. Heterozygous SNPs were counted in 1 million bp windows and scaled to SNP / 1 thousand bp; *b* — estimates of heterozygosity in boxplots for multiple individuals of Sahara oryx from different populations; *c* — violin diagrams of heterozygosity distribution for 2 individuals of Baikal seal, 2 individuals of spotted seal, and 1 individual of gray seal. Original images from [54, 55]

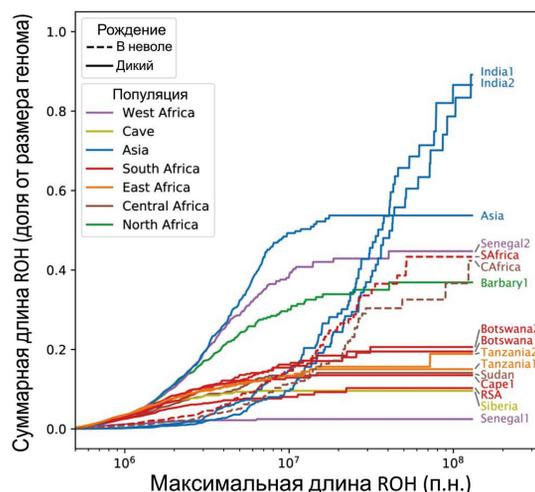
Среди диаграмм распределения стоит выделить график кумулятивной суммы, который является наиболее информативным и частым способом визуализации длины участков ROH и реже ROHnet. Например, данный тип графического представления был использован в работе по изучению львов из различных популяций (рис. 4, *a*) [63]. На графике по оси *x* отложена общая длина ROH, а по оси *y* — доля от всего генома. Таким образом, удастся показать суммарную (кумулятивную) долю генома, содержащегося в ROH, с учетом отсечки по длине. Стоит отметить, что для визуализации распределения длин фрагментов ROH также часто используют диаграммы размаха. Так, в работе по изучению генома оленя Давида смогли показать среднюю, максимальную и минимальную длину участков ROH для исследуемых групп особей [47].

### Линейные графики и тепловые карты

При оценке гетерозиготности в окнах на уровне генома отдельной особи важное значение имеет уровень используемой в качестве референса геномной сборки. Так, при использовании сборки скаффолдного уровня, рассчитанные в окнах оценки можно отобразить в виде протяженного линейного графика. Появление геномных сборок хромосомного уровня открыло новые возможности как для анализа геномных перестроек, так и для изображения гетерозиготности [64, 65]. В частности, стало возможно отрисовывать оценки гетерозиготности в виде тепловых карт (heat map), что позволило более наглядно отображать изменения гетерозиготности на уровне генома каждой особи. Как и в случае диаграмм распределения, исследователям необходимо владеть навыками

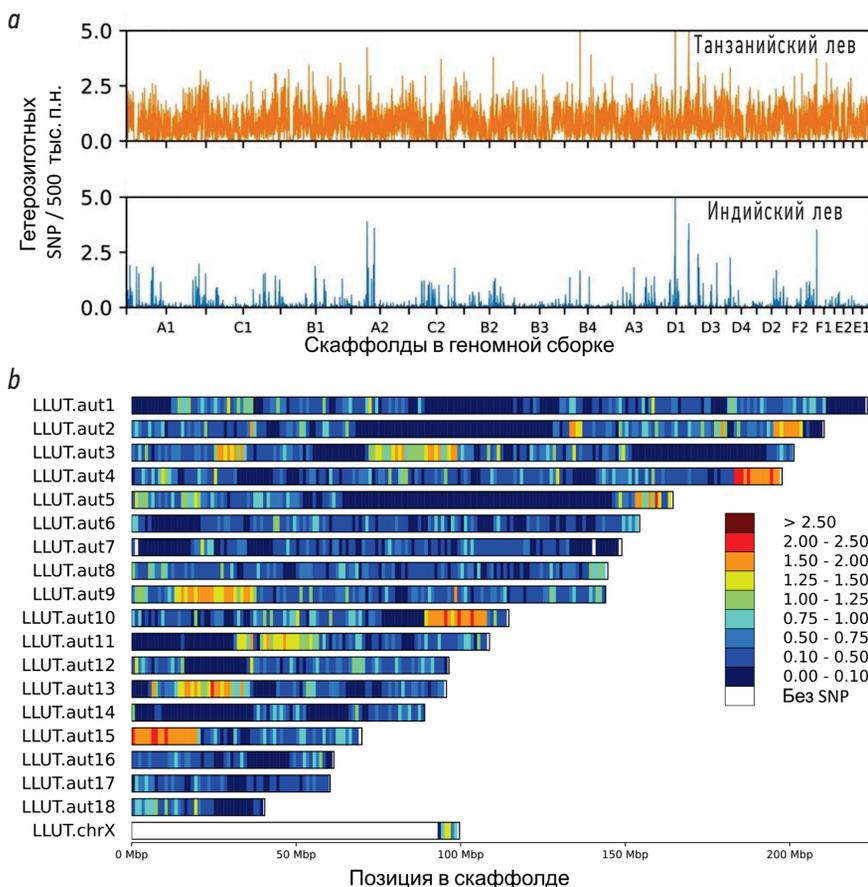
программирования для обработки геномных данных и вычисления оценок средней гетерозиготности в окнах. Помимо этого, стоит учитывать, что визуализация в виде тепловых карт возможна исключительно с использованием специализированных инструментов и библиотек, что обусловлено специфичностью такого способа.

Линейные графики являются частым способом визуализации, поскольку они позволяют отобразить изменения уровня гетерозиготности относительно координат геномной сборки. По оси *x*, согласно ориентации в сборке генома, откладываются последовательно объединенные скаффолды, а по оси *y* — уровень гетерозиготности. Такой подход позволяет проводить оценку как на уровне каждого скаффолда, так и для всего генома. Например, это было выполнено в работе по изучению генетического разнообразия популяций танзанийского и индийского львов (рис. 5, *a*) [63]. Линейные графики также встречаются в исследовании особенностей геномики морской выдры, в публикации геномной сборки хромосомного уровня для домашней собаки и в работе, посвященной



**Рис. 4.** График кумулятивной суммы ROH при различном пороге отсеки по длине для львов из различных популяций. Оригинальное изображение из [63]

**Fig. 4.** Cumulative distribution plot of ROH at different length cut-off threshold for lions from different populations. Original image from [63]



**Рис. 5.** Примеры визуализации средней гетерозиготности в окнах при помощи линейных графиков и тепловой карты: *a* — средняя гетерозиготность в скользящих окнах размером 500 тыс. п. н. у танзанийского и индийского львов по геномной сборке африканского льва. Хромосомы последовательно объединены по оси *x*; *b* — тепловая карта плотности гетерозиготных SNP на основе геномной сборки хромосомного уровня для самца евразийской речной выдры. Гетерозиготные SNP подсчитывались в окнах размером 1 млн п. н. и масштабировались в SNP / 1 тыс. п. н. Оригинальные изображения из [61, 63]

**Fig. 5.** Examples of visualization of mean heterozygosity in windows using line plots and heat maps: *a* — mean heterozygosity in 500 kbp sliding windows in Tanzanian lion (orange plot) and Indian lion (blue plot) based on African lion genome assembly. Chromosomes are sequentially merged along the X axis; *b* — heat map of heterozygous SNP density based on chromosome-level genomic assembly for male Eurasian river otter. Heterozygous SNPs were counted in 1 million bp windows and scaled to SNP / 1 thousand bp. Original images from [61, 63]

исследованию волков с острова Айл-Ройал [44, 48, 59]. Важно учитывать, что данный способ визуализации маскирует локальные отклонения оценок гетерозиготности из-за мелкого масштаба, поскольку весь геном представлен в виде одного непрерывного графика. Тем не менее преимущество подхода заключается в возможности использования в случае, если для исследуемого вида отсутствует геномная сборка хромосомного уровня. Отметим, что данный вариант отлично подойдет для демонстрации резких изменений оценок гетерозиготности при чередовании участков ROH и ROHet.

