

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
«ИНСТИТУТ МОЛЕКУЛЯРНОЙ И КЛЕТОЧНОЙ БИОЛОГИИ СИБИРСКОГО
ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК»

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НОВОСИБИРСКИЙ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»

На правах рукописи

ТОМАРОВСКИЙ АНДРЕЙ АЛЕКСАНДРОВИЧ

**ПОСЛЕДСТВИЯ МЕЖВИДОВОЙ ГИБРИДИЗАЦИИ И
ФИЛОГЕНЕТИЧЕСКИЕ ОТНОШЕНИЯ В РОДЕ *MARTES*
(СЕМ. MUSTELIDAE)**

1.5.7 – Генетика
(биологические науки)

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:
д.б.н., профессор, член-корреспондент РАН
Александр Сергеевич Графодатский

Новосибирск – 2026

Оглавление

Введение.....	4
Цель и задачи исследования.....	6
Научная новизна.....	7
Теоретическая и научно-практическая значимость исследования.....	8
Научные положения, выносимые на защиту.....	9
Личный вклад автора.....	10
Апробация работы.....	10
Структура и объем работы.....	12
Благодарности.....	13
1 Обзор литературы.....	15
1.1 Получение геномных сборок для решения задач природоохранной биологии.....	15
1.1.1 Технологии полногеномного секвенирования.....	15
1.1.2 Геномные сборки хромосомного уровня.....	19
1.2 Оценка гетерозиготности и анализ видовой принадлежности.....	20
1.2.1 Гетерозиготность, как одна из ключевых метрик генетического разнообразия... 20	
1.2.2 Обнаружение геномной интрогрессии.....	30
1.2.3 Анализ происхождения.....	34
1.3 Филогеномный анализ данных.....	38
1.3.1 Методы поиска ортологичных генов.....	39
1.3.2 Подходы множественного выравнивания и фильтрации последовательностей... 40	
1.3.3 Реконструкция филогенетических деревьев.....	41
1.4 Проблематика гибридизации и филогении рода <i>Martes</i>	46
1.4.1 Филогения рода <i>Martes</i> в сем. Mustelidae.....	46
1.4.2 Проблематика положения харзы (подрод <i>Charronia</i>) в роде <i>Martes</i>	52
1.4.3 <i>M. foina</i> как потенциально базальный вид рода <i>Martes</i>	55
1.4.4 Соболь и последствия его реинтродукции.....	57
1.4.5 Гибридизация в роде <i>Martes</i>	62
1.4.5.1 Гибридизация соболя и лесной куницы.....	62
1.4.5.2 Гибридизация американской куницы и тихоокеанской куницы.....	67
1.4.5.3 Японский соболь.....	70
2 Материалы и методы.....	73
2.1 Данные полногеномного секвенирования.....	73
2.2 Морфологический анализ.....	76
2.3 Оценка качества и фильтрация прочтений.....	77
2.4 Получение и оценка качества геномныхборок.....	77
2.5 Предсказание и аннотация белок-кодирующих генов.....	78
2.6 Полногеномные множественные выравнивания и анализ синтении.....	79
2.7 Выравнивание и поиск координат псевдоаутосомного района.....	80
2.8 Поиск, фильтрация и маскирование генетических вариантов.....	81

2.9 Анализ распределений гетерозиготности.....	81
2.10 Обнаружение интрогрессии, анализ происхождения и классификация образцов..	84
2.11 Анализ микросателлитных локусов.....	85
2.12 Выявление протяженных участков гомозиготности (RoH).....	87
2.13 Филогеномный анализ.....	88
2.14 Сборка и аннотация мтДНК. Филогения и сеть гаплотипов.....	90
2.15 Вычисление внутри- и межвидовых генетических расстояний.....	90
2.16 Анализ популяционной истории.....	91
3 Результаты.....	92
3.1 Контроль качества полногеномных данных и размер генома.....	92
3.2 Сборка и аннотация геномов.....	93
3.3 Аннотация повторов, покрытие и псевдоаутосомный район.....	97
3.4 Анализ синтении <i>M. foina</i> с другими видами.....	100
3.5 Синтения между видами <i>Martes</i>	101
3.6 Анализ гетерозиготности.....	103
3.7 Классификация по видовой принадлежности.....	108
3.8 Классификация по видовой принадлежности на основании полученных и общедоступных STR-локусов.....	114
3.9 Анализ участков гомозиготности (RoH).....	115
3.10 Филогения рода <i>Martes</i>	119
3.11 Реконструкция филогении соболей и лесных куниц на основе мтДНК.....	122
3.12 Внутри- и межвидовые генетические расстояния между видами <i>Martes</i>	125
3.13 Анализ популяционной истории.....	127
4 Обсуждение.....	131
4.1 Геномные сборки и синтения.....	131
4.2 Особенности образцов соболей, лесных куниц и их гибридов.....	132
4.3 Концепция анализа компонентов распределения гетерозиготности.....	133
4.4 Гетерозиготность, RoH и интрогрессия соболя и лесной куницы.....	136
4.5 Гетерозиготность исследуемых видов рода <i>Martes</i> в сравнении с другими куньими... 138	
4.6 Кроссинговер и мозаичность хромосом.....	142
4.7 Противоречия классификации по видовой принадлежности.....	145
4.8 Филогения соболя и лесной куницы на основании мтДНК.....	149
4.9 Харза как отдельный род в подсемействе Guloninae.....	151
4.10 Видовая структура <i>M. foina</i> и <i>Ch. (M.) flavigula</i>	152
4.11 Популяционная история рода <i>Martes</i>	153
Заключение.....	157
Выводы.....	159
Список литературы.....	160
Приложения.....	187

Введение

Семейство куньих (сем. Mustelidae) – одно из наиболее богатых видами семейств млекопитающих из отряда хищных (отр. Carnivora), представители которого встречаются практически повсеместно, за исключением Антарктиды и ряда островов. В настоящее время семейство включает по разным оценкам 60-70 видов, относящихся к 22-24 родам [1]. Куньи в значительной степени различаются между собой поведенчески и фенотипически, а также обладают специфическими адаптациями к окружающей среде обитания [2]. Несмотря на интересные биологические особенности животных этой группы, к которым можно отнести эмбриональную диапаузу, многие виды куньих остаются недостаточно изученными с точки зрения генетики, популяционной геномики и эволюционной биологии, вследствие малочисленных или полностью отсутствующих полногеномных данных для особей из различных популяций. Представители рода *Martes* распространены в Европе, Азии и Северной Америке. Большинство животных этой группы характеризуется высоким качеством меха в связи с приспособленностью к жизни при низких температурах в тайге, хвойных и лиственных лесах Северного полушария. Первые исследования по изучению филогении были основаны на фенотипических, морфологических и физиологических признаках, в связи с чем таксономия рода подвергалась многочисленным пересмотрам [3–5]. В более поздних исследованиях филогеографии популяций и филогении видов куньих применялись молекулярно-генетические подходы, основанные как на использовании ядерных и митохондриальных генетических маркеров, так и на полных последовательностях мтДНК [2,6–10]. Последняя масштабная ревизия филогении рода *Martes* включает выделение ильки (*M. pennanti*) в отдельный род *Pekania* [11], а также разделение американской куницы (*M. americana*) на два вида – *M. americana* и *M. caurina*, между которыми происходит активная гибридизация в зоне симпатрии [12]. Однако, несмотря на множество проведенных исследований, до сих пор остаются нерешенными некоторые проблемы, связанные с систематикой видов в роде *Martes*.

На сегодняшний день нет единого мнения о систематическом положении харзы (*M. flavigula*), которая занесена в Красную книгу МСОП (Международного

Союза Охраны Природы) как вид, вызывающий наименьшее опасение [13]. В России харза встречается редко, так как на территорию нашей страны приходится самый север ее ареала [14] и в последние годы рассматривается как угрожаемый вид [15]. В настоящее время харза без сомнений занимает базальное положение в филогении рода *Martes* [7,8], поскольку фенотипически сильно отличается от других представителей рода, имея яркую окраску и крупные размеры [16]. В этой связи в исследовании по таксономии харзы 1995 года В.В. Рожнов предложил выделить этот вид в отдельный род *Lamprogale*, а также поднимает вопрос о необходимости выделений популяций из юговосточной окраины Азии в отдельный вид *L. lasiotis*, на основе морфологических данных и разобщенности ареала [17]. Ранее Е.Н. Матюшкин в работе 1993 года также подчеркивает сложность таксономической структуры харзы, предполагая выделение в отдельный род *Charronia* и обосновывая подразделение харзы как минимум на два вида: *Ch. aterrima* и *Ch. flavigula* [18]. Позднее проводились направленные исследования филогении харзы с использованием митохондриальных маркеров [19,20] и полной последовательности мтДНК [8,21], однако отсутствие полногеномных данных до настоящего времени не позволяло точно определить систематическое положение харзы и поставить точку в этом вопросе.

Другим примером, указывающим на недостаточную изученность, являются самостоятельные виды соболь (*M. zibellina*) и лесная куница (*M. martes*). В зоне симпатрии этих ценных пушных видов на северном Урале и западной Сибири встречаются гибридные особи (кидасы) с промежуточными фенотипическими признаками родительских видов [22,23]. С середины XX века проводились исследования с целью доказать гибридное происхождение кидасов и выявить их отличительные особенности от родительских особей. В результате была установлена частичная плодовитость кидасов и возможность возвратных скрещиваний [24–27]. Тем не менее, все еще существовали обоснованные сомнения и неопределенность касательно гибридного статуса кидасов в естественных условиях [23]. Позднее, гибридное происхождение кидасов в зоне симпатрии было подтверждено с использованием молекулярно-генетических подходов на основе маркеров мтДНК и сателлитных локусов яДНК [28–30]. В ходе проведенных

исследований не удалось в полной мере оценить интенсивность и направление процесса гибридизации, что возможно лишь с использованием качественных полногеномных данных секвенирования большого количества образцов.

Стоит отметить, что у соболя, обитающего в основном на территории России, остается неизвестным уровень генетического разнообразия на полногеномном уровне. Оценка уровня гетерозиготности особей из разных популяций может показать последствия реинтродукции соболя, проводимой в XX веке в СССР [31,32], что впоследствии может способствовать пересмотру нынешнего глобального статуса вида в МСОП. Уровень генетического разнообразия куницы также требует рассмотрения, в связи с расширением, а затем с уменьшением ареала этого вида на востоке Западной Сибири [26]. Дополнительно для изучения эволюции структуры генома важно рассмотреть каменную куницу (*M. foina*), для которой уже опубликованы сравнительные хромосомные карты на основании флуоресцентной *in situ* гибридизации (FISH) [33]. К настоящему времени с помощью набора зондов *M. foina* были изучены кариотипы 12 видов куных [33–36]. Наконец, важно рассмотреть северо-американских куниц – американскую куницу (*M. americana*) и тихоокеанскую куницу (*M. caurina*), охватив тем самым большую часть рода *Martes*.

Цель и задачи исследования

Целью данного исследования является оценка уровня генетического разнообразия изучаемых видов, включая межвидовых гибридов, и проведение филогенетического анализа рода *Martes* на полногеномном уровне.

Для достижения цели работы были поставлены следующие задачи:

1. Оценка качества и фильтрация как полученных в ходе исследования, так и общедоступных полногеномных данных рода *Martes*.
2. Получение и оценка качества референсных геномных сборок хромосомного уровня 4 видов рода *Martes*. Предсказание и аннотация белок-кодирующих генов и повторов в полученных сборках.
3. Поиск и анализ хромосомных перестроек между геномами 4 видов рода *Martes*.

4. Оценка генетического разнообразия исследуемых видов рода *Martes* (включая межвидовых гибридов), и сравнение с другими видами куньих.
5. Поиск интрогрессии, анализ происхождения и классификация образцов по видовой принадлежности с использованием полногеномных данных, STR-маркеров и полных мтДНК.
6. Анализ популяционной истории исследуемых видов рода *Martes*.
7. Филогенетический анализ рода *Martes* на основе полногеномных данных.

Научная новизна

В рамках данной работы были получены геномные сборки хромосомного уровня *M. foina*, *M. zibellina*, *M. martes* и *M. (Ch.) flavigula*. На момент представления этих сборок научному сообществу на сайте консорциума DNA Zoo (2021 год) это были первые сборки этих видов. На текущий момент наша сборка генома *M. zibellina* остается единственной сборкой хромосомного уровня этого вида. Необходимо отметить особое значение именно нашей сборки генома *M. foina*, ведь она была получена именно для той особи, культура клеток которой была использована во множестве цитогенетических исследований эволюции хромосом отряда Carnivora.

Мы впервые провели полногеномный анализ генетического разнообразия *M. martes* и *M. zibellina* и рассмотрели его в контексте гибридизации этих видов. Мы впервые показали присутствие мозаичных рекомбинантных хромосом в гибридах, что говорит о как минимум не подавленном кроссинговере у гибридных особей и непригодности методов низкого разрешения, например STR-типирования, для анализа этого комплекса. Мы впервые показали высокий уровень гетерозиготности современных соболей (*M. zibellina*), что позволяет говорить об успехе программ массовой реинтродукции этого вида в середине XX века. Мы выявили неизвестную ранее низкую (относительно) гетерозиготность лесной куницы (*M. martes*) и показали, что корни этого феномена лежат в древности и связаны с популяционной историей вида.

В рамках филогенетического анализа рода *Martes* мы окончательно решили вопрос о положении харзы в систематике подсемейства Guloninae (является ли она

частью рода *Martes* или нет), который ранее рассматривался только в работах на основе морфологических данных, секвенирования отдельных коротких локусов митохондриальной ДНК и ядерного генома. Наши результаты говорят, что ее необходимо выделить в отдельный род *Charronia*, минимум с тремя видами *Ch. aterrima*, *Ch. flavigula*, и *Ch. gwatkinsii*. Для *M. foina* мы показали неизвестный ранее уровень различий между европейской и азиатской частью ареала как для ядерного, так и для митохондриального геномов, причем для мтДНК он соответствует межвидовому. Мы впервые выявили противоречия в положении *M. martes* на филогенетических деревьях относительно *M. zibellina* и американских куниц (*M. caurina* и *M. americana*), построенных по ядерному и митохондриальному геномам.

Теоретическая и научно-практическая значимость исследования

Полученные хромосомные сборки четырех видов куниц послужат основой дальнейших исследований в области сравнительной геномики. Выявленное разнообразие гибридов, их мозаичные рекомбинантные хромосомы, а значит и неподдавленный кроссинговер у подобных особей – это важные результаты для фундаментального понимания механизмов гибридизации и ее последствий не только для комплекса *M. martes x M. zibellina*, но и для всего семейства куньих. Помимо прочего они четко показали ограничения до сих пор используемого STR-типирования для определения видовой принадлежности, более того это ограничения означающие практическую неприменимость таких маркеров. При включении американских видов рода *Martes* в анализ, мы выяснили, что гибридный комплекс включает и эти виды. Столь сложная гибридная система, а также выявленный эффект референса при выравнивании прочтений требуют создания межвидового пангеномного графа всего комплекса. Использование межвидового пангенома в качестве референса мы рассматриваем как основу будущих исследований по геномике гибридизации.

С практической точки зрения у данной работы есть три основных аспекта. Во-первых, анализ полногеномных данных соболя позволил сделать вывод об успехе массовой реинтродукции этого вида, осуществленной в середине XX века, с

точки зрения современного генетического разнообразия. Таким образом, эта работа дает первую оценку на полногеномном уровне одной из самых масштабных программ реинтродукции в истории. Во-вторых, разрешение филогенетического вопроса с систематическим положением харзы в роде *Martes* имеет практическую значимость для природоохранной биологии в России. Выделение харзы в отдельный род *Charronia* с не менее чем тремя видами, присваивает видовой статус дальневосточному подвиду (*Charronia aterrima*, ранее *Martes flavigula aterrima*), обитающему в том числе и на территории России. Это означает значительно меньшую площадь ареала по сравнению с *Martes flavigula*, а значит является основанием для пересмотра действующих программ охраны и усиления мониторинга численности этого вида на территории России. В-третьих, разработанная в ходе работы концепция анализа компонент гетерозиготности представляет собой новый взгляд на проблему гибридных геномов. После некоторых дополнений и оптимизаций можно будет использовать в качестве дополнительного теста на гибридное происхождение образца, в случае, если нет образцов чистых видов или их количество недостаточно для надежного ADMIXTURE-анализа.

Подводя итог вышеизложенному, можно уверенно заявить, что данное исследование обладает значительной как теоретической, так и научно-практической значимостью. Полученные результаты и сделанные выводы углубляют наше понимание эволюционных и гибридизационных процессов, геномной организации видов рода *Martes*, предлагают новую концепцию анализа данных, диктуют какие подходы необходимо использовать в будущих исследованиях, а также имеют непосредственное практическое значения для природоохранной биологии куньих России.

Научные положения, выносимые на защиту

1. Взаимные интрогрессии между ядерными геномами соболя (*Martes zibellina*) и лесной куницы (*M. martes*) в удаленных от зоны симпатрии регионах указывают на необходимость пересмотра границ ареалов исследуемых видов с возможным расширением зоны симпатрии.

2. Мультимодальное распределение гетерозиготности в геномах особей комплекса *M. zibellina* × *M. martes* является следствием гибридизации видов с различным уровнем гетерозиготности – высокой у *M. zibellina* и низкой у *M. martes*, а также исторических событий.
3. Ревизия филогении рода *Martes* на основе полногеномных данных позволяет выделить харзу (*M. flavigula*) в отдельный род *Charronia*, включающий как минимум три вида – *Ch. aterrima*, *Ch. flavigula* и *Ch. gwatkinsii*.

Личный вклад автора

Дизайн работы, биоинформатической обработки и анализа данных, визуализация и интерпретация полученных результатов были выполнены при непосредственном участии автора. Автором лично реализованы программные скрипты для фильтрации сцепленных прочтений, конвейерного контроля и повышения качества геномных сборок, оценки покрытия и последующего определения координат псевдоаутосомного района X-хромосомы; проведен поиск генетических вариантов, выявление протяженных участков гомозиготности, сравнительная оценка гетерозиготности, оценка степени интрогрессии и анализ популяционной структуры; реализован высокопроизводительный программный конвейер для реконструкции филогении на основании полногеномных данных. Имплементирован как метод конкатенации, так и обобщенно-коалесценцентный подход для реконструкции видовой филогении на основании полногеномных данных. Автором лично проведена биоинформатическая обработка данных на высокопроизводительном вычислительном кластере. Автор активно участвовал как в написании первых черновиков всех публикаций по теме работы, так и в их корректуре, а в случае статей, где он является первым автором, выполнил большую часть этой работы.

Апробация работы

Результаты работы были представлены научному сообществу в следующих публикациях, прошедших peer review либо отправленных в журналы:

1. **Tomarovsky A.A.**, Totikov A.A., Bulyonkova T., Perelman P.L., Abramov A.V., Serdyukova N.A., Yakupova A., Prokopov D., Beklemisheva V.R., Sinding M.-H., Davletshina G., Pobedintseva M., Krasheninnikova K., Foerster D., Mukhacheva A.S., Mironova A., Sidorov M., Nie W., Wang J., Romanenko S.A., Proskurykova A., Ferguson-Smith M., Yang F., Cherkasov N., Balanovska E., Gilbert T.P., Okhlopkov I.M., Zhuk A., Graphodatsky A., Powell R., Koepfli K.-P., Kliver S. (2026) Genomics of sable (*Martes zibellina*) × pine marten (*Martes martes*) hybridization // Genome Biology and Evolution. 2026. DOI: 10.1093/gbe/evag018. Q1, IF=3.3
2. **Tomarovsky A.A.**, Khan R., Dudchenko O., Beklemisheva V.R., Perelman P.L., Totikov A.A., Serdyukova N.A., Bulyonkova T., Pobedintseva M., Abramov A.V., Weisz D., Yakupova A., Zhuk A., Graphodatsky A., Powell R., Aiden E., Koepfli K.-P., Kliver S. Novel chromosome-length genome assemblies of three distinct subspecies of pine marten, sable, and yellow-throated marten (genus *Martes*, family Mustelidae) // Journal of Heredity. 2025. DOI: 10.1093/jhered/esaf101. Q1, IF=2.5
3. **Tomarovsky A.A.**, Khan R., Dudchenko O., Totikov A., Serdyukova N., Weisz D., Vorobieva N., Bulyonkova T., Abramov A., Nie W., Wang J., Romanenko S., Proskuryakova A., Cherkasov N., Ferguson-Smith M., Yang F., Balanovskaya E., Gilbert T., Graphodatsky A., Aiden E., Powell R., Koepfli K., Perelman P., Kliver S. Chromosome length genome assembly of the stone marten (*Martes foina*, Mustelidae): a new view on one of the cornerstones in carnivore cytogenetics // Journal of Heredity. 2025. Vol. 116, № 4. P. 548-557. DOI: 10.1093/jhered/esaf001. Q1, IF=2.5
4. **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: 10.17816/ecogen609552. Q4, IF=0.2
5. Derežanin, L., Blažytė, A., Dobrynin, P., Duchêne, D.A., Grau, J.H., Hofreiter, M., Jeon, S., Kliver, S., Koepfli, K.P., Meneghini, D., Preick, M., **Tomarovsky, A.**, Totikov, A., Fickel, J., Förster, D.W. Multiple types of genomic variation

contribute to adaptive traits in the mustelid subfamily Guloninae // *Molecular Ecology*. 2022. Vol. 31, № 10. P. 2898-2919. DOI: 10.1111/mec.16443. Q1, IF=3.9

Результаты работы была представлены в следующих тезисах научных конференций:

1. **Tomarovsky A.**, Totikov A., Perelman P., Beklemisheva V., Serdyukova N., Bulyonkova T., Sidorov M., Mamaev N., Okhlopov I., Mukhacheva A., Koniaeva K., Abramov A., Graphodatsky A., Kliver S. Whole-genome data confirm the presence of crossover in hybrids of sable (*M. zibellina*) and pine marten (*M. martes*). 14th International Multiconference Bioinformatics of Genome Regulation and Structure / Systems Biology (BGRSSB-2024). August 5–10, 2024, Novosibirsk, Russia. DOI: 10.18699/bgrs2024-abstracts
2. **Томаровский А.А.**, Тотиков А.А., Перельман П.Л., Беклемишева В.Р., Сердюкова Н.А., Бульонкова Т.М., Сидоров М., Мамаев Н., Охлопков И., Мухачева А., Коняева К., Абрамов А.В., Графодатский А.С., Кливер С.Ф. Оценка уровня гетерозиготности соболя (*Martes zibellina*), лесной куницы (*Martes martes*) и их гибридов. Материалы Международной конференции "Хромосома-2023" Новосибирск, 05-10 сентября 2023 года. DOI: 10.25205/978-5-4437-1514-8
3. **Tomarovsky, A.**; Totikov, A.; Beklemisheva, V.; Perelman, P.; Serdyukova, N.; Bulyonkova, T.; Koniaeva, K.; Abramov, A.; Graphodatsky, A.; Koepfli K.; Powell R.; Kliver S. Assembly and annotation of the sable (*Martes zibellina*) and pine marten (*Martes martes*) genomes. Moscow Conference on Computational Molecular Biology (MCCMB-2021). ISBN: 978-5-901158-32-6 (176).

Структура и объем работы

Диссертация включает введение, обзор литературы, описание используемых материалов и методов исследования, полученные результаты и их обсуждение, выводы, список цитируемой литературы (476 ссылок) и приложение. Работа изложена на 186 страницах, содержит 39 иллюстраций и 13 таблиц (без учета приложения).

Благодарности

Автор признателен научному руководителю, главному научному сотруднику Лаборатории разнообразия и эволюции геномов ИМКБ СО РАН, д.б.н., чл.-корр. РАН Александру Сергеевичу Графодатскому за высокий профессионализм, компетентность и возможность выполнения данной работы.

Выражаю особую благодарность сотруднику Center for Evolutionary Hologenomics, The Globe Institute, The University of Copenhagen (Denmark) к.б.н. Сергею Федоровичу Кливеру за дизайн работы, профессиональную помощь и неоценимую поддержку на протяжении всех этапов исследования.

За непосредственный сбор образцов выражаю благодарность Татьяне Михайловне Бульонковой (ИСИ СО РАН), Михаилу Михайловичу Сидорову, Николаю Викторовичу Мамаеву (ИБПК СО РАН, г. Якутск), Светлане Николаевне Верхоланцевой (Новосибирский зоопарк им. Р.А. Шило), Виолетте Робертовне Беклемишевой, Сергею Федоровичу Кливеру (независимый исследователь), Malcolm Ferguson-Smith, Fengtang Yang (Кембриджский университет), Ксении Коняевой (зоопарк “Лесная сказка”, г. Барнаул), Анне Мухачевой (Сихотэ-Алинский заповедник), Александру Мигуру (д. Малый Нарыс), Валентину Слободенюку (с. Перегребное), Валерию Зелепухину (п. Хандыга), Валерию Середкину (п. Невон), Сергею Крехнову (с. Ивашка), Виктору Хазагаеву (с. Наран), Алексею Лахтину (п. Дугда) и Александре Мироновой (г. Усть-Илимск). Также сбору образцов содействовали Ольга Владимировна Шило, Роза Соловьева (Новосибирский зоопарк им. Р.А. Шило), Писарев Сергей Викторович, Павел Владимирович Резниченко (зоопарк “Лесная сказка”, г. Барнаул), Иннокентий Михайлович Охлопков (ИБПК СО РАН г. Якутск), Сергей Цыпленков, Валерий Жданюк (ПЗК "Магистральный"), за что им выражается благодарность.

За выделение ДНК, создание клеточных культур и цитогенетический анализ благодарю сотрудниц ИМКБ СО РАН к.б.н. Виолетту Робертовну Беклемишеву, к.б.н. Полину Львовну Перельман, Наталью Алькуатовну Сердюкову и Марию Побединцеву. Отдельную благодарность выражаю Ольге Дудченко и Ruqayya Khan из консорциума “DNA Zoo” за генерацию данных Hi-C-секвенирования. За

выполнение морфологического анализа выражаю особую благодарность д.б.н. Алексею Владимировичу Абрамову (ЗИН РАН). За иллюстрации соболя, лесной куницы и их гибрида выражаю отдельную благодарность Татьяне Михайловне Бульонковой (ИСИ СО РАН). За выполнение полногеномных множественных выравниваний геномов выражаю благодарность Сергею Федоровичу Кливеру. За выполнение амплификации и последующей фильтрации микросателлитных локусов выражаю благодарность Дмитрию Юрьевичу Прокопову.

Признателен Dr. Klaus-Peter Koepli (Smithsonian-Mason School of Conservation), Dr. Roger Powell (North Carolina State University) и д.б.н. Графодатскому Александру Сергеевичу (ИМКБ СО РАН) за финансирование, организацию работы и ценные рекомендации. Отдельно благодарен к.б.н. Анне Сергеевне Жук, к.т.н. Владимиру Игоревичу Ульянову и к.т.н. Алексею Александровичу Сергушичеву за предоставление доступа к вычислительным ресурсам. За генерацию геномных сборок скаффолдного уровня для американской куницы и тихоокеанской куницы благодарю сотрудника Amedes Medizinische Dienstleistungen GmbH (Берлин, Германия) Jose Horacio Grau Jipoulou.

Диссертационная работа выполнена в Лаборатории разнообразия и эволюции геномов Института молекулярной и клеточной биологии СО РАН, и с использованием Уникальной научной установки “Криобанк культур клеток” ИМКБ СО РАН (Новосибирск, Россия). Выражаю благодарность всем сотрудникам Лаборатории разнообразия и эволюции геномов за внимание и доброжелательность.

Исследование выполнено с использованием вычислительного кластера Университета ИТМО и суперкомпьютерного комплекса ЦКП “Биоинформатика” ИЦИГ СО РАН. На разных этапах, работа была поддержана финансированием со стороны JetBrains Research, гранта РФФИ №20-04-00808, и грантов РНФ: №19-14-00034-П и №22-24-01076.

1 Обзор литературы

Мы начнем повествование с краткого исторического экскурса в развитие технологий полногеномного секвенирования, чтобы показать, как достижения в этой области сделали возможным получение недорогих геномных сборок хромосомного уровня. Поскольку создание таких сборок геномов является первостепенной и одной из ключевых задач данной работы, было принято решение начать именно с этого, соблюдая логическую последовательность изложения.

1.1 Получение геномных сборок для решения задач природоохранной биологии

1.1.1 Технологии полногеномного секвенирования

К настоящему времени существует три основных поколения технологий секвенирования. Первым поколением считается разработанное в 1977 году секвенирование по Сэнгеру или “метод обрыва цепи” [37]. Данный метод обеспечил беспрецедентно высокую точность прочтения каждого нуклеотида (99,9%) с длиной последовательности до ~800-1000 п.н. Первоначальный подход отличался низкой пропускной способностью и относительно высокой стоимостью [38]. Секвенатор АВ370А, представленный в 1986 году компанией Applied Biosystems, увеличил производительность за счет автоматизации капиллярного гель-электрофореза, благодаря чему позволил обрабатывать до 96 образцов одновременно, суммарно прочитывать до 500 тыс оснований в день с максимальной длиной прочтений до 600 оснований. В 1995 году компания представила обновленную версию – секвенатор АВ310, разделение цепей ДНК в котором было основано на линейном полимере, в отличие от первоначально используемого пластинчатого полиакриламидного геля. Автоматизированное секвенирование методом Сэнгера стало основой для завершения проекта “Геном человека” в 2001 году [39], однако в настоящее время уступает более современным подходам из-за невозможности дальнейшего повышения скорости секвенирования [40].

Период между 2005 и 2010 годами ознаменовался появлением второго поколения технологий секвенирования, известных как технологии секвенирования нового поколения (NGS). Платформы основаны на пирофосфатном секвенировании

(Roche 454), ионном полупроводниковом секвенировании (Ion Torrent) и секвенировании путем синтеза цепи (Illumina) [41]. Суть технологии пиросеквенирования, разработанной компанией 454 Life Sciences и приобретенной в 2007 году компанией Roche, заключается в обнаружении пирофосфата, высвобождаемого после включения нуклеотида в синтезируемую цепь ДНК [42]. Отличием Ion Torrent от других технологий NGS стал отказ от использования флуоресцентно меченых нуклеотидов в пользу идентификации высвобождаемых в ходе полимеризации ДНК ионов водорода (многоканальный рН-метр) [43]. Технология секвенирования путем синтеза цепи была предложена в 1998 году и коммерциализирована в 2006 году компанией Solexa. Спустя год компанию Solexa выкупила Illumina, которая на момент написания работы остается лидером в области геномных исследований и успешно развивает эту технологию [41]. Подход заключается в клональной амплификации ДНК фрагментов с закрепленными на поверхности проточной ячейки адаптерами, последовательном встраивании в цепь флуоресцентно меченых оснований и детектировании флуоресценции [44]. Технологии NGS имеют достаточно высокую точность (99,5%) и отличаются от секвенирования по Сэнгеру более короткими прочтениями (75-300 оснований для Illumina, менее 400 п.н. для Ion-Torrent и менее 700 оснований для Roche 454) [45]. Революционным отличием NGS стали исключительно высокая пропускная способность и снижение общей стоимости, что открыло возможности для исследований немодельных видов [46–48].

Третье поколение появилось в период с 2011 по 2014 год. Его отличительной характеристикой стали длинные прочтения (более 2000 п.н.), получаемые путем мономолекулярного секвенирования в реальном времени (SMRT) [47]. Платформы для SMRT-секвенирования на данный момент представлены Pacific Biosystems (PacBio) и Oxford Nanopore Technologies (ONT). SMRT-секвенирование от PacBio заключается в определении последовательности ДНК путем синтеза комплементарной нити с использованием флуоресцентно меченых нуклеотидов. При встраивании в растущую цепь они расщепляются ДНК-полимеразой, что приводит к высвобождению флуорофора, регистрация которого в режиме реального времени позволяет понять, какой нуклеотид встроился ДНК-полимеразой в

комплементарную цепь [49,50]. ONT основан на использовании мембраны с множеством белковых пор, на которую подается напряжение [51]. Идентификация нуклеотидов происходит путем пропускания нитей ДНК через нанометровые поры (нанопоры) и измерения изменений электрического тока, вызванных перемещением нуклеотидов. Конкретная последовательность оснований определяется в режиме реального времени путем анализа уникальной картины тока, создаваемого каждым основанием при прохождении через нанопору [52,53]. Некоторой альтернативой третьему поколению стал подход генерации синтетически длинных прочтений на базе коротких прочтений, полученных на платформах второго поколения [47].

Суть синтетических подходов, в частности платформы 10X Genomics, заключается в особом способе приготовления библиотек и основана на генерации не физически длинных прочтений, а коротких сцепленных прочтений (англ. Linked reads) на базе Illumina [47]. Суть подхода заключается в предварительной маркировке прочтений при помощи уникальных идентификаторов до 16 п.о. (баркодов), благодаря чему удастся сохранить информацию о принадлежности к конкретному крупному фрагменту исходной последовательности (в идеальном случае длиной ~ 100 тыс п.н.) [47]. Полученные прочтения, имеющие идентичный баркод, относятся к одному фрагменту исходной молекулы, благодаря чему могут быть сгруппированы в единые протяженные синтетические прочтения в ходе геномной сборки (Рисунок 1) [54].



Рисунок 1. Схема пробоподготовки для получения сцепленных прочтений 10X Genomics. Адаптировано из [55].

В дополнение к сцепленным прочтениям также используется технология Hi-C, которая является одним из молекулярных протоколов фиксации

пространственной организации хроматина в ядре (англ. Chromosome Conformation Capture или 3C) [56]. Идея Hi-C-секвенирования заключается в секвенировании пространственно близких фрагментов в масштабе всего генома с последующим определением координат контактирующих последовательностей в черновой сборке [57]. Протокол Hi-C-секвенирования включает обработку формальдегидом, фрагментацию при помощи рестриктаз, биотинилирование концов полученных фрагментов ДНК, лигирование сшитых фрагментов, удаление биотинилированных нуклеотидов с их концов. После приготовления библиотеки полученные химерные фрагменты секвенируют подходами высокопроизводительного секвенирования [58].

Технология Hi-C позволяет оценить вероятность взаимодействия двух линейно отдаленных друг от друга участков генома по числу секвенированных парных хромосомных контактов, что открывает возможности для разрешения геномных повторов, неразрешимых при помощи обычных парных прочтений [59]. Ограничением при использовании Hi-C данных в контексте геномной сборки является невозможность оценки фактического расстояния между объединяемыми фрагментами. Это приводит к тому, что в ходе объединения скаффолдов (скаффолдинг) между ними вставляется пробел фиксированной длины. Однако основной проблемой Hi-C остается ориентация мелких фрагментов генома (<1 млн п.н.) [60]. Из-за их небольшого размера генерируется недостаточное количество контактов, что затрудняет корректное определение их положения и ориентации в получаемой сборке. Это ограничение особенно ощутимо при работе с высокофрагментированными геномами, где доля таких мелких фрагментов велика. Таким образом, качество получаемой сборки генома напрямую зависит от степени фрагментации исходной геномной сборки.

Несмотря на это, возможность количественной оценки взаимодействий между пространственно расположенными рядом геномными локусами без ограничений на линейное расстояние между ними позволяет использовать технологию Hi-C-секвенирования для получения геномныхборок хромосомного уровня [61].

1.1.2 Геномные сборки хромосомного уровня

Геномные сборки на уровне хромосом отличаются от обычных геномных сборок за счет наличия скаффолдов хромосомного уровня или С-скаффолдов (англ. C-scaffold (chromosome-scale scaffold)), число которых соответствует числу пар хромосом в организме [62]. Сопоставление хромосомных скаффолдов с реальными хромосомами осуществляется посредством флуоресцентной гибридизации *in situ* (FISH) с использованием искусственных последовательностей или фрагментов ДНК близкородственных видов [63]. Этот подход, используемый для межвидового сравнительного окрашивания хромосом и, в контексте природоохранной биологии, также известный как Zoo-FISH, позволяет высокой точностью сопоставить номера С-скаффолдов с хромосомами другого вида млекопитающих [62,64]. Порог разрешения метода Zoo-FISH составляет около 5 млн п.н. [65]. Как следствие, гомологичные блоки меньшей (а иногда и большей) длины могут быть пропущены при анализе данных Zoo-FISH по причине слабого флуоресцентного сигнала. Важно отметить, что геномные сборки хромосомного уровня также зачастую содержат большое количество коротких скаффолдов, которые не удалось включить в основную последовательность. Хотя скаффолды хромосомной длины не являются идеальным представлением хромосом, поскольку содержат пробелы, вызванные ограничениями подхода Hi-C (описан выше), геномные сборки хромосомного уровня открывают возможности для более полной аннотации генома и оценки генетического разнообразия, что критично для решения актуальных проблем природоохранной биологии [66].

Стремительное развитие и снижение стоимости высокопроизводительного секвенирования в последние годы сделало возможными исследования в области природоохранной биологии с использованием полногеномных данных [67]. Подходы высокопроизводительного секвенирования третьего поколения позволяют получить прочтения рекордной длины, однако отличаются более высокой стоимостью. Использование коротких сцепленных прочтений 10X Genomics с последующим скаффолдингом до уровня хромосом посредством данных Hi-C-секвенирования стало быстрым и недорогим способом генерации высококачественных *de novo* сборок хромосомного уровня у млекопитающих [68].

1.2 Оценка гетерозиготности и анализ видовой принадлежности

1.2.1 Гетерозиготность, как одна из ключевых метрик генетического разнообразия

Появление доступных данных высокопроизводительного секвенирования открыло новые возможности в области природоохранных исследований, направленных на изучение генетического разнообразия, мониторинг численности популяций редких и исчезающих видов, а также разработку стратегий их сохранения и восстановления (реинтродукции) [69]. Использование полногеномных подходов, в отличие от методов низкого разрешения, таких как, например, анализ переменных локусов коротких tandemных повторов (STR), значимых лишь при анализе большого количества образцов [70], позволяют выявлять однонуклеотидные полиморфизмы (SNP) в масштабе практически целого генома исследуемой особи. Последующие анализы на основе SNP позволяют оценить как уровень генетической дифференциации между популяциями, так и изучить генетическое разнообразие внутри каждой из них [71]. Генетическое разнообразие в данном случае определяется генетическими различиями в геномах между отдельными особями и популяциями внутри одного вида, включая не только функционально значимые (влияющие на выживаемость и адаптацию к постоянно изменяющимся условиям среды), но и нейтральные SNP [72]. При этом считается, что оценки нейтрального генетического разнообразия тесно связаны с функциональным разнообразием, в связи с чем они остаются валидными для оценки адаптивного потенциала [73]. Серьезной общеизвестной проблемой является снижение уровня генетического разнообразия – генетическая эрозия. В результате генетической эрозии в геноме может происходить накопление вредных SNP в гомозиготном состоянии с последующим снижением адаптационного потенциала, что в конечном итоге приводит к угрозе вымирания вида [74,75]. Подобные явления классифицируют как инбредную или аутбредную депрессию вида, которая может происходить либо в случае скрещивания близкородственных особей, либо, напротив, особей из отдаленных популяций, соответственно [76,77]. Отметим, что вероятность аутбредной депрессии, как правило, ниже, чем инбредной, в связи с чем ее часто недооценивают [76]. Именно поэтому в ходе организации

мероприятий по реинтродукции видов, находящихся под угрозой исчезновения, важно учитывать вероятность возникновения обоих типов депрессии.

Гетерозиготность – это одна из главных метрик генетического разнообразия, используемая для контроля уровня генетического разнообразия [78]. В контексте природоохранных исследований уровень гетерозиготности является основной мерой оценки эволюционного и адаптивного потенциала популяции [79]. Показано, что находящиеся под угрозой исчезновения виды характеризуются меньшим генетическим разнообразием, в сравнении с видами, чей природоохранный статус не вызывает опасений [80]. В этой связи, генетическое разнообразие, наряду с размерами популяции, ее численностью и протяженностью ареала вида, считается одним из ключевых показателей для принятия решений о включении в Красный список Международного союза охраны природы (МСОП) [80]. Низкие значения уровня гетерозиготности могут указывать на фрагментацию популяции, высокий уровень инбридинга (скрещивание близкородственных особей) и критические сокращения численности (“бутылочные горлышки”) в популяционной истории [81]. Будучи интуитивно простой метрикой для оценки генетического разнообразия, гетерозиготность широко используется для оценки эффективного размера численности исчезающих и находящихся под угрозой исчезновения видов [82]. Как следствие, оценка гетерозиготности играет ключевую роль в разработке программ по сохранению и реинтродукции исчезающих видов, способствуя верному отбору особей для контролируемых скрещиваний и разведения как в неволе, так и в дикой природе [83].

Будучи ключевым ресурсом для поддержания генетического разнообразия, гетерозиготность играет важную роль в эволюции видов. Однако, высокий уровень данной метрики не всегда можно считать однозначным свидетельством благосостояния популяций, изолированных на протяжении долгого периода времени. Так, низкие показатели гетерозиготности, будучи целенаправленно сформированными согласно положениям “генетической революции”, могут являться положительным явлением и способствовать выживанию особей [84]. В изолированных или полуизолированных популяциях длительный отбор в пользу доминантных аллелей приводит к снижению уровня гетерозиготности, что, в свою

очередь, ослабляет негативные последствия инбридинга по сравнению с популяциями, недавно оказавшимися в условиях изоляции [85,86]. Даже если в малочисленной инбредной популяции закрепляются вредные мутации с выраженным фенотипическим эффектом, их воздействие может быть нейтрализовано полигенными модификаторами, регулирующими экспрессию генов и подавляющими проявление нежелательных признаков [87]. Более того, снижение гетерозиготности в субпопуляциях в конечном счете может способствовать увеличению генетической изменчивости как в метапопуляции, так и между популяциями, в случае потока генов между изначально изолированными группами, согласно "теории эволюции со смещающимся равновесием" [88]. Таким образом, уменьшение гетерозиготности в отдельных субпопуляциях может играть роль в поддержании генетического разнообразия на более высоких уровнях организации. И хотя не существует объективного стандарта уровня гетерозиготности, поскольку значение этого показателя у разных таксонов может различаться [89,90], его корректная оценка имеет решающее значение в организации мероприятий по реинтродукции исчезающих видов и способствует общему пониманию эволюционных процессов.

Оценить гетерозиготность можно на различных геномных уровнях, и в первую очередь на уровне отдельной позиции: SNP может быть либо в гомозиготном, либо в гетерозиготном состоянии. Несмотря на то, что последующая функциональная аннотация таких SNP способствует выявлению вредных генетических мутаций и позволяет получить представление об адаптивном потенциале особей [91], в контексте природоохранной биологии гетерозиготность в первую очередь определяется как доля гетерозиготных SNP в диапазоне от 0 до 1 без учета местоположения в геноме [79,92]. При этом выделяют ожидаемую гетерозиготность (H_E), которая оценивается как разница единицы и суммы квадратов частот гомозиготных аллелей, и наблюдаемую гетерозиготность (H_O), которая определяется как доля гетерозигот, усредненная по локусу или целому геному и зависит как от величины генетической изменчивости, так и от уровня инбридинга в популяции [93,94]. Для удобства значение наблюдаемой гетерозиготности часто сводят до усредненного количества SNP на 1 тыс. п.н.,

называя его средней гетерозиготностью. Примечательно, что данные о распределении гетерозиготности по геному могут быть использованы для реконструкции динамики популяционной истории. Общеизвестным является метод Парного последовательного марковского слияния (англ. Pairwise Sequentially Markovian Coalescent, PSMC), который оценивает колебания эффективного размера численности популяции (N_e) в прошлом, анализируя распределение гетерозиготных вариантов по геному [95]. Гетерозиготные варианты в данном случае выступают в роли “молекулярных меток” времен расхождения: чем реже расположены такие позиции, тем древнее событие последнего общего предка. Примечательно, что сопоставление траекторий популяционных историй близкородственных видов позволяет не только выявлять общие “бутылочные горлышки” и увеличение N_e в прошлом, но и делать выводы о времени расхождения видов [96].

Использование полногеномных данных способствовало увеличению точности вышеописанных показателей за счет значительно большего количества участвующих в анализе SNP и открыло новые подходы для оценки гетерозиготности [69,97]. В частности, стал широко использоваться метод скользящих по геному окнах фиксированного размера, который дал возможность работать с локальной гетерозиготностью в масштабе всего генома [98]. Такой подход позволил перейти от рассмотрения оценок гетерозиготности целого генома к анализу их распределения, что позволило не только определить доверительный интервал для среднего значения, но и выявить участки генома с пониженным или повышенным генетическим разнообразием [91]. Если анализ проводится на небольшой выборке особей, что может быть вызвано ограниченной численностью популяции, оценки гетерозиготности могут быть использованы для установления степени родства и оценки генетической структуры популяции, что, в свою очередь, способствует своевременному принятию мер по снижению риска генетической эрозии [99]. Хотя традиционные подходы анализа гетерозиготности на популяционном уровне требуют большого числа образцов, использование данных полногеномного секвенирования сделало возможным проведение анализа даже на небольшой выборке, сохраняя при этом высокую точность благодаря качеству и

покрытию геномных данных [100]. Для выявления генетических вариантов разработано множество биоинформатических инструментов: GATK, FreeBayes, VCFtools [101–103]. Последний является подходящим вариантом для работы с немодельными организмами благодаря гибким настройкам.