Один из наиболее наглядных способов графического представления гетерозиготности для видов с доступной сборкой генома хромосомного уровня — тепловые карты, представляющие собой наглядное изображение плотности гетерозиготных или гомозиготных SNP, подсчитанных в окнах фиксированного размера, при помощи ранжированной цветовой палитры [66]. Значение оценки средней гетерозиготности каждого окна отображается при помощи соответствующего для данного диапазона цвета, то есть происходит так называемое бинирование оценок. Это ключевое отличие тепловых карт от всех вышеперечисленных способов визуализации, поскольку в данном случае происходит переход к кодировке уровня гетерозиготности при помощи цвета, что способствует уменьшению размера графика с минимальной потерей информации.

Очевидно, выполнить визуализацию гетерозиготности по хромосомам можно и в виде линейного графика, представив их не последовательно на оси  $x$ , а в виде обособленных подграфиков. Однако в таком случае ось  $y$  для каждого из них придется значительно увеличить для наглядности, вследствие чего общий график сильно вытянется по вертикали во избежание перекрытий. При использовании геномных сборок хромосомного уровня хромосомные скаффолды поочередно отрисовываются в виде треков — протяженных прямоугольников, каждый из которых представляет собой отдельную хромосому. Как следствие, длина каждого из треков пропорциональна протяженности соответствующего хромосомного скаффолда в геномной сборке. Таким образом, количество треков соответствует гаплоидному набору хромосом исследуемого вида. Полученные треки можно рассматривать как отдельно, так и в совокупности, что позволяет не только определить координаты участков с повышенной или пониженной гетерозиготностью, но и оценить общий уровень гетерозиготности всего генома.

Тепловые карты открывают массу возможностей для последующего анализа, благодаря чему активно используется в современных исследованиях. Так, отображение гетерозиготности подобным способом было выполнено в работе по изучению восьми видов млекопитающих, шесть из которых занесены в Красную книгу Международного союза охраны природы [61]. В результате, в геноме евразийской речной выдры удалось выявить

высокую гетерозиготность лишь в нескольких хромосомных участках (рис. 5, *b*), что позволило усомниться в ранее установленном глобальном статусе этого вида. Стоит обратить внимание, что справа на представленном рисунке располагается легенда с диапазонами оценок гетерозиготности, которая является необходимой в случае визуализации при помощи тепловой карты. Данный подход графического отображения гетерозиготности встречается и в других публикациях [58, 59, 67]. Распространению такого подхода главным образом способствует удешевление технологий по получению качественных полногеномных данных с высоким покрытием и появление геномных сборок хромосомного уровня.

Таким образом, важно подчеркнуть, что способ визуализации зависит, во-первых, от целей исследования, во-вторых, от доступных геномных данных и, в-третьих, от уровня, на котором хочется оценить гетерозиготность. Хотя в представленных выше способах прослеживается развитие, все описанные подходы остаются актуальными и комплементарными друг другу, поскольку позволяют оценивать гетерозиготность на различных геномных уровнях. В современных работах исследователи, как правило, не ограничиваются одним подходом, поскольку каждый из них позволяет посмотреть на гетерозиготность под разными углами. Ключевые особенности вышеизложенных методов представлены в таблице.

## РАЗВИТИЕ ИНСТРУМЕНТОВ ДЛЯ ВИЗУАЛИЗАЦИИ ГЕТЕРОЗИГОТНОСТИ

Чтобы лучше следить за прогрессом подходов для визуализации гетерозиготности, нужно сначала проследить, как новые технологии, вопросы и тенденции визуализации геномных данных формировались на протяжении последних лет. Вплоть до 1990-х годов локальные и глобальные алгоритмы парного и множественного выравнивания последовательностей, такие как алгоритм Смита – Ватермана и алгоритм Нидлмана – Вунша, были в центре внимания при разработке методов биоинформатики [68, 69]. На их основе были созданы многочисленные инструменты для выравнивания последовательностей, такие как MUSCLE, T-Coffee, MAFFT и др. [70–72]. Поскольку данные алгоритмы были не рассчитаны на большие объемы данных, результаты выравнивания сохранялись в интуитивно понятных строках в текстовом формате с ASCII-кодировкой. Многие активно применяемые инструменты (например, BLAST) до сих пор не лишены опций для визуализации результатов выравнивания последовательностей в подобном виде [73]. Позднее многие программные решения все же применяли базовые схемы получения изображения, включающие использование цвета для отображения совпадающих и несовпадающих нуклеотидов или, чаще, аминокислот в выравниваниях белковых последовательностей [74]. Хотя эти инструменты оказались успешными для небольшого количества

**Таблица.** Характерные особенности способов визуализации гетерозиготности

**Table.** Characteristics of heterozygosity visualization approaches

Тип диаграммы	Требуемый тип входных данных	Возможности	Особенности
Столбчатая диаграмма	Количество SNP для множества особей	Сравнение количества гетеро- и гомозиготных SNP и различных метрик гетерозиготности	Универсальный способ визуализации для сравнительного анализа. Не отражает локализацию SNP
Круговая диаграмма		Подходит только для отображения значимых различий в оценках. Не отражает локализацию SNP	
Диаграммы Венна		Визуализация уникальных и общих SNP	Подходит для визуализации множеств SNP для не более 6–10 особей
Гистограмма	1. Множество значений для нескольких особей из одной популяции. 2. Плотности SNP по окнам для одной особи	Визуализация распределений оценок гетерозиготности	Плохо подходит для сравнений распределений нескольких особей. Не отражает локализацию SNP.
Диаграмма размаха		Отображает среднее/медианное значение, нижний и верхний квартили, минимальное и максимальное значение, а также выбросы набора оценок для нескольких особей или плотностей SNP для одной особи	Подходит для визуализации распределений для нескольких особей на одном графике. Не отражает локализацию SNP
Скрипичная диаграмма		Диаграмма аналогична диаграмме размаха. Подходит для отображения мультимодальных распределений. Не отражает локализацию SNP	
Кумулятивный график	Длины участков ROH/ROH <sub>et</sub> для одной особи	Отображает как количество, так и протяженность геномных участков	Подходит для формирования гипотез о состоянии и структуре популяции на основании траектории графика. Не отражает локализацию SNP
Линейный график	Плотности SNP по окнам для одной особи	Отображает оценки гетерозиготности на уровне геномной сборки любого уровня. Подходит в случае, если важно показать чередование регионов ROH и ROH <sub>et</sub>	Из-за низкого разрешения маскирует локальные изменения гетерозиготности в пределах скаффолда
Тепловая карта		Как и линейный график, подходит как для отображения чередования регионов ROH и ROH <sub>et</sub> , так и для общей оценки геномной гетерозиготности. Требуется геномная сборка хромосомного уровня	Подходит для идентификации участков с повышенной или пониженной гетерозиготностью. Плохо подходит для сравнительного анализа гетерозиготности нескольких особей

коротких последовательностей, возник вопрос относительно их применимости для данных полногеномного секвенирования больших геномов.

Проект «1000 геномов», в рамках которого был выполнен всесторонний анализ генетических вариаций человека, и международный проект «НарМар» для выявления общих генетических вариаций среди людей из разных стран, являются лишь несколькими примерами резкого увеличения объема биологических данных после 2005 г. [75, 76]. Такие крупномасштабные наборы геномных данных потребовали мощных инструментов для обработки и графического представления получаемых результатов. Это стало началом для активной разработки геномных браузеров, предназначенных для обработки больших объемов информации, визуализации геномных сборок и интеграции с аннотациями из различных репозиториях [77]. В настоящее время доступно и широко используется множество геномных браузеров: UCSC Genome Browser, IGV, Ensembl [78–80]. Подобные программные решения позволяют изображать как геномные сборки и выравнивания

прочтений, так и геномные перестройки и генетические вариации. Отдельно стоит выделить Uiprgo UGENE — универсальный российский программный пакет с графическим интерфейсом, позволяющий не только наглядно представлять полученные результаты, но и выполнять обработку данных секвенирования: осуществлять аннотацию, выравнивание последовательностей и многое другое, вплоть до реконструкции филогенетических деревьев [81].

Хотя вышеперечисленные программы хорошо рекомендовали себя при работе с большими объемами коротких последовательностей отдельных геномных локусов, они остаются практически неприменимы при работе с большими объемами полногеномных данных. Более того, современные исследования в области природоохранной биологии зачастую подразумевают обработку множества образцов, что внесло значительные изменения в работу исследователя. В частности, оценка гетерозиготности с использованием полногеномных данных осуществляется на высокопроизводительных вычислительных кластерах без графического интерфейса.

И хотя существует множество инструментов по оценке гетерозиготности, реализованных в биоинформатических программных инструментариях по работе с файлами в формате VCF (variant call format), таких как VCFtools, функционал для графического представления полученных результатов остается без внимания [82]. Исключением можно считать BCFtools, который включает в себя не только функцию для поиска ROH, но и скрипт для представления числовой информации в виде линейного графика [83]. Однако подобные решения редко используются исследователями для изображения полученных результатов по причине своей низкой функциональности и гибкости. Наиболее частый сценарий исследовательской работы, это разделение этапа вычисления различных оценок гетерозиготности и этапа визуализации.

Известно, что схема анализа геномных данных включает множество этапов по выявлению, контролю качества и фильтрации SNP. В целях экономии времени и возможности воспроизведения результатов возникла необходимость автоматизации вычислений посредством скриптовых языков программирования, таких как Python и R [84, 85]. Таким образом, в настоящее время наибольшее предпочтение уделяется программным пакетам для визуализации, которые расширяют функционал языков программирования и которые можно добавить в виде обособленного скрипта в общую схему анализа гетерозиготности в качестве финального этапа.

Одним из таких программных пакетов является MACE, основанный на библиотеке Matplotlib языка программирования Python [86]. Данный пакет включает в себя программные скрипты для оценки средней гетерозиготности в окнах и последующим получением изображения в виде диаграмм распределения и тепловых карт. Еще один инструмент — пакет JLOH, используемый для выявления, фильтрации и анализа фрагментов генома с низким уровнем гетерозиготности [87]. Стоит также отметить пакеты для языка R, в частности, пакет karyoploteR, который содержит множество инструментов для визуализации геномных данных, в том числе оценок гетерозиготности [88]. Для графического представления участков ROH и ROHet существуют отдельные пакеты для языка R, к которым можно отнести RZooRoH, HandyCNV, detectRUNS [89–91]. Важно сказать, что несмотря на существование общепринятых форматов VCF и BED, созданных для хранения генетических вариаций и их местоположений в геноме, каждый из вышеперечисленных инструментов имеет свой специфический формат входных данных.

Отметим, что пакеты и библиотеки для визуализации часто используются в универсальных средах разработки, которые позволяют писать код и представлять информацию в интерактивном режиме. В случае Python к таким средам можно отнести Jupyter Notebook, а в случае R — Rstudio [92, 93]. Работа в данных средах разработки напоминает написание отчетов, разделенных на блоки кода, визуализацию и дополнительные комментарии.

## ЗАКЛЮЧЕНИЕ

Область визуализации биологических данных активно развивается по мере увеличения сложности и объема результатов, полученных с использованием полногеномных данных. В последние годы, преимущественно благодаря возросшим объемам таких данных, наблюдается тенденция к изучению генетического разнообразия на всех возможных уровнях. Как следствие, у исследователей возникает потребность в универсальных инструментариях как для оценки, так и для графического представления гетерозиготности. Большинство существующих на данный момент инструментов реализованы в виде программных пакетов, расширяющих возможности общих библиотек для формирования изображения. К сожалению, их установка и использование требует определенных знаний от исследователя. Помимо базовых навыков программирования, зачастую обязательным требованием является умение выполнять вычисления на удаленных высокопроизводительных кластерах, что обусловлено большим объемом геномных данных (десятки терабайт), обрабатывать которые на локальном персональном компьютере не представляется возможным. Проблема также заключается в отсутствии унифицированного формата для хранения оценок гетерозиготности, в особенности рассчитанных по окнам фиксированного размера. И хотя существует расширяемый формат BED в качестве неофициального стандарта, он претерпевает всяческие изменения в зависимости от пакета, для которого он используется в качестве формата входных данных. Это приводит к необходимости самостоятельно разрабатывать инструменты, программные пакеты и даже форматы хранения данных, что невыгодно с точки зрения затраченного времени и усилий. Кроме того, в связи с отсутствием стандартизированных подходов, возникает необходимость изучения многочисленных инструкций под каждый используемый инструмент, что не менее существенно замедляет процесс анализа. Как следствие, в отношении визуализации гетерозиготности наблюдается тенденция к разработке новых или модернизации существующих инструментов под конкретную исследовательскую задачу.

Разработка и популяризация автоматизированных конвейеров по обработке данных о генетическом разнообразии и стандартизация форматов хранения полученных метрик становятся ключевыми задачами в улучшении доступности и обмена информацией между исследователями. Их реализация позволит ускорить научные исследования и улучшить интерпретацию результатов, что в конечном итоге будет способствовать разработке новых подходов по анализу генетической изменчивости. Будущее исследований в области природоохранной биологии зависит от тесного сотрудничества между генетиками и биоинформатиками с целью создания общих стандартов и ресурсов, которые позволят более эффективно исследовать и визуализировать гетерозиготность.

## ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

**Вклад авторов.** Все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией. Личный вклад каждого автора: А.А. Томаровский — обзор литературы, написание текста, обработка, анализ и визуализация литературных данных; А.А. Тотиков — обработка и анализ литературных данных, написание текста; А.Р. Якупова — обзор литературы, визуализация литературных данных; А.С. Графодатский — концепция и дизайн работы; С.Ф. Кливер — концепция и дизайн работы, написание и редактирование текста.

**Источник финансирования.** Исследование проведено при поддержке Российского научного фонда (грант № 19-14-00034-П).

**Конфликт интересов.** Авторы заявляют об отсутствии потенциального конфликта интересов, требующего раскрытия в данной статье.

## ADDITIONAL INFO

**Authors' contribution.** All of the authors made a substantial contribution to the conception of the study, acquisition, analysis, interpretation of data for the work, drafting and revising the article, final approval of the version to be published, and agree to be accountable for all aspects of the study. The contributions of each author: A.A. Tomarovsky — literature review, text writing, processing, analysis and visualization of literary data; A.A. Totikov — processing and analysis of literary data, text writing; A.R. Yakupova — literature review, visualization of literary data; A.S. Graphodatsky — concept and design of the work; S.F. Kliver — concept and design of the work, writing and editing of the text.