Распространенным методом визуализации распределений, в том числе и распределений оценок гетерозиготности, являются гистограммы, диаграммы размаха и скрипичные диаграммы (Рисунок 2) [104]. Эти подходы визуализации хорошо зарекомендовали себя в множестве природоохранных исследований, посвященных млекопитающим [105–109]. Например, в работе, посвященной байкальской нерпе (*Pusa sibirica*), с помощью гистограмм удалось сопоставить гетерозиготность особей этого вида с другими представителями ластоногих (Рисунок 2, А) [105]. В свою очередь, диаграммы размаха были использованы в работе по изучению особей сахарийского орикса (*Oryx dammah*), где с их помощью удалось показать различия в оценках полногеномной гетерозиготности между популяциями (Рисунок 2, В) [110]. Наконец, скрипичные диаграммы, будучи разновидностью диаграмм размаха, подходят для визуализации мультимодальных распределений [111]. Так, с использованием скрипичных диаграмм было выполнено сравнение распределений гетерозиготности близких видов из подсемейства Mustelinae в исследовании, посвященном чернолапому хорьку (*Mustela nigripes*) [106]. Стоит отметить, что скрипичные диаграммы могут быть использованы вместе с диаграммами размаха, как это было сделано в работе по изучению байкальской нерпы (Рисунок 2, С) [105].

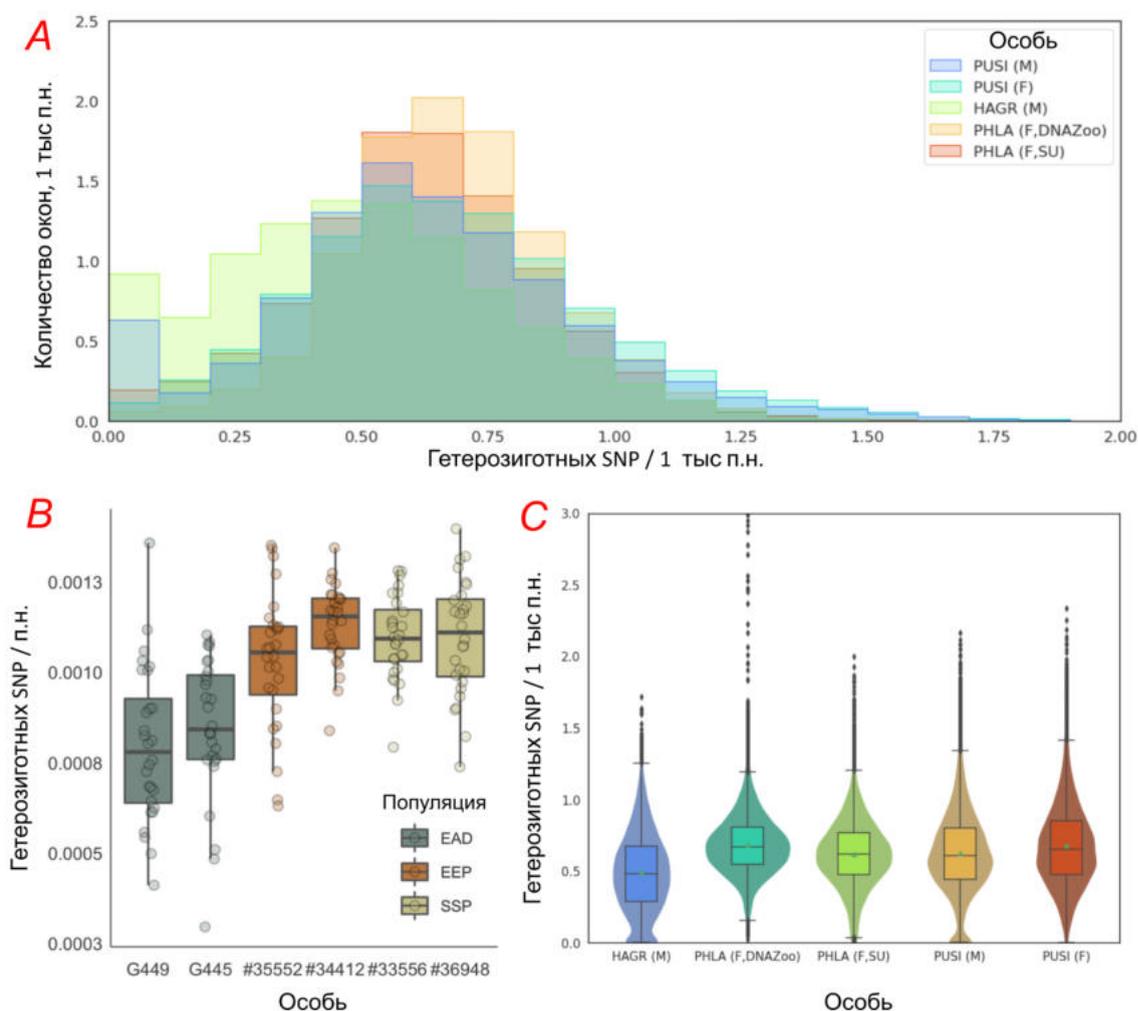


Рисунок 2. Примеры визуализации распределения оценок гетерозиготности.

А – гистограмма распределения оценок гетерозиготности по геному для пяти выборок ластоногих; В – распределение в виде диаграмм размаха для особей сахарийского орикса из различных популяций; С – скрипичные диаграммы распределения гетерозиготности для 2 особей байкальской нерпы, 2 особей пятнистой нерпы и 1 особи серой нерпы. Адаптировано из [104,105,110].

Важным фактором при анализе гетерозиготности является уровень сборки генома, используемой в качестве референса. В случае, если используется геномная сборка скаффолдного уровня, для визуализации распределения уровня гетерозиготности могут быть использованы линейные графики, позволяющие отразить вариацию гетерозиготности относительно координат в сборке генома исследуемого вида. Например, данный подход визуализации использован в

исследовании генетического разнообразия львов (*Panthera leo*), где проводили сравнение уровня гетерозиготности Танзанийских и Индийских популяций (Рисунок 3, А) [112]. Аналогичный подход визуализации гетерозиготности также встречается в работе по сравнительной геномике морской (*Enhydra lutris*) и гигантской (*Pteronura brasiliensis*) выдр [113].

Появление доступных геномные сборки хромосомного уровня, получение которых стало возможно за счет развития как технологий полногеномного секвенирования, так и биоинформатических подходов, открыло новые возможности в контексте природоохранных исследований. Стал возможен анализ вариации уровня гетерозиготности отдельно для каждой хромосомы. Стандартом для визуализации в этом случае стали тепловые карты (*англ.* heat map). К примеру, этот подход был использован в работе, посвященной исследованию восьми видов млекопитающих, где с использованием тепловых карт в геноме евразийской речной выдры (*Lutra lutra*) удалось выявить отдельные регионы как с повышенным уровнем (Рисунок 3, В), так и с пониженным уровнем гетерозиготности, что является наглядным примером инбридинга в зоопарках [66]. Аналогичный подход анализа гетерозиготности на уровне отдельных хромосом встречается в множестве других работ [108,109,114–116].

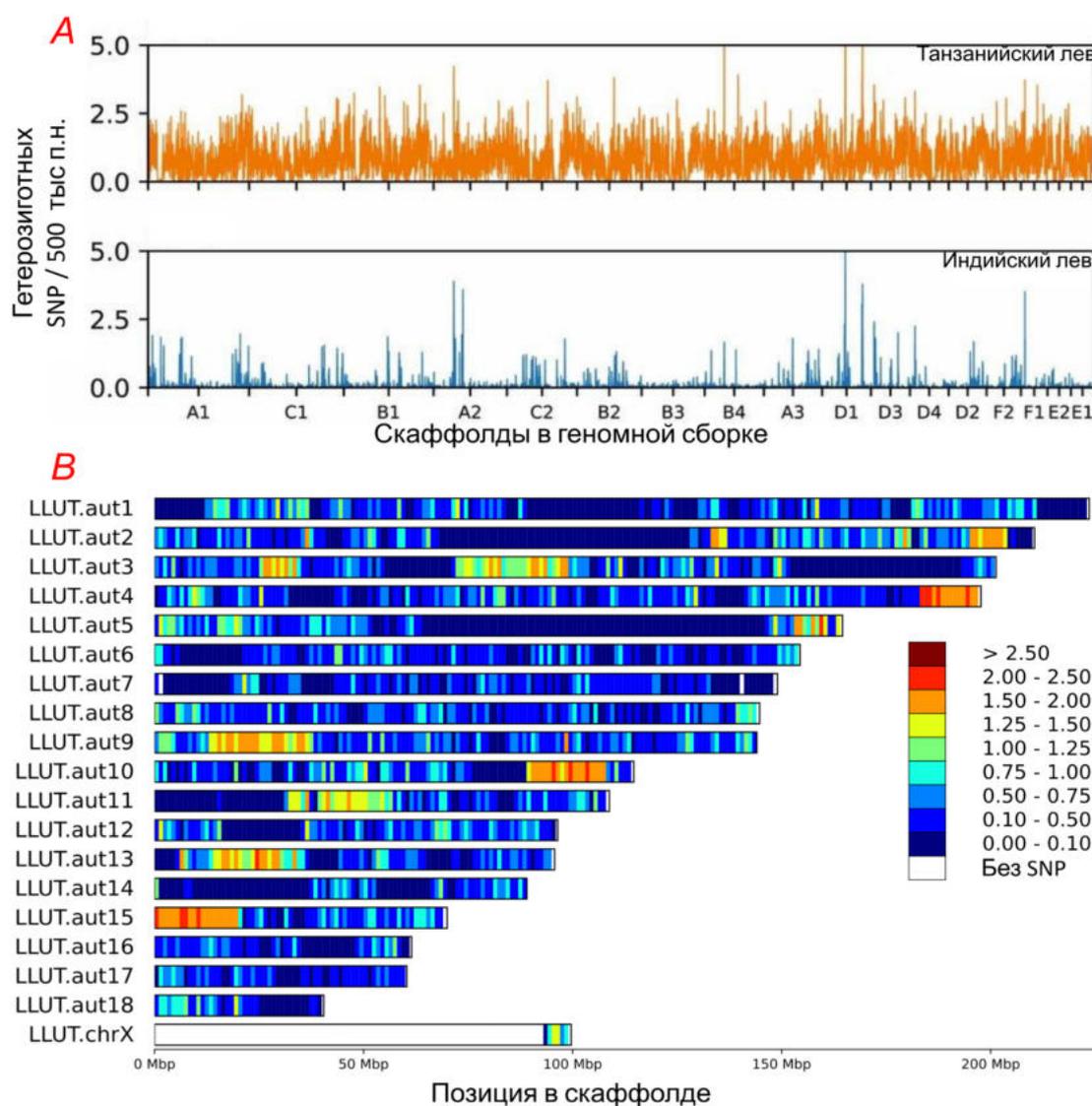


Рисунок 3. Визуализация локальной средней гетерозиготности в окнах.

А – линейный график распределения средней гетерозиготности в скользящих окнах размером 500 тыс п.н. у Танзанийского льва (оранжевый график) и Индийского льва (голубой график). По оси X представлены последовательно объединенные хромосомы; В – тепловая карта плотности гетерозиготных SNP для самца Евразийской речной выдры. Использована геномная сборка хромосомного уровня, гетерозиготные SNP подсчитаны в окнах размером 1 млн. п.н. и масштабированы до SNP на 1 тыс п.н. Оригинальные изображения из [66,104,112].

Анализ гетерозиготности можно условно разделить на изучение регионов с пониженным и повышенным уровнем гетерозиготности. Непрерывным

протяженным участкам с пониженным генетическим разнообразием (англ. Runs of Homozygosity, RoH), уделяется наибольшее внимание [117]. Такие регионы в геноме являются следствием инбридинга, в результате которого наследуются идентичные гаплотипы от обеих родительских особей [117]. При этом длина RoH напрямую связана с давностью событий инбридинга: длинные RoH указывают на недавнее скрещивание близкородственных особей, тогда как короткие и фрагментированные участки свидетельствуют о древних событиях инбридинга. Это связано с тем, что с каждым поколением происходят разрывы протяженных участков RoH посредством кроссинговера [118]. Таким образом, короткие RoH и небольшое их количество характерно для крупных неизолированных популяций с активным потоком генов, тогда как протяженные участки RoH встречаются в изолированных популяциях, в которых поток генов отсутствует. Именно инбридинг способствует увеличению протяженности участков RoH в геноме, что, в свою очередь, может усиливать проявление вредных эффектов рецессивных генетических аллелей, когда они оказываются в гомозиготном состоянии у инбредных особей [119,120]. Это приводит к росту частоты рецессивных заболеваний, и, в конечном итоге, к снижению репродуктивного успеха и жизнеспособности особей в популяции – инбредной депрессии [121]. Помимо этого, идентификация RoH может способствовать выявлению специфических признаков отбора, давление которого может способствовать закреплению одного аллеля в популяции [122]. Такое закрепление (фиксация) определенного аллеля под воздействием отбора приводит к росту уровня гомозиготности в прилегающих геномных регионах [123]. Помимо визуализации RoH непосредственно на хромосомах (при использовании геномных сборок хромосомного уровня), их распределение в геноме часто визуализируют в виде графика кумулятивной суммы. Для примера, этот подход был использован в работе по, который, например, использовался в исследовании генетического разнообразия львов из различных популяций (Рисунок 4) [112].

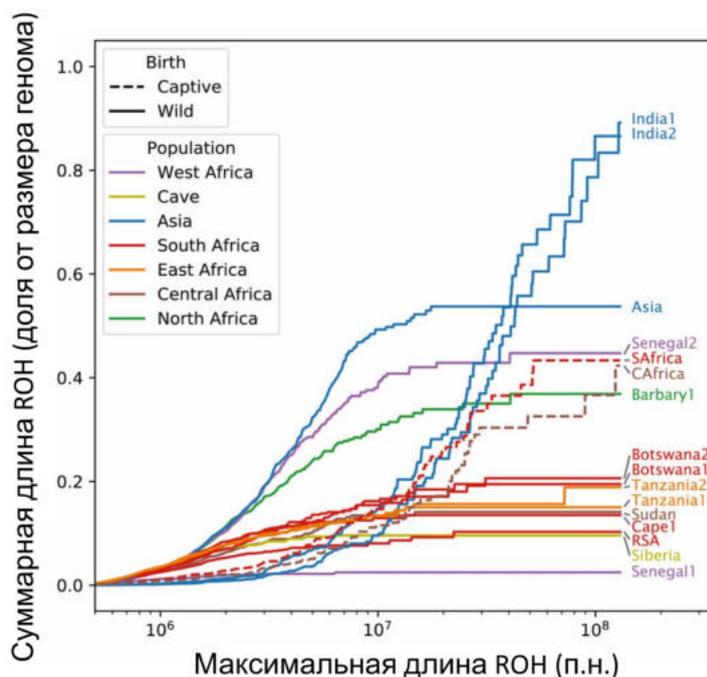


Рисунок 4. Распределение участков RoH в виде графика кумулятивной суммы для львов из различных популяций. Адаптировано из [104,112].

В отличие от RoH, участки с повышенным уровнем гетерозиготности (англ. Runs of Heterozygosity, RoHet) описаны в литературе хуже. Однако их выявление также играет большое значение при изучении эволюционной истории популяции. Например, анализ RoHet позволяет выявлять участки генома, где поддержание высокого уровня гетерозиготности может играть ключевую роль в выживании [124]. Несмотря на то, что для выявления участков RoHet сложнее устанавливать четкие пороговые значения, в отличие от RoH, локализация участков RoHet зачастую выполняется параллельно с анализом RoH, особенно в случае изучения особенностей геномов особей из недавно смешанных близкородственных популяций, геномы которых отличаются чередованием участков RoH и RoHet [125]. Локализация таких участков в геноме и последующий анализ их происхождения позволяет изучить морфологические особенности исследуемых особей [126].

Таким образом, современные технологии секвенирования открывают возможности для детального изучения внутривидовой и межвидовой генетической изменчивости [47]. Геномные сборки на уровне хромосом стали

ключевым ресурсом для корректной оценки и последующего анализа генетического разнообразия, обеспечивая наглядную визуализацию и открывая новые возможности для сравнительного анализа распределений гетерозиготности на хромосомных скаффолдах [66,104]. Кроме того, использование таких сборок геномов вместе с данными полногеномного секвенирования позволяет исследовать закономерности генетического разнообразия, например, при дивергенции между близкородственными видами, позволяя выявлять участки генома, подвергшиеся недавнему или древнему естественному отбору, а также интрогрессии [127,128]. Применение подходов современной геномики открывает новые возможности для решения актуальных вопросов природоохранной биологии, решение которых было недоступно при использовании традиционных молекулярно-генетических подходов, основанных на отдельных молекулярно-генетических маркерах [129].

1.2.2 Обнаружение геномной интрогрессии

Гибридизация и смешение популяций – распространенное следствие вторичного контакта между недавно дивергировавшими видами, играющее важную роль в эволюции, включая как адаптацию, так и видообразование [130,131]. Изучение интрогрессивной гибридизации (или интрогрессии) – процесса переноса генетического материала от одного вида к другому посредством гибридизации и многократных обратных скрещиваний – имеет важное значение для понимания эволюционных процессов и изучения видообразования [132,133]. Поток генов может уменьшать или увеличивать приспособленность особей, в связи с чем идентификация исторического и недавнего потока генов между популяциями или близкородственными видами имеет важное значение для биоразнообразия [91]. В отличие от простой гибридизации, результатом которой является равномерное смешение генетического материала родительских особей, интрогрессия приводит к более сложной и изменчивой смеси генов [132]. При многократных обратных скрещиваниях гибридов F1 с одним из родительских видов терминология «гибрид» и «вид» становится расплывчатой (Рисунок 5).

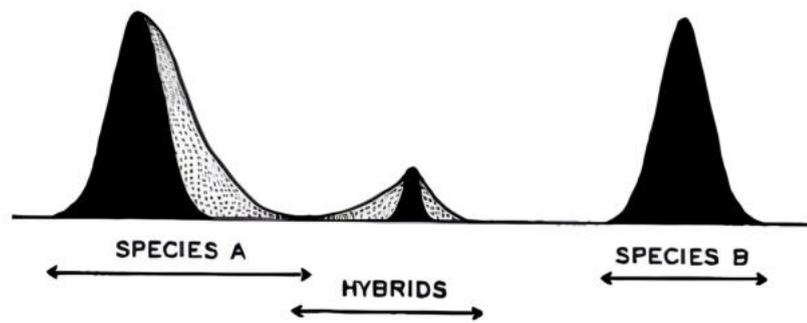


Рисунок 5. Термины «вид» и «гибрид» в случае интрогрессивной гибридизации. Сплошные черные области представляют исходные виды и гибриды первого поколения. Пунктирные области представляют более поздние поколения гибридов и обратные скрещивания. Адаптировано из [133].

Одной из ключевых проблем при анализе интрогрессии является точная количественная оценка степени и направления потока генов. Ниже рассмотрены D-статистика, F4-статистика и F3-статистика – три широко используемых статистических теста для обнаружения потока генов и оценки интрогрессии путем подбора явной модели топологии филогенетического дерева (Рисунок 6) к набору популяций или близкородственных видов [134].

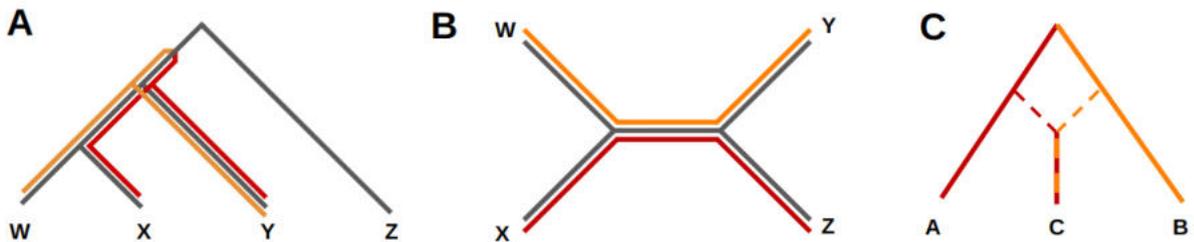


Рисунок 6. Схемы топологий филогенетических деревьев для вычисления статистик, основанных на частотах аллелей.

A – D-статистика; B – F4-статистика; C – F3-статистика. Цветными линиями на A и B обозначены паттерны частот аллелей ABBA (красный) и BAAB (оранжевый). На C пунктирными линиями обозначен поток генов в таксон C от таксонов A и B. Адаптировано из [134].

D-статистика является широко используемым методом для выявления интрогрессии и направления потока генов путем сравнения частот аллелей [134,135]. Метрика основана на тестах АВВА-ВАВА, которые являются распространенным методом анализа моделей распределения аллелей для оценки доказательств потока генов, идея которого заключается в анализе частот предковых ("A") и производных ("B") аллелей в геномах четырех таксонов: W – целевой таксон, X и Y – таксоны с потенциальной гибридизацией и Z – аутгруппа (Рисунок 6, А) [135–137]. При сценарии без интрогрессии два определенных аллельных паттерна АВВА и ВАВА должны встречаться одинаково часто, тогда как избыток одного из паттернов приводит к D-статистике, которая значительно отличается от нуля, свидетельствуя о потоке генов между двумя таксонами. Модели сайтов упорядочены таким образом, что модель ВВАА относится к W и X, разделяющим производный аллель, АВВА – к X и Y, разделяющим производный аллель, и ВАВА – к W и Y, разделяющим производный аллель. D-статистика рассчитывается как:

$$D(W, X, Y, Z) = \frac{\sum_{i=1}^n C_{ВАВА}(i) - C_{АВВА}(i)}{\sum_{i=1}^n C_{ВАВА}(i) + C_{АВВА}(i)}$$

В данном случае, АВВА – число участков, где X и Y имеют общий производный аллель, а W и Y и X и Z – нет, ВАВА – число участков, где W и Y имеют общий производный аллель, а W и X и X и Z – нет. Согласно нулевой гипотезе, которая предполагает отсутствие потока генов, модели АВВА и ВАВА должны возникать с одинаковой частотой, а отклонение от этого ожидания соответствует интрогрессии между Y и W или X [134,136]. Таким образом, положительная D-статистика (т.е. превышение ВАВА) указывает на интрогрессию между W и Y, тогда как отрицательная D-статистика (т.е. превышение АВВА) указывает на интрогрессию между X и Y. Таким образом, отрицательная D-статистика (т.е. превышение АВВА) указывает на интрогрессию между X и Y, тогда как положительная D-статистика (т.е. превышение ВАВА) указывает на интрогрессию между W и Y.

Поскольку D-статистика может быть необъективной, если внешняя группа не является реальной, для подтверждения полученных результатов также применяется F4-статистика (так называемый “тест 4-х популяций”), которая аналогичным

образом позволяет выявить интрогрессию и поток генов на основании частот аллелей среди четырех таксонов [138]. В отличие от D-статистики, F4 предполагает топологию без корней (Рисунок 6, В) и рассчитывается как:

$$F_4(W, X, Y, Z) = \frac{\sum_{i=1}^n C_{BABA}(i) - C_{ABBA}(i)}{\sum_{i=1}^n C_{Sites}(i)}$$

[134]. При отсутствии потока генов разница частот аллелей между W и X полностью независима от разницы частот аллелей между Y и Z, что приводит к нулевой F4-статистике. При условии, что имел место поток генов из Y или Z в W или X, статистика должна быть отличной от нуля. В частности, если F4-статистика отрицательна, это подразумевает поток генов либо между Y и X, либо Z и W, при этом, если она положительна – подразумевается поток генов между W и Y или между X и Z. Стоит отметить, что F4-статистика отличается от D-статистики несколько большей неоднозначностью, за счет измененного знаменателя. Изначально F4 статистика применялась к популяциям одного или родственных видов в целях проверки, являются ли четыре таксона двумя парами сестринских таксонов [138], но она также эффективна для обнаружения недавней интрогрессии между видами, как и D-статистика [139].

F3-статистика применяется в качестве доказательства генетического смешения между двумя популяциями и определяется как среднее от произведения разностей частот аллелей между целевой популяцией C и предковыми A и B, соответственно: $\mu(C - A)(C - B)$ [134]. Следовательно, F3-статистика отрицательна, если частота аллелей целевой популяции (C) находится между частотами аллелей двух предковых популяций (A и B) (Рисунок 6, С). Статистически значимая отрицательная F3-статистика по нескольким позициям подразумевает промежуточные частоты аллелей и указывает на то, что “С” является результатом смешения двух исходных популяций. Статистика F3 полезна в различных сценариях. Во-первых, она может быть использована для определения общего потока генов между двумя популяциями A и B относительно общей аутгруппы C. При таком сценарии более высокое значение F3 указывает на более длительное общее время эволюции между популяциями A и B. Во-вторых,

статистику F_3 можно также использовать для проверки того, является ли целевая популяция C смесью двух предковых популяций A и B [134]. В этом случае значимо отрицательные значения F_3 -статистики будут служить статистическим доказательством генетического смешения. Важно отметить, что положительные значения статистики F_3 не указывают на отсутствие интрогрессии [140].

1.2.3 Анализ происхождения

Анализ происхождения, как способ анализа интрогрессии, позволяет выявить доли предковых популяций в рамках более крупной популяции [141]. Анализ происхождения может дать представление об исторических процессах, таких как миграция и смешение (скрещивание между особями из разных групп, например, между двумя родственными видами), и может помочь в работе по сохранению популяции, выявляя генетически отличные популяции, требующие особого внимания [142]. Существует глобальный анализ происхождения (англ. Global ancestry inference, GAI), представляющий собой оценку доли предковых популяций, усредненную по всему геному, и локальный анализ происхождения (англ. Local ancestry inference, LAI), заключающийся в оценке доли предковых популяций в геномных окнах для каждой хромосомы [143]. Оба варианта анализа активно применяются для изучения структуры популяций и оценки генетической дифференциации близкородственных видов [12,144].

Глобальный анализ происхождения можно осуществить при помощи непараметрических подходов и подходов на основе моделей [145]. Основным непараметрическим подходом является PCA (англ. principal component analysis), который представляет собой математический подход к снижению размерности многомерного набора данных до меньшего набора независимых измерений, называемых главными компонентами [146]. В контексте популяционной генетики, результатом PCA является обобщение генетической изменчивости в виде независимых главных компонент с последующей визуализацией в виде графика в двух- или трехмерном пространстве, что дает наглядное представление о генетическом разнообразии и популяционной структуре. Образование отдельных групп или кластеров на графике является признаком сходства по причине общего

происхождения образцов [147,148]. Как правило, первая и вторая главные компоненты отражают ключевые различия между особями, популяциями или группами близкородственных видов. Оставшиеся компоненты отражают дополнительные различия и могут дать представление о субструктуре внутри популяций. Стоит отметить, что наблюдаемые на графике PCA расстояния между образцами не отражают фактических генетических расстояний вне зависимости от размеров выборки [149]. Ключевым требованием PCA является независимость и отсутствие ложных корреляций частот аллелей, как правило коррелирующих из-за неравновесия по сцеплению (англ. linkage disequilibrium (LD)) [150]. Подход PCA, в применении к молекулярно-генетическим данным, реализован во множестве инструментов: SmartPCA (пакет EIGENSOFT), FlashPCA2, PLINK, Bigstatsr и Bigsnpr (пакеты для языка R), TeraPCA, ProPCA [148,151–155]. Изначально PLINK представлял собой набор инструментов для полногеномных ассоциативных исследований (GWAS) и исследований в области популяционной генетики человека, но позднее был модифицирован для работы с немодельными видами [156]. PLINK является высокоэффективным с вычислительной точки зрения инструментом и предоставляет функционал для фильтрации однонуклеотидных полиморфизмов (SNP) и проведения различных анализов, в том числе и анализа главных компонент (PCA). Несмотря на то, что PCA является детерминированным алгоритмом, незначительные изменения выборки могут приводить к различным результатам, что является фундаментальной проблемой устойчивости, надежности и воспроизводимости [157–159]. Стоит также отметить, что значения дисперсии, объясненные главными компонентами, часто бывают незначительными, что указывает на необходимость подтверждения полученных результатов с использованием альтернативных методов кластеризации [149].

Альтернативные подходы кластеризации, используемые для глобального анализа происхождения, пытаются оценить вероятное количество и пропорции предковых популяций с использованием статистических моделей. Наиболее известными инструментами генетической кластеризации являются STRUCTURE [160] и ADMIXTURE [161], которые моделируют вероятность наблюдаемых генотипов на основании частот аллелей. STRUCTURE и адаптированный для

больших наборов данных fastSTRUCTURE являются инструментами кластеризации на основе байесовской модели, которые позволяют определить генетическую структуру популяции на основе данных о ее мультилокусных генотипах [162]. Метод требует задавать количество предковых популяций и предполагает, что частоты аллелей в каждой популяции находятся в равновесии Харди-Вайнберга. ADMIXTURE использует аналогичную модель, но использует метод максимального правдоподобия для оценки доли предковых популяций. ADMIXTURE также включает в себя алгоритмы оптимизации, что делает его значительно более быстрым, чем STRUCTURE, без потери точности результатов [161]. Оба инструмента оценивают апостериорную вероятность числа предковых популяций и выполняют кластеризацию особей на основании их генетического сходства. Стоит отметить, что хотя подходы на основе моделей отличаются от непараметрических методов, оба этих подхода в теории и на практике на реальных и смоделированных данных выявляют аналогичные паттерны популяционной структуры [163].

Известно, что результаты PCA и STRUCTURE/ADMIXTURE могут быть ошибочно интерпретированы как свидетельство гибридизации в случае, если особи относятся к более чем 2 предковым популяциям [164,165]. Однако, это также может быть результатом изоляции по расстоянию, в результате которой генетическая изменчивость коррелирует с географической близостью, а не с дискретными границами популяции [166]. Одна из стратегий для того, чтобы отличить гибридизацию от изоляции по расстоянию, заключается во включении в кластеризацию географических данных. Такой подход реализован, например, в программе conStruct, который может выявить изоляцию по расстоянию, показывая, что генетические кластеры коррелируют в зависимости от географического расстояния [165]. Другой подход, реализованный, например, в инструменте newhybrids, заключается в классификации особей на основе частот генотипов с целью выявления стадий гибридизации [167]. Наиболее простым и эффективным подходом является визуализация зависимости гибридного индекса и гетерозиготности, реализованная, например, в пакете triangulaR [168]. В случае гибридизации особи с промежуточным (~0.5) гибридным индексом будут

отличаться высоким уровнем гетерозиготности, тогда так при изоляции по расстоянию гетерозиготность не будет возрастать (и может снижаться) при удалении от кластеров родительских популяций [169]. Таким образом, подход позволяет отличить гибридизацию от изоляции по расстоянию и, в случае гибридизации, может использоваться для определения гибридных классов [169,170].

Хромосомы в геномах особей из недавно смешанных популяций представляют собой набор чередующихся сегментов, каждый из которых относится к определенной предковой популяции [125]. Такие хромосомы называют “мозаичными”, поскольку кроссинговер превращает хромосому в мозаику участков хромосом от родительских особей из различных популяций [171]. Распределение этих предковых участков может изменяться и перестраиваться с каждым новым поколением, что представляет собой ценную информацию о популяционной истории. Размеры таких хромосомных участков напрямую зависят от времени генетического смешения – более поздний поток генов приводит в среднем к более протяженным сегментам от разных популяций [172]. Определение координат и последующий анализ происхождения этих сегментов – локальный анализ происхождения или локальный ADMIXTURE-анализ – представляет большой интерес у популяционных генетиков, поскольку дает уникальные возможности для изучения генетических основ как морфологических особенностей, так и различных заболеваний [126]. В отличие от глобального, локальный анализ происхождения подразумевает не только среднюю оценку пропорций предковых популяций для исследуемой особи или группы особей, но и определение координат геномных областей от разных предковых популяций. Локальный анализ происхождения также позволяет получить более точную среднюю оценку пропорций предковых популяций, в связи с чем он может быть использован для подтверждения достоверности результатов глобального анализа происхождения [173].

Одним из первых подходов для оценки происхождения на локальном уровне был вариант, реализованный в инструменте STRUCTURE версии 2.0, идея которого заключалась в определении предковых популяций путем кластеризации гаплотипов на основе паттернов LD [174]. Большинство из существующих на данный момент

инструментов используют скрытые Марковские модели (англ. Hidden Markov model, HMM) для оценки вероятности и моделирования происхождения, основанного на наблюдаемом порядке SNP или гаплотипов, для каждого геномного участка [175,176]. Существуют также альтернативные подходы в качестве дополнения или замены HMM, основанные на PCA (PCAdmix), Марковских цепях Монте-Карло (MCMC), алгоритме случайного леса [176–178]. Важно отметить, что данные подходы для локального анализа происхождения ранее применялись в основном для изучения недавно смешанных человеческих популяций, с целью выявления и анализа генов, ассоциированных с различными заболеваниями (так называемое “картирование примесей”) [179]. Тем не менее, в последние годы аналогичные методы активно применяются в популяционных исследованиях немодельных видов животных [144,180].

1.3 Филогеномный анализ данных

Филогенетический анализ является основополагающим инструментом эволюционной биологии, поскольку он позволяет реконструировать эволюционную историю организмов путем анализа генетических отношений между видами [181]. Благодаря активному развитию высокопроизводительных методов секвенирования, традиционные подходы молекулярной филогенетики, основанные на отдельных молекулярно-генетических маркерах, эволюционировали в “филогеномику”, преимуществом которой является значительное увеличение числа филогенетически информативных регионов [182,183]. Термин “филогеномика” был впервые использован в контексте функциональной аннотации генов [184], однако в настоящее время оно также применяется для обозначения филогенетического анализа на основании полногеномных данных. Основные этапы данного анализа включают идентификацию и получение ортологичных последовательностей из геномных сборок, их выравнивание с последующим исключением гипервариабельных областей и, наконец, реконструкцию филогенетического дерева [181,185].

1.3.1 Методы поиска ортологичных генов

Ортологичные последовательности позволяют проследить эволюционные изменения на молекулярном уровне, однако их анализ может быть затруднен из-за вариабельности отдельных нуклеотидных позиций и различий в длине выравниваний. Переход к уровню генов упрощает интерпретацию данных, поскольку гены представляют собой функционально значимые единицы, что делает их более удобными для филогенетического анализа. Как следствие, реконструкция филогенетического дерева должна основываться на ортологах – генах, которые были унаследованы от общего предка двух или более видов и, как правило, выполняющих аналогичные функции в разных организмах [186–188]. Идентификация ортологичных генов в наборе геномных последовательностей является важнейшей задачей в рамках сравнительной геномики, на результат решения которой может повлиять выбор метода предсказания ортологии [189].

База OrthoDB является ведущим ресурсом, который предоставляет наборы ортологов для различных таксонов [190]. OrthoDB представляет собой иерархический каталог эволюционных и функциональных аннотаций ортологичных генов животных, грибов и бактерий, основанный на филогенетическом алгоритме попарных сравнений последовательностей белков из разных видов для выявления ортологов и паралогов. OrthoDB пополняется ортологичными последовательностями, идентифицированными при помощи инструмента OrthoMCL, в котором реализован подход, основанный на Марковском алгоритме кластеризации (англ. Markov Clustering Algorithm, MCL), для объединения белков в ортологичные группы [191]. Алгоритм сначала формирует матрицу сходства "все против всех" на основе парного сходства последовательностей входных белков, а затем выполняет кластеризацию сходных белков в группы. В дополнение к вычислительным подходам идентификации генов, OrthoDB также включает в себя ручное курирование.

Стандартным инструментом для идентификации и аннотации ортологов в геномных сборках является BUSCO (англ. Benchmarking Universal Single-Copy Orthologs) [192]. Инструмент оценивает полноту и качество сборки генома путем определения набора высококонсервативных однокопийных генов (BUSCOs) из базы

данных ортологов OrthoDB (<https://www.orthodb.org/>). Идентифицированные BUSCO гены объединяются в различные категории (семейства BUSCO) на основе их таксономического положения и эволюционной истории, и могут быть использованы в филогенетическом анализе в качестве маркеров для изучения эволюционных отношений широкого диапазона видов [192]. Было показано, что филогенетический анализ на основе BUSCO устойчив к изменениям в таксономической выборке и может служить надежной филогенетической основой для изучения филогенетических взаимоотношений широкого диапазона видов [193]. Важно отметить, что BUSCO не предоставляет информацию об эволюционных отношениях или функциональном сходстве между генами, но позволяет получить достоверный набор последовательностей ортологов для дальнейшего анализа.

1.3.2 Подходы множественного выравнивания и фильтрации последовательностей

Множественное выравнивание последовательностей (англ. Multiple sequence alignment или MSA) – фундаментальный биоинформатический метод, подразумевающий сравнение трех и более аминокислотных или нуклеотидных последовательностей путем выравнивания [194]. Анализ MSA позволяет выявить структурные и функциональные сходства между последовательностями, которые могут служить основой для гипотез о гомологии и дальнейшего филогенетического анализа. Для выполнения MSA существует множество биоинформатических инструментов: MAFFT, MUSCLE, Kalign, T-Coffee [195–198]. Одним из наиболее предпочтительных является инструмент PRANK, основанный на вероятностном подходе для получения множественного выравнивания последовательностей [199]. В отличие от других инструментов в PRANK реализован алгоритм, использующий филогенетическую информацию, что позволяет учитывать эволюционные расстояния между протяженными последовательностями ДНК, кодонов и аминокислот, а также корректно обрабатывать инсерции и делеции без переоценки этих событий.

Качество полученного MSA оказывает прямое влияние на дальнейшую филогенетическую реконструкцию в той же степени, что и алгоритм выравнивания [200]. Существует множество подходов для удаления отличающихся областей или пробелов из выравнивания. Изначально был предложен подход, основанный на попарных сравнениях для обнаружения консервативных областей, однако этот подход никак не обрабатывал пробелы выравнивания [201]. После был предложен метод удаления только пробелов и находящихся в непосредственной близости с ними различающихся позиций [202]. Позднее был предложен ряд других методов, способных различать как переменные, так и консервативные области выравнивания, но, поскольку эти подходы не были специально разработаны для повышения эффективности филогенетического анализа, они были не способны различать наиболее информативные позиции в выравнивании [203,204]. Gblocks является инструментом для автоматической фильтрации MSA от гиперпеременных участков с минимизированной потерей информативных для последующего филогенетического анализа сайтов [205]. Алгоритм Gblocks основан на сохранении блоков выравнивания исходя из набора заданных параметров в отношении количества смежных консервативных позиций, отсутствия промежутков и высокой степени консервативности, что делает окончательное выравнивание более подходящим для филогенетического анализа. Так, GBlocks был использован, например, в работе Yu et al. 2011, посвященной анализу филогении подсемейств куньих [6].

1.3.3 Реконструкция филогенетических деревьев

Поскольку геномные локусы зачастую отличаются друг от друга скоростью эволюции, топологии филогенетических деревьев отдельных генов часто различаются между собой и не совпадают с деревом видов. Это объясняется тем, что отдельные геномные локусы могут иметь разную эволюционную историю из-за таких процессов, как неполная сортировка линий, прошлая или недавняя интрогрессия и рекомбинация [206]. По этой причине филогенетический анализ на полногеномных данных (филогеномный анализ) обладает наибольшей

разрешающей способностью, поскольку основывается не на одном, а на множестве локусов.

Существует две основных стратегии для реконструкции филогенетических деревьев на основании набора множественных выравнений ортологичных последовательностей. Первая, известная как метод конкатенации или “суперматричный” подход, подразумевает получение единого выравнения (или “суперматрицы”) путем объединения каждого полученного, что позволяет суммировать и усреднять их филогенетических сигнал для получения глобальной оценки полногеномного масштаба для видового дерева (Рисунок 7). Конкатенация рассматривается как подход, который стремится к правильному (усредненному) филогенетическому дереву видов, путем объединения данных по множеству генов. Однако при значительных различиях в топологиях генных деревьев, обусловленных вариациями в моделях и скоростях эволюции генов, а также из-за вышеупомянутых процессов неполной сортировки линий, интрогрессии и рекомбинации [207,208], этот метод может приводить к ошибочным результатам, как было показано, например, при реконструкции филогении 27 видов из семейства кошачьих [206]. Реконструкция филогенетического дерева на основе полученной “суперматрицы”, как правило, осуществляется методами максимального правдоподобия или Байесовского вывода с использованием вероятностных моделей эволюции последовательности [209].

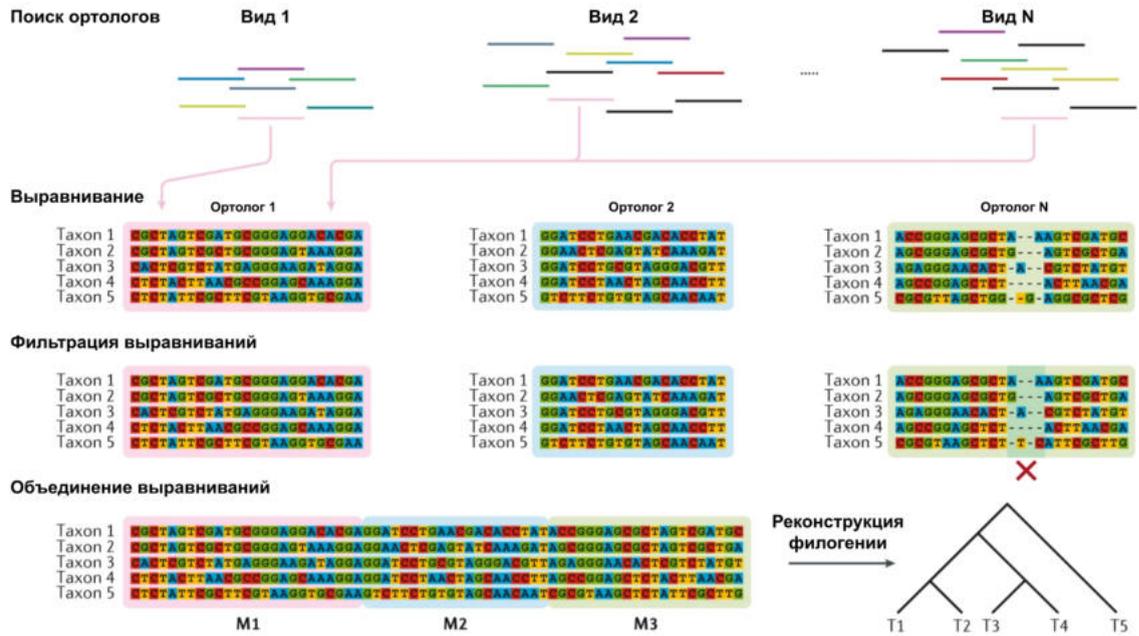


Рисунок 7. Схема реконструкции филогении методом конкатенации. Адаптировано из [181].

Метод максимального правдоподобия (англ. Maximum Likelihood, ML) является широко используемым подходом для реконструкции филогении, который оценивает топологию дерева и длину ветвей, максимизирующую вероятность наблюдаемых данных при заданной модели эволюции. Метод ML предполагает, что последовательности независимо эволюционировали вдоль ветвей дерева и что скорость возникновения замен постоянна во времени. Подход популярен благодаря своей точности и способности работать со сложными эволюционными моделями. Существует множество инструментов для реконструкции филогенетических деревьев методом ML: RAxML-NG, PAML, PhyML и FastTree [210–213]. Одним из предпочтительных является IQ-TREE [214], поскольку он автоматически определяет наиболее подходящую модель замещения при помощи ModelFinder [215]. Метод Байесовского вывода (англ. Bayesian inference, BI) оценивает апостериорное распределение вероятностей деревьев с учетом наблюдаемых данных и предварительного распределения параметров модели эволюции. В отличие от подхода ML, который оценивает топологию одного дерева и длины ветвей, метод BI оценивает все апостериорное распределение деревьев, которое может быть обобщено как консенсусное дерево. Метод BI предполагает задание

предварительного распределения параметров модели эволюции и вычисление апостериорного распределения деревьев с помощью МСМС (англ. Markov Chain Monte Carlo). Байесовский подход для решения задач молекулярной филогенетики был применен в 1990-х годах и позднее реализован в таких программах, как MrBayes, RevBayes, BEAST, PhyloBayes [216–220]. Несмотря на то, что невозможно подобрать однозначно верную модель, подходы ML и BI позволяют проверить предполагаемую модель, оценить и улучшить ее путем включения критических особенностей эволюционного процесса, благодаря чему данные подходы остаются наиболее предпочтительными в области филогенетики [221,222].

Дополнительно важно отметить метод присоединения соседей (англ. Neighbor-Joining, NJ), основанный на минимизации суммы длин ветвей и анализе матрицы попарных эволюционных расстояний между таксонами [223]. В отличие от ML и BI, которые работают напрямую с последовательностями и моделями эволюции, NJ является алгоритмом иерархической кластеризации и основывается на матрице расстояний [223], что делает его вычислительно менее затратным, но менее чувствительным к сложным эволюционным сценариям, происходящих, например, из-за нарушений молекулярных часов. Алгоритм NJ основан на последовательном объединении ближайших (соседних) таксонов, корректируя расстояния с учетом ветвления, пока не будет построено полностью разрешенное дерево [223]. Алгоритм реализован в множестве инструментов, наиболее известными из которых являются PHYLIP, QuickTree и RapidNJ [224–226].

Вторая стратегия реконструкции филогенетических деревьев заключается в первоначальной независимой реконструкции каждого генного дерева (посредством подходов ML или BI) с последующей суммарной оценкой консенсусной видовой топологии (Рисунок 8). Такой подход называют двухэтапным или обобщенно-коалесцентным [227]. В отличие от “суперматричного” метода, такой подход более устойчив к различиям в эволюции разных генов и, как следствие, различиями в топологии по сравнению с деревом видов [228]. Тем не менее, основная критика данного подхода заключается в реконструкции деревьев на порой относительного коротких последовательностях генов, что может исказить конечный результат. Обобщенно-коалесцентные методы оценки видовой топологии

реализованы в таких инструментах, как NJst, ASTRID, DISTIQUE, SVDQuartets [229–232]. Стоит отметить, что оценка видового дерева основана только на ранее реконструированных геновых деревьях, которые рассматриваются как реальные наблюдения без учета стохастических ошибок. Следовательно, неинформативные или ошибочные реконструкции могут приводить к ошибочной топологии из-за неправильного моделирования [233]. ASTRAL III является одним из инструментов реконструкции филогенетического дерева видов из набора геновых деревьев [234]. Подход ASTRAL основан на идее, что дерево видов – это дерево, которое минимизирует расстояние до геновых деревьев. Расстояние измеряется как количество дискордантных расщеплений между геновыми деревьями и видовым деревом. Расщепление является дискордантным, если оно появляется в геномном дереве, но не появляется в видовом дереве, или наоборот. Примечательно, что на полученном филогенетическом дереве удастся оценить длины ветвей только между узлами, тогда как длины к “листьям” дерева остаются не определенными [234].

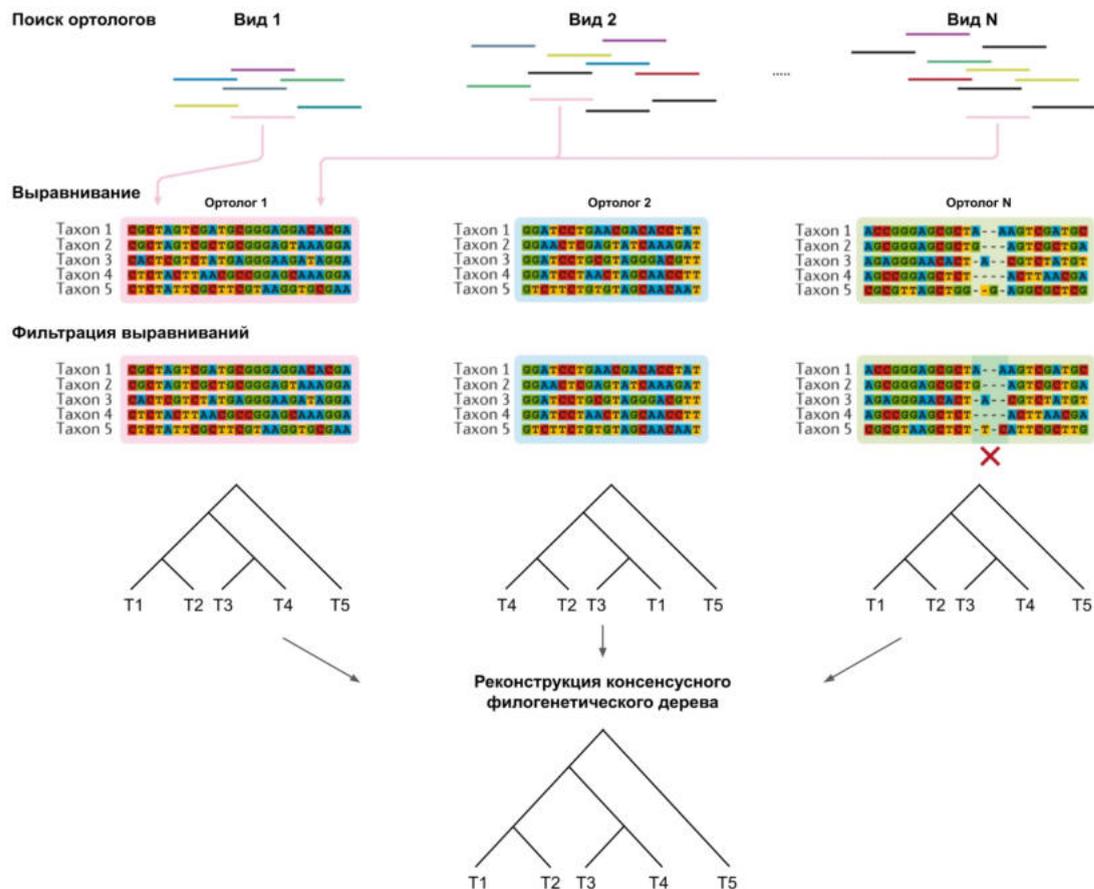


Рисунок 8. Схема реконструкции филогении с использованием обобщенно-коалесцентного подхода. Адаптировано из [181].

Точность метода конкатенации и обобщенно-коалесцентного подхода для реконструкции видовой филогении на основании полногеномных данных до сих пор остается предметом споров [235,236]. Оба метода основаны на предположениях, которые не полностью выполняются, что зачастую может приводить к противоречивым топологиям [237]. Такие несоответствия представляют собой серьезную проблему для получения надежных филогенетических результатов, для решения которой рекомендуется использовать оба подхода (как это было сделано в работе Fulton et al. 2007 [9]), принимая во внимание методологические допущения, лежащие в основе каждого из них, и их потенциальное влияние на полученные филогении [238–240].

1.4 Проблематика гибридизации и филогении рода *Martes*

1.4.1 Филогения рода *Martes* в сем. *Mustelidae*

Куньи (сем. *Mustelidae*) – самое крупное семейство из отряда хищных млекопитающих, по разным оценкам включающее 60-70 видов, относящихся к 22-24 родам [1]. Филогения куньих остается не до конца разрешенной преимущественно в связи с недостатком молекулярно-генетических данных и из-за характерной для представителей данного семейства быстрой эволюции фенотипического и экологического разнообразия (адаптивной радиации) [2,6]. К настоящему времени семейство включает в себя 8 подсемейств: *Taxidiinae* (единственный вид – американский барсук), *Mellivorinae* (единственный вид – медоед), *Melinae* (барсучьи), *Guloninae*, *Helictidinae* (хорьковые барсуки), *Galictinae/Ictonychinae*, *Lutrinae* (выдровые) и *Mustelinae* (хорьки, норки и ласки) [8]. Подсемейство *Guloninae*, к которому относится род *Martes*, также включает роды *Eira*, *Pekania* и *Gulo*. Наконец, род *Martes* на данный момент включает 8 видов, из которых 5 обитают в Евразии: соболь (*M. zibellina*), лесная куница (*M. martes*), каменная куница (*M. foina*), харза (*M. (Charronia) flavigula*) и нилгирская харза (*M. (Charronia) gwatkinsii*); 2 в Северной Америке: американская куница (*M.*

americana) и тихоокеанская куница (*Martes caurina*) и 1 вид в Японии – японский соболь (*M. melampus*) (Рисунок 9) [5].

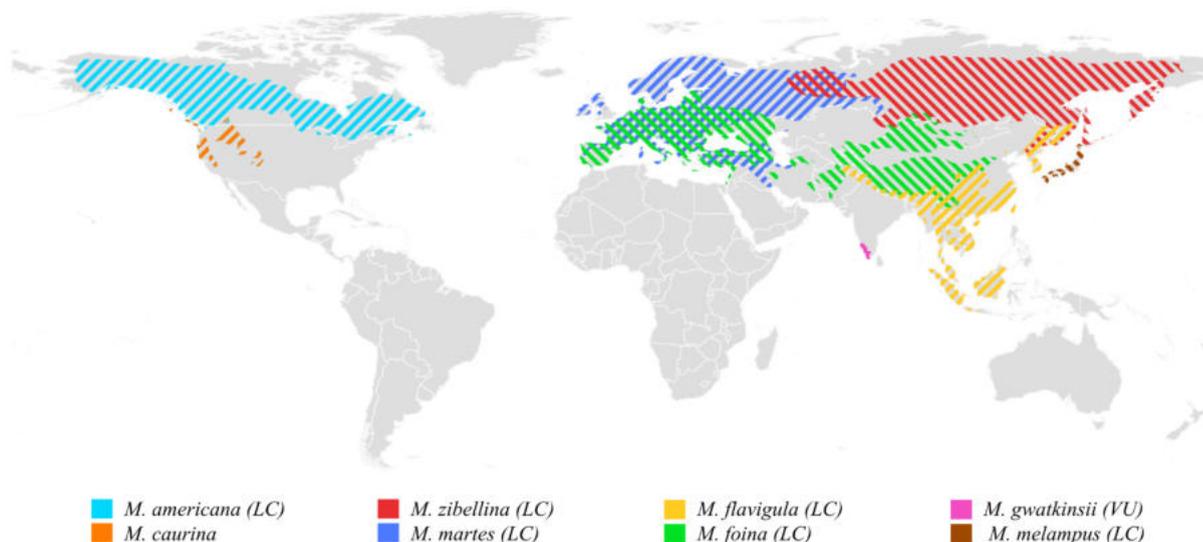


Рисунок 9. Ареалы и природоохранные глобальные статусы видов рода *Martes* (сем. Mustelidae) согласно данным МСОП.

Природоохранные статусы: LC (Least Concern) – «наименьшее беспокойство», VU (Vulnerable) – «Уязвимый» [241–247]. Ареал *M. caurina* отмечен согласно [248].

По причине значительного морфологического разнообразия особей таксономия рода *Martes* подвергалась многочисленным пересмотрам [4]. Долгое время представителей рода *Martes*, а также тайру и росомуху выделяли в отдельное подсемейство *Martinae* [2,6,9,249], которое зачастую ошибочно синонимизировали с подсемейством *Mustelinae* [10,250]. Впоследствии, в соответствии с Международным кодексом зоологической номенклатуры, было решено использовать в качестве наименования подсемейства первоначальное название этой группы видов – *Guloninae* [251]. Последним значительным изменением таксономии рода *Martes* стало выделение ильки (*Pekania penanti*), которая в течение многих лет относилась к роду *Martes*, в отдельный монотипический род *Pekania* [8,11].

Ранее род *Martes* состоял из трех подродов: собственно куниц (*Martes*), который включал соболя, каменную куницу, лесную куницу, американскую куницу, тихоокеанскую куницу и японского соболя; харз (*Charronia*) – харза и нилгирская харза; и *Pekania* – илька, при этом классификация была основана на

морфологических данных [252]. Позднее в работах, основанных на молекулярных данных, было показано, что илька (на тот момент *Martes pekkania*) более тесно связана с тайрой (*Eira barbara*) и росوماхой (*Gulo gulo*), что стало поводом для уточнения таксономии и выделения в отдельный род *Pekkania* [2,6,11]. Таким образом, на момент написания работы род *Martes* включает 2 подрода: собственно куниц (*Martes*) и харз (*Charronia*).

В многочисленных исследованиях филогении рода *Martes*, основанных как на морфологических, так и генетических данных, были продемонстрированы противоречивые систематические взаимосвязи между видами рода. На основании данных о морфологии было показано, что *M. americana*, *M. melampus*, *M. zibellina* и *M. martes* тесно связаны между собой (именно топологии этих четырех видов различаются) и, сильно в меньшей степени, связаны с *M. foina* [252], что было подтверждено в множестве последующих работ, основанных на молекулярных данных.

В 2000 году на основании анализа гена цитохрома b было установлено, что в роде *Martes* базальной группой являются *M. flavigula* и *Pekania pennanti*, на тот момент относящаяся к роду *Martes*. Позднее происходит ответвление *M. foina*, а оставшиеся виды *Martes* остаются тесно связаны между собой. Эти процессы видообразования отчасти объясняются географическим распространением между Евразией и Северной Америкой [20]. В работе Sato et al. 2004 [10] на основании двух генов яДНК с использованием подхода объединенной матрицы было показано, что род *Martes* слабо поддерживается как монофилетический (узел с *M. flavigula* имеет поддержку 57), тогда как остальные, участвующие в реконструкции филогении виды из подрода *Martes* (*M. martes*, *M. americana*, *M. zibellina*, *M. foina*), остаются монофилитичны. В работе Fulton et al. 2006 [9] на основании четырех ДНК-маркеров и одного ядерного экзона была выполнена реконструкция филогении как на основе объединенной матрицы, так и с использованием обобщенно-коалесцентного подхода, по результатам которых харза также была базальной не только по отношению к остальным видам *Martes*, но и к росوماхе. При этом топология видов подрода *Martes* (*M. martes*, *M. americana*, *M. zibellina*, *M. foina* и *M. melampus*) осталась неразрешенной, за исключением *M. martes* и *M.*

americana, которые образовали отдельную кладу. В работе Hosoda et al. 2011 на основании отдельных регионов мтДНК была получена топология, где *M. americana* и *M. melampus*, а также *M. martes* и *M. zibellina*, были сестринскими таксонами, при этом *M. foina* была базальной по отношению к ним. Примечательно, что *M. flavigula* практически сразу отделилась от остальных видов рода, а узел на дереве отличается невысокой поддержкой (42) [253].

В работе Li et al. 2014 [19] на основании полных мтДНК была представлена разрешенная с высокими значениями поддержек топология видов подрода *Martes*: *M. foina* (*M. americana* (*M. melampus* (*M. zibellina*, *M. martes*))). *M. flavigula* занимала базальное положение в роде и была слабо поддержана в зависимости от подхода реконструкции филогении. Стоит заметить, что полученная в данной работе топология как согласовывалась [20,254–256], так и противоречила [253,257,258] множеству ранее опубликованных работ, основанных на отдельных генах мтДНК. Хотя иногда это было вызвано отсутствием некоторых видов, как, например, в работе Yu et al. 2011 [6], топология в которой содержала *M. zibellina* и *M. melampus* в роли сестринских таксонов из-за отсутствия *M. martes*. Позднее, полученная Li et al. 2014 топология была подтверждена работой Hassanin et al 2021 [8], в которой было представлено филогенетическое дерево хищных, также основанное на полных мтДНК (Рисунок 10).

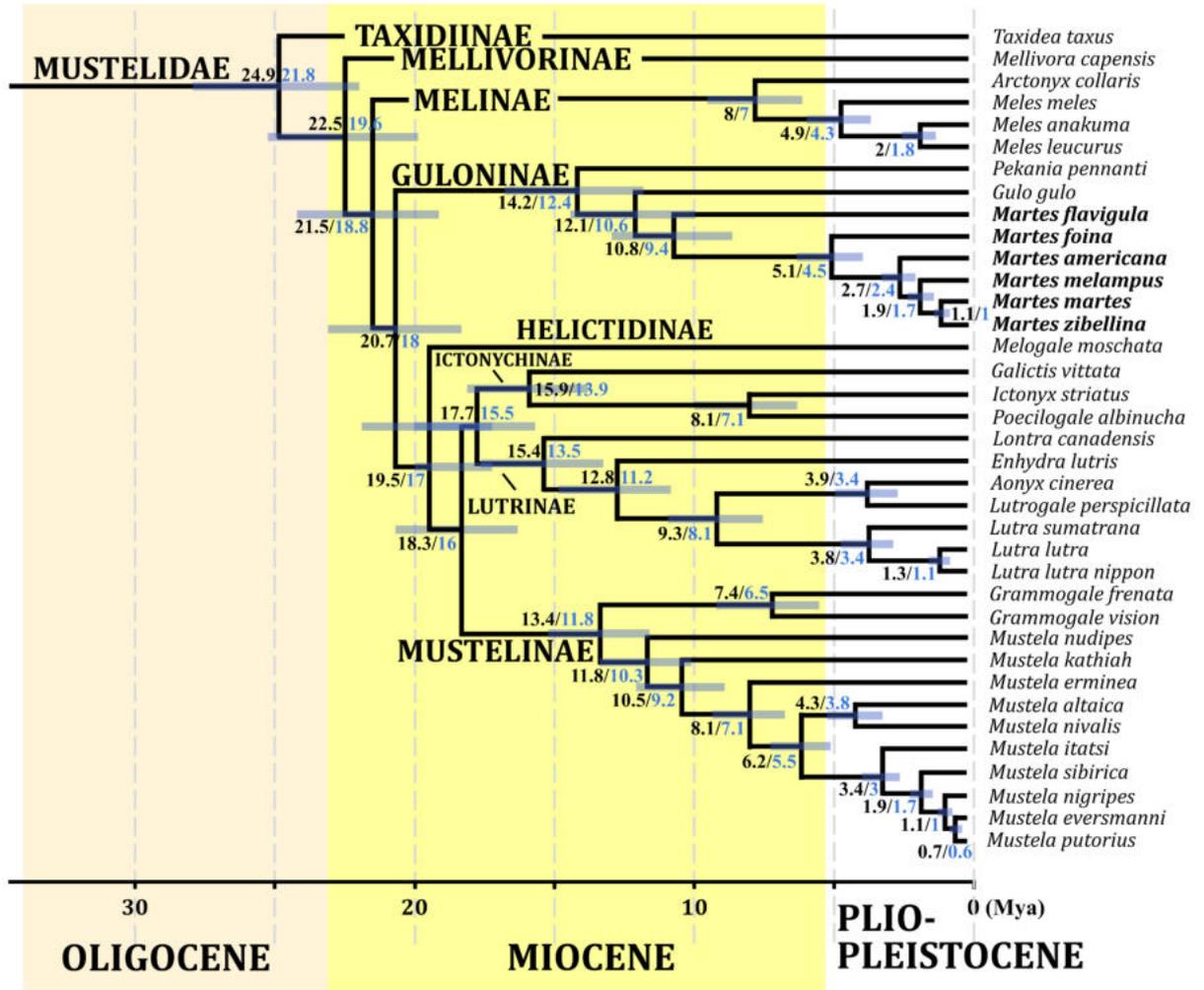


Рисунок 10. Филогения сем. Mustelidae (большинство видов) на основании полных мтДНК с оценками времени расхождения видов. Адаптировано из [8].

В работе Коерфли К.-Р. et al. 2008 [2] на основании 21 участков яДНК и гена цитохрома b на основе объединенной матрицы была практически полностью реконструирована филогения на уровне сем. Mustelidae. По результатам этой работы была подтверждена монофилия подрода *Martes*, при этом харза, будучи представителем подрода *Charronia*, являлась базальной по отношению к ним. При этом топология куниц подрода *Martes* менялась в зависимости от используемого для реконструкции филогении подхода. Так, на дереве, построенном методом максимальной экономии, *M. americana* была базальной по отношению к кладе с *M. melampus* и *M. martes* с *M. zibellina* (в отдельной кладе), а на дереве ML – *M. americana* и *M. melampus* образовывали одну кладу. На реконструированном

методом ВI дереве филогенетические отношения *M. americana*, *M. melampus*, *M. martes* и *M. zibellina* вовсе остались не разрешены. Позднее, в работе Law et al. 2018 [7], была представлена топология, основанная на 4 генах мтДНК и 42 генах яДНК. Это исследование на момент написания работы является в наибольшей степени полным с точки зрения выборки локусов, используемых для реконструкции филогении. На полученном филогенетическом дереве *M. americana* и *M. melampus*, а также *M. zibellina* и *M. martes* образовали сестринские клады (аналогично Hosoda et al. 2011), тогда как *M. foinea* оставалась базальной по отношению к ним. В свою очередь, *M. flavigula* занимает базальное положение в роде (Рисунок 11).

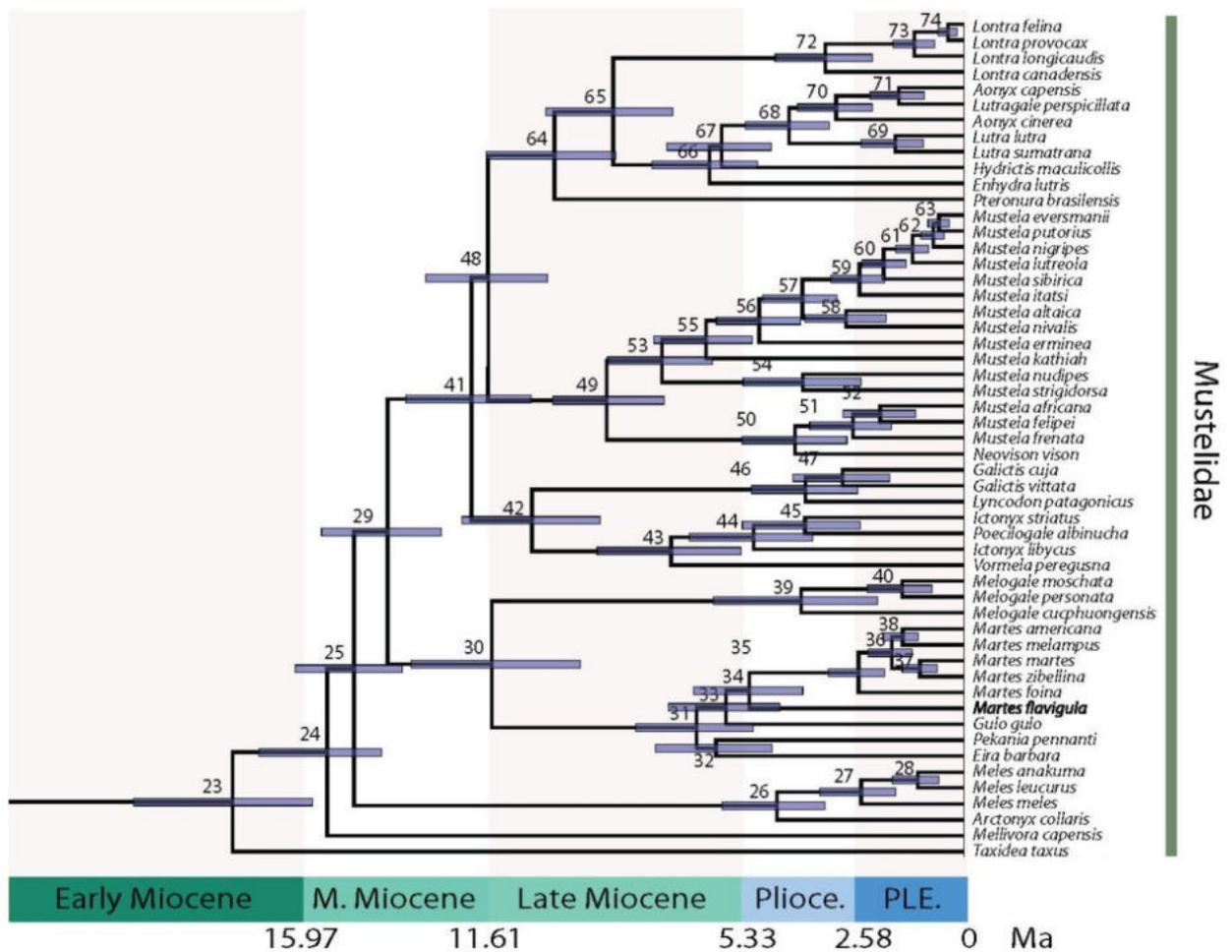


Рисунок 11. Филогения сем. Mustelidae (большинство видов) на основании 4 генов мтДНК и 42 генов яДНК с оценками времени расхождения видов. Адаптировано из [7].

Таким образом, современные представления о филогенетических взаимоотношениях видов рода *Martes* основываются либо на полных мтДНК, либо на комбинации генов мтДНК и отдельных генов яДНК. При этом полученные в разных работах топологии куниц подрода *Martes* различаются в зависимости от набора данных, а также наблюдается явное несоответствие топологий ядерной и митохондриальной ДНК. Ниже мы рассмотрим наиболее интересные особенности каждого из видов рода *Martes*.

1.4.2 Проблематика положения харзы (подрод *Charronia*) в роде *Martes*

Харза́ (*Martes flavigula*), также известная как желтогрудая (или желтогорлая) куница, широко распространена в хвойных и лиственных лесах тропической, субтропической и умеренной Восточной Азии и простирается до северо-восточной Палеарктики [259]. Ее ареал занимает территорию от Афганистана и Пакистана на западе, простирается вдоль Гималаев к югу Китая, занимает острова Ява, Суматра и Калимантан, а также простирается на север от восточного Китая (включая Тайвань) и Кореи до Дальнего Востока России [242]. В отличие от других представителей рода, харзы в основном перемещаются группами по 2-3 особи [260], чем обеспечивают защиту от других хищников и более широкий доступ к пище [261]. Характерной особенностью этого вида является выраженное пятно желтой шерсти на груди и горле [262], причем по последним данным существует до 6 различных типов окраски [263]. Стоит отметить, что диплоидное число хромосом у видов подрода *Charronia* $2n=40$, тогда как у всех остальных видов рода *Martes* – $2n=38$ [264–266]. Еще одним важным признаком родовой принадлежности харзы является строение бакулюма, форма которого (длинный с крючком и четыремя «пальцами-отростками» на конце) отличается от всех остальных представителей рода *Martes* [267,268]. Наконец, согласно В.В. Рожнову (1995) ключевым отличительным признаком харзы от других видов *Martes* является “наличие лишённого волос участка кожи позади плантарной мозоли” [17]. Несмотря на вышеперечисленные отличительные от остальных куниц особенности, положение харзы в роде *Martes* остается предметом дискуссий, что обусловлено прежде всего высокой географической изменчивостью особей.

Известно, что харза занимает базальное положение в таксономии рода *Martes* [7,8], будучи при этом в отдельном подроде *Charronia*. Самостоятельность рода *Charronia* (именно рода, а не подрода) и сохранение этого наименования, предложенного Gray в 1865 году [269], поддерживается рядом авторов [18,264]. Альтернативным синонимичным наименованием является *Lamprogale*, первоначально предложенное R.I. Росcock (1922), использованное С.И. Огнёвым (1928) и позднее В.В. Рожновым (1995) [17,270]. Проведенные исследования подчеркивают сложность таксономической структуры харзы, предполагая выделение в отдельный род на основании отличительных от остальных видов рода морфологических признаков, при этом упоминается оба наименования, как *Charronia*, так и (реже) *Lamprogale*, в значении как подрода, так и отдельного рода с разными видами и подвидами [17,18,267,270,271]. Важно отметить, что самостоятельность рода *Charronia* и, как следствие, выделение харзы из рода *Martes* оспаривается А.А. Лисовским и И.Я. Павлиновым [272]. И хотя в работе Kim et al. 2024, основанной на полных мтДНК, подтвердили наличие двух монофилетических подродов в роде *Martes*, филогенетическое положение харзы остается спорным [21].

В.В. Рожнов пишет о существовании на территории Вьетнама как минимум двух морфологически различных групп харз: северную и южную, обосновывая выделение форм юго-восточной окраины Азии в отдельный вид *L. lasiotis* (с подвидами *L. l. indochinensis*, *L. l. peninsularis* и *L. l. lasiotis*) [17]. Примеру В.В. Рожнова последовали, например, Абрамов и др. (2007) [14]. Матюшкин, со своей стороны, высказывал мнение о разделении материковой части ареала и о возможной видовой самостоятельности северной формы – амурской или уссурийской харзы *Ch. aterrima* [18]. В своем исследовании он обосновывал подразделение харзы *M. flavigula* на два вида рода *Charronia*: *Ch. aterrima* (с включением подвидов *Ch. f. aterrima*, *Ch. f. koreana*, *Ch. f. hainana*) и *Ch. flavigula* с включением подвидов южной части ареала, не отнесенных к *lasiotis* [18]. Такое заключение Е.Н. Матюшкин обосновывает распространением амурской харзы исключительно в северной части ареала (Приморье и Приамурье, Корейский полуостров и северо-восточный Китай) [18]. Примечательно, что популяции харзы

на территории России отличаются характерной белой окраской передних и реже задних лап – от нескольких пальцев до белой кисти, причем как на одной, так и на обоих лапах [16].

Тем не менее, внутривидовая систематика харзы остается противоречивой и запутанной, а большинство выделяемых зоологами подвидов (около 30) впоследствии сводятся в синонимы [15]. Признанными на момент написания работы считаются следующие подвиды харзы, как вида рода *Martes*, считаются: *M. f. flavigula* (восточные Гималаи, южная территория Китая), *M. f. aterrima* (северо-восточный Китай, п-ов Корея, Приамурье и Приморье), *M. f. chrisospila* (о-в Тайвань), *M. indochinensis* (п-ов Индостан), *M. f. peninsularis* (п-ов Малакка) и *f. M. saba* (о-в Калимантан) [267]. Примечательно, что статус харзы на Дальнем Востоке России рассматривается как угрожаемый, с сильно сокращающимися численностью и ареалом в последние годы [15].

Нилгирийская харза, *Martes gwatkinsii* – редкий и малоизученный вид рода *Martes*, эндемик горного массива на юго-западе Индии (горы Нилгири) [273,274]. Именно этот вид, относящийся на данный момент к подроду *Charronia*, также отличается неоднозначным положением в таксономии рода *Martes*. Отличительной особенностью нилгирийской харзы является окраска – в отличие от других харзы, особи этого вида имеют четкую границу темным туловищем и между лимонно-желтыми горлом, грудью и шеей [16]. Примечательно, что Согласно данным МСОП вид признан уязвимым в связи с предполагаемой низкой численностью, не превышающей 1000 половозрелых особей в дикой природе [243]. Обитает нилгирийская харза преимущественно в вечнозеленых лесах и мозаичных ландшафтах горных лесов и лугов, может встречаться на плантациях чая и кофе [243,275]. Встречи с этим видом регистрировались на высотах от 300 до 2600 м, однако наиболее часто его обнаруживают немного выше – на высотах около 800-2600 м [276]. Хотя в прошлом нилгирийскую харзу иногда рассматривали как подвид харзы, большинство авторов признают ее самостоятельным видом, поскольку ее окраска значительно отличается от *M. flavigula*, подчеркивая видовую самостоятельность [17,277]. Частота встреч *M. gwatkinsii* значительно ниже по сравнению с *M. flavigula*, что может свидетельствовать о низкой плотности её

популяции в природе. На юго-западе Индии нилгирийская харза рассматривается некоторыми исследователями как подвид *Martes flavigula* [278].

1.4.3 *M. foina* как потенциально базальный вид рода *Martes*

Каменная куница (*Martes foina*) – преимущественно ночной всеядный представитель рода, по размерам сопоставимый с лесной куницей. Это активные хищники, которые успешно приспособились к обитанию вблизи человека, причем антропогенное изменение ландшафта (вырубка лесов), на удивление, только ускорили экспансию ареала [241,279]. Это объясняется тем, что каменные куницы используют здания в качестве укрытий от холода. Как следствие, развитие городских и сельских районов не создает барьеров для популяций этого вида, а в совокупности с потеплением климата в последние годы, даже способствует расширению северной границы ареала. При этом крупные лесные массивы, наоборот, препятствуют потоку генов между популяциям, способствуя их генетической дифференциации [279]. Примечательно, что несмотря на способность каменной куницы сосуществовать рядом с человеком, она предпочитает лесные местообитания и избегает открытых участков, при этом её численность в лесах ограничивается в связи с конкуренцией с лесной куницей [280]. Важно отметить, что каменная куница практически не страдала от пушного промысла, в связи с более низким качеством меха в сравнении с лесной куницей и соболем [281].

Ареал распространения этого вида расположен в умеренных широтах Евразии и охватывает территорию от Пиренейского полуострова на западе до Центральной Азии и северных районов Мьянмы на юго-востоке [241]. Примечательно, что ареал заметно разделяется на Европейскую часть (Центральная и Южная Европа до Крыма и Кавказа) и фрагментированную Азиатскую часть (Ближний Восток, Малая и Центральная Азия) [282] (Рисунок 12). На данный момент на основании морфометрических и краниометрических данных для каменной куницы выделяют 11 подвидов: *M. f. foina*, *M. f. bosniaca*, *M. f. bunites*, *M. f. intermedia*, *M. f. kozlovi*, *M. f. mediterranea*, *M. f. nehringi*, *M. f. rosanowi*, *M. f. syriaca*, *M. f. toufoeus* [5]. Примечательно, что гибридизация каменной куницы и

лесной куницы не выявлена (либо о ней, по крайней мере, ранее не сообщалось), несмотря на значительное перекрывание их ареалов в Европе.

Особенности филогеографии каменной куницы тесно связаны с климатическими колебаниями позднего плейстоцена и голоцена. Дивергенция этого вида от других представителей рода *Martes* произошла около 2.5-3.5 миллионов лет назад, с последующим распространением особей этого вида из юго-западной Азии в леса Европы в раннем голоцене [2,7,283]. Во время последнего ледникового максимума (около 26–19 тыс лет назад) из-за суровых климатических условий каменная куница сохранилась в рефугиумах на Балканском полуострове и в южной Анатолии. С началом потепления, ослабление климатических барьеров позволило каменной кунице распространиться сначала в Западную Турцию и Европу, а затем – в Центральную Азию [283]. На основании анализа контрольного региона мтДНК было показано, что современная филогеография *M. foina* определяется двумя крупными гаплогруппами: евразийской, охватывающей ареал от Европы до Центральной Азии, и восточно-турецкой, ограниченной горными районами восточной Турции. При этом каменная куница, согласно анализу контрольного региона мтДНК, демонстрирует относительно высокое разнообразие линий мтДНК по всему ареалу, несмотря на исторические периоды изоляции [283]. На основании 22 STR-маркеров была показана слабая генетическая дифференциация между популяциями Польши, подтверждая слабую изолированность современных популяций из-за антропогенных факторов [279]. Исследования генетического разнообразия с использованием полногеномных данных на момент написания работы не проводились.

Каменная куница выделяется на фоне других видов рода *Martes* своей ценностью в контексте цитогенетических исследований. Кариотип каменной куницы $2n=38$, аналогично соболу и лесной кунице [34,265,284]. Согласно результатам проточной цитометрии, кариотип каменной куницы отличается хорошо разрешенными пиками, в следствие чего ZooFISH-зонды, созданные из хромосом этого вида, показывали чистый и сильный флуоресцентный сигнал в экспериментах по окраске хромосом. По этой причине каменная куница оказалась крайне ценным видом для изучения эволюции геномов хищных в контексте цитогенетики. Так, в

работе Nie et al. 2002 с использованием набора из 20 зондов (18 аутосом, X и Y) каменной куницы удалось целиком картировать хромосомные наборы шести видов хищных [34]. В общей сложности на момент написания работы с использованием зондов каменной куницы изучены кариотипы 18 видов из 9 семейств отряда хищных [33,34,63,285–289]. В ходе анализа геномных перестроек и филогении хищных было установлено, что хромосомный набор каменной куницы консервативен и представляет собой предковый кариотип *Carnivora* [63].

1.4.4 Соболь и последствия его реинтродукции

Соболь (*M. zibellina*) – ценный представитель рода *Martes*, преимущественно распространенный в таежной зоне Евразии и отличающийся обширным ареалом от западных склонов Уральского хребта до Дальневосточного побережья, включая полуостров Камчатка, Сахалин, Хоккайдо и Курильские острова [246]. Преобладающее большинство популяций этого вида расположены на территории России. Морфология этого всеядного представителя отряда хищных схожа с близкими представителями рода: он характеризуется меньшими размерами, в сравнении с харзой, и немного большими, в сравнении с лесной куницей и американской куницей [252,290]. Соболь демонстрирует значительную внутривидовую дифференциацию. Для соболя выделяют до 17 подвидов: *M. z. angarensis*, *M. z. arsenjevi*, *M. z. averini* (он же *altaica*, *jurgensoni*), *M. z. brachyura*, *M. z. ilimpiensis*, *M. z. jakutensis*, *M. z. kamtschadalis* (*kamtschatica*, *hamgyenensis*), *M. z. linkouensis*, *M. z. obscura*, *M. z. princeps* (*baicalensis*, *vitimensis*), *M. z. sahalinensis*, *M. z. sajanensis*, *M. z. schantaricus*, *M. z. tomensis*, *M. z. tungussensis*, *M. z. yeniseensis* и *M. z. zibellina* (*alba*, *asiatica*, *fusco-flavescens*, *maculata*, *ochracea*, *rupestris*, *sylvestris*) [5]. Разнообразие соболя определяется широкой географической изменчивостью окраски меха, диапазон которой варьирует от черной до светло-коричневой, при этом голова, как правило, отличается более светлым окрасом в сравнении с туловищем. Стоит также отметить наличие характерного белого, кремового или желто-оранжевого горлового пятна [291,292].

Соболь бесспорно занимает важное место в пушном промысле и звероводстве России [293]. Однако интенсивная его добыча, обусловленная

высоким спросом на высококачественный мех, стала одной из причин значительного сокращения природных популяций этого вида. Нерегулируемая охота, начавшаяся еще в средневековый период, в совокупности с распашкой земель, нашествием кедрового шелкопряда и масштабными лесными пожарами, привели к существенному сокращению численности Сибирских популяций соболя к концу XVIII – началу XIX веков [294,295]. К концу XIX – началу XX века антропогенное давление возросло по причине активной индустриализации Сибири, что привело не только к значительному сокращению мест обитания соболя, но и к разрозненности природных территорий, по причине активного использования их приграничных участков [295,296]. В результате соболь был практически полностью истреблен с территории Восточной и Западной Сибири, что стало причиной значительной фрагментации ареала с образованием в труднодоступных регионах малочисленных изолированных популяций, характеризующихся низким генетическим разнообразием [294,297]. Таким образом, фрагментация ареала усилила изолированность групп соболей и подорвала генетическое разнообразие вида. Позднее, особи из этих малочисленных групп расселились по территории Евразии, сформировав современные смешанные популяции, частично утратившие при этом бывшее генетическое разнообразие [293].

С начала XX века проводились многочисленные мероприятия, направленные на восстановление численности соболей [31,32]. Так, в 1935-1941 года был введен мораторий на охоту и была организована масштабная программа по реинтродукции [298]. В ходе многочисленных мероприятий по реинтродукции в период с 1901 по 1970 год в разные районы ареала было переселено по разным оценкам в общей сложности около 19-20 тыс зверьков [298,299]. Переселение осуществлялось преимущественно за счет баргузинских/витимских популяций (*M. z. princeps*) из Байкальского региона, соболя которых обладают наиболее ценным и высококачественным тёмным мехом [295,300]. Камчатская популяция (*M. z. kamtschadalica*) также использовалась в качестве донорской для восстановления практически исчезнувших соболей в районах Восточной Сибири (Магаданская область, часть Якутии и т.д.) [301]. Усилия по восстановлению природных

популяций в целом считаются успешными [300]. По данным на 2021 год Российская часть ареала насчитывает около 1,6 миллионов особей [302].

Согласно репрезентативному генетическому анализу, основанному на 15 микросателлитных локусов 474 особей соболя из 20 популяций Среднесибирского плоскогорья, который был проведен в работе Filimonov et al. (2024), было показано сохранение высокого уровня генетического разнообразия как у аборигенных, так и у реинтродуцированных соболей [295]. Это подтверждает, что запрет на промысел в середине XX века и последующая реинтродукция не привели к вытеснению местных подвидов, несмотря на их географическую близость. Было также показано, что аборигенные популяции восстановились естественным путем из изолированных очагов, а реинтродуцированные особи, происходящие преимущественно от витимской популяции, адаптировались к новым условиям благодаря высокой гетерозиготности донорской группы [303]. При этом полученная генетическая структура популяций соболя состояла из пяти генетических кластеров, тесно связанных с филогеографией, благодаря естественным барьерам (реки, болота), которые ограничили миграцию соболей и способствовали сохранению целостности исходных аборигенных популяций [295]. Примечательно, что современная популяция соболя в бассейне Ваха, сформированная интродуцированными особями, фенотипически отличается от близлежащих автохтонных популяций [304]. Это также подтверждается проявлением так называемого “эффекта основателя”, который проявился в связи с пространственной изоляцией первичных групп интродуцированных особей [305]. Вышеперечисленные данные согласуются с ранее проведенными популяционно-генетическими исследованиями на основании микросателлитных маркеров [29,304,306–308].

Несмотря на успех мероприятий по реинтродукции, современные популяции соболя демонстрируют и потенциальные риски для биоразнообразия, связанные с видоспецифической инвазивностью [299,309,310]. В последние годы наблюдается экспансия соболя в зону тундры, где он конкурирует с главным хищником этих местообитаний – песцом (*Alopex lagopus*), что вызывает опасения касательно дестабилизации тундровых сообществ [311]. Это подтверждается данными о более

успешном расселении соболя в северных направлениях, нежели в южных, что объясняется тем, что соболь – холодолюбивый вид, который плохо переносит повышенную температуру летом в сочетании с высокой влажностью [299]. Тем не менее, соболь также неоднократно был замечен в Свердловской области, южнее юго-западной границы ареала [312].

Известно, что инвазивные виды способны напрямую влиять на экологию местных видов, что впоследствии может приводить к изменениям их численности и ареала обитания [313]. Такое расширение ареала соболя, обусловленное высоким адаптивным потенциалом к различным местообитаниям, мобильностью и плодовитостью, а также с глобальными изменениями климата [294,314], имеет параллели с другими случаями видовой инвазии среди куньих. В частности, американская норка (*Neogale vison*), завезенная из Северной Америки для пушного звероводства в начале XX века, успешно распространилась в 28 европейских странах после случайных побегов и преднамеренных выпусков [315,316]. Этот вид также укоренился в некоторых регионах Южной Америки и Азии [317,318]. Подобно соболю, американская норка демонстрирует высокую адаптивность к новым условиям и оказывает давление на местные экосистемы. Таким образом, экспансия северной границы ареала соболя в последние годы вызывает опасения по поводу дестабилизации биоценоза тундровой экосистемы [311,319]. Примечательно, что наблюдаемое увеличение численности соболей, при чем как “чистых”, так и предполагаемых особей с интрогрессией от лесной куницы, также связывают с их участвовавшей гибридизацией [320].

Анализ внутривидовой филогенетической структуры соболя на основании полных мтДНК, проведенные Li et al. (2021), показал наличие трех основных гаплогрупп (А, В и С) [321], подтверждая результаты предыдущих исследований [322,323]. Предполагается, что три основные клады могли возникнуть в период теплого микулинского (казанцевского) межледниковья около 128-117 тыс лет назад, что совпало с периодом расширения таежных экосистем [324]. При этом некоторые гаплотипы одной из клад (А), согласно предположениям Li et al. (2021), достигли Северо-Восточной Азии в более поздний каргинский период плейстоцена около 50-25 тыс лет назад [325], включая популяции полуострова Камчатка и южных

Курильских островов, где сохранилась их изоляция [321]. Это подтверждает гипотезу о поэтапной колонизации периферийных регионов в ответ на климатические изменения. Кроме того, популяция соболя на острове Хоккайдо, мигрировала с континента через остров Сахалин [322,326], что объясняет частичную монофилию этой линии и ее генетическую дифференциацию. Было также подтверждено увеличение генетического разнообразия от границ к центральной части ареала [293,321]. При этом на основании мтДНК у соболя показано отсутствие филогеографической структуры в Евразии, за исключением частично монофилетических линий в горах Чанбай (*M. z. hamgyenensis*) и на п-ве Камчатка (*M. z. kamtschadalis*), где популяции демонстрируют признаки длительной изоляции [321]. Это стало подтверждением результатов более ранних исследований, согласно которым на северо-востоке Китая, Камчатке, а также о. Сахалин, о. Хоккайдо и южных Курил существуют обособленные линии соболиной мтДНК [293,322,323]. Предполагается, что общие предки островных и континентальных соболей существовали в позднем плейстоцене [322].

Генетическая структура соболей с Камчатки представляет особый интерес. Было показано, что Камчатская популяция отличается низким уровнем генетического разнообразия (нуклеотидного и гаплотипического), а также в значительной степени генетически дифференцирована от всех остальных континентальных популяций [293]. Примечательно, что филогенетически близкие соболя с о. Хоккайдо также отличаются пониженным генетическим разнообразием [327]. Исходя из этого по Камчатской популяции был сделан вывод, что она образована одной из небольших филогенетически обособленных предковых групп соболей, изолированных в различных рефугиумах в периоды плейстоценовых оледенений [293]. Позднее генетическая обособленность камчатской популяции, наряду с низкой дифференциацией центральносибирских популяций, была подтверждена с использованием 8 микросателлитных локусов [298]. Интересным предположением является то, что особи в этих группах, которые впоследствии расселились сформировали современные смешанные популяции, могли различаться по способности к гибридизации с лесной куницей [328]. В случае Камчатской популяции, перемещениям соболей через перешеек полуострова препятствует

тундровая зона, даже несмотря на их высокую миграционную активность, подтвержденную в том числе с использованием прямого мечения животных [329]. При этом, многочисленные мероприятия по реинтродукции на территории Камчатки не проводились, и, как следствие, не нарушили обособленность Камчатской популяции [293].