**Funding source.** The study has been supported by Russian Science Foundation (grant No. 19-14-00034-П).

**Competing interests.** The authors declare that they have no competing interests.

## СПИСОК ЛИТЕРАТУРЫ

1. Soulé M.E. What is conservation biology? A new synthetic discipline addresses the dynamics and problems of perturbed species, communities, and ecosystems // *BioSci*. 1985. Vol. 35, No. 11. P. 727–734. DOI: 10.2307/1310054
2. Fuentes-Pardo A.P., Ruzzante D.E. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations // *Mol Ecol*. 2017. Vol. 26, No. 20. P. 5369–5406. DOI: 10.1111/mec.14264
3. Hoban S., Kelley J.L., Lotterhos K.E., et al. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions // *Am Nat* the University of Chicago Press. 2016. Vol. 188, No. 4. P. 379–397. DOI: 10.1086/688018
4. Hoban S., da Silva J.M., Mastretta-Yanes A., et al. Monitoring status and trends in genetic diversity for the Convention on Biological Diversity: An ongoing assessment of genetic indicators in nine countries // *Conserv Lett*. 2023. Vol. 16, No. 3. ID e12953. DOI: 10.1111/cont.12953
5. Ng P.C., Kirkness E.F. Whole genome sequencing. In: Genetic variation: methods and protocols / M.R. Barnes, G. Breen, editors. Totowa, NJ: Humana Press, 2010. P. 215–226. DOI: 10.1007/978-1-60327-367-1\_12
6. Breed M.F., Harrison P.A., Blyth C., et al. The potential of genomics for restoring ecosystems and biodiversity. 10 // *Nat Rev Genet*. 2019. Vol. 20, No. 10. P. 615–628. DOI: 10.1038/s41576-019-0152-0
7. Кливер С.Ф. Полногеномный подход в природоохранной биологии и его перспективы // *Экологическая генетика*. 2021. Т. 19, № 3. С. 281–298. DOI: 10.17816/ecogen65152
8. Joop Ouborg N., Angeloni F., Vergeer P. An essay on the necessity and feasibility of conservation genomics // *Conserv Genet*. 2010. Vol. 11, No. 2. P. 643–653. DOI: 10.1007/s10592-009-0016-9
9. Dudchenko O., Shamim M.S., Batra S.S., et al. The Juicebox assembly tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000 // *bioRxiv*. 2018. ID 254797. DOI: 10.1101/254797
10. Durand N.C., Robinson J.T., Shamim M.S., et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom // *Cell Syst*. 2016. Vol. 3, No. 1. P. 99–101. DOI: 10.1016/j.cels.2015.07.012
11. Luikart G., England P.R., Tallmon D., et al. The power and promise of population genomics: from genotyping to genome typing. 12 // *Nat Rev Genet*. 2003. Vol. 4, No. 12. P. 981–994. DOI: 10.1038/nrg1226
12. Campbell M.R., Vu N.V., LaGrange A.P., et al. Development and Application of single-nucleotide polymorphism (SNP) genetic markers for conservation monitoring of burbot populations // *Trans Am Fish Soc*. 2019. Vol. 148, No. 3. P. 661–670. DOI: 10.1002/tafs.10157
13. Bijlsma R., Loeschcke V. Genetic erosion impedes adaptive responses to stressful environments // *Evol Appl*. 2012. Vol. 5, No. 2. P. 117–129. DOI: 10.1111/j.1752-4571.2011.00214.x
14. Leroy G., Carrol E.L., Bruford M.W., et al. Next-generation metrics for monitoring genetic erosion within populations of conservation concern // *Evol Appl*. 2018. Vol. 11, No. 7. P. 1066–1083. DOI: 10.1111/eva.12564
15. Frankham R., Ballou J.D., Eldridge M.D., et al. Predicting the probability of outbreeding depression // *Conserv Biol*. 2011. Vol. 25, No. 3. P. 465–475. DOI: 10.1111/j.1523-1739.2011.01662.x
16. Charlesworth D., Willis J.H. The genetics of inbreeding depression. 11 // *Nat Rev Genet*. 2009. Vol. 10, No. 11. P. 783–796. DOI: 10.1038/nrg2664
17. Майр Э. Популяции, виды и эволюция. Москва: Мир, 1968. 398 с.
18. Tomimatsu H., Ohara M. Genetic diversity and local population structure of fragmented populations of *Trillium camschatcense* (Trilliaceae) // *Biol Conserv*. 2003. Vol. 109, No. 2. P. 249–258. DOI: 10.1016/S0006-3207(02)00153-2
19. Hanski I. The Shrinking world: Ecological consequences of habitat loss // *Excell Ecol*. 2005. Vol. 14.
20. Lande R., Barrowclough G. Effective population size, genetic variation, and their use in population management. In: Soulé M., editor. Viable populations for conservation. Cambridge: Cambridge University Press, 1987. P. 87–124. DOI: 10.1017/CBO9780511623400.007
21. Wright S. Random drift and the shifting balance theory of evolution. In: Kojima K., editor. Mathematical topics in population genetics. Berlin, Heidelberg: Springer, 1970. P. 1–31. DOI: 10.1007/978-3-642-46244-3\_1
22. Nevo E. Genetic variation in natural populations: Patterns and theory // *Theor Popul Biol*. 1978. Vol. 13, No. 1. P. 121–177. DOI: 10.1016/0040-5809(78)90039-4
23. Левонтин П. Генетические основы эволюции. Москва: Мир, 1978. 351 с.