1.4.5 Гибридизация в роде *Martes*

В сравнении с другими млекопитающими, семейство куньих и, в частности, представители рода *Martes*, отличаются низкими значениями генетических дистанций несмотря на значительные географические расстояния между ареалами, что подчеркивает их высокую мобильность [330]. Явление межвидовой гибридизации широко распространено среди представителей семейства: хорьков, горностаев, норок, барсуков и куниц [29,248,331–333]. В случае рода *Martes*, гибридизация известна как среди евразийских (*M. zibellina* и *M. martes*), так и среди североамериканских куниц (*M. americana* и *M. caurina*).

1.4.5.1 Гибридизация соболя и лесной куницы

Одним из ближайших родственных видов к соболю является лесная куница, дивергенция с которым по разным оценкам произошла в среднем-позднем плейстоцене порядка 0,4-1,1 млн лет назад [2,8,255,334]. В сравнении с соболем, лесная куница в меньшей степени ценится как пушной зверь и отличается более крупными размерами, более длинным хвостом, меньшей опушенностью подошв, а также характеризуется выраженным светлым пятном, которое занимает часть шеи, грудь и передние лапы животного [281]. Ареал лесной куницы расположен преимущественно на территории Европы и дополнительно включает леса Скандинавии, Англию, Ирландию и северную часть Греции. Восточная часть ареала располагается вплоть до Западной Сибири, включая территорию Уральских гор [244]. У лесной куницы выделяют до 5 подвидов: *M. m. martes*, *M. m. sabaneevi*, *M. m. ruthena*, *M. m. uralensis* и *M. m. lorenzi* [22,281]. Восточная граница ареала лесной куницы расположена в Западной Сибири, где это вид обитает преимущественно в зоне тайги [335], однако с 50-х годов ареал лесной куницы

активно расширяется в юго-восточном направлении, захватывая лесостепную зону Западной Сибири и Урала [26,336].

Антропогенная фрагментация среды обитания (дороги, вырубка лесов, города) снижает миграционную активность и увеличивает генетическую дифференциацию популяций, со временем приводя к изоляции, инбридингу и потере генетического разнообразия [76]. Известно, что безлесные участки и автотрассы длиной более 5 км практически полностью блокируют миграцию, хотя иногда животные используют дренажные трубы для пересечения дорог [337,338]. В случае лесной куницы, ареал которой расположен преимущественно в Европе, отмечают значительную генетическую дифференциацию популяций [339]. Однако влияние антропогенных факторов на этот вид неоднозначно: в то время как одни исследования указывают на слабую связь между изменением среды под действие антропогенных факторов и распространением лесной куницы [340], другие демонстрируют серьезные последствия фрагментации из-за потери местообитаний, отлова и охоты [341,342]. Тем не менее, согласно последним данным из Испании, известно, что смертность во время перебежания дорог – основная причина гибели Европейских куниц, как лесной куницы, так и каменной куницы [343]. Стоит также сказать про довольно частое поражение электрическим током, зафиксированное в этом же исследовании среди каменных куниц (21,2%) [343]. Здесь хочется отметить, что, например, каменные куницы дважды (сперва в 2016 году, и другая особь повторно в 2017) стали причиной остановки Большого адронного коллайдера – зверьки влезали в трансформаторные будки и перегрызали электрические кабели, что и послужило их мгновенной гибели [344].

Восточная часть ареала лесной куницы и западная часть ареала соболя частично перекрываются, образуя зону симпатрии, которая с северо-запада на юго-восток охватывает несколько регионов России: от Республики Коми на северо-западе Урала через Ханты-Мансийский автономный округ, Тюменскую, Омскую, Томскую и Новосибирскую области [328]. В зоне симпатрии, посредством скрещивания и интрогрессивной гибридизации этих биологически и экологически близкородственных видов, происходит образование гибридов, так называемых кидасов (или кидусов), упоминания о которых берут начало в период с XVIII по

XIX века [23]. По данным на 2006 год, в некоторых районах Ханты-Мансийского автономного округа наблюдается высокая частота встречаемости особей, классифицируемых охотниками как “кидасы”, доля которых составляет более половины от всех пойманных особей [345]. Примечательно, что хотя соболя и кидасы чаще питаются полевками, лесная куница опережает их по потреблению белок и даже птиц – глухарей и рябчиков. Это указывает на ее преимущество в добыче корма в кронах деревьев, тогда как соболь явно уступает в подвижности на высоте [346]. Это подтверждает конкуренцию соболя и лесной куницы в зоне симпатрии.

Гибридизация как родительских видов, так и отдельных гибридных особей была экспериментально подтверждена, хотя результаты были противоречивыми. Плодовитое потомство удалось получить только от скрещивания от пары самец соболя × самка куницы в неволе, при этом потомства от реципрокного скрещивания (самец куницы × самка соболя) получено не было [24]. При этом утверждалось, что невозможно получить плодовитое потомство при скрещивании самки кидаса и самца соболя, самки кидаса и самца кидаса, а также самки соболя и самца кидаса [24]. При этом ранее Портновой (1941) удалось получить потомство от скрещивания самца лесной куницы и самки кидаса [25]. Как следствие, несмотря на выявленную стерильность самцов кидасов, возможность интрогрессивной гибридизации, согласно проведенным экспериментам, сохраняется благодаря самкам. Интересным предположением является то, что межвидовая гибридизация в зоне симпатрии соболя и лесной куницы служит барьером, препятствующим расширению ареалов этих видов [328]. Это связано с тем, что межвидовые гибриды, как правило, страдают из-за нарушений биохимических процессов, что может приводить к их стерильности или пониженной жизнеспособности [347].

Первый сравнительный анализ морфологии лесной куницы, соболя и их гибридов в естественных условиях обитания выполнил П.Б. Юргенсон в 1947 году [22]. Было показано, что одним из отличительных краниологических признаков соболя от лесной куницы является более узкая и плоская переносица, что было позднее подтверждено в множестве других работ [22,28,348,349]. В литературе также обсуждались такие признаки, как особенности зубов и морфология нижней

челюсти, расположение сонных ямок [23]. Для классификации гибридных особей ряд морфологических характеристик был описан в работе Б.Ю. Кассала и Г.Н. Сидорова 2013 года (Рисунок 12) [328]. Однако, как показала практика, достоверно выявить кидасов сложно, поскольку разные особи в различной степени демонстрируют характерные гибридам морфологические признаки. Павлинин В.Н. в 1963 году вовсе поставил под сомнение гибридное происхождение части ранее выявленных кидасов, аргументируя это значительным разнообразием морфологических признаков родительских видов [23]. Однако было показано, что в сравнении с обычным разнообразием лесных куниц, кидасы могут демонстрировать более или менее выраженные морфологические признаки, характерные для соболя, в различных комбинациях [328].



Рисунок 12. Внешние отличительные признаки гибридов соболя (*Martes zibellina*) и лесной куницы (*Martes martes*). Иллюстрации Бульонковой Т. М. Подписи адаптированы из [328].

Традиционно идентификация кидасов выполнялась по относительной длине хвоста и окраске меха [22]. Хвост у лесной куницы длиннее и выступает за концы вытянутых задних конечностей более чем на $\frac{1}{4}$ (число хвостовых позвонков по разным оценкам 15-22), тогда как хвост соболя короче и лишь слегка выходит за пределы задних лап (число хвостовых позвонков 15-16) [348,350]. В свою очередь, гибриды отличаются промежуточной длиной хвоста. Окраска меха гибридных особей в большинстве случаев сходна с лесной куницей: голова темнее спины (у соболя, напротив, светлее), тело равномерно окрашено, а горловое пятно либо отсутствует, либо представлено мелкими разрозненными пятнами [23]. В работе В. Г. Монахова 2013 года на основании морфометрических признаков было показано, что кидасы занимают промежуточное положение между лесными куницами и соболями, что соответствует морфологическим ожиданиям для гибридов первого поколения (F1). Однако, точность выявления гибридов составила лишь 62,5% [351]. Примечательно, что соболь в данной работе рассматривался как вид, потенциально угрожаемый из-за активной интрогрессивной гибридизации. Позднее в 2021 году В. Г. Монахов предложил новый краниометрический признак Δ (дельта) для классификации соболя и лесной куницы, определяемый как расстояние от заглазничного сужения до линии, проведенной между заглазничными отростками в сагиттальной плоскости, что позволило выполнять идентификацию родительских видов с точностью более 97% [352]. По его результатам, среднее значение Δ у соболей значимо меньше, чем у лесных куниц. Однако, оставалось не ясно, возможно ли выявление гибридов и оценка степени их интрогрессии на основании анализа краниометрических данных. В попытке ответить на этот вопрос использовались молекулярно-генетические подходы.

В работе Рожнова и др. 2010 года на основании контрольного региона мтДНК была подтверждена взаимная двунаправленная интрогрессия между соболем и лесной куницей в популяциях Северного Урала, причем было показано,

что видоспецифичные мтДНК не коррелируют с казалось бы однозначно установленным фенотипом [28]. Различия митотипов соболя и лесной куницы при этом не вызывают сомнений [28,328]. Примечательно, что подтвердилась в том числе интрогрессия мтДНК от соболя к лесной кунице, что считалось невозможным согласно экспериментам по скрещиванию в неволе [28]. Свидетельства интрогрессии мтДНК между видами были представлены и в других работах [349,353]. Это стало окончательным подтверждением плодовитости самок кидасов, а также успешных возвратных скрещиваний с обоими родительскими видами [349].

Позднее, классификация гибридных особей была выполнена в ряде исследований с использованием не более 11 микросателлитных маркеров, что позволило подтвердить постоянство гибридизации в зоне симпатрии [29,349,354]. Было показано, что фенотипически “чистые” соболь и лесная куница значительно отличаются по ряду STR-локусов, сохраняя генетическую обособленность, при этом фенотипические гибриды генетически были ближе к соболю, чем к лесной кунице, тем самым ускользя от внимания исследователей [29]. Важно отметить, что гибридные особи, как правило, обнаруживались только в зоне симпатрии и отсутствовали за ее пределами [354]. При этом стало очевидно, что на основании лишь одного STR маркера выявить гибридов невозможно из-за низкой разрешающей способности [320]. Таким образом, масштабы геномной интрогрессии с учетом возвратных скрещиваний остались не до конца ясными по причине недостаточной разрешающей способности отдельных STR-маркеров.

1.4.5.2 Гибридизация американской куницы и тихоокеанской куницы

Американская (*Martes americana*) и тихоокеанская (*Martes caurina*), куницы являются экономически ценными парапатрическими сестринскими видами рода *Martes*, обитающими в Северной Америке [355,356]. *Martes americana* обладает большей приспособленностью к различным климатическим условиям и распространена в бореальных лесах Канады, на Аляске, на Среднем Западе и северо-востоке США, включая острова вдоль северо-западного побережья Тихого океана, куда вид вторгся естественным образом, а также был намеренно

интродуцирован [254,357–359]. Примечательно, что популяции американской куницы характеризуются слабой генетической дифференциацией согласно исследованиям на основании микросателлитных маркеров, может объясняться высокими значениями эффективной численности популяций [330]. Подобная картина генетической структуры наблюдается, например, у россомахи (*Gulo gulo*) [360]. К этому также располагает способность *M. americana* мигрировать через фрагментированный ландшафт, вопреки ранним представлениям о строгой привязанности к специфическим местообитаниям [361]. Всё это подтверждает устойчивость американской куницы как вида к антропогенным изменениям среды, в отличие от многих других североамериканских хищников, например, волков (*Canis lupus*) и бурых медведей (*Ursus arctos*) [362,363]. В частности, это сильно контрастирует с ситуацией у близкородственной ильки (*Pekania pennanti*), которая ранее также относилась к роду *Martes*: ее популяции в США и Канаде показывают выраженную генетическую раздробленность и ограниченный поток генов, что связывают с антропогенным воздействием [364]. Важно сказать, что изменение климата может значительно повысить генетическую дифференцированность популяций и снизить поток генов в популяциях американской куницы, что может привести к снижению генетического разнообразия.

Ареал *M. caurina* характеризуется фрагментированными популяциями горных регионов на северо-западном побережье Тихого океана от юга Британской Колумбии до Калифорнии, а также в Скалистых горах вплоть до Мексики, при этом Великие равнины образуют барьер для распространения вида на восток [357,365]. Оба вида имеют ряд сходств по размеру тела, предпочитают леса с плотной кроной и густым подлеском, избегая открытых пространств, имеют сходные характеристики жизненного цикла и пищевые привычки [357,359,366–368].

При этом классификация американских куниц претерпевала значительные изменения по причине гибридизации особей. Первоначально два вида были описаны как отдельные – *M. americana* и *M. caurina*, однако позднее тихоокеанская куница была понижена до подвида *M. americana* на основании морфологических исследований [369,370]. Впоследствии североамериканские куницы долгое время рассматривались как один вид, однако позднее первоначальная двухвидовая

таксономическая классификация была подтверждена [12,248,355]. Так, наличие двух монофилетических клад куниц (*M. americana* и *M. caurina*) было подтверждено с использованием гена цитохрома *b* мтДНК, причем для *M. americana* отсутствует четкая популяционно-генетическая структура для исследуемой части ареала [254], что было также подтверждено с использованием микросателлитных маркеров [371]. При этом стоит отметить, что межпопуляционный поток генов для *M. americana* в северной части Скалистых гор слабо связан с географическим расстоянием и зависит от перепадов высот [372].

На данный момент для североамериканских куниц выявлены две зоны симпатрии – на северо-западе Северной Америки в Скалистых горах и на прибрежных островах (Куиу и Кургеапоф) вдоль тихоокеанского побережья [248,365,372]. В зоне симпатрии в Скалистых горах два вида куниц преимущественно разграничены за счет рельефа – *M. americana* встречается на более низких высотах (около 1110 м), тогда как *M. caurina* встречается на более высоких высотах (около 2000 м). Гибридные особи встречаются на разных высотах, но преимущественно на промежуточных [372]. Для североамериканских куниц проводились программы по реинтродукции с целью увеличения численности популяций и улучшения качества меха, в результате чего *M. americana* и *M. caurina* неоднократно непреднамеренно смешивались, что впоследствии усложнило управление популяциями и их генетическим разнообразием [358,373].

Так, в середине 1900-х годов переселение *M. americana* на прибрежные острова вдоль тихоокеанского побережья привели к непреднамеренной интрогрессивной гибридизации с *M. caurina* [373]. Стоит отметить, что природоохранный статус некоторых островных популяций *M. caurina* вызывают беспокойство в связи с вырубкой лесов [254,374]. Исторически сложившиеся практики выполнения мероприятий по интродукции видов зачастую приводили к непреднамеренному смешиванию американских и тихоокеанских куниц, что объясняет их неоднозначный успех [254,359,375]. Так, в 1980-х годах тихоокеанские куницы из Онтарио, Миннесоты и Колорадо были интродуцированы в Висконсин, где ранее обитали американские куницы. Однако генетических следов присутствия тихоокеанской куницы после переселения обнаружено не было, что

может свидетельствовать о том, что менее многочисленные особи могли быть вытеснены либо из-за генетического затопления, либо в результате стохастических событий, таких как высокая смертность в неподходящих климатических условиях среды [376]. Напротив, интродукция десяти американских куниц на остров Принца Уэльского с юго-восточной Аляски в 1930-х годах привело к значительному увеличению численности популяции, что, вероятно, было вызвано интрогрессивной гибридизацией с местными тихоокеанскими куницами [357].

1.4.5.3 Японский соболь

Наконец, стоит сказать про японского соболя (*Martes melampus*), которого также называют японской куницей (*Martes melampus*) — представителя рода *Martes*, эндемика Японских островов Хонсю, Сикоку, Кюсю и Цусима, за исключением острова Хоккайдо, на котором обитает соболь (*M. zibellina*) [377,378]. Вид предпочитает широколиственные леса, но также может встречаться в пригородных и более открытых участках, где достаточно деревьев для обеспечения среды обитания [379]. *M. melampus* ведет преимущественно плотоядный образ жизни, хотя значительную часть рациона составляют фрукты. В этой связи вид играет важную экологическую роль как один из ключевых распространителей семян крупных плодовых деревьев в субальпийских кустарниковых зарослях [380]. Филогеографический анализ на основании 10 микросателлитных локусов позволил выявить различия между двумя подвидами: *M. m. melampus*, распространённым на основных островах, и *M. m. tsuensis*, обитающим на Цусиме [378,381,382]. Различия в окрасе особей также подкрепляют это разделение: желтый фенотип встречается на Хонсю и Кюсю, тогда как темно-коричневый фенотип локализован на острове Цусима, Сикоку и части полуострова Кии [383]. Интересно, что в прошлом два фенотипа ошибочно считались разными подвидами: темно-коричневый классифицировался как *M. m. bedfordi*, а желтый – как *M. m. melampus*. Но в настоящее время оба отнесены к *M. m. melampus* [252,384]. При этом цусимская популяция *M. m. tsuensis* признана единственным отдельным подвидом и, согласно данным МСОП, считается уязвимой [385]. Таксономический статус цусимской популяции обусловлен морфологическими и митохондриально-генетическими

различиями [252,386,387]. Генетическая отличительность и однородность цусимской популяции являются следствием длительной изоляции [385]. Высота над уровнем моря, как показали исследования в районе Акая, не оказывает существенного влияния на выбор местообитания [388]. Природный ареал простирается от уровня моря до высот около 2 000 метров [378]. Японская куница интродуцирована на острова Садо и Хоккайдо, где была завезена с целью развития меховой промышленности [378]. Стоит отметить, что согласно анализу *M. melampus* с островов Хонсю, Сикоку и Кюсю, проведенному на основании гена цитохрома b, популяции этого вида демонстрируют значительно более высокую генетическую изменчивость по сравнению с *M. zibellina* с Хоккайдо [386].

Для сохранения этого вида особенно важно защищать широколиственные леса, которые обеспечивают необходимые условия среды [389]. *M. melampus* возник в результате географической изоляции на Японских островах от материковой части Азии. Ареалы *M. melampus* и близкородственного *M. zibellina* разделены проливом Цугару, расположенным между Хоккайдо и Хонсю [390]. Пролив Цугару служит барьером и для других родственных видов млекопитающих, например, бурого медведя (*Ursus arctos*) и азиатского черного медведя (*Selenarctos thibetanus*). Палеонтологические данные свидетельствуют о первом появлении двух видов *Martes* в позднем плейстоцене [252]. Андерсон предположил, что *M. melampus* произошел от *M. zibellina* в южной части японского архипелага, что подтверждается как морфологическими данными, так и на основании анализа последовательностей цитохрома b митохондриальной ДНК, что согласуется с межвидовыми различиями в других представителях семейства куньих [252,391,392].

Молекулярно-генетическая изменчивость японского соболя на полногеномном уровне остается неизученной из-за отсутствия полногеномных данных на момент написания работы, несмотря на высокую значимость для понимания эволюционных процессов в роде *Martes*. Опубликованные исследования основываются главным образом на анализе изменчивости гена цитохрома b и других генах мтДНК. При этом особый интерес представляет вопрос о возможной гибридизации между *M. melampus* и *M. zibellina*. Так, в 2004 году в ходе анализа контрольного и других регионов мтДНК было установлено, что гаплотипы обоих

видов группируются в два кластера, причем в каждом из них присутствуют последовательности как *M. zibellina*, так и *M. melampus* [393]. Но это позволило лишь предположить гибридизацию между этими видами. При этом первоначальная классификация соболей и японских куниц в данной работе происходила по длине и цвету хвоста: хвост японской куницы длиннее и отличается кончиком светлее, чем окраска меха туловища [393]. Наконец, в работе Inoue et al. (2010) молекулярные генетические особенности японской куницы и соболя рассматривались на основании гипервариабельного региона D-петли мтДНК и гибридизация между видами была опровергнута [394]. Однако, несмотря на это, вопрос гибридизации остается открытым.

Таким образом, изучение геномики куньих и, в частности, рода *Martes*, представляет интерес как для фундаментальной науки, так и для практического применения. Современные технологии секвенирования позволяют оценить масштабы гибридизации соболя и лесной куницы, исследовать генетическое разнообразие соболя для изучения последствий реинтродукции и разрешить вопрос филогенетического положения харзы в систематике рода *Martes*.

2 Материалы и методы

2.1 Данные полногеномного секвенирования

Клеточные культуры и образцы ДНК для полногеномного секвенирования были предоставлены УНУ “Криобанк культур клеток” ИМКБ, Новосибирск, Россия. В рамках работы были получены собственные данные полногеномного секвенирования (англ. Whole genome sequencing, WGS). Для *de novo* сборки референсных геномов четырех видов куниц (каменной куницы, лесной куницы, соболя и харзы) были получены как сцепленные риды 10X Genomics, так и данные Hi-C секвенирования, (Таблицы 1, 2 и 4). Кроме того были ресеквенированы 6 образцов лесной куницы (Таблица 1), 12 образцов соболя (Таблица 2), а также 12 образцов предполагаемых гибридов соболя и лесной куницы (Таблица 3). Секвенирование было выполнено на платформе Illumina Novaseq 6000 с использованием парных ридов длиной 150 п.н. Дополнительно в работе использованы опубликованные полногеномные прочтения тихоокеанских куниц (4 образца, SRA ID: SRR11575343, SRR11575346, SRR11575348, SRR11575351) и американских куниц (1 образец, SRA ID: SRR11575352) [248] и опубликованные данные секвенирования китайского подвида соболя (1 образец, SRA ID: PRJNA495455) (Таблица 5) [395]. Карта ареалов и мест сбора образцов соболя, лесной куницы и их предполагаемых гибридов представлена на Рисунке 13. Идентификаторы каждого образца в базе данных Национального центра биотехнологической информации США (англ. National Center for Biotechnology Information, NCBI) перечислены в Приложении 1.

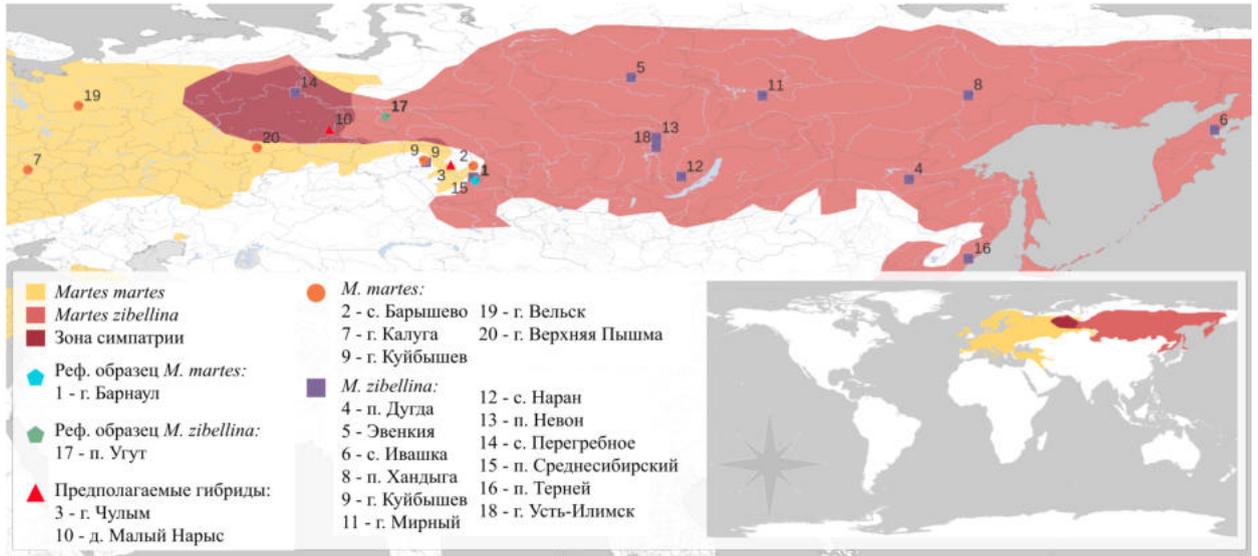


Рисунок 13. Ареалы и места сбора образцов соболя (*M. zibellina*) и лесной куницы (*M. martes*).

Ареал лесной куницы показан желтым цветом, ареал соболя – красным, а зона их симпатрии – темно-красным. Различные маркеры указывают локализацию образцов.

Таблица 1. Описание собственных полногеномных образцов лесной куницы (*M. martes*).

Тип данных*	ID	Происхождение особи	Пол**
10X, Hi-C	10xmmar	г. Барнаул, зоопарк «Лесная сказка»	F
PE	S44	Свердловская область, г. Верхняя Пышма	-
PE	S46	Свердловская область, г. Верхняя Пышма	-
PE	S49	Архангельская область, г. Вельск	-
PE	S50	Калужская область, г. Калуга	-
PE	T149	Новосибирская область, с. Барышево	M
PE	T151	Новосибирская область, г. Куйбышев	M

* Тип данных: 10X – сцепленные прочтения 10X Genomics; PE – данные Paired-End Illumina; Hi-C – данные Hi-C секвенирования

** Пол: F (Female) – самка; M (Male) – самец.

Таблица 2. Описание собственных полногеномных образцов соболя (*M. zibellina*).

Тип данных*	ID образца	Происхождение особи	Пол**
-------------	------------	---------------------	-------

10X, Hi-C	10xmzib	Ханты-Мансийский Автономный округ, п. Угут, Россия	F
PE	S26	Красноярский край, Эвенкия, Россия	F
PE	T8	Племенной звероводческий комплекс "Магистральный", Алтайский край, Россия	-
PE	T18	Ханты-Мансийский автономный округ, с. Перегребное, Россия	M
PE	T26	Якутия, Томпонский район, п. Хандыга, Россия	-
PE	T50	Иркутская область, Усть-Илимский район, п. Невон, Россия	-
PE	T72	Камчатский край, Карагинский район, с. Ивашка, Россия	M
PE	T90	Иркутская область, Качугский район, с. Наран, Россия	-
PE	T104	Приморский край, п. Терней, Россия	M
PE	T118	Амурская область, Зейский район, п. Дугда, Россия	-
PE	T148	Иркутская область, г. Усть-Илимск, Россия	-
PE	T150	Новосибирская область, г. Куйбышев, Россия	-
PE	T194	Якутия, Мирнинский район, Россия	F

* Тип данных: 10X – сцепленные прочтения 10X Genomics; PE – данные Paired-End Illumina; Hi-C – данные Hi-C секвенирования.

** Пол: F (Female) – самка; M (Male) – самец.

Таблица 3. Описание собственных полногеномных образцов предполагаемых гибридов соболя (*M. zibellina*) и лесной куницы (*M. martes*), классифицированных по фенотипическим признакам.

Тип данных*	ID образца	Происхождение особи	Пол**
PE	T24	Новосибирская область, Чулымский район, Россия	M
PE	T76	Тюменская область, Уватский район, д. Малый Нарыс, Россия	F
PE	T77	Тюменская область, Уватский район, д. Малый Нарыс, Россия	M
PE	T78	Тюменская область, Уватский район, д. Малый Нарыс, Россия	F
PE	T79	Тюменская область, Уватский район, д. Малый Нарыс, Россия	F
PE	T81	Тюменская область, Уватский район, д. Малый Нарыс, Россия	M
PE	T82	Тюменская область, Уватский район, д. Малый Нарыс, Россия	F
PE	T83	Тюменская область, Уватский район, д. Малый Нарыс, Россия	M
PE	T84	Тюменская область, Уватский район, д. Малый Нарыс, Россия	M
PE	T85	Тюменская область, Уватский район, д. Малый Нарыс, Россия	F
PE	T86	Тюменская область, Уватский район, д. Малый Нарыс, Россия	M
PE	T87	Тюменская область, Уватский район, д. Малый Нарыс, Россия	M

* Тип данных: 10X – сцепленные прочтения 10X Genomics; PE – данные Paired-End Illumina.

** Пол: F (Female) – самка; M (Male) – самец.

Таблица 4. Описание собственных полногеномных образцов харзы (*M. flavigula*) и каменной куницы (*M. foina*).

Тип данных	Вид	ID образца	Происхождение особи	Пол*
10X, Hi-C	<i>M. foina</i>	10xmfoi	провинция Ганьсу, Китай	М
10X, Hi-C	<i>M. flavigula</i>	10xmfla	Новосибирский зоопарк имени Р.А. Шило, Россия (предки с Дальнего Востока)	М

* Тип данных: 10X – сцепленные прочтения 10X Genomics; Hi-C – данные Hi-C секвенирования.

** Пол: F (Female) – самка; M (Male) – самец.

Таблица 5. Описание общедоступных полногеномных образцов соболя (*M. zibellina*), тихоокеанской куницы (*M. saurina*) и американской куницы (*M. americana*) [248,395].

Тип данных	Вид	ID образца	Происхождение особи	Пол*
PE	<i>M. zibellina</i>	china	провинция Хэйлуцзян, Китай	М
PE	<i>M. saurina</i>	SRR11575343	о. Грейам, США	М
PE	<i>M. saurina</i>	SRR11575346	штат Колорадо, США	М
PE	<i>M. saurina</i>	SRR11575348	о. Адмиралтейский, США	М
PE	<i>M. saurina</i>	SRR11575351	штат Вашингтон, США	F
PE	<i>M. americana</i>	SRR11575352	Центральная Аляска, США	М

* Тип данных: PE – данные Paired-End Illumina.

** Пол: F (Female) – самка; M (Male) – самец.

2.2 Морфологический анализ

Морфологический анализ 10 особей предполагаемых гибридов из зоны симпатрии соболя и лесной куницы (Т76 - Т79, Т81 - Т85, Т87) выполнен д.б.н. Алексеем Владимировичем Абрамовым (ЗИН РАН). Черепа отобранных особей сравнивались с образцами *M. martes* и *M. zibellina* из аллопатических популяций, хранящихся в коллекции Зоологического института Российской академии наук (Санкт-Петербург, Россия). Схема измерений описана в работе Monakhov et al. 2021 [352].

2.3 Оценка качества и фильтрация прочтений

Баркоды из сцепленных прочтений 10X Genomics были удалены при помощи EMerAld (EMA) v.0.6.2 [396]. Оценка качества исходных прочтений была выполнена при помощи FastQC v.0.11.9 [397] и KrATeR v.2.5 [398]. Базу распределения k-меров (длиной 23 нуклеотида) была сгенерирована с использованием JellyFish v.2.3.0 [243] и затем использована для оценки размера генома при помощи GenomeScope v.2.0 [400].

Удаление адаптеров и фильтрация данных по качеству была проведена в два этапа. Вначале, крупные фрагменты адаптеров были удалены с помощью Cookiecutter [401], затем оставшиеся мелкие фрагменты адаптеров и участки с низким качеством были обрезаны при помощи Trimmomatic v.0.36 [402], с параметрами “ILLUMINACLIP:TruSeq2-PE.fa:2:30:10:1 SLIDINGWINDOW:8:20 MINLEN:50”. В качестве последовательностей адаптеров была использована библиотека TruSeq2-PE, встроенная в Trimmomatic.

2.4 Получение и оценка качества геномных сборок

Первичные сборки геномов на основе сцепленных прочтений 10X Genomics были получены при помощи Supernova v.2 [54]. Hi-C данные были выровнены на первичные сборки, отфильтрованы и дедулицированы при помощи Juicer v.2019 [403]. Затем, конвейер 3D-DNA v.2019 был использован для получения черновых гаплоидных геномных сборок хромосомного уровня [404]. Наконец, финальные сборки были вручную скорректированы при помощи Juicebox Assembly Tools v.2019 [68].

Необходимо отметить, что в геномных сборках аутбредных видов с высоким уровнем гетерозиготности часто присутствуют ошибочно сохраненные дублицированные гаплотипы, которые в конечном итоге препятствуют корректному поиску генетических вариантов и аннотации генов. Выявление гаплотипических дубликатов было проведено при помощи purge_dups v.1.2.5 [405]. В ходе фильтрации геномных сборок были удалены все мелкие скаффолды, дублицированные последовательности которых были обнаружены в хромосомных

скаффолдах. В случае, когда обе последовательности находились в нехромосомных скаффолдах, удалялся наименьший по длине из них.

Оценки качества и целостности геномных сборок исследуемых видов были получены при помощи скрипта `quastcore.py` из пакета `Biocrutch` (<https://github.com/tomarovsky/Biocrutch>). Оценка качества геномных сборок по ортологичным группам была проведена при помощи `BUSCO v.5.4.2` [192] с использованием базы ортологов `OrthoDB v.10` [190] для млекопитающих `Mammalia_odb v.10` (2021-02-19). В качестве инструментов для поиска и аннотации ортологичных генов в ходе были использованы `Metaeuk v.6.a5d39d9` [406] и `HMMER v.3.3.2` [407]. Тандемные повторы в сборках геномов были выявлены и “мягко” маскированы (англ. `softmasked`) с использованием `TRF v.4.09.1` и `WindowMasker v.1.0.0` [408,409]. Для аналогичной процедуры в отношении диспергированных повторов был использован `RepeatMasker v.4.1.5` с базой повторов хищных `Dfam 3.7` [410,411].

2.5 Предсказание и аннотация белок-кодирующих генов

Модели генов были предсказаны при помощи конвейера `BRAKER v.3.0.8` [412], который включает набор программных инструментов, перечисленных ниже. В качестве исходных данных для конвейера были использованы все доступные на момент написания работы публичные данные РНК-секвенирования, полученные для видов рода *Martes* (Таблица 6) [395,413–415], базу данных `Mammalia_odb10` (2024-01-08) и белковые подсказки (англ. `protein hints`) из базы `Metazoa (OrthoDB v.11)`, сгенерированные при помощи конвейера `orthodb-clades` (<https://github.com/tomasbruna/orthodb-clades>) [416]. Программа `GeneMark-ETP v1.02` [417] была использована для обучения и предсказания генов на основе данных РНК-секвенирования и белков, а программа `AUGUSTUS v.3.5.0` [418,419] обеспечила дальнейшее предсказание генов, подкрепленное внешними данными. Мы присвоили названия генов предсказанным моделям при помощи `eggNOG-mapper v.2.1.12` и базы данных `EggNOG v5.0` (подмножество `Mammalia`) [420,421]. Линейные диаграммы Венна, отражающие перекрывание множеств

именованных генов, были визуализированы при помощи пакета *supervenn v.0.5.0* [422].

Таблица 6. Данные РНК-секвенирования, использованные для предсказания и аннотации генов

Вид	ID	Секвенатор	Происхождение особи	Ткань
<i>M. martes</i>	ERR11872609	Illumina NovaSeq 6000	Великобритания	печень
<i>M. zibellina</i>	SRR13013010	Illumina HiSeq 2500	Китай	жировая ткань
<i>M. zibellina</i>	SRR13013011	Illumina HiSeq 2500	Китай	жировая ткань
<i>M. zibellina</i>	SRR8074161	Illumina HiSeq 2500	Китай, пров. Хэйлунцзян, Большой Хинган	селезёнка
<i>M. zibellina</i>	SRR8074163	Illumina HiSeq 2500	Китай, пров. Хэйлунцзян, Большой Хинган	легкое
<i>M. zibellina</i>	SRR8074165	Illumina HiSeq 2500	Китай, пров. Хэйлунцзян, Большой Хинган	мышца
<i>M. zibellina</i>	SRR8074168	Illumina HiSeq 2500	Китай, пров. Хэйлунцзян, Большой Хинган	почка
<i>M. zibellina</i>	SRR8074169	Illumina HiSeq 2500	Китай, пров. Хэйлунцзян, Большой Хинган	сердце
<i>M. zibellina</i>	SRR31089886	Illumina HiSeq X	Россия, г. Барнаул, звероводческий комплекс "Магистральный"	печень
<i>M. zibellina</i>	SRR31089885	Illumina HiSeq X	Россия, г. Барнаул, звероводческий комплекс "Магистральный"	полушария
<i>M. flavigula</i>	SRR21460068	Illumina NovaSeq 6000	Китай, пров. Сычуань	кишечник
<i>M. flavigula</i>	SRR21460069	Illumina NovaSeq 6000	Китай, пров. Сычуань	селезенка
<i>M. flavigula</i>	SRR21460070	Illumina NovaSeq 6000	Китай, пров. Сычуань	сердце
<i>M. flavigula</i>	SRR21460071	Illumina NovaSeq 6000	Китай, пров. Сычуань	поджелудочная железа
<i>M. flavigula</i>	SRR21460072	Illumina NovaSeq 6000	Китай, пров. Сычуань	почка
<i>M. flavigula</i>	SRR21460074	Illumina NovaSeq 6000	Китай, пров. Сычуань	яичко

2.6 Полногеномные множественные выравнивания и анализ синтении

Сборки геномов хромосомного уровня для домашней собаки (*Canis lupus familiaris*, CanFam3.1), домашней кошки (*Felis catus*, *Felis_catus_9.0*) и человека (*Homo sapiens*, GRCh38.p13) были загружены из базы данных NCBI Genome. Маскирование повторов было выполнено по схеме, описанной выше в разделе "Получение и оценка качества геномных сборок". Последующее множественное полногеномное выравнивание геномов было проведено при помощи *Progressive Cactus v.2.8.0* [423]. Синтенные блоки были извлечены из полученного выравнивания при помощи *halSynteny v.2.2* [424] с опциями: "--minBlockSize 50000 --maxAnchorDistance 50000". Последующая фильтрация была выполнена, как

описано в работе Romanenko et al., 2023 [425]. Фильтрация и визуализация блоков синтении была проведена скриптом `draw_synteny.py` из пакета MACE. Гены, расположенные в пределах инверсий, были выявлены при помощи Bedtools v.2.31.1 [426].

2.7 Выравнивание и поиск координат псевдоаутосомного района

Случайная подвыборка прочтений (даунсемплинг) с целью получения примерно одинакового покрытия во всех образцах была выполнена при помощи инструментария BBmap v.38.96 [427]. Выравнивание прочтений на геномные сборки было проведено при помощи BWA v.0.7.11 [428]. Сортировка выравниваний, фильтрация по качеству и маркировка дубликатов были выполнены при помощи программного пакета Samtools v1.15.1 [429]. Подсчет покрытия генома был осуществлен при помощи Mosdepth v.0.3.0 с отсечкой по качеству выравнивания равной 10 (`--mapq 10`) [430].

Для поиска координат псевдоаутосомного района (англ. pseudoautosomal region, PAR) мы реализовали алгоритм, основанный на вычислении медианного, среднего, максимального и минимального значений покрытия для окон размером 10 тыс п.н. в пределах половой хромосомы. На первом этапе в качестве кандидатов из полученного набора окон отбираются только те, чье медианное покрытие в которых составляет не менее чем 70% от значения медианного значения покрытия всего генома. Затем, первичный список участков формируется из не менее чем 10 подряд идущих окон, подходящих по заданному условию. На третьем и финальном этапе, алгоритм объединяет разрозненные участки в единые области, если медианное значение покрытия между ними составляет не менее 70% от медианного значения покрытия для всего генома. Вышеописанный алгоритм был реализован в виде скриптов, доступных в составе набора программных инструментов Biocrutch (<https://github.com/tomarovsky/Biocrutch>). Визуализацию оценок покрытия всего генома в виде тепловых карт была выполнена при помощи набора программных инструментов MACE (<https://github.com/mahajrod/MACE>).

2.8 Поиск, фильтрация и маскирование генетических вариантов

Поиск генетических вариантов в образцах был проведен при помощи Vcftools v1.15.1 [101] со следующими параметрами: "-d 250 -q 30 -Q 30 --adjust-MQ 50 -a AD,INFO/AD,ADF,INFO/ADF,ADR,INFO/ADR,DP,SP,SCR,INFO/SCR" для Vcftools mpileup и "--group-samples -m -O и -v -f GQ,GP" для Vcftools call. Генетические варианты низкого качества были удалены при помощи Vcftools filter с опциями ('QUAL < 20.0 || (FORMAT/SP > 60.0 || FORMAT/DP < 5.0 || FORMAT/GQ < 20.0)'). Для каждого образца был сформирован трек маскирования на основе данных покрытия, который включал участки со слишком высоким (>2.5 от медианного покрытия всего генома) или слишком низким (<0.33 от медианного покрытия всего генома) покрытием. Затем, индивидуальные треки были попарно пересечены, а полученные пересечения объединены в единый финальный трек маскирования, который и был использован для удаления недостоверных вариантов при помощи Bedtools v.2.31.1 [426]. Описанная процедура маскирования удаляет только те варианты, которые находятся в участках имеющих слишком низкое или слишком высокое покрытие не менее чем в двух образцах, что позволяет избежать избыточного маскирования и потери информации. Фильтрованные и маскированные генетические варианты были разделены на однонуклеотидные полиморфизмы (SNP) и инсерции/делеции (indel). Все последующие анализы были основаны исключительно на SNP. Количество гомозиготных и гетерозиготных SNP было подсчитано в скользящих окнах размером 1 млн п.н. с шагом в 100 тыс п.н. Распределения гетерозиготных SNP для ряда других кунных: морской выдры (*Enhydra lutris*), речной выдры (*Lutra lutra*), восточной бескоготной выдры (*Aonyx cinereus*) и тайры (*Eira barbara*), были взяты из работ [66,115]. Визуализацию распределений плотности генетических вариантов в виде тепловых карт и скрипичных диаграмм была получена при помощи набора программных инструментов MACE.

2.9 Анализ распределений гетерозиготности

Для анализа гетерозиготности мы рассчитали количество гетерозиготных SNP (независимо для каждого образца) в скользящих окнах размером 1 млн п.н. с

шагом в 100 тыс п.н, при этом X-хромосомы были исключены для всех образцов. Для получения эмпирических распределений данные были разделены на бины с шагом в 0,1 SNP/тыс п.н. Первые два бина (соответствующие RoH) были удалены для упрощения теоретической модели.

Мы аппроксимировали полученные эмпирические данные при помощи смешанного распределения, а именно линейной комбинации отрицательных биномиальных распределений, для чего был использован пакет `mixdist v.0.5-5` [431]. При использовании отрицательных биномиальных распределений в качестве компонент смешанного распределения, `mixdist` требует задать только стартовые значения для среднего значения (μ) и стандартного отклонения (σ) каждой компоненты. Выбор именно отрицательного биномиального распределения обусловлен несколькими следующими причинами. Во-первых, в первом приближении присутствие гетерозиготного сайта в окне можно рассмотреть как редкое событие (частота гетерозиготных сайтов в большинстве окон составляла менее 2 на 1000 п.н.) происходящее с некоторой средней частотой (средней гетерозиготностью по геному), распределение которых должно соответствовать распределению Пуассона. Однако, распределение Пуассона требует независимости событий, что строго говоря не выполняется в данном случае, например из-за сцепления некоторых SNP. Для распределения Пуассона характерно $\mu = D \approx \sigma^2$, однако в наших данных дисперсия значительно превышала среднее значение (Таблица 8). Хорошо известным и опробованным решением данной проблемы является замена распределения Пуассона на отрицательное биномиальное распределение [432], что позволяет учесть повышенную вариабельность за счет снятия требования на совпадение дисперсии и среднего значения, что в свою очередь вводит дополнительный параметра. Подобный подход был использован, например, при разработке пакета `GenomeScore` [400] для аппроксимации распределения покрытия k-меров при оценке размера генома где авторы также столкнулись с проблемой повышенной дисперсии.

Мы протестировали шесть различных моделей (Таблица 7), начиная с полной трехкомпонентной, включающей компоненты соболя (S), лесной куницы (P) и гибридного происхождения (H), и заканчивая наиболее простыми

однокомпонентными моделями “чистых” видов. Стартовые значения среднего значения и стандартного отклонения каждой компоненты были выбраны согласно модам эмпирических распределений и ширине соответствующих пиков (Таблица 8). Для каждого образца наиболее оптимальная модель была выбрана независимо в два этапа. В первую очередь мы отфильтровали модели по конечному среднему значению (μ); если значение μ одной из компонент выходило за пределы диапазона начального $\mu \pm$ стандартное отклонение (σ) – модель отбрасывалась. Этот этап потребовался из-за невозможности наложить ограничения на конечные значения во время оптимизации аппроксимации. Наконец, наилучшая модель была выбрана для каждого образца по наименьшему значению квадратичного отклонения, которое рассчитывалось как сумма квадратичных разностей между эмпирическим и аппроксимированным распределением (QDE).

Таблица 7. Модели, использованные для анализа распределений гетерозиготности, и их компоненты.

Модель	Формула *	Компоненты		
		Лесная куница (P)	Соболь (S)	Гибрид (H)
Полная модель	$a_p * Nb_p + a_s * Nb_s + a_h * Nb_h$	+	+	+
Модель без гибридной гетерозиготности	$a_p * Nb_p + a_s * Nb_s$	+	+	–
“Чистые” лесные куницы	$a_p * Nb_p$	+	–	–
“Чистые” соболя	$a_s * Nb_s$	–	+	–
Лесные куницы с гетерозиготной интрогрессией	$a_p * Nb_p + a_h * Nb_h$	+	–	+
Соболь с гетерозиготной интрогрессией	$a_s * Nb_s + a_h * Nb_h$	–	+	+

* Коэффициенты:

- a – вес отдельной компоненты (коэффициент в линейной комбинации).
- Nb – отрицательное биномиальное распределение со средним значением (μ) и стандартным отклонением (σ).
- $p \ s \ h$ – идентификаторы отдельных компонент: лесные куницы, соболя и гибриды, соответственно.