24. Steiner C.C., Putnam A.S., Hoeck P.E.A., Ryder A. Conservation genomics of threatened animal species // *Annu Rev Anim Biosci*. 2013. Vol. 1, No. 1. P. 261–281. DOI: 10.1146/annurev-animal-031412-103636
25. Weir B.S. Genetic data analysis II: Methods for discrete population genetic data. Oxford, New York: Oxford University Press, 1996. 445 p.
26. Ritland K. Estimators for pairwise relatedness and individual inbreeding coefficients // *Genet Res*. 1996. Vol. 67, No. 2. P. 175–185. DOI: 10.1017/S0016672300033620
27. Nei M., Li W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases // *PNAS*. 1979. Vol. 76, No. 10. P. 5269–5273. DOI: 10.1073/pnas.76.10.5269
28. Wright S. The interpretation of population structure by  $F$ -statistics with special regard to systems of mating // *Evolution*. 1965. Vol. 19, No. 3. P. 395–420. DOI: 10.2307/2406450
29. Shafer A.B.A., Wolf J.B.W., Alves P.C., et al. Genomics and the challenging translation into conservation practice // *Trends Ecol Evol*. 2015. Vol. 30, No. 2. P. 78–87. DOI: 10.1016/j.tree.2014.11.009
30. Hoffmann A., Griffin P., Dillon S., et al. A framework for incorporating evolutionary genomics into biodiversity conservation and management // *Clim Change Responses*. 2015. Vol. 2, No. 1. ID 1. DOI: 10.1186/s40665-014-0009-x
31. Benestan L.M., Ferchaud A.-L., Hohenlohe P.A., et al. Conservation genomics of natural and managed populations: building a conceptual and practical framework // *Mol Ecol*. 2016. Vol. 25, No. 13. P. 2967–2977. DOI: 10.1111/mec.13647
32. Hoban S., Gaggiotti O., ConGRESS Consortium, Bertorelle G. Sample planning optimization tool for conservation and population genetics (SPOTG): a software for choosing the appropriate number of markers and samples // *Methods Ecol Evol*. 2013. Vol. 4, No. 3. P. 299–303. DOI: 10.1111/2041-210x.12025
33. Nazareno A.G., Bemmels J.B., Dick C.W., Lohmann L.G. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species // *Mol Ecol Resour*. 2017. Vol. 17, No. 6. P. 1136–1147. DOI: 10.1111/1755-0998.12654
34. Gibson J., Morton N.E., Collins A. Extended tracts of homozygosity in outbred human populations // *Hum Mol Genet*. 2006. Vol. 15, No. 5. P. 789–795. DOI: 10.1093/hmg/ddi493
35. McQuillan R., Leutenegger A.-L., Abdel-Rahman R., et al. Runs of homozygosity in European populations // *Am J Hum Genet*. 2008. Vol. 83, No. 3. P. 359–372. DOI: 10.1016/j.ajhg.2008.08.007
36. Darwin C. The effects of cross and self fertilisation in the vegetable kingdom. Ams Press Inc, 1877. DOI: 10.5962/bhl.title.104481
37. Ceballos F.C., Joshi P.K., Clark D.W., et al. Runs of homozygosity: windows into population history and trait architecture: 4 // *Nat Rev Genet*. 2018. Vol. 19, No. 4. P. 220–234. DOI: 10.1038/nrg.2017.109
38. Hoffman J.I., Simpson F., David P., et al. High-throughput sequencing reveals inbreeding depression in a natural population // *PNAS*. 2014. Vol. 111, No. 10. P. 3775–3780. DOI: 10.1073/pnas.1318945111
39. Muir W.M., Wong G.K.-S., Zhang Y., et al. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds // *PNAS*. 2008. Vol. 105, No. 45. P. 17312–17317. DOI: 10.1073/pnas.0806569105
40. Urbinati I., Stafuzza N.B., Oliveira M.T., et al. Selection signatures in Canchim beef cattle // *J Anim Sci Biotechnol*. 2016. Vol. 7, No. 1. ID 29. DOI: 10.1186/s40104-016-0089-5
41. Samuels D.C., Wang J., Ye K., et al. Heterozygosity ratio, a robust global genomic measure of autozygosity and its association with height and disease risk // *Genetics*. 2016. Vol. 204, No. 3. P. 893–904. DOI: 10.1534/genetics.116.189936
42. Li N., Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data // *Genetics*. 2003. Vol. 165, No. 4. P. 2213–2233. DOI: 10.1093/genetics/165.4.2213
43. Rife D.C. Populations of hybrid origin as source material for the detection of linkage // *Am J Hum Genet*. 1954. Vol. 6, No. 1. P. 26–33.
44. Robinson J.A., Rääkkönen J., Vucetich L.M., et al. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction // *Sci Adv*. 2019. Vol. 5, No. 5. ID eaau0757. DOI: 10.1126/sciadv.aau0757
45. Koepfli K.-P., Tamazian G., Wildt D., et al. Whole genome sequencing and re-sequencing of the sable antelope (*Hippotragus niger*): A resource for monitoring diversity in *ex situ* and *in situ* populations // *G3 Genes Genomes Genetics*. 2019. Vol. 9, No. 6. P. 1785–1793. DOI: 10.1534/g3.119.400084
46. Большая Советская Энциклопедия. Т. 20. 3-е изд. 1974. С. 25.
47. Zhu L., Deng C., Zhao X., et al. Endangered Père David's deer genome provides insights into population recovering // *Evol Appl*. 2018. Vol. 11, No. 10. P. 2040–2053. DOI: 10.1111/eva.12705
48. Beichman A.C., Koepfli K.-P., Li G., et al. Aquatic adaptation and depleted diversity: A Deep dive into the genomes of the sea otter and giant otter // *Mol Biol Evol*. 2019. Vol. 36, No. 12. P. 2631–2655. DOI: 10.1093/molbev/msz101
49. Abascal F., Corvelo A., Cruz F., et al. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx // *Genome Biol*. 2016. Vol. 17, No. 1. ID 251. DOI: 10.1186/s13059-016-1090-1
50. Cho Y.S., Hu L., Hou H., et al. The tiger genome and comparative analysis with lion and snow leopard genomes: 1 // *Nat Commun*. 2013. Vol. 4, No. 1. ID 2433. DOI: 10.1038/ncomms3433
51. Miller W., Schuster S.C., Welch A.J., et al. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change // *PNAS*. 2012. Vol. 109, No. 36. P. E2382–E2390. DOI: 10.1073/pnas.1210506109
52. Venn J.I. On the diagrammatic and mechanical representation of propositions and reasonings // *Lond Edinb Dublin Philos Mag J Sci*. 1880. Vol. 10, No. 59. P. 1–18. DOI: 10.1080/14786448008626877
53. Miller W., Hayes V.M., Ratan A., et al. Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil) // *PNAS*. 2011. Vol. 108, No. 30. P. 12348–12353. DOI: 10.1073/pnas.1102838108
54. Humble E., Dobrynin P., Senn H., et al. Chromosomal-level genome assembly of the scimitar-horned oryx: Insights into diversity and demography of a species extinct in the wild // *Mol Ecol Resour*. 2020. Vol. 20, No. 6. P. 1668–1681. DOI: 10.1111/1755-0998.13181
55. Yakupova A., Tomarovsky A., Totikov A., et al. Chromosome-length assembly of the baikal seal (*Pusa sibirica*) genome reveals a historically large population prior to isolation in Lake Baikal: 3 // *Genes*. 2023. Vol. 14, No. 3. ID 619. DOI: 10.3390/genes1403061
56. Kliver S., Houk M.L., Perelman P.L., et al. Chromosome-length genome assembly and karyotype of the endangered black-footed ferret (*Mustela nigripes*) // *J Hered*. 2023. Vol. 114, No. 5. P. 539–548. DOI: 10.1093/jhered/esad035
57. Li R., Fan W., Tian G., et al. The sequence and *de novo* assembly of the giant panda genome: 7279 // *Nature*. 2010. Vol. 463, No. 7279. P. 311–317. DOI: 10.1038/nature08696