Таблица 8. Начальные значения среднего значения (μ) и стандартного отклонения (σ) для каждой компоненты.

Компонента	μ	σ
Лесные куницы (P)	700	200
Соболя (S)	1700	500
Гибриды (H)	4500	700

Для каждого образца мы определили средние значения компонент наиболее оптимальной модели. Затем мы объединили полученные значения и кластеризовали их. Для этого мы использовали метод иерархической агломеративной кластеризации при помощи функции `linkage()` из пакета `Scipy v1.13.1` с методом «average». Кластеры были сформированы при помощи функции `fcluster()` `Scipy` на основе критерия *maxclust* с заданным числом кластеров равным 3 (компоненты соболя, лесной куницы и гибридов). Описанная процедура была проведена в качестве дополнительной проверки и подтвердила точность аппроксимации, поскольку отклонений от ожидаемой кластеризации (каждая компонента в своем кластере без перекрытия) не было выявлено. Значения медиан для каждого из кластеров были в дальнейшем использованы как характерные значения соответствующих компонент. Наконец, распределения гетерозиготных SNP относительно обоих референсов с обозначением полученных медиан были визуализированы в виде псевдо “скрипичных” гистограмм с помощью набора программных инструментов `MACE`.

2.10 Обнаружение интрогрессии, анализ происхождения и классификация образцов

Анализ происхождения (`ADMIXTURE`-анализ) был проведен на основе маскированных SNP аутосом и псевдоаутосомного района. Полученный набор SNP был дополнительно отфильтрован с помощью `PLINK v.1.9` [156] по минимальной частоте минорного аллеля ("`--maf 0,03`"), с сохранением вариантов только со 100% частотой генотипирования для всех образцов ("`--geno 0`"). Фильтрация сцепленных SNP с высоким уровнем неравновесия по сцеплению (англ. *Linkage disequilibrium*, *LD*) была выполнена в скользящих окнах по 50 SNP с шагом в 10 SNP, с порогом r^2 равный 0,7 ("`--indep-pairwise 50 10 0,7`"). Анализ главных компонент (`PCA`) на

основе отфильтрованного набора SNP был также осуществлен с помощью PLINK v.1.9 ("--rca"). Для дополнительного подтверждения гибридизации и исключения сценария изоляции по расстоянию мы построили треугольные диаграммы зависимости межгрупповой гетерозиготности (определяется как отношение средней гетерозиготности между группами к их общей гетерозиготности, термин введен Colella et al, 2024) и гибридного индекса с использованием *triangulaR* v.0.0.1 [168].

Оценка уровня генетического разнообразия и определение степени генетического смешения образцов из разных популяций были выполнены при помощи ADMIXTURE v.1.3.0 [161] для 1-6 предполагаемых генетических кластеров исходных популяций (K). Анализ был проведен трижды с вычислением ошибок перекрестной валидации (англ. Cross-validation errors, CV errors) и значений среднего попарного сходства. Локальный ADMIXTURE анализ был выполнен путем отдельных запусков ADMIXTURE (K=2) с использованием скользящих перекрывающихся окон размером 1 млн п.н. с шагом в 100 тыс п.н.

Для статистического подтверждения гибридизации и доказательства наличия потока генов между идентифицированными группами образцов мы провели повторный поиск генетических вариантов с добавлением образца каменной куницы (*Martes foina*) в качестве аутгруппы. Маскирование и фильтрацию SNP были осуществлены аналогично предыдущим этапам. Оценка степени гибридизации была проведена с помощью инструмента HyDe v.0.4.3 [433]. Вычисление D-, F4- и F3-статистик было выполнено при помощи AdmixTools v.7.0.2 [134], с использованием *M. foina* в качестве аутгруппы. Протокол был реализован в виде программного конвейера AdmixPipe (<https://github.com/tomarovsky/AdmixPipe>).

2.11 Анализ микросателлитных локусов

Для сравнительного анализа результатов классификации по видовой принадлежности, полученных на основании полногеномных данных, мы выполнили кластеризацию исследуемых особей соболя, лесной куницы и их гибридов с использованием 3-х наборов STR-маркеров:

1. Набор из исследования Рожного и др. 2013 [29]. Оригинальный набор включает 9 STR-маркеров: Mel10, Ma-1, Ma-3, Ma-8, Ma-15, Ma-18, Ma-19, Mvis072 и Mer041. Семь из них, за исключением Ma-3 и Mer041, прошли описанную ниже фильтрацию и использованы в дальнейшем анализе.
2. Набор из исследования Каштанова и др. 2022 [311]. Оригинальный набор включает 16 STR-маркеров:
 - a. Восемь STR-маркеров из предыдущего исследования: Ma-1, Ma-3, Ma-8, Ma-15, Ma-18, Ma-19, Mvis072, Mer041. Каждый из них прошел описанную ниже фильтрацию и использован в анализе.
 - b. Восемь дополнительных STR-маркеров: Mf3.7, Mar08, Mar21, Mar36, Mar43, Mar58 и Mar64. Все из них, за исключением Mar53, прошли фильтрацию и использованы в работе.
3. Полный набор, включающий как прошедшие фильтрацию STR-локусы как из двух вышеперечисленных наборов, так и STR-маркеры, полученные из других публикаций по генетике куньих, а именно: лесной куницы (*Martes martes*), каменной куницы (*Martes foina*), американской куницы (*Martes americana*), россомахи (*Gulo gulo*), американского барсука (*Taxidea Taxus*), европейского барсука (*Meles meles*), американской норки (*Neovison vison*) и горностая (*Mustela erminea*) [434–439].

Для праймеров, соответствующих отобранному локусам, мы провели *in silico* ПЦР при помощи Simulate_PCR v1.2 [285]. Было разрешено не более 4 различий между целевой последовательностью и каждым праймером, а длина ампликона была ограничена интервалом 50-1000 п.н. Данный этап работы был выполнен Дмитрием Юрьевичем Прокоповым. Дальнейшую фильтрацию ампликонов проводили, как описано ранее в [66]. Затем отобранные STR-локусы были вручную проверены на присутствие/отсутствие повторяющегося мотива, длину повтора и количества повторов. Локусы с STR длиной более 100 п.н. (ограничение связанное с длиной ряда 150 п.н.), с 2 и более STR и без STR, а также STR-локусы, сцепленные с X хромосомой, были исключены из последующего анализа.

Для ускорения дальнейших расчетов, для каждого образца мы отобрали из ранее полученных выравниваний прочтения, выровненные на финальный набор

STR-локусов, включая прилегающие регионы длиной 1 тыс п.н., при помощи Samtools v.1.19.2 [101]. Проверка парности выделенных прочтений была выполнена и при помощи Bazam v1.0.1 [441]. Повторное выравнивание было выполнено при помощи BWA v.0.7.17 с учетом инделов с использованием IndelRealigner из GATK v.4.4.0.0 [102,428]. Генотипирование STR-локусов было проведено при помощи hipSTR v0.6.2 [442]. Анализ происхождения на основе STR был выполнен при помощи STRUCTURE v.2.3.4 с параметром K от 2 до 6 в трехкратной повторности [160]. Последующее выравнивание кластеризации между разными запусками было реализовано при помощи Clumpp v.1.1.2 [443]. Визуализация результатов была проведена в при помощи программного пакета MACE.

2.12 Выявление протяженных участков гомозиготности (RoH)

Поиск RoH мы провели на полученном ранее подсчете гетерозиготных SNP в скользящих окнах размером 100 тыс п.н. с шагом в 10% от размера окна. Алгоритм основан на фильтрации окон по значению уровня гетерозиготности: окно сохранялось, если значение уровня гетерозиготности не превышало 0,05. В случае, если значение находилось в пределах между 0,05 и 0,1, окна сохранялись только в том случае, если не более 5 из них были расположены подряд. Окна, значение гетерозиготности в которых превышало 0,1, отбрасывались. В последствии, полученные окна с низким уровнем гетерозиготности объединялись в единые протяженные участки, если расстояние между ними (концом предыдущего окна и началом следующего) было меньше размера половины окна. Координаты RoH для каждого образца были определены на всех хромосомах, за исключением X хромосомы самцов, поскольку в этом случае поиск генетических вариантов в диплоидном режиме происходил только в области PAR. Наконец, мы классифицировали RoH по трем группам: короткие RoH (англ. Short RoH, S) – < 1 млн п.н., длинные RoH (англ. Long RoH, L) – ≥ 1 млн п.н. и ультра-длинные RoH (англ. Ultra Long RoH, UL) – ≥ 10 млн п.н.. Визуализацию была выполнена при помощи программного пакета MACE.

2.13 Филогеномный анализ

В качестве входных данных для филогеномного анализа мы использовали геномные сборки представителей следующих родов: *Martes*, *Pekania*, *Gulo*, *Eira* (сем. Mustelidae, подсемейство Guloninae), *Meles* (сем. Mustelidae, подсемейство Melinae), *Mustela*, *Neogale* (сем. Mustelidae, подсемейство Mustelinae), *Enhydra*, а также *Canis* (сем. Canidae, подсемейство Caninae, в качестве аутгруппы). Идентификаторы всех используемых в анализе геномных сборок, опубликованных в базе данных GenBank NCBI, перечислены в Таблице 9.

В целях исключения “эффекта референса” для *M. caurina* и *M. americana* мы получили сборки геномов получали с использованием сборщика W2rap [444] из образцов SRR11575351 и SRR11575352, соответственно, с последующим скаффолдингом при помощи RagTag v.2.1.0 [445] с использованием полученной сборки генома *M. zibellina* в качестве референса (отмечены в Таблице 9 как “W2rap + RagTag”). Геномную сборку *P. pennanti* (GCA_047676605.1) мы дополнительно проскаффолдили до уровня хромосом при помощи RagTag v.2.1.0 [445] с использованием сборки “DNAzoo” этого вида в качестве референса.

Таблица 9. Описание геномных сборок, используемых в построении филогенетического дерева.

Подсемейство	Латинское название	Вид	Идентификатор*	Уровень сборки**
<i>Guloninae</i>	Харза	<i>Martes flavigula</i>	DNAzoo	Chromosome
	Харза	<i>Martes flavigula</i>	GCA_029410595.1	Chromosome
	Лесная куница	<i>Martes martes</i>	DNAzoo	Chromosome
	Лесная куница	<i>Martes martes</i>	GCA_963455335.1	Chromosome
	Каменная куница	<i>Martes foina</i>	DNAzoo	Chromosome
	Каменная куница	<i>Martes foina</i>	GCA_964304585.1	Chromosome
	Соболь	<i>Martes zibellina</i>	DNAzoo	Chromosome
	Американская куница	<i>Martes americana</i>	W2rap + RagTag	Chromosome
	Тихоокеанская куница	<i>Martes caurina</i>	W2rap + RagTag	Chromosome
	Илька	<i>Pekania pennanti</i>	DNAzoo	Chromosome
	Илька	<i>Pekania pennanti</i>	GCA_047676605.1	Chromosome
	Росомаха	<i>Gulo gulo</i>	GCA_024509225.1	Scaffold
	Тайра	<i>Eira barbara</i>	GCA_020311275.1	Scaffold

<i>Mustelinae</i>	Горноста́й	<i>Mustela erminea</i>	GCA_009829155.1	Chromosome
	Обыкновенная ласка	<i>Mustela nivalis</i>	GCA_019141155.1	Scaffold
	Домашний хорек	<i>Mustela putorius furo</i>	DNAzoo	Chromosome
	Степной хорек	<i>Mustela eversmanii</i>	GCA_963422785.1	Scaffold
	Американская норка	<i>Neogale vison</i>	GCA_020171115.1	Chromosome
<i>Lutrinae</i>	Морская выдра	<i>Enhydra lutris</i>	DNAzoo	Chromosome
	Евразийская речная выдра	<i>Lutra lutra</i>	DNAzoo	Chromosome
<i>Melinae</i>	Обыкновенный барсук	<i>Meles meles</i>	GCA_922984935.1	Chromosome
<i>Canidae</i>	Волк	<i>Canis lupus</i>	GCA_905319855.2	Chromosome

* Идентификатор: DNAzoo – сборка получена в сотрудничестве с консорциумом DNAzoo; W2rap + RagTag – сборка, полученная на основании коротких прочтений (W2rap) с последующим скаффолдингом (RagTag) с использованием геномной сборки соболя.

** Уровень сборки: Chromosome – хромосомный уровень; Scaffold – скаффолдный уровень.

Реконструкция филогенетического дерева была выполнена на основе объединенной матрицы последовательностей ортологичных консервативных генов, полученных для каждого вида при помощи BUSCO v.5.4.2 на основе базы ортологов OrthoDB v.10 для млекопитающих Mammalia_odb v.10 (2021-02-19). Множественное выравнивание общих объединенных нуклеотидных последовательностей консервативных ортологов было выполнено при помощи PRANK v.170427 [199] с параметрами “-codon”, с последующей фильтрацией при помощи GBLOCKS v.0.91b с параметрами “-t=Codons” [205]. Построение филогенетического дерева было выполнено методом максимального правдоподобия при помощи IQ-Tree v.2.2.0 [214] с автоматическим подбором наилучшей модели с помощью ModelFinder [215] и оценкой поддержки ветвей на основе 1000 bootstrap-репликатов, и MrBayes v.3.2.6 [217] с параметрами: модель GTR, 2000000 генераций MCMC, 10 запусков, 4 цепи для каждого, 15% деревьев были отброшены как отжигаемая часть. Альтернативную филогенетическую реконструкцию с использованием обобщенно-коалесцентного подхода мы провели при помощи ASTRAL-III v.5.7.1 [234]. Независимые деревья каждого гена были получены с использованием IQ-Tree v.2.2.0 с автоматическим подбором наилучшей модели и

оценкой поддержки ветвей на основе 1000 bootstrap-репликатов. Полученные филогенетические деревья были визуализированы при помощи ETE Toolkit v.3.1.2 [446]. Вышеописанный протокол филогенетического анализа на полногеномных данных был реализован в виде программного конвейера BuscoClade (<https://github.com/tomarovsky/BuscoClade>).

2.14 Сборка и аннотация мтДНК. Филогения и сеть гаплотипов.

Выделение, сборку и аннотацию мтДНК для каждого образца соболя, лесной куницы и их гибридов мы осуществили на случайной выборке в 20 млн. пар прочтений с помощью MitoZ v.2.3 [447] и Mitos v.2 [448]. Случайная подвыборка прочтений была выполнена при помощи BBMap v.38.18 [427]. Дополнительно в ходе работы были собраны мтДНК для образцов соболей из статьи Liu et al. 2020 (PRJNA495455) [395] и Manakhov et al. 2021 (SRR13213810, SRR13213811, SRR13213812) [449]. Все общедоступные полные последовательности мтДНК для *M. zibellina* и *M. martes* (на момент обращения 01.09.2022) также были включены в анализ [6,321,323,450–452]. Полученные полные мтДНК (всего 140) были выровнены с помощью MAFFT v.7.490 [195]. Фильтрация выравнивания была осуществлена с помощью TrimAl v.1.2rev59 [453]. Кластеризация гаплотипов по сборкам мтДНК исследуемых образцов была проведена при помощи CD-HIT v.4.8.1 [454]. Построение митохондриального филогенетического дерева было выполнено при помощи IQ-Tree v.2.2.0 [214] с автоматическим подбором модели и оценкой поддержки ветвей на основе 1000 bootstrap-репликатов. Построение и визуализация сети гаплотипов была выполнена при помощи PopArt v.1.7 [455] с использованием модели “Median joining networks”.

2.15 Вычисление внутри- и межвидовых генетических расстояний

Для вычисления внутри- и межвидовых генетических расстояний мы реконструировали геномные сборки для всех ресеквенированных образцов на основе SNP и референсных геномных сборок исследуемых видов. Для этого мы использовали инструмент FastaAlternateReferenceMaker из пакета GATK v4.4.0.0 [102] с параметром "--use-iupac-sample" для учета гетерозиготных SNP. Инделлы при

этом не учитывались. Поиск однокопийных ортологов в полученных геномных сборках был выполнен как описано выше в разделе “Филогеномный анализ” Материалов и методов. Выравнивания как ядерных генов, так и последовательностей митохондриальных геномов, были получены с помощью MAFFT v.7.490 [195], с последующей фильтрацией гипервариабельных областей при помощи TrimAl v.1.2rev59 [453]. Матрицы расстояний между как ядерными, так и митохондриальными геномами, были сгенерированы при помощи RapidNJ v2.3.3 [225].

2.16 Анализ популяционной истории

Для реконструкции исторической динамики эффективного размера численности популяции (N_e) мы использовали PSMC v0.6.5 с параметрами “-N25 -t15 -r5 -p ‘4+25*2+4+6’” [95]. Консенсусные диплоидные последовательности (входные данные PSMC) были получены при помощи Samtools v.0.1.19 и скрипта vcfutils.pl [429], при этом X хромосома была исключена. Минимальный и максимальный пороги покрытия (параметры -d и -D инструмента vcf2fq) были установлены на 33% и 250% от медианного покрытия соответствующего генома. Варианты с меньшим или большим покрытием были исключены из анализа. Для масштабирования результатов PSMC мы использовали значение времени генерации (g) в 5 лет [248] и частоту замен на поколение (μ) равную $4,64 \times 10^{-9}$ [456]. В качестве нижней и верхней границы доверительного интервала для μ были использованы значения μ равные $2,94 \times 10^{-9}$ и $7,37 \times 10^{-9}$, соответственно [456]. Для каждого образца анализ был повторен для 100 бутстреп-репликатов (-b 100).

3 Результаты

3.1 Контроль качества полногеномных данных и размер генома

Контроль качества полногеномных прочтений для всех исследуемых образцов до и после фильтрации не выявил признаков контаминации. На основании анализа распределения 23-меров (Рисунок 14) у нескольких образцов соболя (T50, T8, T18, T118, T72 и, в меньшей степени, T104), лесной куницы (T151) и их предполагаемых гибридов (T87, T84, T24, T86) мы обнаружили дополнительный пик на половинном покрытии (отмечен для образца T87). Этот пик может свидетельствовать о высоком уровне гетерозиготности, а также может быть связан с полом особи, и может наблюдаться у самцов млекопитающих в менее выраженной форме. Наиболее выраженный пик выявлен у предполагаемого гибридного образца T87.

Анализ размеров геномов исследуемых особей выявил незначительную вариацию в пределах вида (Приложение 1). Среди предполагаемых соболей наименьший размер генома выявлен у образца T72 (2,39 млрд п.н.), а наибольший – у S26 (2,53 млрд п.н.). У лесных куниц наименьший геном у референсного образца 10xmtar (2,45 млрд п.н.), а наибольший – у S50 (2,54 млрд п.н.). Среди предполагаемых гибридов минимальный размер генома отмечен у образцов T87, T86 и T84 (2,45 млрд п.н.), тогда как максимальный – у T76 (2,5 млрд п.н.). Таким образом, для перечисленных групп особей вариация в размерах геномов составила менее 5%. Наибольшая вариация в размере генома (~15%) установлена среди четырех особей тихоокеанских куниц: наименьший размер генома выявлен у образца SRR11575343 (2,43 млрд п.н.), а наибольший – у SRR11575348 (2,83 млрд п.н.). Образцы каменной куницы и харзы показали оценки размера геномов равные 2,37 млрд п.н. и 2,46 млрд п.н., соответственно.

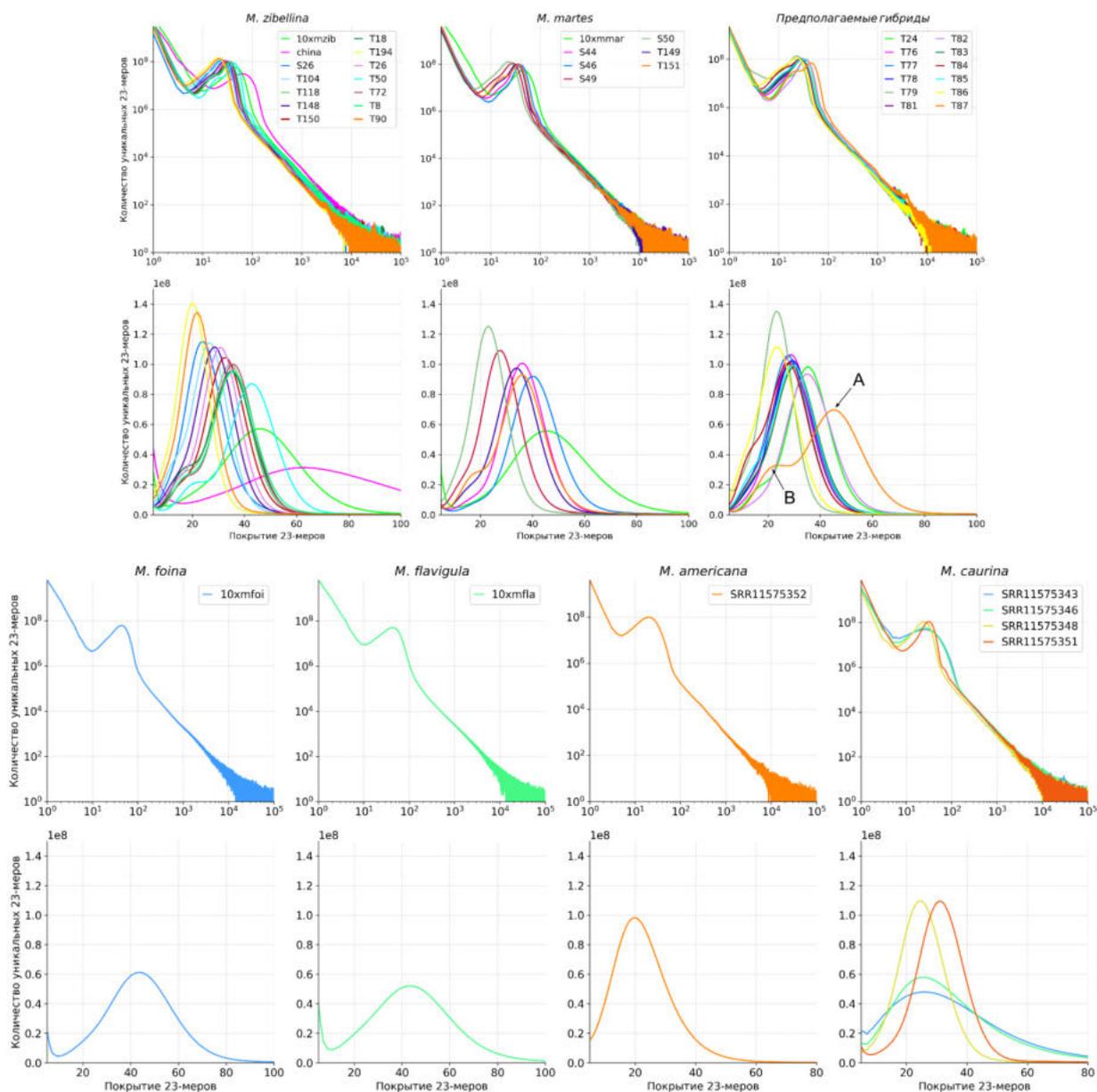


Рисунок 14. Распределения 23-меров для особей соболя, лесной куницы и их гибридов.

Верхние графики даны в логарифмическом масштабе по обеим осям, а нижние – в линейном. Для образца T87 стрелками обозначен основной (A) и дополнительный (B) пик.

3.2 Сборка и аннотация геномов

Мы впервые получили сборки хромосомного уровня для каменной куницы и соболя. Основные характеристики геномных сборок представлены в Таблице 10. Оцененные размеры геномов референсных образцов соболя, лесной куницы, харзы

и каменной куницы, полученные на основании распределения 23-меров (Приложение 1), согласуются с размерами полученных геномных сборок. Наименьшая длина в 2,39 млрд п.н. у сборки генома соболя, а наибольшая – 2,45 млрд п.н. у харзы. Сборки геномов характеризуются значениями N50 от 137,4 млн п.н. (у харзы) до 144,6 млн п.н. (у лесной куницы). Отметим, что значение N50 в данном случае обусловлено длиной хромосомных скаффолдов. GC-контент совпадает для сборок соболя и лесной куницы (41,27%), а у каменной куницы и харзы составляет 41,1% и 41,26%, соответственно.

Геномные сборки соболя, лесной куницы и каменной куницы имеют 19 хромосомных скаффолдов, что соответствует количеству пар хромосом в кариотипе этих видов ($2n=38$). В свою очередь, геномная сборка харзы содержит 20 хромосомных скаффолдов, поскольку кариотип этого вида отличается бóльшим количеством хромосом ($2n=40$). Отметим, что сборка харзы характеризуется более чем в три раза бóльшим количеством контигов (22884) по сравнению с другими полученными геномными сборками. Сборка каменной куницы, в свою очередь, отличается наибольшим количеством пробелов, общей протяженностью в 33,8 млн п.н. Это связано с тем, что изначальные черновые сборки этих видов на основании сцепленных прочтений получились значительно более фрагментированные, и, как следствие, меньшее число черновых скаффолдов удалось включить в хромосомные на этапе Hi-C скаффолдинга.

Таблица 10. Основные характеристики полученных сборок геномов исследуемых видов.

Вид	Кол-во контигов * (≥ 1000 п.н.)	Размер сборки, млрд п.н.	Ns **, млн п.н.	Макс. длина контига, млн п.н.	GC, %	N50, млн п.н.	L50
<i>Martes zibellina</i>	5965	2,39	24,9	216,8	41,27	143,6	7
<i>Martes martes</i>	6354	2,40	24,9	216,9	41,27	144,6	7
<i>Martes foina</i>	8115	2,41	33,8	216,4	41,1	144,1	7
<i>Martes flavigula</i>	22884	2,45	25,9	214,0	41,26	137,4	7

* Количество контигов протяженностью более 1 тыс п.н., включая хромосомные скаффолды;

** Ns – количество пропущенных нуклеотидов.

Для оценки качества геномных сборок важно учитывать не только общие метрики, но и полноту, которую можно оценить по представленности и целостности консервативных генов. С этой целью мы использовали метрики BUSCO (Таблица 11). Результаты BUSCO с использованием базы данных однокопийных консервативных ортологов Mammalia_odb v.10 (2021-02-19) показали, что более 96% однокопийных ортологов (BUSCOs) были успешно собраны. Лишь менее 3% генов были пропущены, за исключением сборки генома харзы, где доля пропущенных генов составила 4,3%. Также у харзы наблюдалось большее количество дублицированных (1,5%) и фрагментированных (2,1%) BUSCOs в сравнении с другими сборками (Рисунок 15). Несмотря на более низкие показатели качества, сборка генома харзы остается сборкой хромосомного уровня, пригодной для выполнения поставленных задач. В целом, значения метрик BUSCO сопоставимы с полученными ранее сборками геномов хищных [105,106,457]. Отметим, что в полученных сборках геномов дополнительно выполнено удаление гаплотипических дубликатов, что несколько сократило количество дублицированных BUSCOs.

Таблица 11. Оценки качества геномных сборок по консервативным ортологичным группам (BUSCOs).

BUSCO метрики	<i>M. zibellina</i>		<i>M. martes</i>		<i>M. flavigula</i>		<i>M. foina</i>	
Цельные BUSCOs	8870	96,1%	8895	96,4%	8636	93,6%	8881	96,3%
Цельные однокопийные BUSCOs	8816	95,6%	8843	95,8%	8501	92,1%	8819	95,6%
Цельные дублицированные BUSCOs	54	0,6%	52	0,6%	135	1,5%	62	0,7%
Фрагментированные BUSCOs	100	1,1%	86	0,9%	195	2,1%	95	1%
Отсутствующие BUSCOs	256	2,8%	245	2,7%	395	4,3%	250	2,7%
Всего BUSCOs	9226	100%	9226	100%	9226	100%	9226	100%



Рисунок 15. Диаграмма на основании абсолютных значений оценки качества по консервативным ортологичным группам (BUSCOs) полученных геномных сборок исследуемых видов.

В результате предсказания белок-кодирующих генов в геномных сборках мы получили от 20087 (*M. foina*) до 20521 (*M. flavigula*) моделей генов, из которых для >90% были успешно присвоили имена (Таблица 12). Больше всего именованных генов у *M. flavigula* (18595 генов), а наименьшее – у *M. martes* (18248 генов). Количество общих именованных генов между сборками составило 12626 генов, при этом присутствует до ~1,4 тыс уникальных генов для каждого вида (Рисунок 16). Оценка предсказания на основе BUSCO (Mammalia_odb10) показала присутствие более 97% цельных BUSCOs (за исключением харзы, у которой 95,1%), что свидетельствует о высокой полноте аннотации.

Таблица 12. Результаты предсказания и аннотации генов.

Вид	Количество моделей генов	Количество транскриптов	Количество именованных генов	Цельные BUSCOs (%)
<i>M. zibellina</i>	20158	30837	18300	97,7%
<i>M. martes</i>	20259	31116	18248	97,7%
<i>M. flavigula</i>	20521	30087	18595	95,1%
<i>M. foina</i>	20087	30708	18283	97,8%

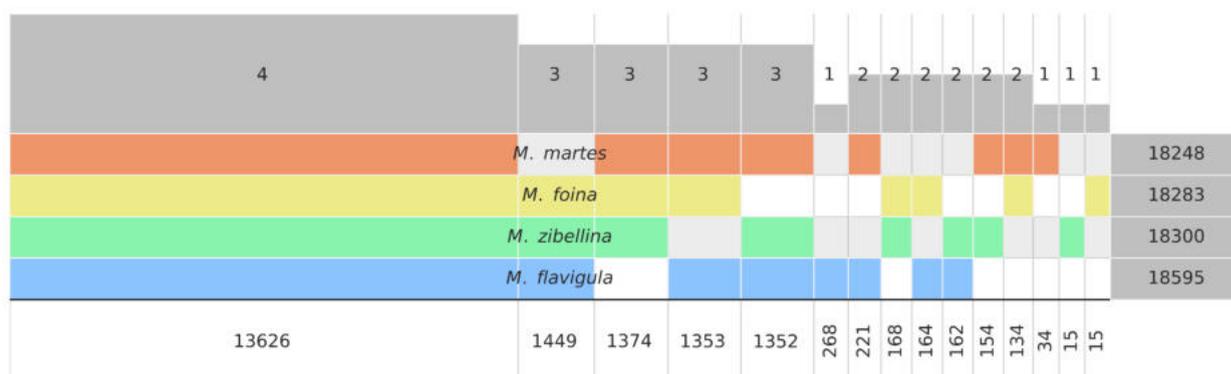


Рисунок 16. Перекрывание именованных генов в полученных геномных сборках видов рода *Martes* (линейная диаграмма Венна).

Цветными блоками обозначены множества именованных генов. Сверху показано у какого количества видов пересекаются множества именованных генов, справа –

суммарное количество генов для сборки каждого вида, а снизу – количество генов в перекрывающихся подмножествах.

3.3 Аннотация повторов, покрытие и псевдоаутосомный район

В геномных сборках исследуемых видов мы выявили сопоставимые доли (от 39,07% генома каменной куницы до 39,70% генома харзы) повторяющихся элементов (Приложение 2). Большинство повторов (36,19% - 36,86%) относятся к группе ретроэлементов, что является обычным для хищных видов животных. Среди них от 21,57% (*M. foina*) до 22,1% (*M. flavigula*) составляют LINE (длинные рассеянные повторы), за ними следуют от 4,77% (*M. martes* и *M. foina*) до 4,8% (*M. flavigula*) LTR-элементов (длинные концевые повторы) и от 9,84% (*M. foina*) до 9,96% (*M. flavigula*) SINE (коротких рассеянных повторов). Только от 2,82% (*M. flavigula*) до 2,88% (*M. zibellina*) генома представлено ДНК-транспозонами. Профили Кимуры также имеют одинаковое количество и расположение пиков для всех трехборок (Рисунок 17).

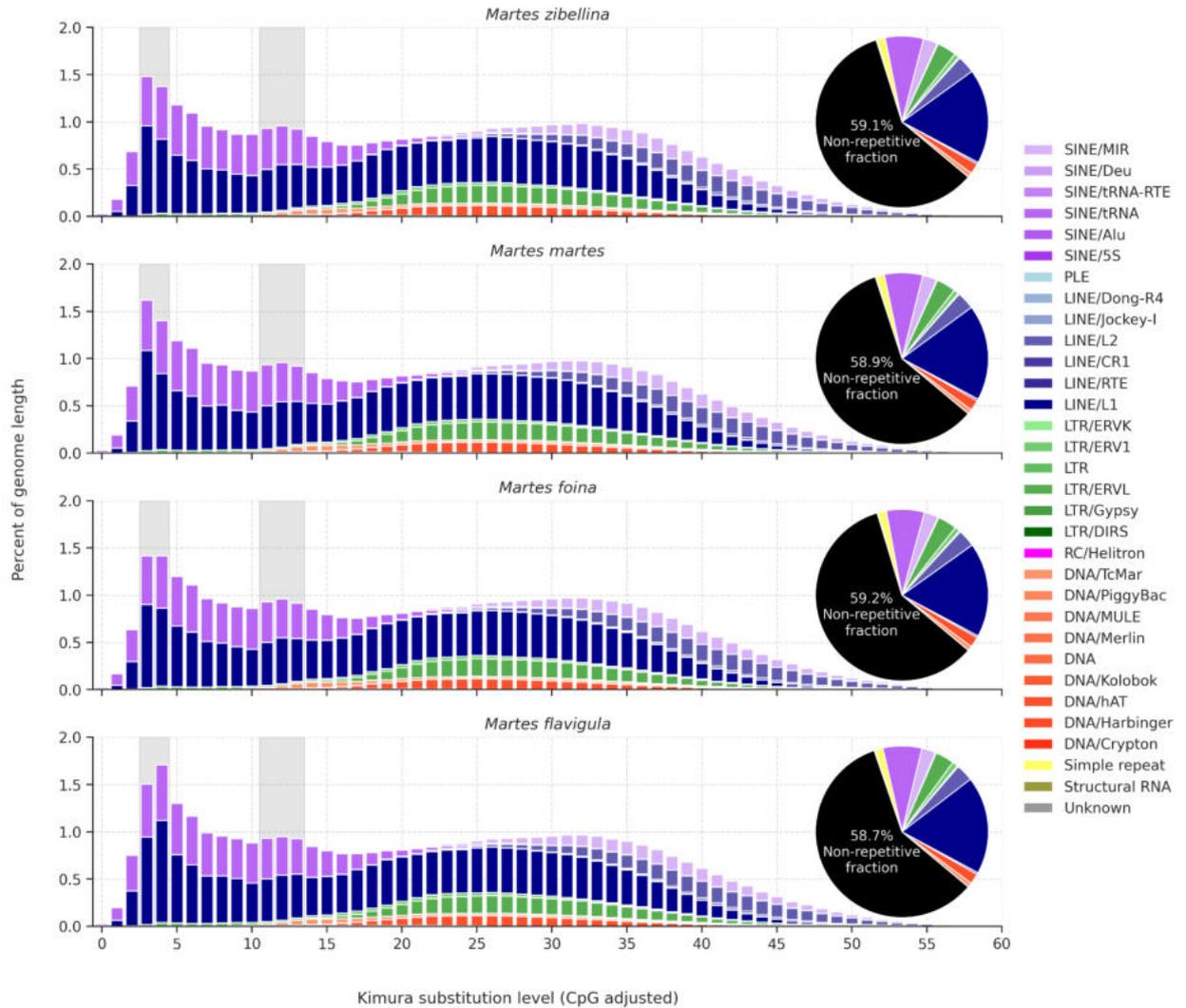


Рисунок 17. Анализ расхождения копий повторенных элементов в геномных сборках *M. zibellina*, *M. martes*, *M. foina* и *M. flavigula* на основе расстояний Кимуры.

Графики представляют покрытие генома для каждого суперсемейства повторов в анализируемых геномах.

Дальнейшие этапы анализа, основанные на генетических вариантах, требуют примерно одинакового покрытия во всех образцах для получения сопоставимых результатов. В связи с различным покрытием генома, которое варьировалось от 21,24x (T194) до 76,78x (china), для всех образцов мы выполнили случайную подвыборку прочтений для достижения примерно одинакового покрытия в 22x (Приложение 1).

После маскирования регионов со слишком высоким (более 250% медианного покрытия всего генома) или низким (менее 33%) покрытием (Приложения 3 и 4) мы выявили значимые различия в покрытии при выравнивании образцов соболей, лесных куницы и их гибридов относительно референсных геномов (*M. zibellina* и *M. martes*) (Приложение 5 и 6). Это стало понятно только после классификации этих образцов по видовой принадлежности (см. раздел Методов “*Оценка степени интрогрессии и классификация по видовой принадлежности*”). Выявленные различия в покрытии между группами затрудняли выбор подходящего референса (сборки соболя или лесной куницы) для последующего анализа. Таким образом, по причине наличия “эффекта референса”, все анализы для образцов соболей, лесных куниц и их гибридов проводились дважды – с использованием каждой изборок.

Дополнительно в ходе анализа покрытия (Приложения 7 и 8) был определен (либо подтвержден фенотипически установленный) пол каждого образца (Приложение 1). Китайский образец соболя (china), изначально идентифицированный как самец, на самом деле оказался самкой.

В целях корректного поиска диплоидных генетических вариантов для геномныхборок определены координаты псевдоаутосомного района (PAR) с помощью подхода, основанного на покрытии. Важность определения координат псевдоаутосомного района (PAR) обусловлена необходимостью корректно указывать координаты гаплоидной части X хромосомы у самцов на этапе поиска генетических вариантов. Как и ожидалось, PAR был обнаружен в конце X-хромосомы и имеет примерно одинаковую длину (~6,5 млн п.н.) с небольшими различиями междубороками (Приложения 9-12). В случае американской куницы длина PAR составила 6,48 млн п.н. (с 200 тыс п.н. до 6,680 тыс п.н.), тогда как для образцов тихоокеанской куницы наблюдалась незначительная вариация, вызванная неравномерным покрытием: наименьшая длина PAR в 6,48 млн п.н. (с 200 тыс п.н. до 6,680 тыс п.н.) была у SRR11575348, а наибольшая – 6,74 млн п.н. (с 0,2 млн п.н. до 6,94 млн п.н.) у SRR11575346.

3.4 Анализ синтении *M. foina* с другими видами

Каменная куница зарекомендовала себя в качестве эталонной модели в множестве цитогенетических исследований кариотипов млекопитающих (см. раздел “*M. foina* как потенциально базальный вид рода *Martes*” Обзора литературы) [458]. Мы выполнили сравнительный анализ синтении ее генома (MFOI) с геномом человека (*Homo sapiens*, HSAP), домашней кошки (*Felis catus*, FCAT) и собаки (*Canis familiaris*, CFAM) на основе полногеномных выравниваний (Рисунок 18). Полученная синтения согласуется с опубликованными картами сравнительного хромосомного окрашивания Zoo-FISH между каменной куницей и этими видами [34,63]. Тем не менее, нам удалось выявить 3 синтенных блока длиной чуть более 5 млн п.н., отсутствующие в ранее опубликованной [63] карте окраски CFAM против MFOI: CFAM 33 на MFOI 2, CFAM 27 на MFOI 9 и CFAM 26 на MFOI 12 (Рисунок 18B, Приложение 13). В опубликованной карте человека и каменной куницы [63] отсутствуют 4 подобных синтенных блока, один из которых (HSAP 1 на MFOI 12) имеет длину около 9,7 млн п.н. (Рисунок 18C, Приложение 14). Стоит отметить, что для каждой сравниваемой пары видов были также обнаружены короткие синтенные блоки менее 5 млн п.н., отсутствующие в опубликованных картах хромосомного окрашивания (Приложения 13-15). Так, для синтении между FCAT и MFOI оказалось 5 таких блоков, для CFAM и MFOI – 28 блоков, и для HSAP и MFOI – 11 блоков.

Синтения между CFAM и MFOI (Рисунок 18B, Приложение 13) оказалась наиболее полезной по причине наибольшего числа хромосом и большого количества перестроек. Это позволило нам приблизительно (поскольку сами центромеры отсутствуют в геномной сборке) определить координаты центромер для каждой хромосомы (Приложение 16). Практически все центромеры были расположены в местах разрыва синтении между участками, соответствующими разным хромосомам в геноме CFAM, что значительно упростило определение их координат.

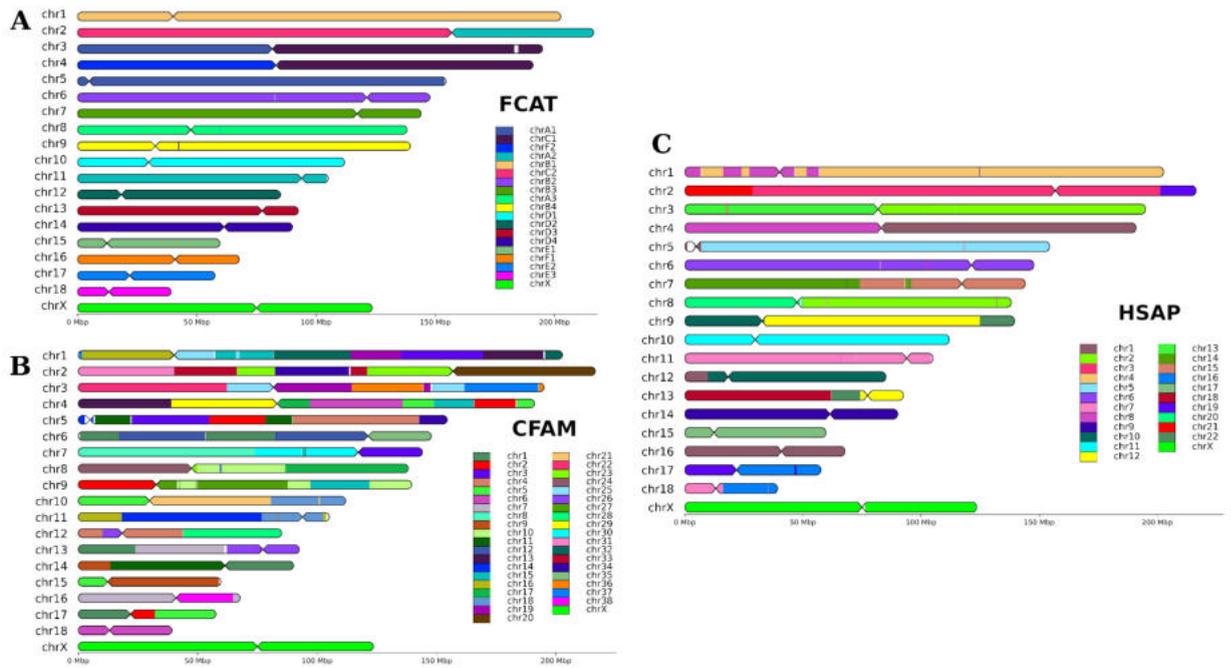


Рисунок 18. Синтения между каменной куницей, домашней кошкой, домашней собакой и человеком.

А – синтеные блоки с домашней кошкой (*Felis catus*, FCAT) на хромосомах каменной куницы; В – с домашней собакой (*Canis familiaris*, CFAM) и С – с человеком (*Homo sapiens*, HSAP), соответственно.

3.5 Синтения между видами *Martes*

Для изучения синтении между исследуемыми видами рода *Martes* мы провели множественное полногеномное выравнивание (Рисунок 19). Известно, что каменная куница, лесная куница и соболь имеют одинаковое число хромосом $2n=38$, тогда как харза отличается дополнительной парой хромосомы 19 [265]. Основываясь на этом, в ходе анализа синтении нам удалось установить соответствие 1 к 1 между хромосомными скаффолдами и хромосомами в кариотипах исследуемых видов (Приложение 17), а также определить примерные координаты центромер лесной куницы, соболя и харзы (Приложение 16).

Поскольку некоторые хромосомные скаффолды в геномных сборках соболя, лесной куницы и харзы имеют разную ориентацию относительно скаффолдов в геномной сборке каменной куницы (Приложение 17), они были инвертированы для удобства анализа. Скаффолды в геномной сборке каменной куницы были

использованы в качестве референса для ориентации, в связи с чем координаты всех инверсий определены относительно сборки генома этого вида (Таблица 13). Всего мы выявили 4 инверсии, некоторые из которых являются специфичными для отдельных видов. Инверсии на хромосомах 11 и 14 являются специфичными для соболя и харзы, соответственно. На хромосоме 15 (у харзы хромосома 12) обнаружили инверсию между парами видов: между харзой и каменной куницей (1-ая пара) и между соболем и лесной куницей (2-ая пара). Инверсия в этом случае произошла как у лесной куницы, так и у соболя, относительно каменной куницы, а значит она, вероятно, произошла у их общего предка. Наконец, на хромосоме 12 (у харзы хромосома 13) мы выявили общую инверсию для других пар видов: между соболем и каменной куницей (1-ая пара) и между харзой и лесной куницей (2-ая пара). Эта инверсия является гомоплазией, поскольку она не согласуется с филогенией видов. Это, предположительно, может быть результат двух независимых инверсий: либо у лесной куницы и харзы, либо у соболя и каменной куницы. Отметим, что инверсии на хромосомах 11, 12 и 15 расположены на р-плечах акроцентрических хромосом. При этом инверсия на хромосоме 11 полностью охватывает р-плечо этой хромосомы и была подтверждена сотрудницами ИМКБ СО РАН к.б.н. Виолеттой Робертовной Беклемишевой и к.б.н. Полиной Львовной Перельман на основе сравнения G-бэндинга кариотипов соболя и лесной куницы (Приложение 18).