58. Dobrynin P., Liu S., Tamazian G., et al. Genomic legacy of the African cheetah, *Acinonyx jubatus* // *Genome Biol.* 2015. Vol. 16, No. 1. ID 277. DOI: 10.1186/s13059-015-0837-4
59. Lindblad-Toh K., Wade C.M., Mikkelsen T.S., et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog: 7069 // *Nature.* 2005. Vol. 438, No. 7069. P. 803–819. DOI: 10.1038/nature04338
60. Benjamini Y. Opening the box of a boxplot // *Am Stat.* 1988. Vol. 42, No. 4. P. 257–262. DOI: 10.2307/2685133
61. Totikov A., Tomarovsky A., Prokopov D., et al. Chromosome-level genome assemblies expand capabilities of genomics for conservation biology: 9 // *Genes.* 2021. Vol. 12, No. 9. ID 1336. DOI: 10.3390/genes12091336
62. Hintze J.L., Nelson R.D. Violin plots: A box plot-density trace synergism // *Am Stat.* 1998. Vol. 52, No. 2. P. 181–184. DOI: 10.1080/00031305.1998.10480559
63. de Manuel M., Barnett R., Sandoval-Velasco M., et al. The evolutionary history of extinct and living lions // *PNAS USA.* 2020. Vol. 117, No. 20. P. 10927–10934. DOI: 10.1073/pnas.1919423117
64. Burton J.N., Adey A., Patwardhan R.P., et al. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions: 12 // *Nat Biotechnol.* 2013. Vol. 31, No. 12. P. 1119–1125. DOI: 10.1038/nbt.2727
65. Lewin H.A., Graves J.A.M., Ryder O.A., et al. Precision nomenclature for the new genomics // *GigaScience.* 2019. Vol. 8, No. 8. ID giz086. DOI: 10.1093/gigascience/giz086
66. Wilkinson L., Friendly M. The history of the cluster heat map // *Am Stat.* 2009. Vol. 63, No. 2. P. 179–184. DOI: 10.1198/tas.2009.0033
67. de Ferran V., Figueiro H.V., de Jesus Trindade F., et al. Phylogenomics of the world's otters // *Curr Biol.* 2022. Vol. 32, No. 16. P. 3650–3658.e4. DOI: 10.1016/j.cub.2022.06.036
68. Needleman S.B., Wunsch C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins // *J Mol Biol.* 1970. Vol. 48, No. 3. P. 443–453. DOI: 10.1016/0022-2836(70)90057-4
69. Smith T.F., Waterman M.S. Identification of common molecular subsequences // *J Mol Biol.* 1981. Vol. 147, No. 1. P. 195–197. DOI: 10.1016/0022-2836(81)90087-5
70. Katoh K., Standley D.M. MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability // *Mol Biol Evol.* 2013. Vol. 30, No. 4. P. 772–780. DOI: 10.1093/molbev/mst010
71. Magis C., Taly J.-F., Bussotti G., et al. T-Coffee: tree-based consistency objective function for alignment evaluation. In: Russell D.J., editor. *Multiple sequence alignment methods.* Totowa, NJ: Humana Press, 2014. P. 117–129. DOI: 10.1007/978-1-62703-646-7\_7
72. Edgar R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity // *BMC Bioinformatics.* 2004. Vol. 5, No. 1. ID 113. DOI: 10.1186/1471-2105-5-113
73. Altschul S.F., Gish W., Miller W., et al. Basic local alignment search tool // *J Mol Biol.* 1990. Vol. 215, No. 3. P. 403–410. DOI: 10.1016/S0022-2836(05)80360-2
74. Thompson J.D., Gibson T.J., Higgins D.G. Multiple sequence alignment using ClustalW and ClustalX // *Curr Protoc Bioinforma.* 2003. Vol. 1. P. 2–3. DOI: 10.1002/0471250953.bi0203s00
75. Altshuler D., Donnelly P.; The International HapMap Consortium. A haplotype map of the human genome: 7063 // *Nature.* 2005. Vol. 437, No. 7063. P. 1299–1320. DOI: 10.1038/nature04226
76. Durbin R.M.; The International HapMap Consortium, et al. A map of human genome variation from population-scale sequencing: 7319 // *Nature.* 2010. Vol. 467, No. 7319. P. 1061–1073. DOI: 10.1038/nature09534
77. Nusrat S., Harbig T., Gehlenborg N. Tasks, techniques, and tools for genomic data visualization // *Comput Graph Forum.* 2019. Vol. 38, No. 3. P. 781–805. DOI: 10.1111/cgf.13727
78. Karolchik D., Baertsch R., Diekhans M., et al. The UCSC genome browser database // *Nucleic Acids Res.* 2003. Vol. 31, No. 1. P. 51–54. DOI: 10.1093/nar/gkg129
79. Thorvaldsdóttir H., Robinson J.T., Mesirov J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration // *Brief Bioinform.* 2013. Vol. 14, No. 2. P. 178–192. DOI: 10.1093/bib/bbs017
80. Yates A.D., Achuthan P., Akanni W., et al. Ensembl 2020 // *Nucleic Acids Res.* 2020. Vol. 48, No. D1. P. D682–D688. DOI: 10.1093/nar/gkz966
81. Okonechnikov K., Golosova O., Fursov M., et al. Unipro UGENE: a unified bioinformatics toolkit // *Bioinformatics.* 2012. Vol. 28, No. 8. P. 1166–1167. DOI: 10.1093/bioinformatics/bts091
82. Danecek P., Auton A., Abecasis G., et al. The variant call format and VCFtools // *Bioinformatics.* 2011. Vol. 27, No. 15. P. 2156–2158. DOI: 10.1093/bioinformatics/btr330
83. Narasimhan V., Danecek P., Scally A., et al. BCftools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data // *Bioinformatics.* 2016. Vol. 32, No. 11. P. 1749–1751. DOI: 10.1093/bioinformatics/btw044
84. Ihaka R., Gentleman R. R: a language for data analysis and graphics // *J Comput Graph Stat.* 1996. Vol. 5, No. 3. P. 299–314. DOI: 10.1080/10618600.1996.10474713
85. van Rossum G. Python reference manual // *Dep Comput Sci.* 1995. ID R9525.
86. Hunter J.D. Matplotlib: A 2D Graphics environment // *Comput Sci.* 2007. Vol. 9, No. 3. P. 90–95. DOI: 10.1109/MCSE.2007.55
87. Schiavinato M., del Olmo V., Muya V.N., Gabaldon T. JLOH: Inferring loss of heterozygosity blocks from sequencing data // *bioRxiv.* 2023. ID 2023.05.04.539368. DOI: 10.1101/2023.05.04.539368
88. Gel B., Serra E. KaryoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data // *Bioinformatics.* 2017. Vol. 33, No. 19. P. 3088–3090. DOI: 10.1093/bioinformatics/btx346
89. Bertrand A.R., Kadri N.K., Flori L., et al. RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments // *Methods Ecol Evol.* 2019. Vol. 10, No. 6. P. 860–866. DOI: 10.1111/2041-210X.13167
90. Zhou J., Liu L., Lopdell T.J., et al. HandyCNV: Standardized summary, annotation, comparison, and visualization of copy number variant, copy number variation region, and runs of homozygosity // *Front Genet.* 2021. Vol. 12. ID 731355. DOI: 10.3389/fgene.2021.731355
91. Biscarini F., Cozzi P., Gaspa G., Marras G. detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network), 2018. Режим доступа: <https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html>
92. Allaire J. RStudio: integrated development environment for R // Boston MA. 2012. Vol. 770, No. 394. P. 165–171.
93. Kluyver T., Ragan-Kelley B., Pérez F., et al. Jupyter Notebooks—a publishing format for reproducible computational workflows // *EL-pub.* 2016. Vol. 2016. P. 87–90. DOI: 10.3233/978-1-61499-649-1-87

## REFERENCES

1. Soulé ME. What is conservation biology? A new synthetic discipline addresses the dynamics and problems of perturbed species, communities, and ecosystems. *BioSci*. 1985;35(11):727–734. DOI: 10.2307/1310054
2. Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Mol Ecol*. 2017;26(20):5369–5406. DOI: 10.1111/mec.14264
3. Hoban S, Kelley JL, Lotterhos KE, et al. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am Nat The University of Chicago Press*. 2016;188(4):379–397. DOI: 10.1086/688018
4. Hoban S, da Silva JM, Mastretta-Yanes A, et al. Monitoring status and trends in genetic diversity for the Convention on Biological Diversity: An ongoing assessment of genetic indicators in nine countries. *Conserv Lett*. 2023;16(3):e12953. DOI: 10.1111/conl.12953
5. Ng PC, Kirkness EF. Whole genome sequencing. In: Barnes MR, Breen G, editors. *Genetic variation: methods and protocols*. Totowa, NJ: Humana Press, 2010. P. 215–226. DOI: 10.1007/978-1-60327-367-1\_12
6. Breed MF, Harrison PA, Blyth C, et al. The potential of genomics for restoring ecosystems and biodiversity. *Nat Rev Genet*. 2019;20(10):615–628. DOI: 10.1038/s41576-019-0152-0
7. Kliver SF. Whole genome approach in conservation biology and its perspectives. *Ecological genetics*. 2021;19(3):281–298. DOI: 10.17816/ecogen65152
8. Joop Ouborg N, Angeloni F, Vergeer P. An essay on the necessity and feasibility of conservation genomics. *Conserv Genet*. 2010;11(2):643–653. DOI: 10.1007/s10592-009-0016-9
9. Dudchenko O, Shamim MS, Batra SS, et al. The Juicebox assembly tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv*. 2018;254797. DOI: 10.1101/254797
10. Durand NC, Robinson JT, Shamim MS, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016;3(1):99–101. DOI: 10.1016/j.cels.2015.07.012
11. Luikart G, England PR, Tallmon D, et al. The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet*. 2003;4(12):981–994. DOI: 10.1038/nrg1226
12. Campbell MR, Vu NV, LaGrange AP, et al. Development and application of single-nucleotide polymorphism (SNP) genetic markers for conservation monitoring of burbot populations. *Trans Am Fish Soc*. 2019;148(3):661–670. DOI: 10.1002/tafs.10157
13. Bijlsma R, Loeschcke V. Genetic erosion impedes adaptive responses to stressful environments. *Evol Appl*. 2012;5(2):117–129. DOI: 10.1111/j.1752-4571.2011.00214.x
14. Leroy G, Carrol EL, Bruford MW, et al. Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evol Appl*. 2018;11(7):1066–1083. DOI: 10.1111/eva.12564
15. Frankham R, Ballou JD, Eldridge MD, et al. Predicting the probability of outbreeding depression. *Conserv Biol*. 2011;25(3):465–475. DOI: 10.1111/j.1523-1739.2011.01662.x
16. Charlesworth D, Willis JH. The genetics of inbreeding depression. *Nat Rev Genet*. 2009;10(11):783–796. DOI: 10.1038/nrg2664
17. Mayr E. *Populations, species and evolution*. Bknap Press. 453 p.
18. Tomimatsu H, Ohara M. Genetic diversity and local population structure of fragmented populations of *Trillium camtschaticense* (Trilliaceae). *Biol Conserv*. 2003;109(2):249–258. DOI: 10.1016/S0006-3207(02)00153-2
19. Hanski I. The Shrinking world: Ecological consequences of habitat loss. *Excell Ecol*. 2005;14.
20. Lande R, Barrowclough G. Effective population size, genetic variation, and their use in population management. In: Soulé M, editor. *Viable populations for conservation*. Cambridge: Cambridge University Press, 1987. P. 87–124. DOI: 10.1017/CBO9780511623400.007
21. Wright S. Random drift and the shifting balance theory of evolution. In: Kojima K, editor. *Mathematical topics in population genetics*. Berlin, Heidelberg: Springer, 1970. P. 1–31. DOI: 10.1007/978-3-642-46244-3\_1
22. Nevo E. Genetic variation in natural populations: Patterns and theory. *Theor Popul Biol*. 1978;13(1):121–177. DOI: 10.1016/0040-5809(78)90039-4
23. Lewontin R. *The genetic basis of evolutionary change*. Columbia University Press, 1974. 346 p.
24. Steiner CC, Putnam AS, Hoeck PEA, Ryder A. Conservation genomics of threatened animal species. *Annu Rev Anim Biosci*. 2013;1(1):261–281. DOI: 10.1146/annurev-animal-031412-103636
25. Weir BS. *Genetic data analysis II: Methods for discrete population genetic data*. Oxford, New York: Oxford University Press, 1996. 445 p.
26. Ritland K. Estimators for pairwise relatedness and individual inbreeding coefficients. *Genet Res*. 1996;67(2):175–185. DOI: 10.1017/S0016672300033620
27. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS*. 1979;76(10):5269–5273. DOI: 10.1073/pnas.76.10.5269
28. Wright S. The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*. 1965;19(3):395–420. DOI: 10.2307/2406450
29. Shafer ABA, Wolf JBW, Alves PC, et al. Genomics and the challenging translation into conservation practice. *Trends Ecol Evol*. 2015;30(2):78–87. DOI: 10.1016/j.tree.2014.11.009
30. Hoffmann A, Griffin P, Dillon S, et al. A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Clim Change Responses*. 2015;2(1):1. DOI: 10.1186/s40665-014-0009-x
31. Benestan LM, Ferchaud A-L, Hohenlohe PA, et al. Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Mol Ecol*. 2016;25(13):2967–2977. DOI: 10.1111/mec.13647
32. Hoban S, Gaggiotti O, ConGRESS Consortium, Bertorelle G. Sample planning optimization tool for conservation and population genetics (SPOTG): a software for choosing the appropriate number of markers and samples. *Methods Ecol Evol*. 2013;4(3):299–303. DOI: 10.1111/2041-210x.12025
33. Nazareno AG, Bemmels JB, Dick CW, Lohmann LG. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Mol Ecol Resour*. 2017;17(6):1136–1147. DOI: 10.1111/1755-0998.12654
34. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet*. 2006;15(5):789–795. DOI: 10.1093/hmg/ddi493

35. McQuillan R, Leutenegger A-L, Abdel-Rahman R, et al. Runs of homozygosity in European populations. *Am J Hum Genet.* 2008;83(3):359–372. DOI: 10.1016/j.ajhg.2008.08.007
36. Darwin C. *The effects of cross and self fertilisation in the vegetable kingdom.* Ams Press Inc, 1877. DOI: 10.5962/bhl.title.104481
37. Ceballos FC, Joshi PK, Clark DW, et al. Runs of homozygosity: windows into population history and trait architecture: 4. *Nat Rev Genet.* 2018;19(4):220–234. DOI: 10.1038/nrg.2017.109
38. Hoffman JI, Simpson F, David P, et al. High-throughput sequencing reveals inbreeding depression in a natural population. *PNAS.* 2014;111(10):3775–3780. DOI: 10.1073/pnas.1318945111
39. Muir WM, Wong GK-S, Zhang Y, et al. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *PNAS.* 2008;105(45):17312–17317. DOI: 10.1073/pnas.0806569105
40. Urbinati I, Stafuzza NB, Oliveira MT, et al. Selection signatures in Canchim beef cattle. *J Anim Sci Biotechnol.* 2016;7(1):29. DOI: 10.1186/s40104-016-0089-5
41. Samuels DC, Wang J, Ye K, et al. Heterozygosity ratio, a robust global genomic measure of autozygosity and its association with height and disease risk. *Genetics.* 2016;204(3):893–904. DOI: 10.1534/genetics.116.189936
42. Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics.* 2003;165(4):2213–2233. DOI: 10.1093/genetics/165.4.2213
43. Rife DC. Populations of hybrid origin as source material for the detection of linkage. *Am J Hum Genet.* 1954;6(1):26–33.
44. Robinson JA, Räikkönen J, Vucetich LM, et al. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Sci Adv.* 2019;5(5):eaau0757. DOI: 10.1126/sciadv.aau0757
45. Koepfli K-P, Tamazian G, Wildt D, et al. Whole genome sequencing and re-sequencing of the sable antelope (*Hippotragus niger*): A resource for monitoring diversity in *ex situ* and *in situ* populations. *G3 Genes Genomes Genetics.* 2019;9(6):1785–1793. DOI: 10.1534/g3.119.400084
46. Big Soviet Encyclopedia. Vol. 20. 3<sup>rd</sup> ed. 1974. P. 25. (In Russ.)
47. Zhu L, Deng C, Zhao X, et al. Endangered Père David's deer genome provides insights into population recovering. *Evol Appl.* 2018;11(10):2040–2053. DOI: 10.1111/eva.12705
48. Beichman AC, Koepfli K-P, Li G, et al. Aquatic adaptation and depleted diversity: A Deep dive into the genomes of the sea otter and giant otter. *Mol Biol Evol.* 2019;36(12):2631–2655. DOI: 10.1093/molbev/msz101
49. Abascal F, Corvelo A, Cruz F, et al. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biol.* 2016;17(1):251. DOI: 10.1186/s13059-016-1090-1
50. Cho YS, Hu L, Hou H, et al. The tiger genome and comparative analysis with lion and snow leopard genomes: 1. *Nat Commun.* 2013;4(1):2433. DOI: 10.1038/ncomms3433
51. Miller W, Schuster SC, Welch AJ, et al. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *PNAS.* 2012;109(36):E2382–E2390. DOI: 10.1073/pnas.1210506109
52. Venn JI. On the diagrammatic and mechanical representation of propositions and reasonings. *Lond Edinb Dublin Philos Mag J Sci.* 1880;10(59):1–18. DOI: 10.1080/14786448008626877
53. Miller W, Hayes VM, Ratan A, et al. Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *PNAS.* 2011;108(30):12348–12353. DOI: 10.1073/pnas.1102838108
54. Humble E, Dobrynin P, Senn H, et al. Chromosomal-level genome assembly of the scimitar-horned oryx: Insights into diversity and demography of a species extinct in the wild. *Mol Ecol Resour.* 2020;20(6):1668–1681. DOI: 10.1111/1755-0998.13181
55. Yakupova A, Tomarovsky A, Totikov A, et al. Chromosome-length assembly of the baikal seal (*Pusa sibirica*) genome reveals a historically large population prior to isolation in Lake Baikal: 3. *Genes.* 2023;14(3):619. DOI: 10.3390/genes1403061
56. Kliver S, Houk ML, Perelman PL, et al. Chromosome-length genome assembly and karyotype of the endangered black-footed ferret (*Mustela nigripes*). *J Hered.* 2023;114(5):539–548. DOI: 10.1093/jhered/esad035
57. Li R, Fan W, Tian G, et al. The sequence and *de novo* assembly of the giant panda genome: 7279. *Nature.* 2010;463(7279):311–317. DOI: 10.1038/nature08696
58. Dobrynin P, Liu S, Tamazian G, et al. Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol.* 2015;16(1):277. DOI: 10.1186/s13059-015-0837-4
59. Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog: 7069. *Nature.* 2005;438(7069):803–819. DOI: 10.1038/nature04338
60. Benjamini Y. Opening the box of a boxplot. *Am Stat.* 1988;42(4):257–262. DOI: 10.2307/2685133
61. Totikov A, Tomarovsky A, Prokopov D, et al. Chromosome-level genome assemblies expand capabilities of genomics for conservation biology: 9. *Genes.* 2021;12(9):1336. DOI: 10.3390/genes12091336
62. Hintze JL, Nelson RD. Violin plots: A box plot-density trace synergism. *Am Stat.* 1998;52(2):181–184. DOI: 10.1080/00031305.1998.10480559
63. de Manuel M, Barnett R, Sandoval-Velasco M, et al. The evolutionary history of extinct and living lions. *PNAS USA.* 2020;117(20):10927–10934. DOI: 10.1073/pnas.1919423117
64. Burton JN, Adey A, Patwardhan RP, et al. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions: 12. *Nat Biotechnol.* 2013;31(12):1119–1125. DOI: 10.1038/nbt.2727
65. Lewin HA, Graves JAM, Ryder OA, et al. Precision nomenclature for the new genomics. *GigaScience.* 2019;8(8):giz086. DOI: 10.1093/gigascience/giz086
66. Wilkinson L, Friendly M. The history of the cluster heat map. *Am Stat.* 2009;63(2):179–184. DOI: 10.1198/tas.2009.0033
67. de Ferran V, Figueiro HV, de Jesus Trindade F, et al. Phylogenomics of the world's otters. *Curr Biol.* 2022;32(16):3650–3658.e4. DOI: 10.1016/j.cub.2022.06.036
68. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.* 1970;48(3):443–453. DOI: 10.1016/0022-2836(70)90057-4
69. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* 1981;147(1):195–197. DOI: 10.1016/0022-2836(81)90087-5
70. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–780. DOI: 10.1093/molbev/mst010