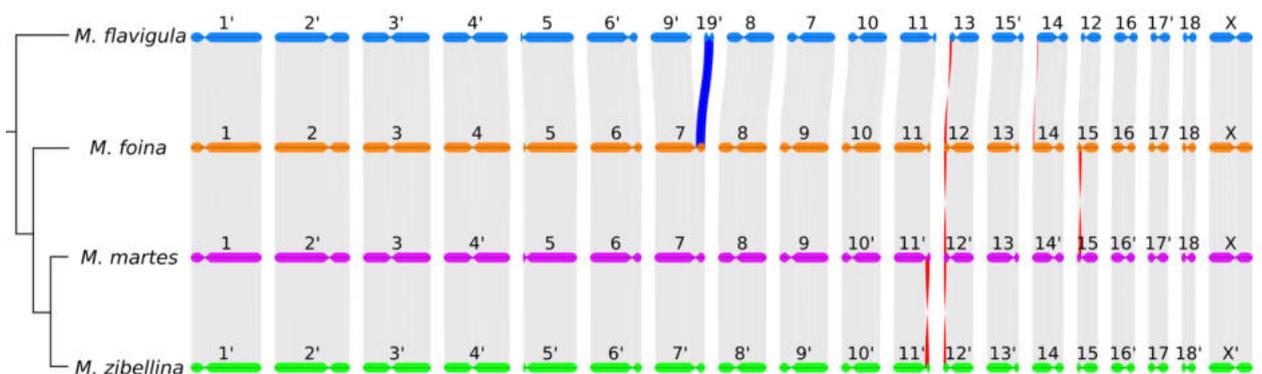


Рисунок 19. Синтенные блоки между исследуемыми представителями рода *Martes*: харзой (*M. flavigula*), каменной куницей (*M. foina*), лесной куницей (*M. martes*) и соболем (*M. zibellina*).

Инверсии и транслокации размером от 1 млн п.н. выделены красным и синим цветом, соответственно. Разнонаправленные хромосомные скаффолды помечены штрихом.

Внутри выявленных инверсий мы обнаружили небольшое количество генов: от 34 (Chr11) до 77 (Chr15) (Таблица 13), в связи с чем анализ обогащения (англ. GO enrichment) не показал статистически значимых результатов. Поэтому мы проанализировали описания и известные функции генов, расположенных в этих участках геномов. Нас особо интересовало сравнение соболя и лесной куницы в контексте различия их фенотипов, а также гибридизации. Мы нашли два гена, расположенные внутри инверсии на хромосоме 11 и участвующие в репродуктивных процессах: *SPMIP7* и *ZPBP*, а также ген *GLI3*, связанный с развитием конечностей и скелета.

Таблица 13. Координаты инверсий относительно сборки генома *M. foina* и количество генов в них.

Хромосома	Координаты				Протяженность, млн п.н	Количество генов
	Скаффолд	NCBI ID	Начало	Конец		
Chr11	HiC_scaffold_11	CM081894.1	90919984	102388530	11,5	34
Chr12	HiC_scaffold_14	CM081895.1	616761	7205853	6,6	47
Chr14	HiC_scaffold_13	CM081897.1	1113377	3565510	2,5	39
Chr15	HiC_scaffold_16	CM081898.1	4271212	12359108	8,1	77

3.6 Анализ гетерозиготности

Для анализа распределения гетерозиготности по хромосомам мы провели подсчет гомозиготных и гетерозиготных SNP в перекрывающихся окнах размером 1 млн п.н. с шагом 100 тыс п.н. Общее число гетерозиготных SNP существенно варьировало между образцами соболей, лесных куниц и их гибридов: от 2,01 млн (10xmmar) до 9,01 млн (T87) SNP, при этом наибольшее количество гетерозиготных SNP наблюдалось у предполагаемых гибридов (Приложение 19).

В ходе анализа распределений гетерозиготных SNP (Рисунок 20 и Приложение 20) мы обнаружили четыре основных моды распределений гетерозиготности, присутствующих в различных комбинациях у разных особей.

Аппроксимируя распределения при помощи линейной комбинации отрицательных биномиальных распределений (см. раздел Методов “*Анализ распределений генетических вариантов*”), мы выделили четыре компоненты гетерозиготности для образцов, визуализировали распределения гетерозиготных SNP (Приложения 21 и 22) и определили соответствующие моды и средние значения каждой компоненты (Приложение 23 и 24). Первая компонента со значением 0 SNP/тыс п.н. соответствует регионам с низкой гетерозиготностью (потенциальные участки RoH) и в разной степени проявляется практически у всех образцов. Вторая компонента со средним значением 0,7 SNP/тыс п.н. характерна для образцов, априори относимых к лесным куницам, но также присутствует у особей из зоны симпатрии. Третья компонента (среднее значение 1,7 SNP/тыс п.н.) была обнаружена во всех образцах соболей, а также у образцов T84 (сопоставимый уровень с соболями) и T87 (значительно более низкий уровень). Наконец, последняя компонента (среднее значение 4,5 SNP/тыс п.н.) в основном связана с куницами из зоны симпатрии (T78, T79, T85, T86, T84, T87) и особенно явно проявляется в образцах T84 и T87. Однако она также присутствует у нескольких соболей (T18) и лесных куниц (S50, T151). Этот первый шаг анализа показал, что наш набор образцов охватывает как “чистых” соболей и лесных куниц, так и широкий спектр их предполагаемых гибридов. Высокое соответствие аппроксимированных кривых эмпирическим данным (Приложения 21 и 22) подтвердило адекватность использования отрицательного биномиального распределения для компенсации наблюдаемой сверхдисперсии распределений гетерозиготности.

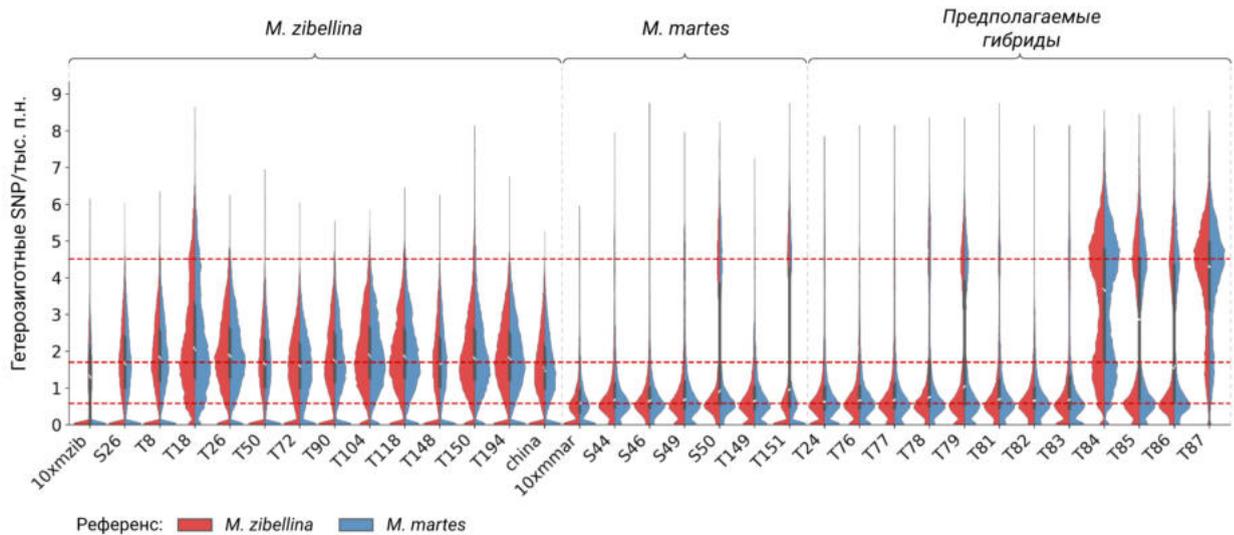


Рисунок 20. Распределения гетерозиготности (гетерозиготные SNP) в скользящих окнах размером 1 млн п.н. с шагом в 100 тыс п.н. для образцов соболей, лесных куниц и их предполагаемых гибридов.

В качестве референса использованы геномная сборка соболя (красный) и сборка лесной куницы (синий). Каждое распределение масштабировано относительно оси X. Горизонтальные красные линии обозначают 4 уровня гетерозиготности: 0, 0,7, 1,7 и 4,5 SNP/тыс п.н.. Особи сгруппированы в соответствии с предварительной классификацией.

При этом группы окон из различных компонент гетерозиготности образуют на хромосомах протяженные участки с примерно одинаковым уровнем гетерозиготности. Так, для референсных образцов соболя (10xmzib) и лесной куницы (10xmmar) окна нулевой компоненты формируют протяженные участки гомозиготности (RoH) на хромосомах 3, 4, 6, 7, 8, 11 и 12 и хромосомах 1, 3 и 4, соответственно (Рисунок 21G, 21H). В свою очередь, участки высокой гетерозиготности соответствуют гибридной компоненте. Например, для образца S50 такие участки расположены на хромосомах 3, 4, 7, 8, 15, 16, 18 (Рисунок 21J), для образца T79 – на хромосомах 2, 4, 6, 8, 9, 10, 12 (Рисунок 21K), а для образца T85 – на хромосомах 1, 4, 5, 6-9, (Рисунок 21L). Отметим, что такие протяженные участки могут занимать как целую хромосому, так и небольшую ее часть.

Среди общедоступных образцов *M. saurina* выделяются 2 группы особей: материковые (Рисунок 21А и 21В) и островные (Рисунок 21D и 21Е). Материковые особи из штата Колорадо (SRR11575346) и штата Вашингтон (SRR11575351) показали среднюю гетерозиготность равную 0,49 SNP/тыс п.н. и 0,58 SNP/тыс п.н., соответственно. В свою очередь, островные особи с о. Грейам (SRR11575343) и о. Адмиралтейский (SRR11575348) отличались крайне низким уровнем гетерозиготности, равным лишь 0,11 SNP/тыс п.н. и 0,02 SNP/тыс п.н., соответственно. При этом для общедоступного образца американской куницы (SRR11575352) характерна более высокая средняя гетерозиготность равная 0,83 SNP/тыс п.н. (Рисунок 21С и Приложение 25). Наконец, анализ гетерозиготности референсных образцов харзы и каменной куницы показал среднее значение гетерозиготности в 0,82 SNP/тыс п.н. и 0,42 SNP/тыс п.н., соответственно (Рисунок 21F и 21I, Приложение 25). Распределения гетерозиготных и гомозиготных SNP (SNP/тыс п.н.) для всех исследуемых образцов в виде тепловых карт в Приложении 26.

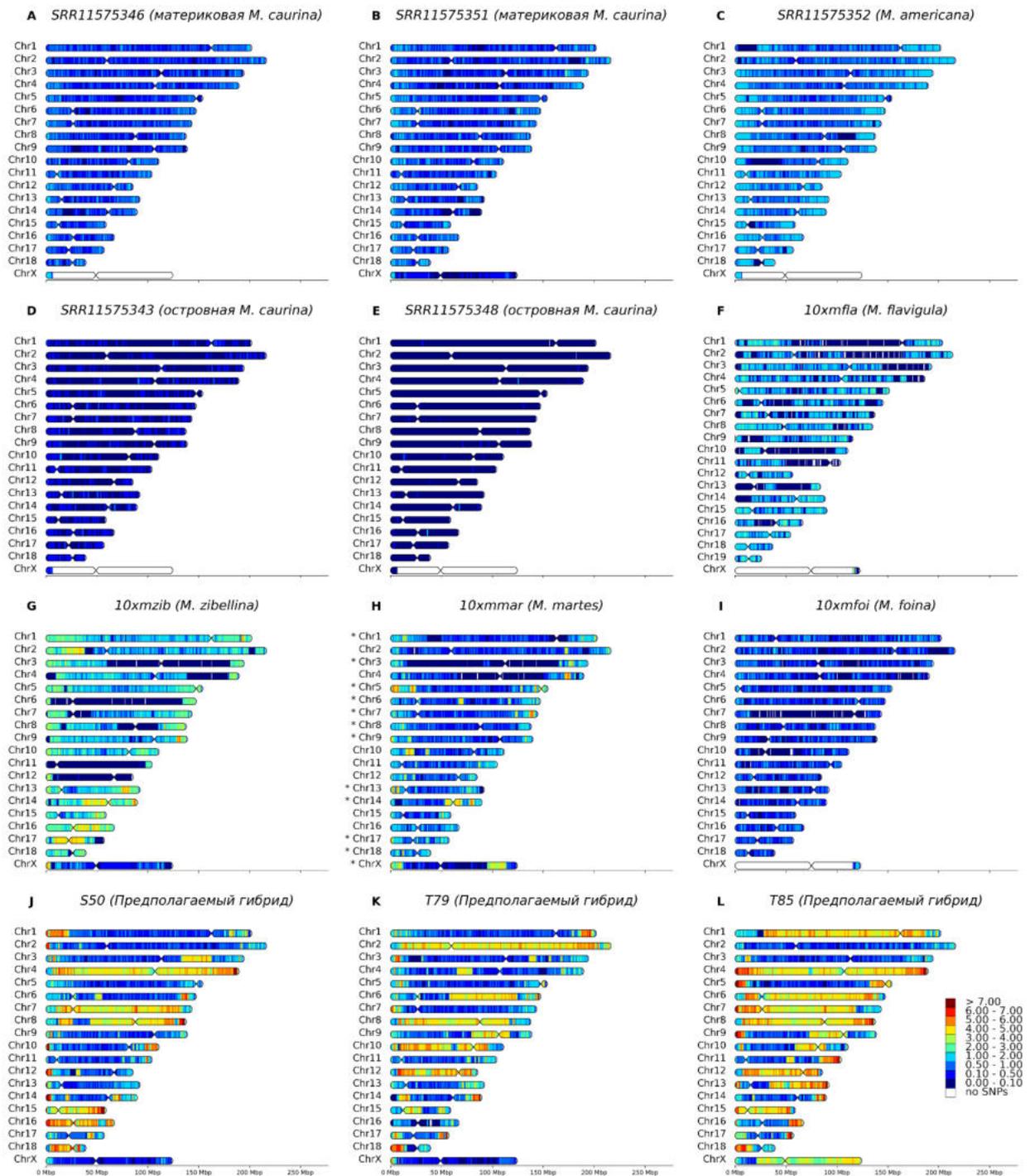


Рисунок 21. Тепловые карты гетерозиготности (гетерозиготные SNP) в скользящих окнах размером 1 млн п.н. с шагом в 100 тыс п.н.

А, В, D, E – материковые (SRR11575346 и SRR11575351) и островные (SRR11575343 и SRR11575348) образцы *M. saurina*, соответственно. Референс: *M. zibellina*; С – образец *M. americana* (SRR11575352); F, G, H, I – референсные образцы харзы (10xmfla), соболя (10xmzib), лесной куницы (10xmmar) и каменной куницы (10xmfoi), соответственно. Инvertированные хромосомы 10xmmar

обозначены '*'. В качестве референсов использованы сборки геномов соответствующих видов; J, K, L – образцы предполагаемых гибридов соболя и лесной куницы S50, T79 и T85, соответственно. Референс: *M. zibellina*.

3.7 Классификация по видовой принадлежности

Для классификации особей по видовой принадлежности был выполнен PCA (Рисунок 22А), в результате которого образцы сформировали 2 отчетливых кластера относительно 1-й главной компоненты (горизонтальная ось): кластер лесных куниц (слева) и кластер соболей (справа). Образцы T87 и T84 расположены почти посередине этих двух групп, что подтверждает гибридный статус этих образцов, потенциально наиболее близких к поколению F1. Таким образом, 1-ая главная компонента (ось X), объясняющая 30,27% дисперсии образцов, что, как мы выяснили в ходе последующего анализа, связано с различной долей интрогрессии от соболя, а также, возможно, с вариациями внутри лесной куницы. С другой стороны, 2-ая главная компонента (ось Y) имеет низкий процент объясненной дисперсии (5,59%) и, вероятно, связана с разнообразием соболей. Важно также отметить еще два образца на графике. Первый, T18, находится между гибридными образцами и кластером соболей, а второй, T72, входит в кластер соболей, но сильно дистанцируется от остальных особей.

Последующий полногеномный анализ ADMIXTURE (анализ происхождения) показал наличие примесей (>5%) в следующих 9 образцах (включая образцы родительских видов): T18, S50, T151, T78, T79, T84, T85, T86 и T87 (Рисунок 22D). Анализ был проведен для предполагаемого количества кластеров (K) в диапазоне от 1 до 6, но, как видно из графиков средних ошибок перекрестной проверки и графиков среднего попарного сходства (Рисунок 22С), наиболее достоверным является K = 2. Доля примеси в образцах колеблется от 3,9% (T78) до 40,8% (T87), в то время как остальные образцы не содержат следов смешения. Локальный анализ ADMIXTURE, выполненный в перекрывающихся окнах размером 1 млн п.н. с шагом 100 тыс п.н., позволил обнаружить примесь еще в образцах T81 и T83 равную 6,17% и 7,86%, соответственно (Рисунок 22D). Кроме того, обнаружены следы интрогрессии от 1% до 5% в нескольких других образцах,

включая референсные: 10xmzib, 10xmmar, T150, S44, S46, S49, T149, T24, T76, T77, T82. Именно референсный образец соболя 10xmzib отделяется от остальных соболей в сторону лесных куниц на графике PCA (Рисунок 22А). Важно отметить, что незначительные следы интрогрессии (<1%) были обнаружены и в остальных образцах. Схожие результаты мы получили и при использовании геномной сборки *M. martes* в качестве референса (Приложение 27).

Поскольку в литературе ранее не приводились точные доли интрогрессии для классификации особей как “гибридных”, мы определили условный порог минимальной доли интрогрессии в 5%. В результате, образцы были разделены на 3 группы:

1. “Чистые” соболя (13 особей): 10xmzib, S26, T8, T26, T50, T72, T90, T104, T118, T148, T150, T194, china;
2. “Чистые” лесные куницы (9 особей): 10xmmar, S44, S46, S49, T149, T24, T76, T77, T82;
3. Гибридные особи с не менее чем 5% интрогрессии, которые мы дополнительно разбили на классы исходя из доли интрогрессии другого вида (11 особей):
 - a. F1-подобные (35-50%): T87;
 - b. Backcross-подобные (15-35%): S50, T79, T151, T86, T85, T84, T18;
 - c. “Нетипичные” (5-15%): T81, T83, T78.

Для исключения сценария изоляции по расстоянию выполнили визуализацию зависимости межгрупповой гетерозиготности и гибридного индекса (Рисунок 22, В). Расположение образцов, сопоставимое с результатами PCA-анализа, формирует “арку”, подтверждая сценарий нейтральной диффузии. Общее количество информативных предковых маркеров для выбранного порога разности частот аллелей (δ) = 0,75 составило 502893.

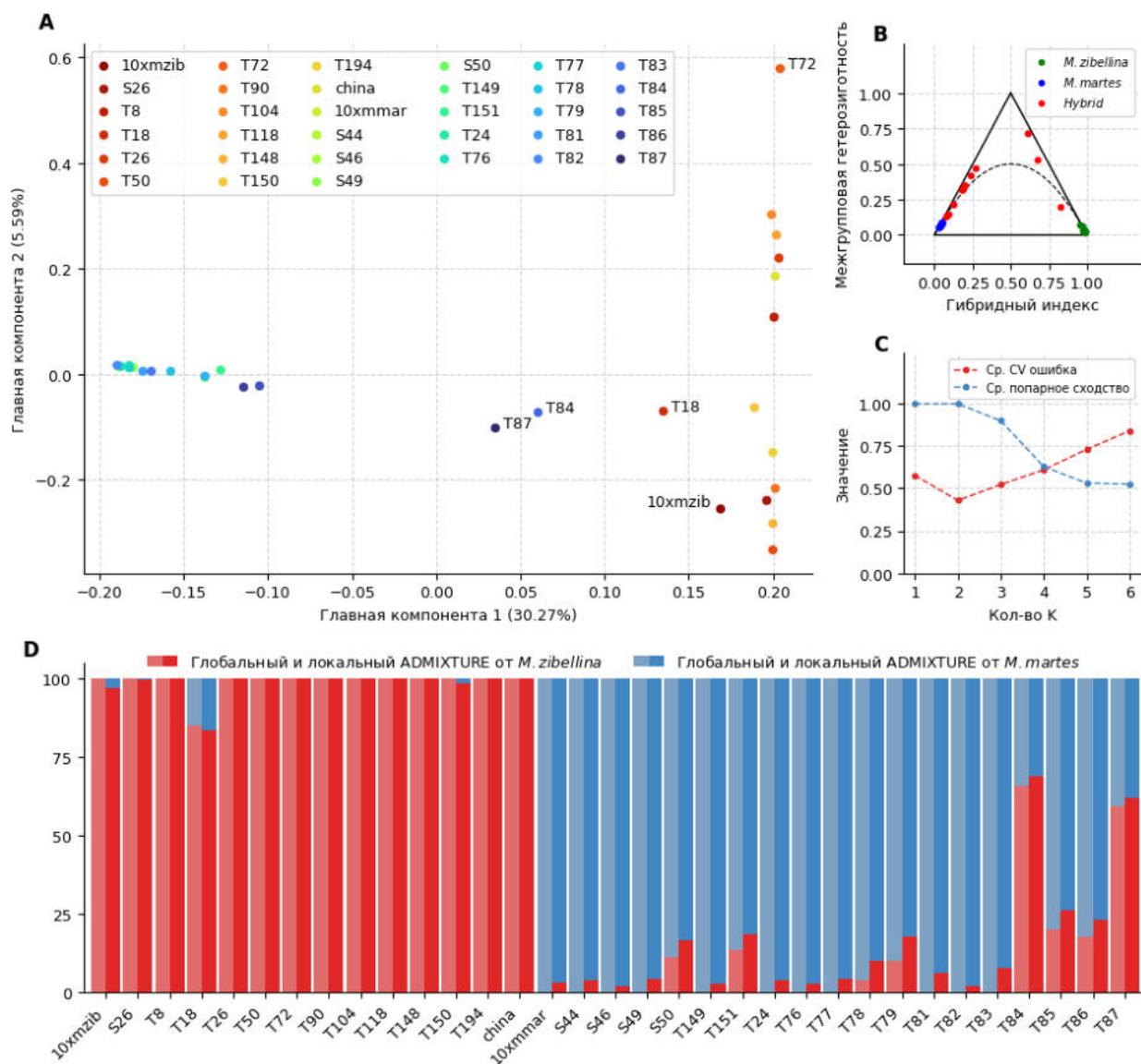


Рисунок 22. Кластеризация и смешение проанализированных образцов соболя, лесной куницы и предполагаемых гибридов.

А – Анализ главных компонент (PCA) на основе SNP из аутосом и псевдоаутосомного региона (PAR) сборки *M. zibellina*. Пунктирной линией обозначены кластеры лесных куниц (слева), соболей (справа) и их гибридов (посередине). Образец T18 находится вне кластеров. Первая и вторая главные компоненты представлены на осях X и Y соответственно. В скобках указан процент дисперсии, объясняемой соответствующей компонентой; В – Треугольная диаграмма зависимости индекса гибридизации (ось X) и гетерозиготности (ось Y). Гибридные образцы отмечены красным, лесные куницы – синим, а соболя – зеленым (Референс: *M. zibellina*); С – Средние значения ошибки перекрестной

валидации (CV) (синий) и среднего парного сходства (красный) на основе глобальной ADMIXTURE (Референс: *M. zibellina*); D – Глобальный ADMIXTURE (левая половина) и локальный ADMIXTURE (правая половина), определенный в скользящих окнах размером 1 млн п.н. с шагом 100 тыс п.н. (Референс: *M. zibellina*). Образцы обозначены по оси X, а по оси Y – средняя вероятность происхождения.

Анализ HyDe подтвердил интрогрессию между соболями и лесными куницами, выявив значимые групповые ($Z\text{-score} = 224,47$, $\gamma \approx 0,19$) и индивидуальные значения γ от 0,02 (T81) до 0,74 (T18) (Приложение 28). Образцы T84 и T87 наиболее близки к соотношению 50/50 ($\gamma = 0,53$ и 0,47, соответственно), что подтверждает наибольшую близость этих образцов к гибриднему поколению F1. При этом образцы T78, T81 и T83 отличаются $\gamma < 0,05$, что согласуется с тем, что эти образцы имеют наименьшую долю примеси по результатам локального ADMIXTURE анализа.

Значения F3-, D- и F4-статистик согласуются с полученными результатами, подтверждая недавний поток генов. Значение статистики F3, рассчитанное для классифицированных групп особей, составило -0,163824 ($Z\text{-score} = -90,896$, $\text{stderr} = 0,001802$), а при анализе каждого образца (Приложение 29), T87 показал самый высокий уровень генетического смешения ($F3 = -0,21$, $Z\text{-score} = -45,03$). Образцы T81 и T83 показали самые высокие значения F3 (близкие к 0) с соответствующими высокими (близкими к 0) значениями $Z\text{-score}$ по сравнению с остальными образцов, подтверждая низкий уровень интрогрессии от другого вида. Результаты D- и F4 статистик между группами особей показали аналогичные результаты и подтвердили наличие потока генов как от соболей, так и от лесных куниц (Приложения 30). При тестировании на предмет потока генов из группы соболей образцы T18, T84 и T87 имели наиболее значимые значения D-статистики: 0,6211, 0,6014 и 0,5991, соответственно (Приложения 31). T18 имел самый высокий $Z\text{-score}$ 79,898, что указывает на наибольшее отклонение от нулевой гипотезы об отсутствии примеси от соболя. Статистика F4 (Приложения 31) подтвердила результаты D-статистики – те же три образца показали самые высокие значения F4: 0,053196 (T18), 0,043018 (T84) и 0,040847 (T87). T18 также отличается самым

высоким значением Z -score – 42,691. Образцы T81 и T83 также отличаются наименьшими значениями D - (0,13 и 0,19, соответственно) и F -статистик (0,003 и 0,004, соответственно), при этом согласно значениям Z -score лишь образец T81 отличается наименее выраженным паттерном гибридизации с низкой степенью достоверности, поскольку его значение $F4$ -статистики имеет Z -score равный 4,352.

После классификации образцов по видовой принадлежности мы решили вернуться к анализу гетерозиготности наших образцов. Ожидаемо, гетерозиготность у соболей оказалась статистически значимо выше, чем у лесной куницы (односторонний тест Манна-Уитни: p -value составило 0,038 и 0,047 для соболя и лесной куницы в качестве референса, соответственно). В ходе анализа распределений и сравнения образцов мы обнаружили, что из наших образцов можно выстроить ряд, отражающий переход от “чистых” лесных куниц к “чистым” соболям через гибридов (Рисунок 23А). При этом, компонента «чистых» лесных куниц уменьшается, а гибридная возрастает, достигая максимума у гибридов, близких к поколению F1 (например, T87), после чего доля «чистых» соболей увеличивается, сопровождаясь снижением гибридной компоненты. Отметим, что распределения гетерозиготных и гомозиготных SNP (Рисунок 23В и 23С) показывают, что геномные окна, принадлежащие к определенной компоненте, формируют длинные непрерывные участки, что подтверждается результатами локального ADMIXTURE-анализа (Рисунок 23D). Результаты локального ADMIXTURE-анализа для всех образцов соболей, лесных куниц и их гибридов в Приложении 32.

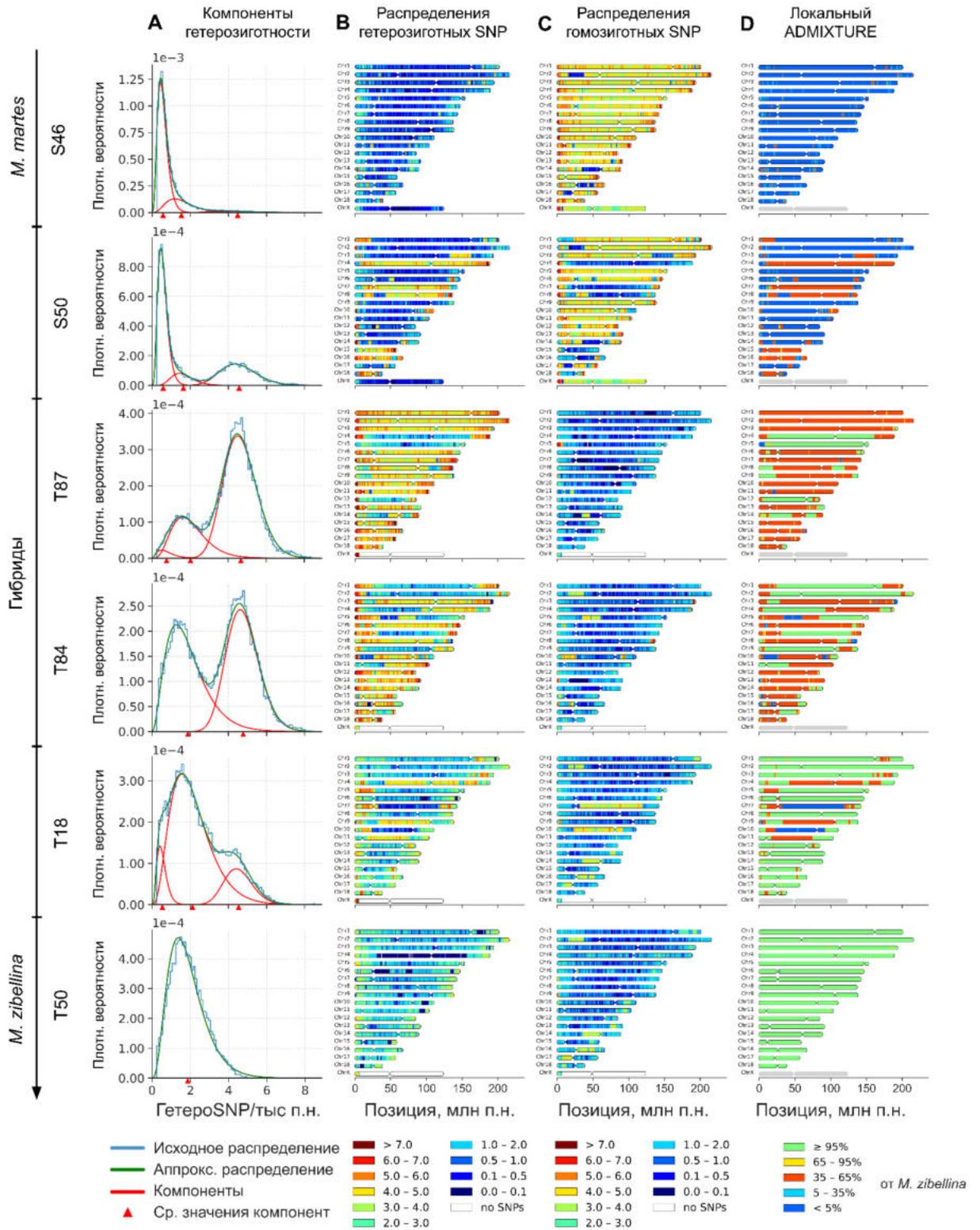


Рисунок 23. Межвидовой переход от *M. martes* к *M. zibellina*.

А – Компоненты распределений гетерозиготности (только SNP), В – тепловые карты средней гетерозиготности (гетерозиготные SNP) в скользящих окнах размером 1 млн п.н. с шагом в 100 тыс п.н. (Референс: *M. zibellina*), С – тепловые

карты средней гомозиготности(гомозиготные SNP) в скользящих окнах размером 1 млн п.н. с шагом в 100 тыс п.н. (Референс: *M. zibellina*), D – локальный ADMIXTURE.

3.8 Классификация по видовой принадлежности на основании полученных и общедоступных STR-локусов

Помимо классификации особей по видовой принадлежности на основании полногеномных данных, с целью сравнения с ранее опубликованными работами мы дополнительно выполнили классификацию на основании STR-локусов от разных наборов STR-маркеров (см. Методы “Анализ микросателлитных локусов”). Подробное описание финального набора из 36 STR-маркеров, используемых в анализе, представлено в Приложении 33. По результатам генотипирования, полный набор маркирует все хромосомы, кроме Chr11, Chr16 и ChrX (Рисунок 24А), со средней плотностью 1,9 STRs/Chr (STR-маркеры на хромосому). Подмножества из исследования Рожнова и др. 2013 и Каштанова и др. 2022 имеют значительно меньший охват: первое маркирует только 6 из 19 хромосом (0,4 STRs/Chr, Рисунок 24В), а второе – 11 из 19 (0,7 STRs/Chr, Рисунок 24С), соответственно.

Анализ ADMIXTURE показал разницу между наборами для классификации некоторых гибридов (Рисунок 24D). Например, в образце Т85 по полному набору STR было обнаружено 27,8% соболя, но по подмножествам Каштанова и др. 2022 и Рожнова и др. 2013 эти значения значительно ниже – 2,9% и 15,5%, соответственно. Более того, для образца Т84 разница между полным набором (63,4%) и подмножествами (95,8% и 93,1%) составила ~30%. Для остальных гибридов разница не столь значительна. При использовании лесной куницы в качестве референса мы получили аналогичные результаты (Приложение 34).

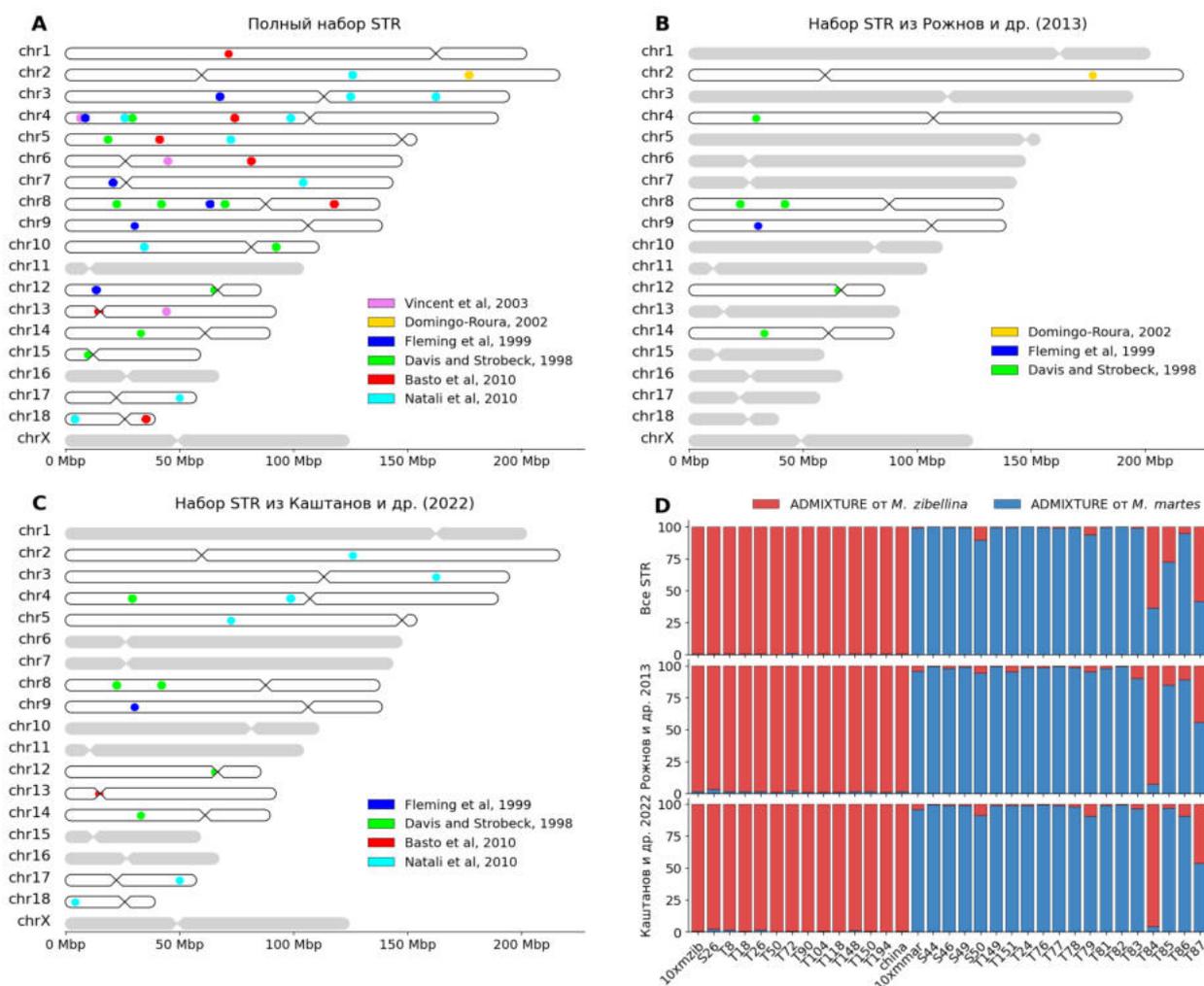


Рисунок 24. Локализация STR- локусов из разных наборов на хромосомах *M. zibellina* и соответствующий ADMIXTURE-анализ.

A – полный набор, включающий все STR-локусы; B – набор из исследования Каштанова и др. 2022; C – набор Рожнова и др. 2013. Референс: *M. zibellina*.

3.9 Анализ участков гомозиготности (RoH)

Анализ участков гомозиготности (RoH) показал различия в их кумулятивном распределении (Рисунок 25A) и доле генома (Рисунок 25B) среди исследованных групп образцов: соболей (*M. zibellina*), лесных куниц (*M. martes*) и гибридов. Среди соболей наблюдается значительная вариация в доле и спектре длин RoH. Например, кумулятивная длина RoH для референсного образца соболя (10xmzib) составила 681,3 млн п.н., что соответствует 30,5% генома, при этом большая часть приходится на ультра-длинные RoH ($UL \geq 10$ млрд п.н., 24,1%) и длинные ($10 \text{ млрд п.н.} > L \geq$

1 млрд п.н., 3,1%). Именно у этого образца мы обнаружили наибольшую долю ультра-длинных RoH, в то время как у остальных соболей доля таких участков варьирует от 1,2% (T104) до 8% (S26). Образец T72 (Камчатский соболь) выделяется наибольшим числом коротких RoH (9,3%) и наименьшей долей длинных (6,1%) и ультра-длинных RoH (2,2%). Также, у этого образца мы выявили наибольшее среди соболей суммарное количество RoH (769). Однако, они охватывают “лишь” 17,6% генома. Как следствие, траектория кумулятивного распределения RoH для образца T72 значительно отличается от других соболей и напоминает траектории лесных куниц (Рисунок 25А). Для остальных образцов соболей (за исключением 10xmzib и T72) наблюдается незначительная вариация в доле коротких (от 1,2% в T118 до 3,7% в S26) и длинных RoH (от 1,5% в T26 до 5,8% в T148).

Для референсного образца лесной куницы (10xmmap) кумулятивная длина RoH составила 476,06 млн п.н. (21,32%). При этом доля ультра-длинных RoH примерно вдвое меньше (11,8%) чем у референсного соболя. У лесных куниц участки RoH в целом характеризуются меньшей протяженностью по сравнению с соболями. Например, доля ультра-длинных RoH не превышает от 0,8% (T24) до 4,7% (T77), за исключением референсного образца 10xmmap (11,8%) и образца S46, у которого ультра-длинных RoH отсутствуют. Короткие RoH занимают от 5,5% (10xmmap) до 7,4% (T76) генома, а длинные RoH – от 4,9% (S44) до 6,7% (T24).

Гибридные образцы продемонстрировали широкий диапазон долей RoH. Наиболее низкие кумулятивные значения, как и ожидалось, отмечены у гибридных образцов, близких к поколению F1. Так, для образца T87 отсутствуют ультра-длинных RoH, а доли длинных и коротких RoH составляют 0,5% и 1,2%, соответственно. Отметим, что ультра-длинных RoH также отсутствуют в образцах T84-T86 и S50, а оставшиеся гибридные образцы характеризуются долей от 0,5% (T18 и T78) до 2,4% (T79). Наибольшие доли ультра-длинных RoH 8,3% и 5,6% характерны для образцов T151 и T83, соответственно. Сопоставимые значения наблюдаются и при использовании сборки генома лесной куницы в качестве референса (Приложение 35). Участки RoH для исследуемых образцов соболя и

лесной куницы относительно обоих референсов были также визуализированы в виде тепловых карт (Приложение 36).

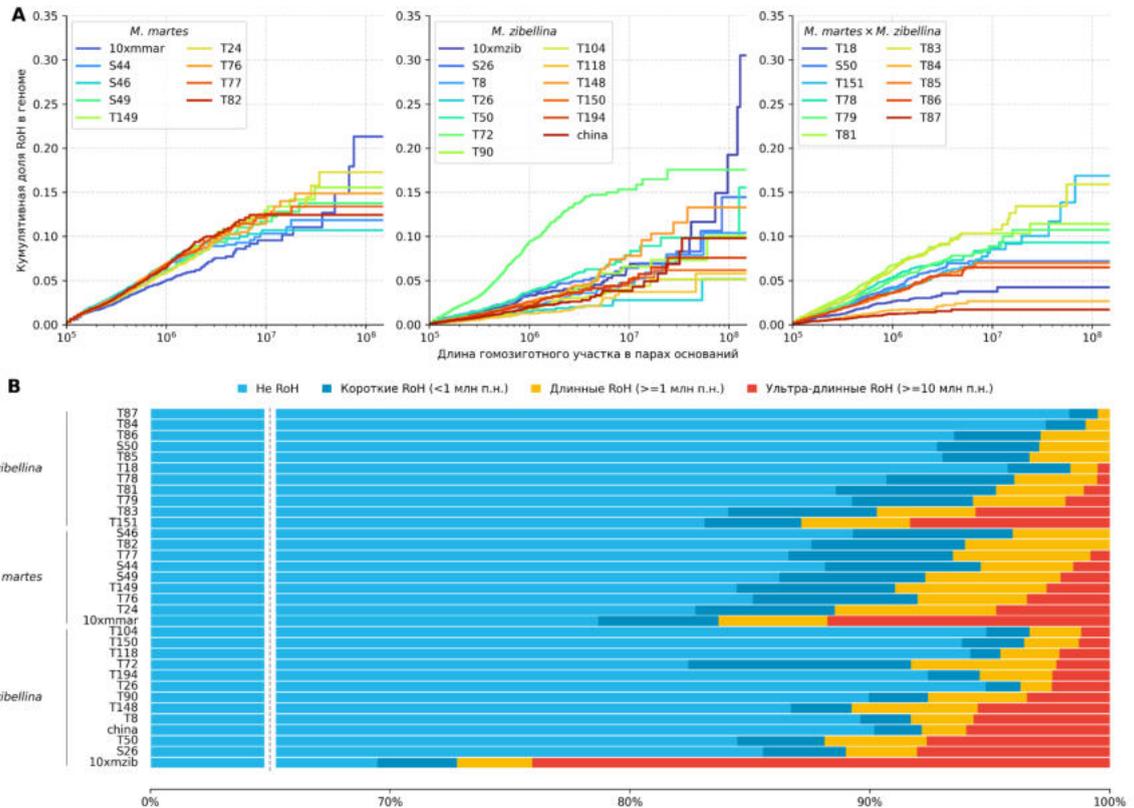


Рисунок 25. Анализ RoH соболей, лесных куниц и их гибридов.

А – Кумулятивные распределения участков с пониженным генетическим разнообразием (RoH) для образцов (слева направо) лесных куниц, соболей и их гибридов с интрогрессией более 5%; В – Доля генома в RoH. В качестве референса использована геномная сборка соболя (*M. zibellina*), хромосома X исключена для всех образцов.

Анализ участков RoH для образцов каменной куницы, харзы, американской куницы и тихоокеанской куницы выявил значительные различия как на межвидовом, так и на внутривидовом (только для *M. caurina*) уровне (Рисунок 26). Для референсного образца каменной куницы (10xmfoi) кумулятивная доля генома в RoH составила 51,7%, из которых 24,7% составляют короткие RoH, 15,6% — длинные RoH и 11,4% — ультра-длинные RoH. Харза (10xmfla) показала схожий общий уровень гомозиготности (55,1% генома в RoH), но с более высокой долей ультра-длинных RoH (24,2%). Американская куница (SRR11575352) в данном

случае отличается самым высоким уровнем гетерозиготности: доля генома в RoH составила всего 13,1% с преобладающим количеством коротких RoH (6,8%) над длинными и ультра-длинными RoH (2,8% и 3,5%, соответственно). Среди образцов тихоокеанской куницы особенно выражены различия между островными и материковыми особями. У изолированных островных образцов SRR11575348 и SRR11575343 наблюдается крайне высокая доля генома в RoH: у SRR11575348 почти весь геном представлен ультра-длинными RoH (99,4%), а у SRR11575343 кумулятивная доля RoH составляет 77,9%, которую составляют 11,1% коротких, 45,3% длинных, 21,6% ультра-длинных RoH. У образцов с материка (SRR11575346 и SRR11575351) уровень гомозиготности значительно ниже: доля генома в RoH составила 19,6% и 23,5% соответственно, при этом ультра-длинные RoH для обоих образцов занимают менее 3% генома.