71. Magis C, Taly J-F, Bussotti G, et al. T-Coffee: tree-based consistency objective function for alignment evaluation. In: Russell DJ, editor. *Multiple sequence alignment methods*. Totowa, NJ: Humana Press, 2014. P. 117–129. DOI: 10.1007/978-1-62703-646-7\_7
72. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5(1):113. DOI: 10.1186/1471-2105-5-113
73. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–410. DOI: 10.1016/S0022-2836(05)80360-2
74. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinforma*. 2003;1:2–3. DOI: 10.1002/0471250953.bi0203s00
75. Altshuler D, Donnelly P; The International HapMap Consortium. A haplotype map of the human genome: 7063. *Nature*. 2005;437(7063):1299–1320. DOI: 10.1038/nature04226
76. Durbin RM; The International HapMap Consortium, et al. A map of human genome variation from population-scale sequencing: 7319. *Nature*. 2010;467(7319):1061–1073. DOI: 10.1038/nature09534
77. Nusrat S, Harbig T, Gehlenborg N. Tasks, techniques, and tools for genomic data visualization. *Comput Graph Forum*. 2019;38(3):781–805. DOI: 10.1111/cgf.13727
78. Karolchik D, Baertsch R, Diekhans M, et al. The UCSC genome browser database. *Nucleic Acids Res*. 2003;31(1):51–54. DOI: 10.1093/nar/gkg129
79. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–192. DOI: 10.1093/bib/bbs017
80. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. *Nucleic Acids Res*. 2020;48(D1):D682–D688. DOI: 10.1093/nar/gkz966
81. Okonechnikov K, Golosova O, Fursov M, et al. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28(8):1166–1167. DOI: 10.1093/bioinformatics/bts091
82. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–2158. DOI: 10.1093/bioinformatics/btr330
83. Narasimhan V, Danecek P, Scally A, et al. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. 2016;32(11):1749–1751. DOI: 10.1093/bioinformatics/btw044
84. Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comput Graph Stat*. 1996;5(3):299–314. DOI: 10.1080/10618600.1996.10474713
85. van Rossum G. Python reference manual. *Dep Comput Sci*. 1995; R9525.
86. Hunter JD. Matplotlib: A 2D Graphics environment. *Comput Sci*. 2007;9(3):90–95. DOI: 10.1109/MCSE.2007.55
87. Schiavinato M, del Olmo V, Muya VN, Gabaldon T. JLOH: Inferring loss of heterozygosity blocks from sequencing data. *bioRxiv*. 2023;2023.05.04.539368. DOI: 10.1101/2023.05.04.539368
88. Gel B, Serra E. KaryoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*. 2017;33(19):3088–3090. DOI: 10.1093/bioinformatics/btx346
89. Bertrand AR, Kadri NK, Flori L, et al. RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods Ecol Evol*. 2019;10(6):860–866. DOI: 10.1111/2041-210X.13167
90. Zhou J, Liu L, Lopdell TJ, et al. HandyCNV: Standardized summary, annotation, comparison, and visualization of copy number variant, copy number variation region, and runs of homozygosity. *Front Genet*. 2021;12:731355. DOI: 10.3389/fgene.2021.731355
91. Biscarini F, Cozzi P, Gaspa G, Marras G. detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network), 2018. Available at: <https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html>
92. Allaire J. RStudio: integrated development environment for R. *Boston MA*. 2012;770(394):165–171.
93. Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. *Elpub*. 2016;2016:87–90. DOI: 10.3233/978-1-61499-649-1-87

## ОБ АВТОРАХ

**Андрей Александрович Томаровский;**

ORCID: 0000-0002-6414-704X; eLibrary SPIN: 6727-8664;

e-mail: andrey.tomarovsky@gmail.com

**Азамат Альбертович Тотиков;** ORCID: 0000-0003-1236-631X;

eLibrary SPIN: 9767-3971; e-mail: a.totikov1@gmail.com

**Алия Рафиковна Якупова;** ORCID: 0000-0003-1486-0864;

eLibrary SPIN: 4292-0609; e-mail: aliyah.yakupova@gmail.com

**Александр Сергеевич Графодатский,** д-р биол. наук;

ORCID: 0000-0002-8282-1085; eLibrary PIN: 4436-9033;

e-mail: graf@mcb.nsc.ru

**\*Сергей Федорович Кливер,** независимый исследователь;

ORCID: 0000-0002-2965-3617; eLibrary SPIN: 8635-4259;

e-mail: mahajrod@gmail.com

## AUTHORS' INFO

**Andrey A. Tomarovsky;**

ORCID: 0000-0002-6414-704X; eLibrary SPIN: 6727-8664;

e-mail: andrey.tomarovsky@gmail.com

**Azamat A. Totikov;** ORCID: 0000-0003-1236-631X;

eLibrary SPIN: 9767-3971; e-mail: a.totikov1@gmail.com

**Aliya R. Yakupova;** ORCID: 0000-0003-1486-0864;

eLibrary SPIN: 4292-0609; e-mail: aliyah.yakupova@gmail.com

**Alexander S. Graphodatsky,** Dr. Sci. (Biology);

ORCID: 0000-0002-8282-1085; eLibrary SPIN: 4436-9033;

e-mail: graf@mcb.nsc.ru

**\*Sergei F. Kliver,** independent researcher;

ORCID: 0000-0002-2965-3617; eLibrary SPIN: 8635-4259;

e-mail: mahajrod@gmail.com

\* Автор, ответственный за переписку / Corresponding author