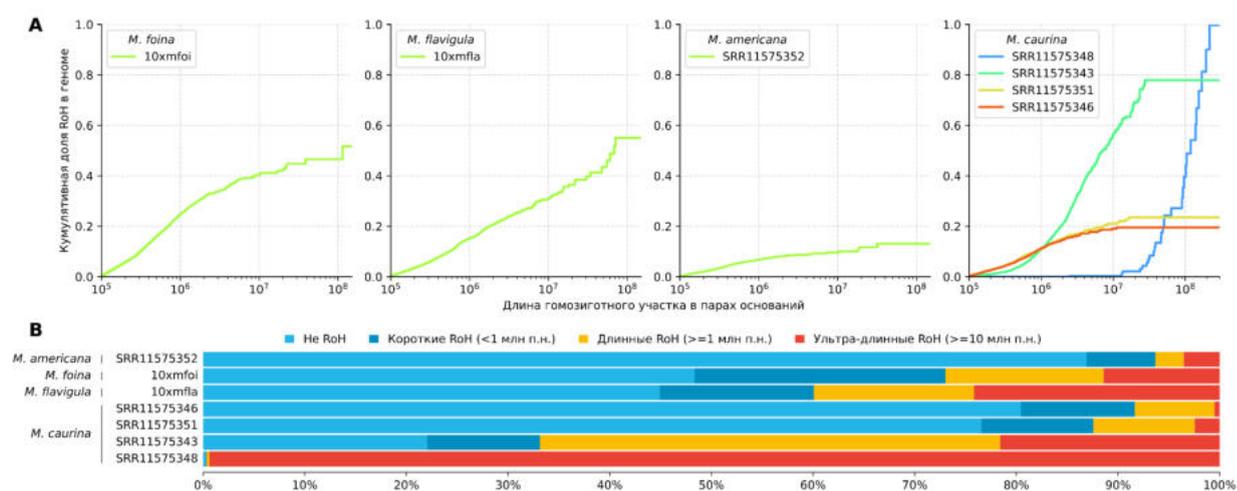


Рисунок 26. Анализ RoH американской куницы, каменной куницы, харзы и тихоокеанских куниц.

А – Кумулятивные распределения участков с пониженным генетическим разнообразием (RoH) для образцов (слева направо) *M. foinea*, *M. flavigula*, *M. americana* и *M. caurina*; В – Доля генома в RoH. В качестве референса для *M. americana* и *M. caurina* использована геномная сборка *M. zibellina*. Для *M. foinea* и *M. flavigula* использованы сборки геномов соответствующих видов. Хромосома X исключена для всех образцов.

3.10 Филогения рода *Martes*

В ходе анализа в геномных сборках исследуемых видов мы выявили 6535 общих последовательностей ортологичных генов. В сравнении с другими геномными сборками в сборке харзы наибольшее количество пропущенных генов – 458 (4,3%), однако это не помешало проведению филогеномного анализа. Полученная геномная сборка каменной куницы выгодно отличается от общедоступной, поскольку имеет меньше пропущенных генов (2,8% против 3,8%) и больше однокопийных (95,6% против 94,5%). Наименьшее количество пропущенных генов в опубликованной геномной сборке *Gulo gulo* – 222 или 2,4% от общего количества (Рисунок 26).



Рисунок 26. Диаграмма на основании абсолютных значений оценки качества по консервативным ортологичным группам (BUSCOs) как полученных (выделены жирным), так и общедоступных геномных сборок, используемых далее для реконструкции филогении.

Суммарная длина последовательностей в получившемся кодоновом выравнивании после фильтрации составила 10'902'231 нуклеотид. Построение филогенетических деревьев выполняли с использованием подхода объединенной матрицы методами ML и BI (Рисунок 27А и 27В), а также с использованием обобщенно-коалесцентного подхода (Рисунок 27). Все полученные

филогенетические деревья показали идентичную топологию. Харза отделилась от остальных представителей рода *Martes* практически сразу после расхождения с линией россомахи (*Gulo gulo*). При этом на дереве, полученном с использованием обобщенно-коалесцентного подхода (Рисунок 28), узел с харзой имеет долю деревьев лишь 31,19%, то есть только 2119 из 6535 генных деревьев содержат этот узел. Апостериорная вероятность всех узлов при этом составила 100%. Стоит отметить, что барсук (*Meles meles*) на всех полученных деревьях оказался в одной кладе с представителями подродов *Mustelinae* и *Lutrinae*, что, возможно, связано с эффектом “притяжения длинных ветвей”.

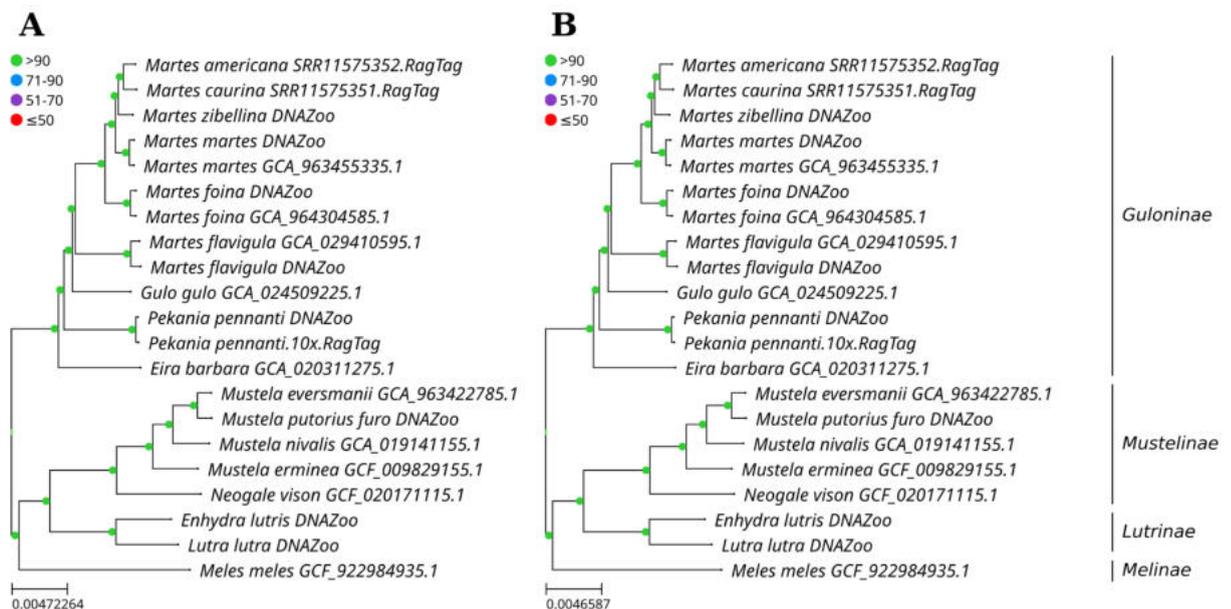


Рисунок 27. Реконструкция филогенетического дерева Guloninae и Mustelinae на основе полногеномных данных.

А – дерево построенное методом ML; В – методом BI, соответственно. Поддержка каждого узла показана цветом: зеленый = 100. Используемая в качестве аутгруппы сборка волка *Canis lupus* (GCA_905319855.2) удалена с дерева.

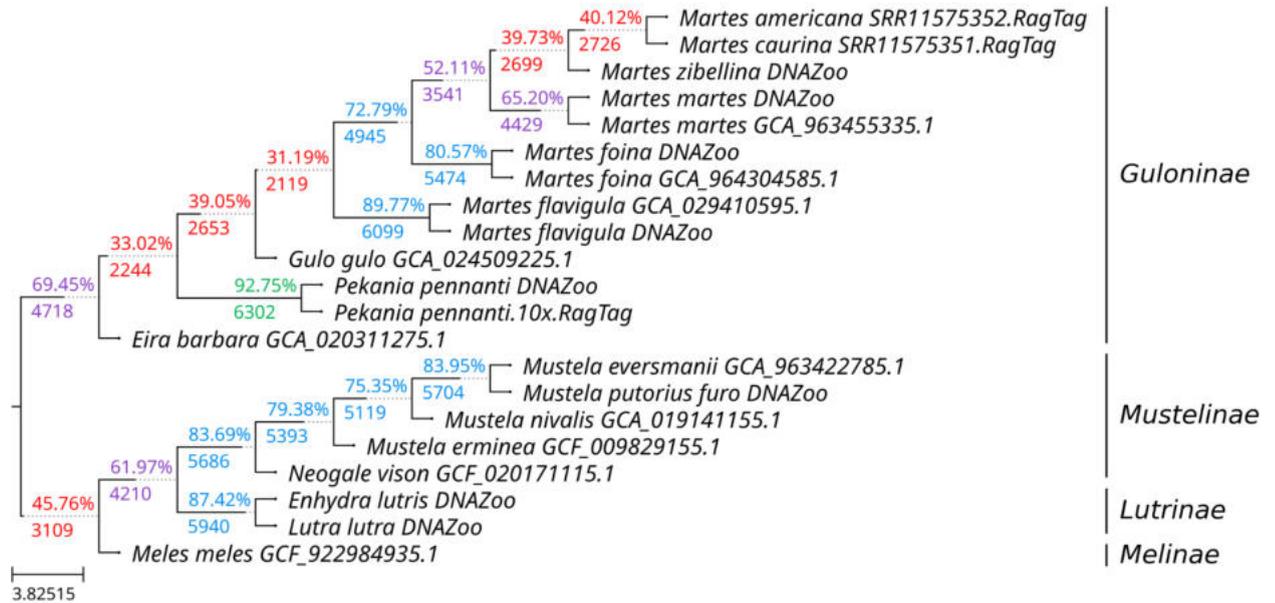


Рисунок 28. Филогенетическое дерево, построенное с использованием обобщенно-коалесцентного подхода.

Доля деревьев, где присутствует конкретный узел, показана цветом: зеленым – > 90%; синим – > 70%; фиолетовым – > 50% и красным менее 50% генных деревьев имеют этот узел. Используемая в качестве аутгруппы сборка волка *Canis lupus* (GCA_905319855.2) удалена с дерева.

Дополнительно мы выполнили реконструкцию филогении методом ML на основании 308 общедоступных мтДНК видов рода *Martes*, с добавлением митохондриальных геномов, полученных в ходе исследования (Рисунок 29, Приложение 37). Полученная на основании мтДНК топология видов противоречит топологии, полученной по данным яДНК (выше), в случае которой *M. zibellina* является базальным таксоном по отношению к *M. americana* и *M. caurina* (Рисунки 27 и 28), тогда как на митохондриальном дереве базальной группой к *M. zibellina*, *M. martes* и *M. melampus* являются североамериканские куницы (Рисунок 29).

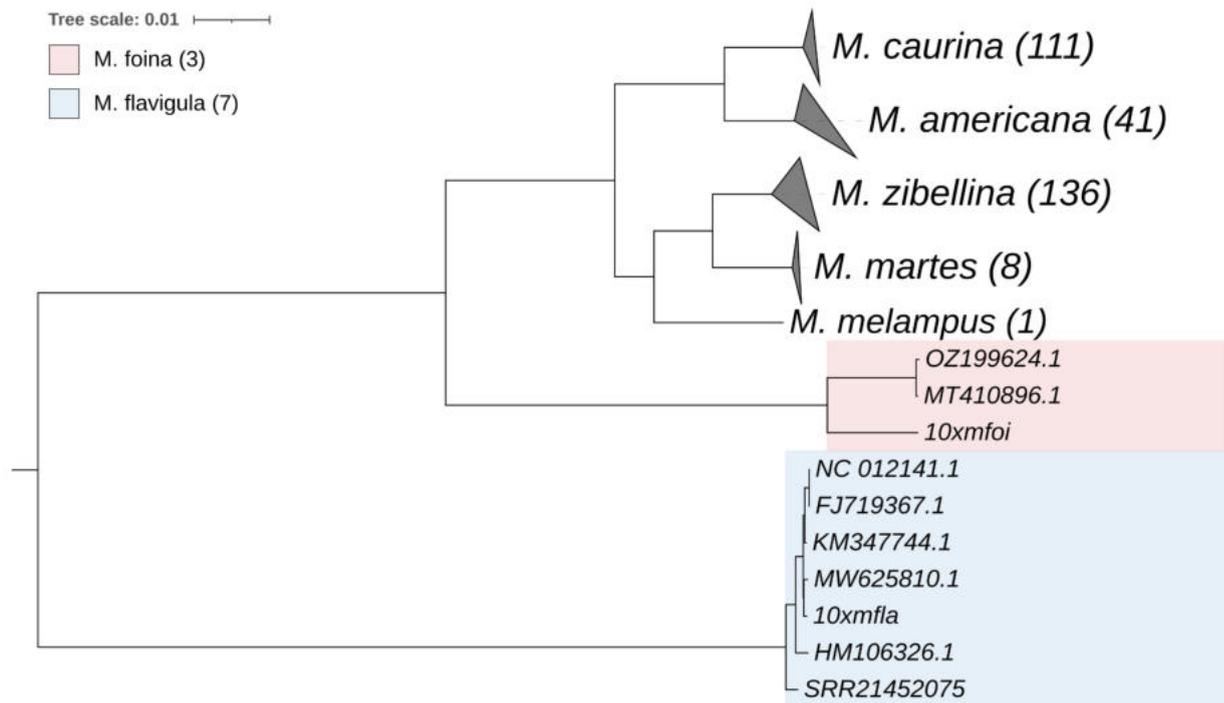


Рисунок 29. Филогенетическое дерево рода *Martes* на основании мтДНК.

В скобках указано количество мтДНК каждого вида. В качестве аутгруппы (удалена с дерева) использован митохондриальный геном россомахи *Gulo gulo* (NC_009685.1).

3.11 Реконструкция филогении соболей и лесных куниц на основе мтДНК

На основании 104 общедоступных и 36 полученных в ходе работы полных последовательностей мтДНК соболя и лесной куницы мы реконструировали филогенетическое дерево (Рисунок 30) и сеть гаплотипов (Рисунок 31). Полученное филогенетическое дерево демонстрирует топологию, в которой 136 гаплотипов дивергировали на 3 основные клады (А, В и С) с максимальной поддержкой (100), а также выделилась клада с 4 мтДНК лесных куниц (аутгруппа). Аналогичные филогенетические отношения также поддерживаются и при построении сети гаплотипов, разделение между А и С кладами на которой составляет 83 нуклеотида, тогда как между В и С – 64. Используемые при построении 136 гаплотипов образовали 98 гаплогрупп.

Ранее изучение филогении этих видов было выполнено в работе Li et al. 2021 [321], где провели реконструкцию сети гаплотипов и филогенетического дерева и

отметили, что "только гаплотипы в пределах трех групп (A2, B2, C2) всегда отдельно кластеризуются вместе и образуют три субклады в филогенетическом анализе с сильной поддержкой". Полученная на нашем случае часть второстепенных клад являются немонафилетичными. Основываясь на наших результатах, мы считаем, что каждая из гаплогрупп B1 и C1 должна быть разделена на две меньшие монафилетические гаплогруппы, B3 и B4, и C3 и C4, соответственно, а гаплогруппа A1 должна быть разделена на 3 части: A3, A4 и A5.

Интересно, что мтДНК соболя обнаружена в 7 из 9 образцов "чистых" лесных куниц (включая референсный 10хmmar) за исключением S44 и S46. Важно отметить, что в "чистых" куницах обнаружена мтДНК соболей из разных клад: для T24 мтДНК из клады A5; для S49 и T77 – из клады A5; для 10хmmar, T149 и T76 – из клады B2 и для T82 из C4. Большинство образцов гибридов также имеют мтДНК от различных подвидов соболя, за исключением T78 с мтДНК от лесной куницы.

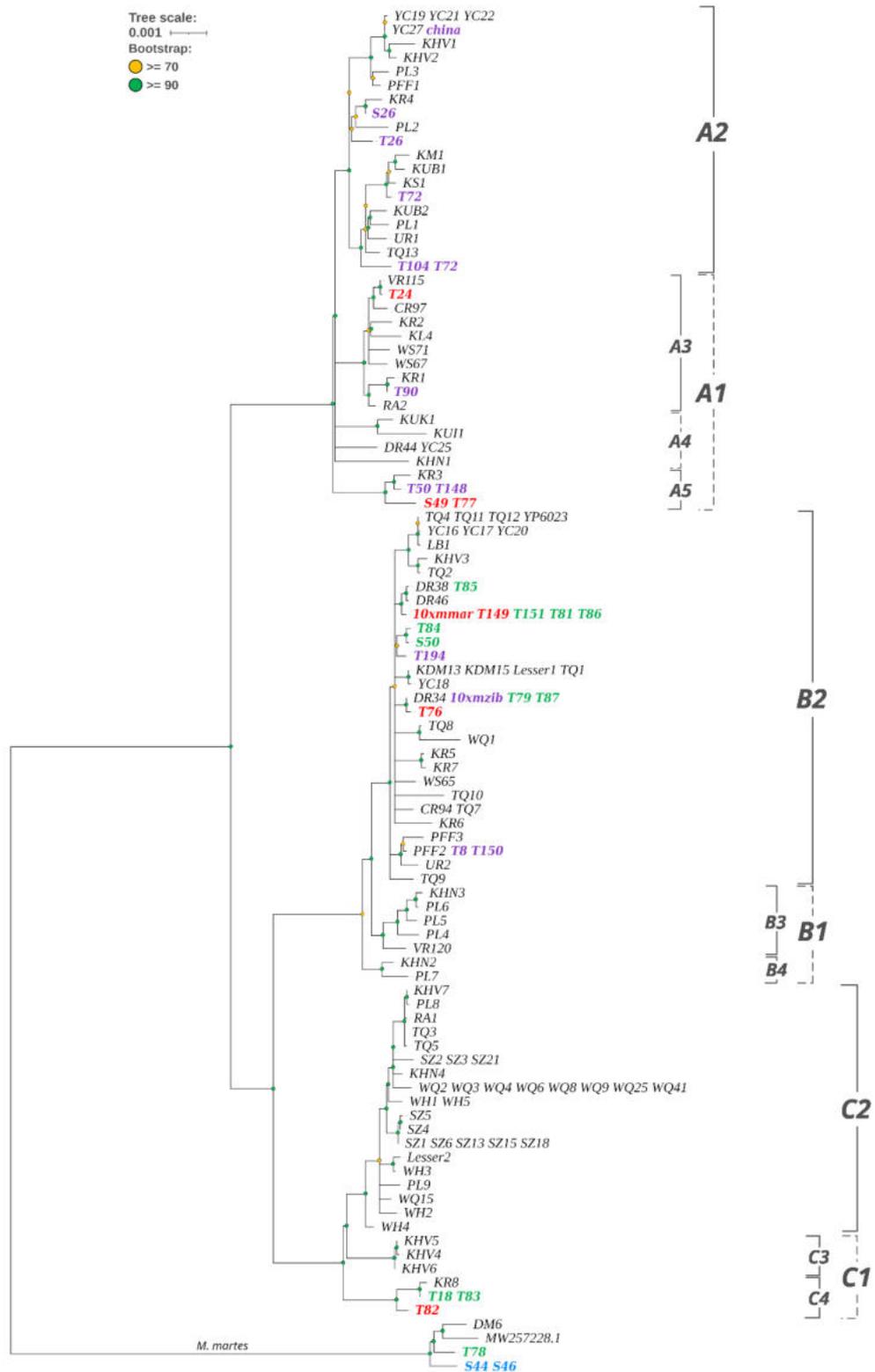


Рисунок 30. Филогенетическое дерево на основе мтДНК соболя, лесной куницы и особей из зоны симпатрии.

Узлы с поддержкой более 90 отмечены зеленым, а узлы от 70 до 90 – желтым цветом. Узлы со значениями поддержки менее 70 были удалены. Выделены 3

основные клады (А, В и С), включающие субклады (А1-5, В1-4 и С1-4). Исследуемые образцы показаны цветом: образцы *M. zibellina* с мтДНК от *M. zibellina* – фиолетовым; образцы *M. martes* с мтДНК от *M. martes* – синим; образцы *M. martes* с мтДНК от *M. zibellina* – красным; гибриды с мтДНК от *M. zibellina* – зеленым.

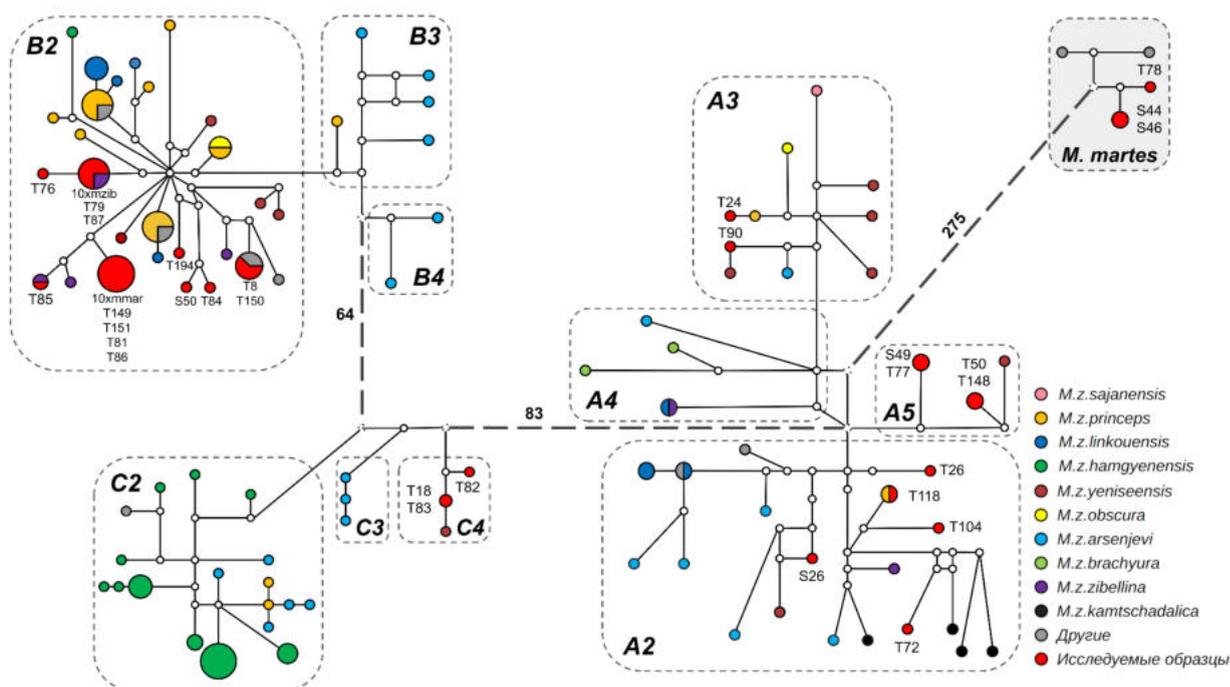


Рисунок 31. Сеть гаплотипов на основе последовательностей мтДНК соболя и лесной куницы.

Размеры кружков пропорциональны частоте гаплотипов, а расстояния между гаплотипами – примерно соответствуют количеству замен между ними. Подвиды отмечены разными цветами. Число замен между основными кладами (А, В и С) отмечено цифрой.

3.12 Внутри- и межвидовые генетические расстояния между видами *Martes*

Мы вычислили попарные генетические расстояния для ядерных геномов с учетом гетерозиготных SNP (Рисунок 32А), а также для 308 последовательностей мтДНК рода *Martes*, включая как полученные в ходе работы, так и опубликованные на момент исследования (Рисунок 32В). Важно отметить, что для тех же пар особей *M. foina* и *M. flavigula*, для которых доступны сборки ядерных геномов, также были

использованы мтДНК. Для *M. foina* из п. Лазе, Словения (GCA_964304585.1) мтДНК уже была доступна – OZ199624, тогда как для *M. flavigula* из п. Сычуань, Китай (GCA_029410595.1) мтДНК была собрана нами в ходе работы.

Расстояния между мтДНК каменных куниц (Рисунок 32В) распадаются на две группы: нижнюю (между образцами из Словении и Дании) и верхнюю (между образцом из провинции Ганьсу Китая и европейскими). Расстояния в верхней группе значительно превосходят внутривидовые расстояния других видов и сопоставимы с межвидовыми, например, с *M. zibellina* - *M. martes* и *M. americana* - *M. caurina*. Однако, расстояния между ядерными геномами *M. foina* (Китай и Словения) также больше внутривидовых, но все же существенно ближе к ним, чем к межвидовым (Рисунок 32А).

Для харз мы получили противоположный паттерн. Генетические расстояния между харзами (Дальний Восток, Россия и п. Сычуань, юго-западный Китай) близки с межвидовыми расстояниями *M. americana* - *M. caurina* (Рисунок 32А). При этом расстояние между мтДНК этих же особей близко к расстояниям между митохондриальными геномами других харз (Южная Корея, п. Хэйлунцзян, северо-восточный Китай) и сопоставимо с внутривидовыми расстояниями других видов.

Примечательно, что в ходе анализа выяснилось, что шесть опубликованных мтДНК *M. americana* (NC_020642.1, MK320894.1, MK320876.1, MK320867.1, MK320849.1, NM106324.1) на самом деле являются *M. caurina*.

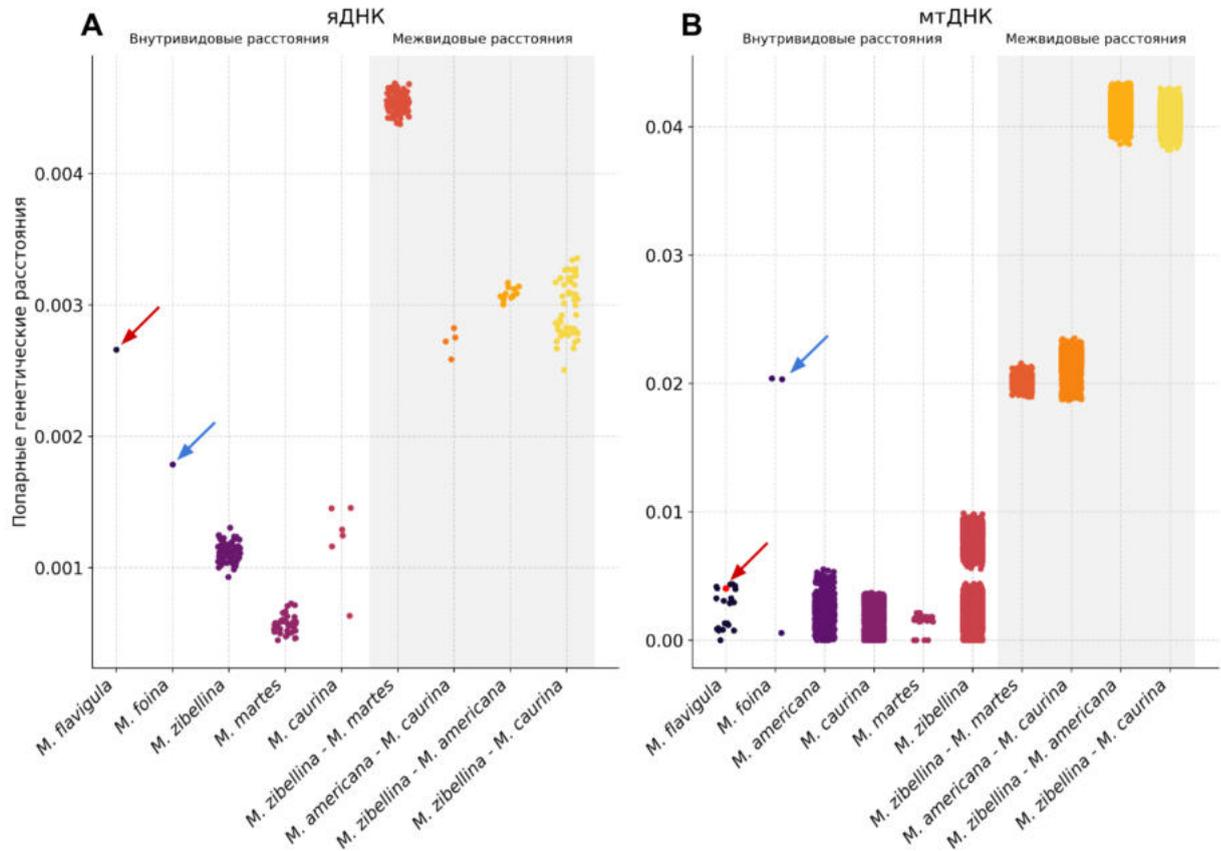


Рисунок 32. Внутри- и межвидовые генетические расстояния по ядерным (ядНК) (А) и митохондриальным (мтДНК) данным (В).

Стрелками обозначены расстояния между одними и теми же особями, красной – для харзы, синей – каменной куницы, соответственно.

3.13 Анализ популяционной истории

Реконструкцию популяционной истории мы выполнили при помощи PSMC с использованием опорного значения частоты замен на поколение (μ) равного $4,64 \times 10^{-9}$ ($2,94 \times 10^{-9}$; $7,37 \times 10^{-9}$ – значения μ для поздней и ранней границы доверительного интервала, соответственно). В случае образцов соболей, лесных куниц и их гибридов, траектории популяционных историй разделились на 3 различающиеся группы (Рисунок 33). Для траекторий популяционной истории соболей (Рисунок 33А) наблюдаются колебания эффективного размера численности популяции (N_e) между тремя небольшими пиками: $\sim 2,5$ (3,94; 1,57), $\sim 0,5$ (0,79; 0,31) и $\sim 0,1$ (0,15; 0,06) млн лет назад. При этом, наиболее значительные колебания между образцами в основном наблюдаются позднее ~ 100 (158; 63) тыс лет назад,

когда траектории отдельных соболей начинают расходиться. Популяционная история лесных куниц сильно отличается от соболиной (Рисунок 33В): начиная с ~ 5 (7,89; 3,15) млн лет назад их численность росла и достигала максимума $\sim 1,25$ (1,97; 0,79) млн лет назад, после чего последовало резкое снижение. Результатом этого “бутылочного горлышка” стало пятикратное снижение численности $\sim 0,7$ (1,1; 0,44) млн лет назад. И хотя после этого значения N_e вышли на плато, тренд к снижению сохранился и уже ~ 50 (78; 31) тыс лет назад численность упала еще вдвое. Отметим, что ~ 125 (197; 79) тыс лет назад было небольшое возрастание численности, однако это не переломило основной тренд снижения.

Группа гибридов отличается наиболее разнообразными траекториями популяционных историй (Рисунок 33С), разительно отличаясь от “чистых” родительских видов, демонстрируя характерный “популяционный взрыв” $\sim 1,5$ (2,37; 0,94) млн лет назад. Как оказалось, высота пика этого “популяционного взрыва” сильно коррелирует (τ Кендалла = 0,67, p-value = 0,003; ρ Спирмена = 0,76, p-value = 0,006; r Пирсона = 0,778, p-value = 0,005) с уровнем интрогрессии в геномах гибридов (Приложение 38). Примечательно, что до 3,5 (5,52; 2,20) млн лет назад траектории гибридов оставались очень схожими как между собой, так и с траекториями родительских видов.

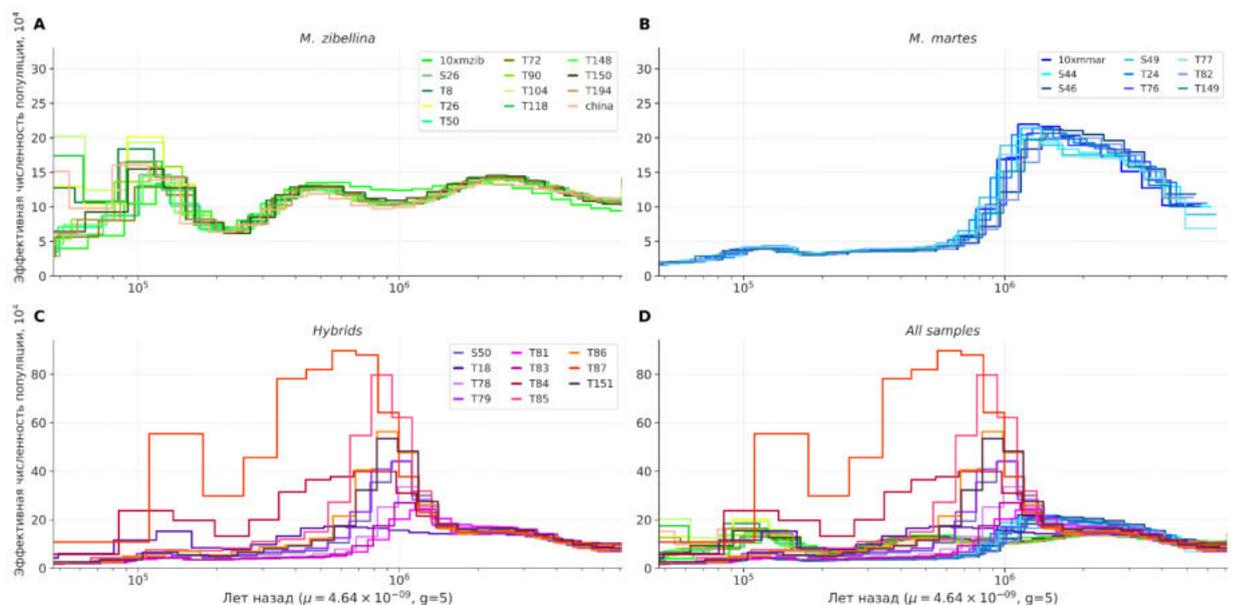


Рисунок 33. Реконструкция популяционной истории образцов: А – соболей, В – лесных куниц, С – их гибридов, D – траектории для всех образцов.

Частота замен на поколение (μ) = $4,64 \times 10^{-9}$. Время генерации = 5 для всех графиков. В качестве референса использована геномная сборка соболя.

Для сравнительного анализа динамики численности популяций исследуемых видов рода *Martes* мы восстановили траектории популяционной истории каменной куницы, харзы, американской куницы и тихоокеанской куницы (Рисунок 34). Траектория каменной куницы схожа с соболиной (за исключением самого начала), но характеризуется меньшим значением N_e , при этом сильно отличаясь от траектории лесной куницы. Отметим, что она имеет два пика примерно одинаковой высоты $\sim 0,55$ (0,87; 0,35) млн лет назад и ~ 150 (237; 94) тыс лет назад, разделенных промежутком с примерно 30%-ным снижением N_e . Траектория популяционной истории харзы, в сравнении с другими исследуемыми видами рода *Martes*, в целом демонстрирует рост и не характеризуется сокращениями численности в прошлом. Так, с ~ 5 (7,89; 3,15) до ~ 1 (1,57; 0,63) млн лет назад происходил рост, далее наблюдалось плато с сохранением возрастающего тренда, и после – снова возрастание N_e в период с ~ 250 (395; 157) до ~ 150 (237; 94) тыс лет назад. Наконец, траектории Североамериканских куниц схожи между собой, отличаясь при этом от траекторий евразийских видов. Для обоих видов наблюдается обоюдное снижение N_e до $\sim 0,8$ (1,26; 0,5) млн лет назад. После этого, N_e американской куницы плавно возрастает, достигая пика ~ 125 (1,97; 0,79) тыс лет назад, аналогично евразийским видам. В свою очередь, траектория тихоокеанской куницы продолжила плавное снижение, с незначительным возрастанием N_e лишь позднее ~ 90 (142; 57) тыс лет назад. Интересно, что позднее ~ 150 (237; 94) тыс лет назад все исследуемые виды, за исключением тихоокеанской куницы, характеризуются значительным снижением N_e до менее чем 100 тыс особей.

Шлю

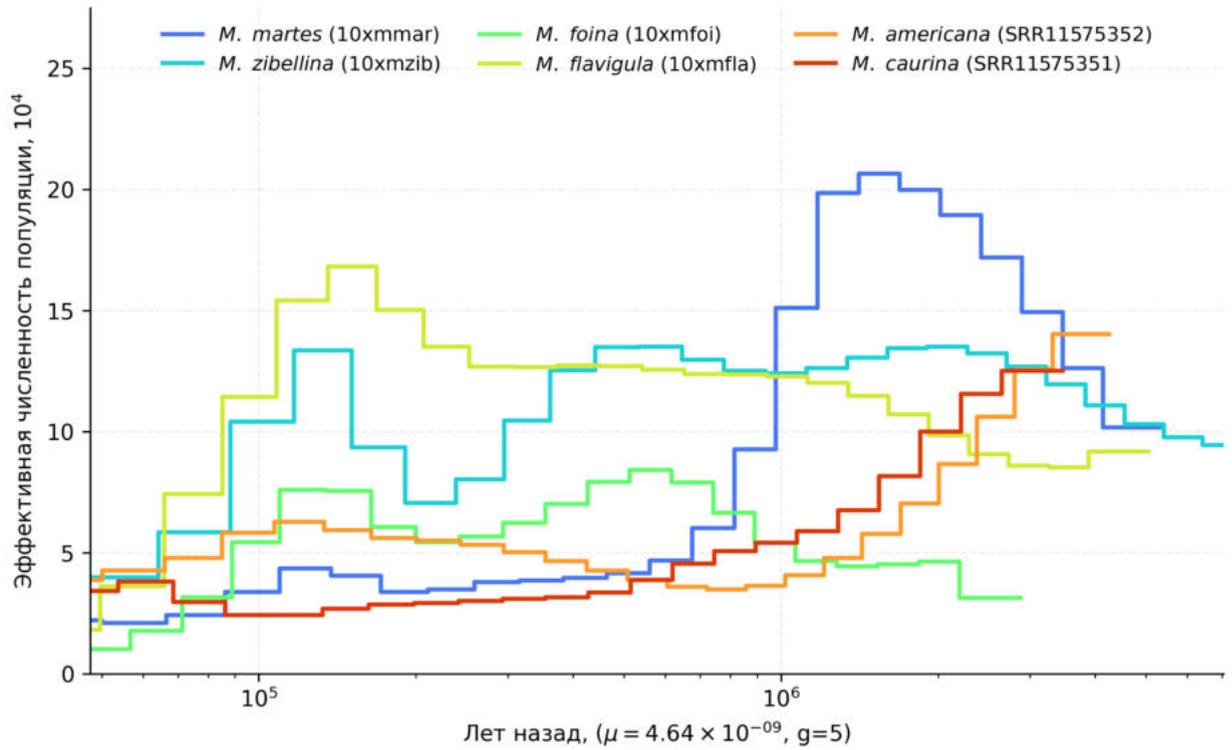


Рисунок 34. Реконструкция популяционной истории референсных образцов соболя, лесной куницы, каменной куницы, харзы, американской куницы и тихоокеанской куницы.

Частота замен на поколение (μ) = $4,64 \times 10^{-9}$. Время генерации $g = 5$ лет для всех видов. В качестве референсов использованы соответствующие видам геномные сборки.

4 Обсуждение

4.1 Геномные сборки и синтения

Геномные сборки хромосомного уровня обеспечивают высокое качество представления генома, позволяя включать его основную часть в длинные хромосомные скаффолды, соответствующие целым хромосомам. Такие сборки в настоящее время уже стали стандартом как в популяционной геномике, так и в природоохранной биологии [66]. В рамках данной работы мы получили геномные сборки хромосомного уровня для 4-х видов рода *Martes*: соболя, лесной куницы, каменной куницы и харзы. Оценки качества полученных геномныхборок с использованием метрик BUSCO сопоставимы с полученными ранее сборками геномов других хищных [105,106,457], что подтверждает высокий уровень представленности и целостности консервативных ортологов. В ходе аннотация геномныхборок с использованием общедоступных транскриптомных данных нам удалось предсказать более 20 тыс белок-кодирующих генов для каждого вида. Последующая оценка предсказания на основе BUSCO подтвердила высокое качество и полноту аннотации. Сборка хромосомного уровня для соболя получена впервые, а сборки геномов лесной куницы и каменной куницы превосходят общедоступные сборки согласно полученным оценкам BUSCO. Геномная сборка харзы хотя и уступает по качеству опубликованной сборке этого вида, представляет ценный ресурс для дальнейших исследований.

Как и ожидалось, PAR четырех исследуемых видов рода *Martes* расположен на конце X-хромосомы и имеет обычную длину ~6 млн п.н., характерную для хищных [66,105]. Различия в длине PAR, а также то, что он начинается не от самого конца хромосомы, вызвано трудностями точного определения границ PAR и смещением, связанным с референсной сборкой генома и покрытием. Так, вариация размера PAR в случае *M. saurina* вызвана неравномерностью в покрытии всего генома, в том числе и хромосомы X. Хотя используемый алгоритм поиска PAR хорошо зарекомендовал себя [66], важно учитывать, что точность определения координат напрямую зависит от величины и равномерности покрытия в районе PAR и всего генома.

Анализ синтении между каменной куницей домашней кошкой, домашней собакой и человеком (Рисунок 18), а также между всеми исследуемыми видами рода *Martes* (Рисунок 19), показал практически полное соответствие полученных результатов с опубликованными картами сравнительного хромосомного окрашивания [34,63,265]. Незначительные противоречия, выявленные в ходе исследования, скорее всего связаны с ограниченной разрешающей способностью цитогенетических подходов (удается достоверно различать перестройки протяженностью не менее 5 млн п.н.) [65]. Анализ синтении исследуемых видов рода *Martes* позволил нам обнаружить 4 видоспецифичных инверсии, размер которых также находится на пределе возможностей цитогенетических методов. Однако только инверсия на хромосоме 11 между соболем и лесной куницей была подтверждена нашими коллегами на основе сравнения G-бэндинга кариотипов. Остальные инверсии вполне могут быть артефактами геномных сборок, что объясняется распространенностью ошибок ориентации при скаффолдинге с использованием Hi-C данных.

4.2 Особенности образцов соболей, лесных куниц и их гибридов

Наш набор включал 33 полногеномных образца (Таблицы 1-3), 32 из которых были секвенированы в рамках данной работы. До начала анализа, 14 из них были описаны как соболя, 7 – как лесные куницы и 12 – как особи из зоны симпатрии с гибридным фенотипом (кидасы), соответственно. После реклассификации (см. разделы ниже) мы получили 9 чистых лесных куниц, 13 чистых соболей и 11 гибридных особей с интрогрессией более 5%. Таким образом, итоговые группы “чистых” родительских видов имели сопоставимые размеры.

Для соболя мы секвенировали по крайней мере одну особь из крупных популяций, охватив большую часть ареала от Ханты-Мансийского округа на западе до полуострова Камчатка на востоке. Для лесной куницы было доступно значительно меньше образцов, которые ограничивались восточной частью ее ареала. Образцы гибридов были получены в основном (8 из 11, за исключением образцов T18, T151, S50) из одного места (Тюменская область, Уватский район, д. Малый Нарыс) в зоне симпатрии. Гибридный образец T18, изначально

классифицированный как “чистый” соболь, получен из другого места также в зоне симпатрии (Ханты-Мансийский автономный округ, с. Перегребное). T151 получен в самой восточной части ареала лесной куницы (Новосибирская область, г. Куйбышев), но вне зоны перекрывания с ареалом соболя согласно данным МСОП (2024-02) [244,246]. Происхождение последнего гибридного образца (S50) загадочно. Он является самым западным из всех образцов и был собран в глубине ареала лесной куницы (Калужская область, г. Калуга). Учитывая высокую долю интрогрессии от соболя (16,47%), эта особь может считаться потомком (по крайней мере, второго или высшего поколения или возвратного скрещивания), что не характерно для столь отделенных от зоны симпатрии регионов. Эта особь может быть либо доказательством исторически бóльшего ареала соболя, либо может быть одним из потомков соболей, сбежавших со звероферм. Ранее сообщалось, что такие беглецы добирались даже до Финляндии, в связи с чем соболь рассматривался там как вид с инвазивными признаками [459]. Ареалы соболя и лесной куницы были оценены МСОП в 2015 году и, вероятно, уже устарели, поскольку в последние годы сообщается, что соболь увеличил численность и встречается за пределами ареала, по крайней мере, за северной [311] и юго-западной границей [312]. Это подтверждает возможность естественного происхождения и неуникальность гибридного образца S50.

4.3 Концепция анализа компонентов распределения гетерозиготности

На среднюю и медианную гетерозиготность существенно влияют RoH и интрогрессия, в связи с чем они дают лишь суммарную и порой смещенную оценку. Обычный способ преодолеть эту проблему – использовать подход с геномными окнами, который позволяет получить распределение оценок в каждом из окон вместо одного значения (см. раздел “*Гетерозиготность, как одна из ключевых метрик генетического разнообразия*” в Обзоре литературы) (Рисунки 20 и 23). В случае использования геномной сборки хромосомного уровня в качестве референса, такой подход позволяет изучать распределение гетерозиготности по отдельным хромосомам (Рисунок 23В, Приложение 26).

Мы решили пойти дальше и представить концепцию нового метода – анализа компонентов распределения гетерозиготности. Вдохновением для него послужили распределения гетерозиготности наших гибридных образцов и инструмент Genomescope2 [400], в котором реализован похожий подход, но для другой цели (оценка размера генома). В основе нашего метода лежит аппроксимация линейной комбинации отрицательных биномиальных распределений к эмпирическому распределению гетерозиготности (см. раздел “Анализ распределений гетерозиготности” в Материалах и методах).

Для всех исследуемых образцов соболей, лесных куниц и их гибридов мы успешно разложили распределения гетерозиготности на комбинацию отдельных компонент: чистых лесных куниц (P), чистых соболей (S) и гибридов (H) (Рисунок 35, Рисунок 23А). Среднее значение P-компоненты было очень близким среди всех содержащих его образцов (0,536 - 0,763, $\sigma = 0,047$), в то время как для чистых лесных куниц (0,559 - 0,576, $\sigma = 0,008$) было близко к глобальным медианным оценкам гетерозиготности (0,56 - 0,67, оранжевые столбцы гистограммы на Рисунке 35). Однако заметное различие со средними глобальными оценками (0,84 - 1,07, светло-оранжевые столбцы на Рисунке 35) легко объяснить небольшой интрогрессией от соболя в чистые образцы лесных куниц. Из-за высоких значений H-компоненты (4,17 - 4,51, $\sigma = 0,11$) и низких значений P-компоненты (зеленые вертикальные черточки) глобальные средние значения гетерозиготности были заметно смещены в сторону увеличения за счет присутствия небольшого числа “гибридных” окон (Рисунок 35). Примечательно, что присутствие небольшой H-компоненты (красные вертикальные черточки на Рисунке 35) у чистых лесных куниц было подтверждено локальным (но не глобальным) ADMIXTURE-анализом (Рисунок 22D). Таким образом, мы подтвердили устойчивость нашего метода и еще раз подчеркнули, что одно глобальное среднее значение гетерозиготности является плохой метрикой даже в случае низкой интрогрессии.

После некоторых дополнений (автоматическое определение начальных параметров) и оптимизаций для улучшенной аппроксимации наш метод можно будет использовать в качестве дополнительного теста на гибридное происхождение образца, в случае, если нет образцов чистых видов или их количества недостаточно

для надежного ADMIXTURE-анализа. Принципиальным ограничением для такого применения является требование заметного числа окон с существенно различающейся гетерозиготностью, поэтому этот подход, например, не будет работать для F1-гибридов, но должен работать для F1-подобных, таких как образец T87 (Рисунок 23А). В качестве слабой стороны нашего метода следует отметить, что пока неясно, в какой степени неточность аппроксимации распределения влияет на конечный результат и сопоставим ли ее вклад с биологической вариацией между образцами.

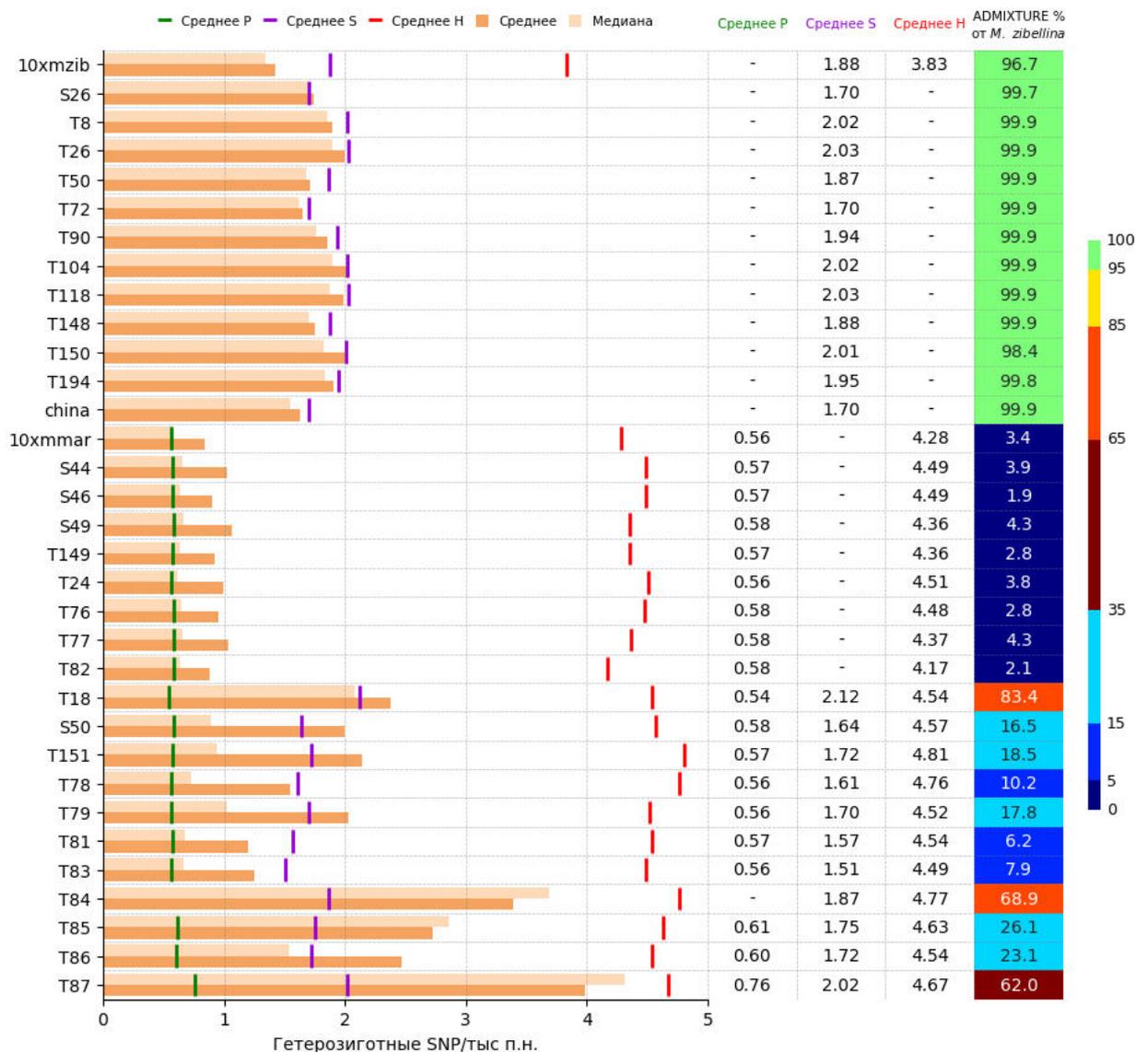


Рисунок 35. Анализ компонент гетерозиготности.

На гистограмме показаны среднее (оранжевый) и медианное (светло-оранжевый) значения глобальной гетерозиготности по выборкам. Вертикальные пунктиры

обозначают средние значения компонент: зеленый (лесная куница, P), фиолетовый (соболь, S) и красный (гибрид, H). Абсолютные значения компонент P, S и H, а также доля интрогрессии (%) соболя показаны справа.

4.4 Гетерозиготность, RoH и интрогрессия соболя и лесной куницы

Первоначальная классификация наших образцов основывалась на простых фенотипических и морфологических признаках, отличающих соболя, лесную куницу и кидаса, которые были предоставлены охотниками: длина хвоста, длина и качество меха, горловое пятно и другие (см. раздел “Гибридизация соболя и лесной куницы” Обзора литературы). Однако уже первый анализ гетерозиготности показал, что первоначальные предположения не отражают реальный генетический фон некоторых образцов (Рисунок 20). Как следствие, потребовалась переклассификация особей, в ходе которой мы выявили несколько интересных закономерностей.

Во-первых, мы обнаружили, что глобальный ADMIXTURE-анализ (анализ всего генома за один прогон) не чувствителен к низкому уровню интрогрессии. Для наших образцов он полностью пропустил примесь ниже 7%, в то время как локальный ADMIXTURE-анализ (расчеты на основе окон с последующим усреднением) ее обнаружил. Интрогрессия была подтверждена анализом компонентов гетерозиготности (см. выше), который мы представили здесь впервые, а также HuDe, F3-, D- и F4-статистиками. Исходя из полученных результатов мы предполагаем, что локальный ADMIXTURE-анализ должен быть предпочтительнее глобального. Геномные сборки на уровне хромосом сегодня широко распространены во всех областях биологии, благодаря чему фрагментация сборки на фрагменты короче размера окна больше не является препятствием.

Во-вторых, образцы соболя на графике PCA (Рисунок 22А) были распределены в основном вертикально вдоль оси Y (PC2, 5,59% дисперсии). Мы обнаружили, что образцы, собранные от Восточной Якутии до Камчатки (включая образец из Китая), расположены в верхней части графика ($Y > 0$) и что их положение коррелирует с долготой: чем дальше на восток был собран образец, тем выше он был расположен на графике. Остальные образцы ($Y < 0$), из Средней

Сибири и Западной Якутии, не отличались такой закономерностью. Мы предполагаем, что подобная картина является следствием широкого ареала соболя (более 4700 км с запада на восток), а также его истории и мероприятий по реинтродукции (см. раздел “Соболь и последствия его реинтродукции” в Обзоре литературы).

Удивительно, но мы обнаружили, что лесная куница значительно менее гетерозиготна, чем соболь (Рисунок 20, Рисунок 35), при сравнении как глобальной медианной гетерозиготности (0,64 против 1,73, p -value одностороннего теста Манна-Уитни = 0,00005), так и соответствующих компонентов гетерозиготности (среднее $P=0,57$ против среднего $S=1,9$, p -value Манна-Уитни = 0,00005). До начала исследования мы ожидали противоположной картины, основываясь на истории соболя. Хотя на лесную куницу также велась охота, она была менее интенсивной по сравнению с соболем, преимущественно из-за меха более низкого качества. Вероятно, разрушение мест обитания этого вида в Европе посредством антропогенного давления стало одной из причин не столько сокращения численности, сколько ограничения возможности скрещивания, тем самым способствуя снижению генетического разнообразия.

Мы выявили заметную разницу в паттернах RoH между соболем и лесной куницей. У соболя меньшую часть генома занимают короткие и длинные RoH, в то время как для сверхдлинных (> 10 млн п.н.) мы не обнаружили различий с лесными куницами. Мы подозреваем, что смешивание с реинтродуцированными особями, в основном с баргузинскими и витимскими (*M. z. princeps*), способствовало разрушению большей части RoH, что является однозначным подтверждением успешности мероприятий по реинтродукции. С другой стороны, в Сибири, от Уральских гор до берегов Тихого океана, никогда не было высокой плотности населения, поэтому у соболя было множество локальных убежищ от охотников в глубине тайги, вдали от городов, деревень и крупных рек. Естественное переселение и скрещивание во время запрета охоты в середине XX века и позднее общее снижение популярности натурального меха также могут объяснить наблюдаемую закономерность.

Сейчас соболь считается экспансивным и даже инвазивным видом [311], но, возможно, это просто восстановление исторического ареала и ответ на климатические изменения. Тем не менее, активное распространение соболя вызывает беспокойство и нуждается в наблюдении. Среди соболей необходимо выделить образец с Камчатки (Т72), кумулятивная траектория распределения RoH которого отличается от других соболей и сходна с траекторией лесных куниц (Рисунок 25А). Этот образец по всей видимости является представителем автохтонной популяции, которая была меньше подвержена чрезмерному отлову или значительному притоку реинтродуцированных особей. Это дополнительно подтверждается кластеризацией его мтДНК с другими камчатскими образцами (Рисунок 31). Таким образом, наши результаты подтверждают, что Камчатская популяция соболя оставалась в более стабильном состоянии в сравнении с другими популяциями Сибири на протяжении всей депрессии вида. Однако, поскольку мы секвенировали лишь один образец, состояние Камчатской популяции необходимо будет рассмотреть подробнее в будущих исследованиях.

4.5 Гетерозиготность исследуемых видов рода *Martes* в сравнении с другими куньими

Соболь показал неожиданно высокую среднюю гетерозиготность (1,42-2,02 SNP/тыс п.н.) с учетом своей видовой истории. Его оценки сопоставимы с гетерозиготностью (1,78 SNP/тыс п.н.) тайры (*Eira barbara*) – тропического южно- и центральноамериканского вида куньих [460], для которого никогда не сообщалось о том, что на него как-то интенсивно охотились, если вообще охотились. Стоит отметить, что представленная здесь для сравнения особь (тайра) относится лишь ко второму поколению, рожденному в неволе [115], что не позволяет делать выводы относительно свободно живущих популяций. Тем не менее, можно предполагать, что природные популяции тайры также имеют высокую гетерогенность по причине минимального антропогенного давления и обширным естественным местообитаниям. Учитывая, что мы рассматриваем образец Т72 (1,65 SNP/тыс п.н.) как автохтонную особь из популяции, избежавшей чрезмерной охоты, сходная гетерозиготность соболей из Западной и Средней Сибири подтверждает

успешность проводимых мероприятий по реинтродукции. Однако, скорее всего, ценой стала контаминация местных популяций реинтродуцированным поголовьем, представленного преимущественно баргузинским и витимским соболем [294].

Исторически сложилось так, что мех соболя и лесной куницы считался предметом роскоши и на протяжении веков служил ценной природной валютой. Однако на рубеже XX века стоимость соболиного меха достигла пика из-за сочетания снижения объемов добычи и высокого спроса, созданного индустрией моды как в Европе, так и в Северной Америке. Например, при расчете натурального налога в 1921 году в СССР, шкурка одного темного соболя приравнивалась к 18'000 шкурок сусликов, 25-ти росомах, 15-ти барсов или 4-х белых медведей [461]. При этом на динамику численности видов, по всей видимости, влияла не только добыча пушнины. Для лесной куницы, значительная часть ареала которой охватывает территории, густо населенные людьми, заметную роль мог сыграть иной тип давления, связанный с развитием сельского хозяйства и деградацией основных лесных местообитаний [462]. Для соболя роль этого фактора была ниже, поскольку бóльшая часть его ареала находится в отдаленных от густонаселенных человеком районах. Таким образом, соболь и лесная куница, испытывали значительное давление со стороны человека, однако степень и характер антропогенного давления на эти два вида трудно сравнивать, и заметные различия в гетерозиготности, о которых здесь сообщается, трудно объяснить.

Лесные куницы с гетерозиготностью (медианной) $\sim 0,56-0,67$ SNP/тыс п.н. (Рисунок 20, Рисунок 35, Приложение 19) менее разнообразны, чем некоторые другие исчезающие виды, такие как, например, красная панда (*Ailurus fulgens*, 1,05 SNP/тыс п.н.) и азиатская бескоготная выдра (*Aonyx cinereus*, 1,18 SNP/тыс п.н.). При этом, лесная куница имеет гетерозиготность лишь на $\sim 40\%$ выше таких знаковых для природоохранной биологии и геномики видов, как гепард (*Acinonyx jubatus*, 0,44 SNP/тыс п.н.) и дымчатый леопард (*Neofelis nebulosa*, 0,45 SNP/тыс п.н.). Наша реконструкция популяционной истории (см. раздел “Популяционная история рода *Martes*” в Обсуждении) предполагает, что это, вероятно, древняя черта вида, а не современная особенность. Учитывая ареал, численность и отсутствие инбредной депрессии, мы рассматриваем этот вид как успешный,

который, вероятно, вытеснил большинство рецессивных вредных аллелей из своего пангенома. Однако для подтверждения этой гипотезы необходим расширенный набор данных и дальнейшие исследования. Кроме того, все наши выводы и предположения относятся к лесным куницам восточной части ареала, тогда как состояние западных популяций остается неизвестным.

Гибриды соболя и лесной куницы значительно отличаются от всех остальных видов куньих за счет высокой гетерозиготности с выраженной мультимодальностью, которая стала следствием гибридизации видов с различным уровнем гетерозиготности (рассмотрели выше). Харза демонстрирует уровень гетерозиготности (0,78 SNP/тыс п.н.), сопоставимый с показателями лесной куницы и американской куницы. Однако, будучи особью из зоопарка, исследуемая особь не характеризует свободноживущие популяции этого вида. Каменная куница в сравнении с другими видами отличается относительно низким уровнем гетерозиготности (0,42 SNP/тыс п.н.), немного превышая показатели для речной выдры (*Lutra lutra*, 0,33 SNP/тыс п.н.), при этом около половины генома (51,7%) составляют участки RoH различной протяженности. Несмотря на то, что результаты анализа всего одной особи, добытой еще в 90-х годах [284], не отражают текущее генетическое разнообразие вида, полученный результат позволяет поднять вопрос о необходимости более подробной оценки состояния генетического разнообразия каменной куницы, по крайней мере, в пределах провинции Ганьсу, Китай.

Особое внимание стоит уделить североамериканским куницам. Между образцами SRR11575348 (о. Адмиралтейский, *M. caurina (is)*) и SRR11575351 (штат Вашингтон, *M. caurina (con)*) наблюдается значительная разница в среднем уровне гетерозиготности и в доле участков RoH. Хотя в работе Colella et al. 2020 отмечалось, что островные формы отличаются высокими значениями коэффициента инбридинга (Fis), в сравнении с материковыми, выявление и анализ участков RoH остался без внимания [248]. Критически низкая гетерозиготность у островных *M. caurina*, в отличие от континентальных популяций, свидетельствуют об инбридинге и отсутствии потока генов, что в дальнейшем может привести к сокращению популяции. Гетерозиготность островных *M. caurina* значительно ниже,

чем у морской выдры (*Enhydra lutris*), снижение генетического разнообразия для которой последовало после неконтролируемой охоты в XVIII и XIX веках [463,464]. Отметим, что *M. caurina* и *M. americana* на момент написания работы МСОП до сих пор считает одним видом (*M. americana*), несмотря на опубликованные работы об их генетических и морфологических различиях [12,248], что маскирует потенциальные проблемы островных популяций тихоокеанских куниц. При этом, даже в случае внесения *M. caurina* как отдельного вида, его глобальный статус не должен интерпретироваться как полная и достаточная оценка по той причине, что ситуация между материковыми и островными популяциями значительно отличается. Примечательно, что близкородственная *M. americana*, будучи исключительно материковым видом, не испытывает проблем с гетерозиготностью.

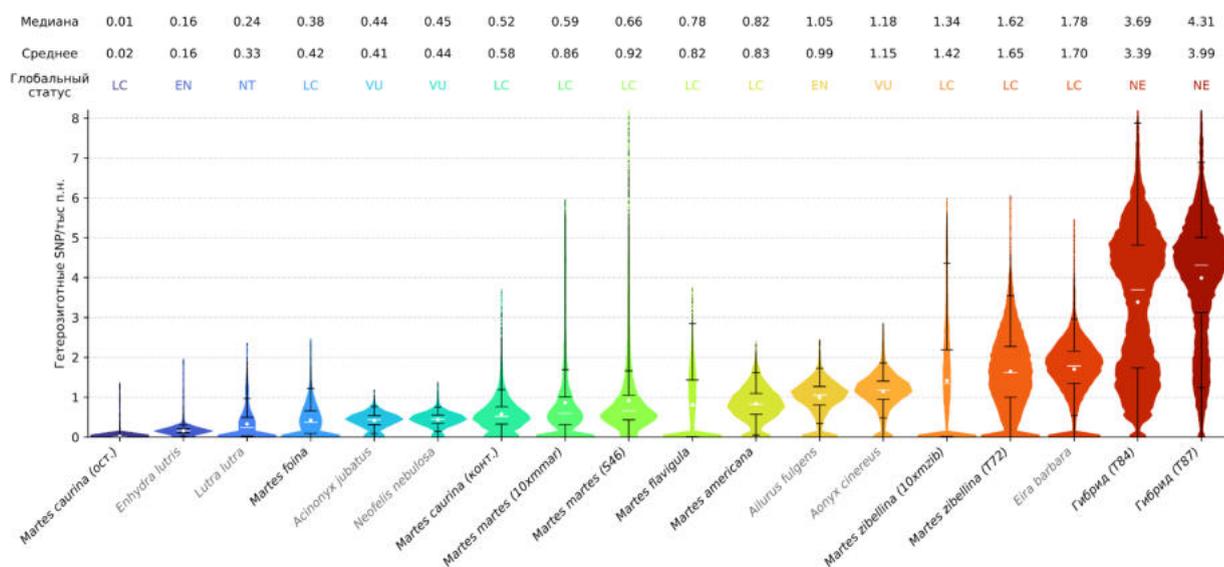


Рисунок 36. Сравнение гетерозиготности исследуемых представителей рода *Martes* с другими видами куньих.

Распределения средней гетерозиготности (гетерозиготные SNP) в скользящих окнах размером 1 млн п.н. с шагом в 100 тыс п.н. Серым обозначены образцы дополнительных видов хищных из [66]. Каждое распределение масштабировано относительно оси X. Средние значения обозначены белыми точками, медианы – белыми линиями. Особи отсортированы по среднему значению гетерозиготности. Сверху подписан глобальный природоохранный статус согласно МСОП: EN

(Endangered) – находящийся под угрозой, VU (Vulnerable) – уязвимый, LC (Least Concern) – не вызывающий беспокойства, NE – не оценен.

4.6 Кроссинговер и мозаичность хромосом

Является ли гибридное потомство плодовитым или нет – ключевой вопрос для гибридизующихся видов. Охотники часто используют термины “соболеподобный” или “куницеподобный” для оценки и разграничения шкурок кидасов по окрасу и качеству меха. В публикациях также встречались упоминания о “нетипичных” особях обоих видов [320]. Это подтверждает высокую изменчивость кидасов и подразумевает наличие широкого спектра варьирующих признаков, которые могут быть как промежуточными между соболем и лесной куницей, так и полностью относиться к одному из родительских видов. Однако плодовитость кидасов долгое время оставалась предметом споров: возникали трудности как при скрещивании соболя и лесной куницы, так и в ходе возвратных скрещиваний с гибридами F1 [24–27]. Позднее генетические исследования подтвердили наличие в дикой природе отличных от F1 гибридов [28–30,293]. Однако данные работы основывались на отдельных STR-маркерах, вследствие чего были ограничены небольшим числом (10 или менее) локусов. Кроме того, расположение используемых STR-маркеров на хромосомах оставалось неизвестно. Согласно полученным результатам оказалось, что в некоторых исследованиях STR-маркерами было покрыто даже менее половины хромосом (Рисунок 24В и 24С) [29]. Как следствие, оставалось не до конца ясно, подавляется ли кроссинговер в геномах гибридов или нет.

В случае гибридов соболя и лесной куницы, локальный ADMIXTURE-анализ каждой хромосомы позволил выявить участки недавней интрогрессии, что проявилось в характерной мозаичности разной степени для множества рекомбинантных хромосом (Рисунок 37А). Обнаруженная мозаичность хромосом у гибридов проявляется в разных пропорциях и локализуется на различных хромосомных участках. Все рекомбинантные хромосомы, представленные на этом рисунке ("adm" – каждая третья хромосома), имеют участки, унаследованные как от соболя (зеленый цвет) и от лесной куницы (темно-синий), так и от обоих

родительских видов, то есть гибридные (красный). В случае подавленного кроссинговера, мы наблюдали бы картину, при которой хромосомы соболя и лесной куницы оставались бы разделенными в потомках (Рисунок 38В), но наши результаты подтверждают обратное (Рисунок 38А). Возможные варианты фазирования для отдельных хромосом (Рисунок 37В) предполагают, что по крайней мере в некоторых парах оба гомолога рекомбинантны и, более того, состоят более чем из двух сегментов, то есть являются продуктами множественных событий кроссинговера. Такая картина может возникнуть только после многократных последовательных скрещиваний в различных комбинациях чистых видов и гибридов разных типов, и только если кроссинговер, по крайней мере, подавлен не полностью. Стоит сказать, что интрогрессированные регионы разного размера выявлены не только в предполагаемых гибридных образцах, но и в предполагаемых образцах “чистых” видов. В частности, интрогрессия от соболя в разной степени была обнаружена во всех образцах лесной куницы.

Примечательно, что среди 11 выявленных гибридов ни один всецело не принадлежал поколению F1 (Рисунок 39). Наиболее близкими к F1 оказались образцы Т84 (68,85 % от соболя) и Т87 (62,03 %), но даже они имеют множество рекомбинантных хромосом (Рисунок 23). Исходя из полученного результата можно выдвинуть предположение, что межвидовой поток генов не ограничивается общеизвестной зоной симпатрии родительских видов, но происходит в том числе за ее пределами, образуя условную “зону интрогрессии” (Рисунок 38С). Границы видов, в этом случае, оказываются полупроницаемыми и неоднородными, с постепенным переходом от одного вида к другому, что обусловлено межвидовым потоком генов [132]. Из-за ограниченного количества образцов (для видов с такими большими ареалами) мы не можем обозначить границы зоны интрогрессии, но она вполне может распространяться до восточной границы Ханты-Мансийского автономного округа на востоке и, возможно, до центральных районов европейской части России на западе.

Выявленная мозаичность хромосом в том числе в обоих референсных образцах соболя и лесной куницы поднимает вопрос о создании пангеномов для этих близкородственных видов. В отличие от традиционных линейных референсов,

пангеномы включают информацию из геномов сразу нескольких особей, что позволяет избежать смещения относительно одной из них и охватить разнообразие аллелей, присутствующих в различных популяциях [465]. Такой подход особенно важен в контексте активно гибридизующихся близкородственных видов, где существует сложность выявления особей “чистых” видов исключительно по морфологическим признакам. Более того, использование пангеномов, аккумулирующих информацию из различных популяций, позволит с большей точностью проводить локальную оценку их генетического разнообразия. Данной исследование послужит основой будущих работ по сравнительной геномике кунных в том числе в этом направлении.

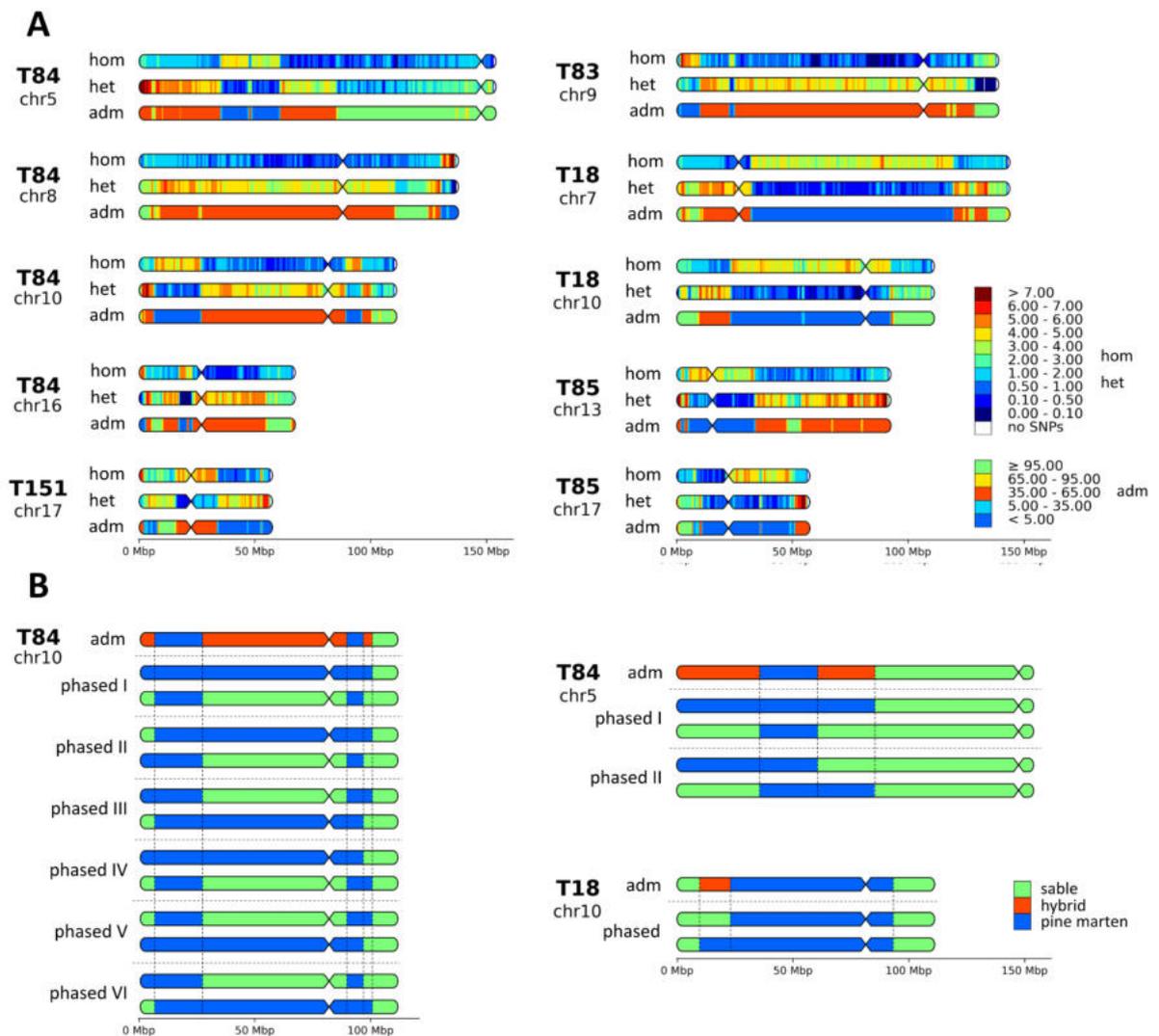


Рисунок 37. Мозаичные рекомбинантные хромосомы.

А – мозаичные хромосомы: hom – тепловые карты гомозиготных SNP; het – тепловые карты гетерозиготных SNP; adm – локальный ADMIXTURE (% *M. zibellina*), В – схема возможных вариантов фазирования для трех выделенных хромосом. Т** – идентификаторы образцов.

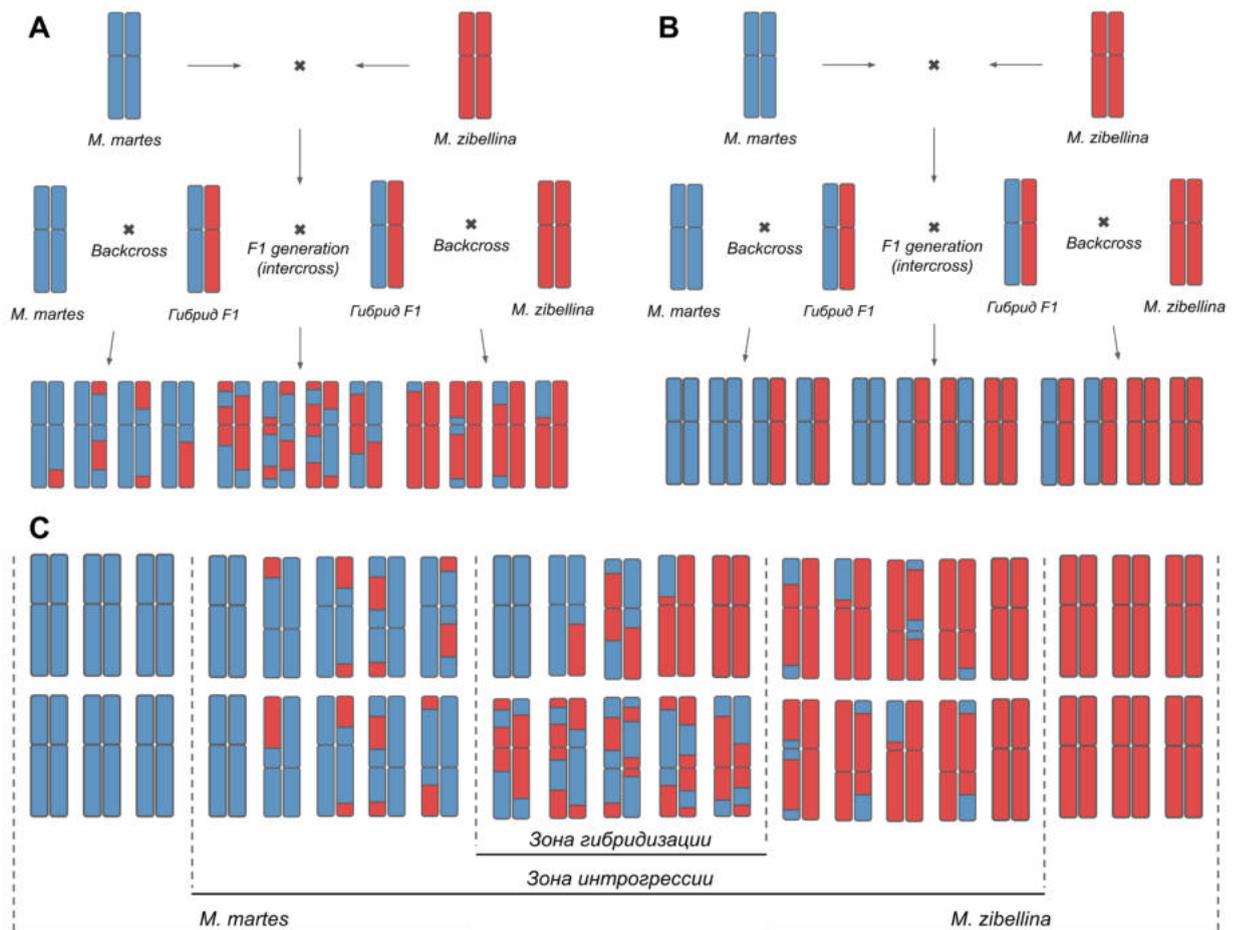


Рисунок 38. Гибридизация, кроссинговер и интрогрессия.

А – случай неподдавленного кроссинговера между гомологичными хромосомами разного происхождения; В – гипотетический случай полностью подавленного кроссинговера; С – градиент интрогрессии между чистыми видами в случае неподдавленного кроссинговера.

4.7 Противоречия классификации по видовой принадлежности

Для видовой классификации особей соболей, лесных куниц и их гибридов мы использовали различные подходы и данные: глобальный и локальный ADMIXTURE-анализы на полногеномных данных, ADMIXTURE на основании разных наборов STR-маркеров, реконструкция филогении с использованием

полных мтДНК, морфологический анализ гибридов. Полученные нами результаты различаются в зависимости от подхода и входных данных (Рисунок 39).

Как было сказано выше (см. раздел “Гетерозиготность, *RoH* и видовая принадлежность”), мы считаем локальный ADMIXTURE-анализ на основе полногеномных данных наилучшим способом классификации особей. Нам удалось выявить 11 особей с интрогрессией более 5%. Из рассмотренных 12-ти образцов потенциальных кидасов подтвержденными оказались только 8 из них, тогда как оставшиеся 4 относились к лесным куницам с интрогрессией со стороны соболя менее условного порога в 5%. При этом 1 из 14 образцов фенотипически "чистых" соболей и 2 из 7 образцов фенотипически "чистых" лесных куниц также оказались гибридами. Из рассмотренных 33 образцов соболя, лесной куницы и их предполагаемых гибридов выявлено 11 особей с интрогрессией более 5%, при этом ни одна из них не оказалась гибридом поколения F1.

Однако существует множество исследований гибридизации между соболем и лесной куницей, а также разнообразия соболей, которые проводились с использованием различных наборов STR и митохондриальных маркеров [29,298,311,321,466]. Применение широко используемых STR-маркеров для классификации по видовой принадлежности позволило выявить значительное расхождение с результатами на основании полногеномных данных для гибридных образцов, в основном в виде недооцененной интрогрессии (Рисунок 39). Например, для образца T84, доля лесной куницы варьирует от 4,2% до 36,6% в зависимости от набора STR-маркеров. Это ожидаемый результат из-за малого количества маркеров, которые даже в объединенном наборе не покрывают все хромосомы (Рисунок 24А).

Выявленный в данном исследовании неподавленный кроссинговер у гибридов соболя и сосновой куницы предполагает, что для правильной классификации образцов по видовой принадлежности требуются тысячи маркеров, равномерно распределенных по всем хромосомам. Это означает, что в будущих исследованиях гибридизации следует избегать использования STR-маркеров и других маркеров с низкой плотностью, и что предыдущие количественные оценки [29] нуждаются в пересмотре. Только в случае полностью подавленного кроссинговера и доказанного покрытия всех хромосом, такой тип маркеров может

рассматриваться как надежный подход. С другой стороны, STR-типирование использовалось как относительно дешевый метод, позволяющий исследователям охватить сотни образцов, в том числе с деградировавшей ДНК. В качестве замены мы предлагаем использовать панели обогащения (англ. capture arrays) или микрочипы (англ. microarray technology), когда полногеномное секвенирование недоступно из-за качества образцов или ограниченного бюджета. Оба метода уже успешно применяются в исследованиях по сохранению биоразнообразия как более дешевая альтернатива полногеномного секвенирования [467].

Отметим, что термин “кидас”, используемый для определения “гибридных особей из зоны симпатрии” в случае интрогрессивной гибридизации становится неоднозначным и расплывчатым с генетической точки зрения, что подтверждается выявленными противоречиями видовой классификации между фенотипическими, митохондриальными и ядерными данными. Безусловно, особи близкие к F1 могут называться “кидасами”, однако потомство от первого и последующих обратных скрещиваний становится сложно (и часто невозможно) отличить от “чистых” родительских особей на основании фенотипических и морфологических признаков. Особей с интрогрессией, которых нельзя было однозначно классифицировать ни как гибридных, ни как чистых, в литературе называют “нетипичными” [320], однако вопрос точного определения остается открытым.

По результатам морфологического анализа, все образцы из зоны симпатрии были отнесены к лесной кунице, за исключением образца T87, который изначально был отмечен как гибрид из-за характерного цвета горлового пятна. По результатам классификации с использованием недавно предложенного краниометрического признака Δ [352] все образцы предполагаемых гибридов были ближе к лесной кунице, чем к соболу. Учитывая, что большинство этих образцов были классифицированы как гибриды разных типов, включая образцы T84 и T87, близкие к F1, мы не можем считать этот маркер надежным для видовой классификации. На момент исследования, краниометрические признаки гибридов среди куньих остаются слабо изучены, однако, как правило, гибриды F1 характеризуются признаками одного из родительских видов. Например, сравнительный анализ черепов лесного хорька (*Mustela putorius*), европейской

норки (*Mustela lutreola*), и их гибридов с северо-запада России показал, что большинство гибридных особей краниологически очень похожи на *M. putorius* [468]. Черепа гибридов европейского и азиатского барсуков, *Meles meles* и *M. leucurus*, более похожи на *M. leucurus* [469].

Митохондриальная ДНК, очевидно, не может быть использована для классификации гибридов, но в случае наших образцов она даже не коррелирует с ядерным геномом чистых куниц (Рисунок 39). Поэтому мтДНК соболя и куницы, а также, вероятно, и других активно гибридирующих видов куньих, следует с осторожностью использовать даже в филогенетических исследованиях.

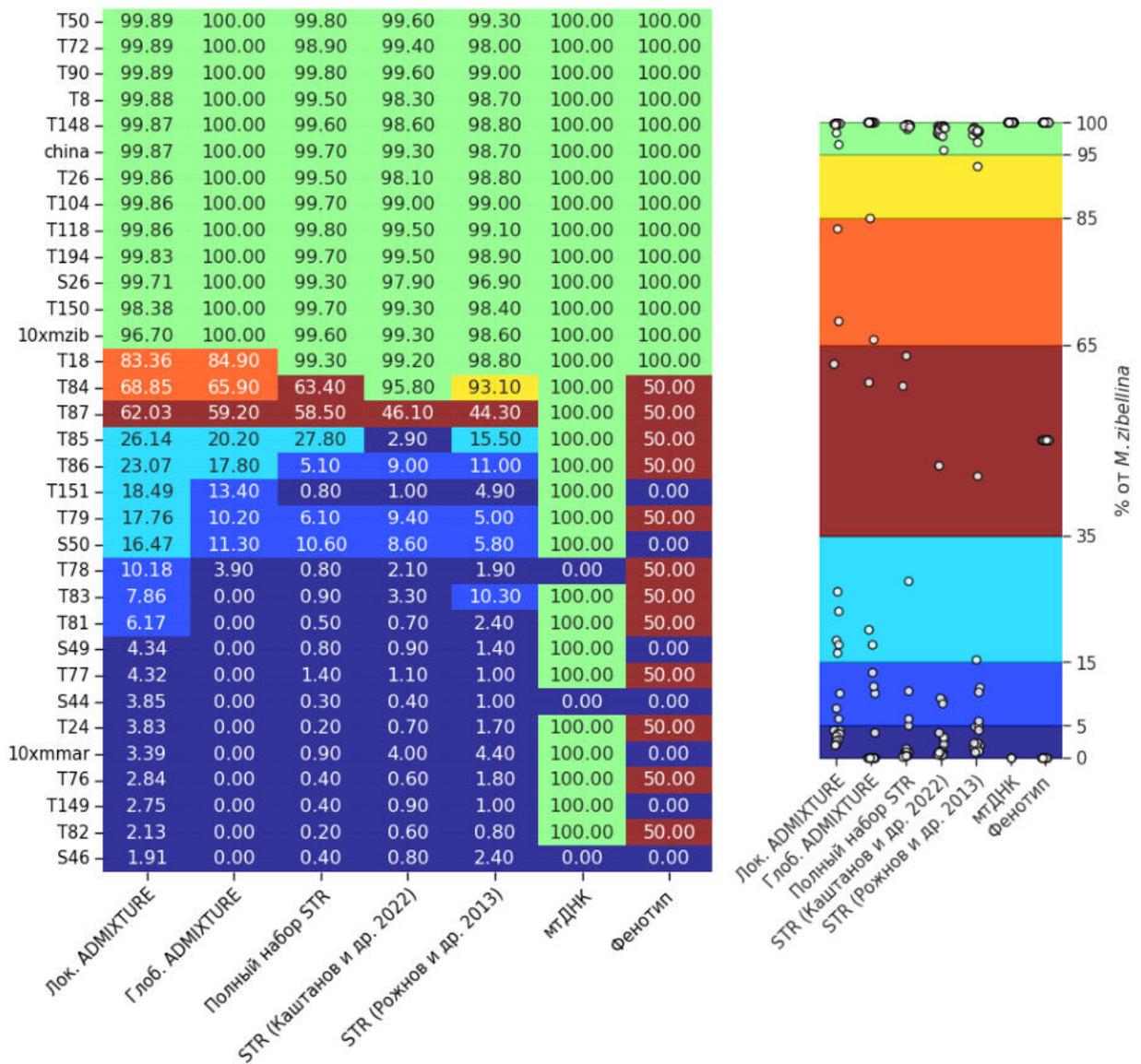


Рисунок 39. Противоречия между классификацией особей на основе различных данных: локальный ADMIXTURE, глобальный ADMIXTURE, все STR-маркеры,

STR-маркеры из работы Каштанов и др. 2022, STR-маркеры из работы Рожнов и др. 2013, мтДНК и фенотип (первоначальная классификация).

4.8 Филогения соболя и лесной куницы на основании мтДНК

Реконструкция филогении и анализ сети гаплотипов с использованием полных мтДНК подтверждают результаты Li et al. 2021 о том, что четкая филогеографическая структура подвидов соболя отсутствует на всей территории Евразии [321]. Предыдущая работа основывалась на около 100 геномах мтДНК и лишь нескольких лесных куницах в качестве аутгруппы. Мы дополнили этот набор нашими 32 образцами, а также митохондриальной ДНК, собранной из 4 опубликованных образцов соболя (PRJNA495455, SRR13213810, SRR13213811, SRR13213812), в результате чего общее число образцов достигло 140. Добавление новых данных улучшило представление некоторых ранее слабо охваченных митохондриальных линий (Рисунок 30) внутри кластеров C1 и A1. Гаплогруппы A1, B1 и C1 были описаны Li et al. как немонофилетические. Наш анализ расширенного набора данных показал, что гаплогруппы B1 и C1 следует разделить на две более мелкие монофилетические гаплогруппы: B3 и B4, а также C3 и C4, соответственно. Гаплогруппу A1 также следует подразделить на три линии: A3, A4 и A5 (Рисунок 31). Аналогично Li et al., мы не обнаружили четкой корреляции между гаплогруппами и формальными подвидами (определяемыми по месту отбора проб). Примечательно, что только образец T72 (Камчатский) является единственным образцом, мтДНК которого кластеризуется с соответствующим формальным подвидом – *M. z. kamtschadalica*. Этот факт поддерживает гипотезу о том, что Камчатка может быть местом обитания автохтонной популяции (или популяций), не затронутой интрогрессией от других подвидов. Особенности камчатских соболей ранее были отмечены в ряде работ [298,321,322,466].

Реконструированное филогенетическое дерево и построенная сеть гаплотипов демонстрируют отсутствие видоспецифичности последовательностей мтДНК, в результате чего мтДНК соболя обнаружена в 7 из 9 образцов “чистых” лесных куниц (включая референсный 10хmmag), а также у одного гибрида. Интрогрессия в данном контексте, вероятно, происходила неоднократно, поскольку

у “чистых” лесных куниц обнаружена мтДНК из различных немонафилетичных клад соболей: А3, А5, В2 и С4. Такое явление можно объяснить биологическими и поведенческими особенностями видов. Считается, что соболь лучше приспособлен к жизни в условиях глубокого снега, на что указывают ряд морфологических признаков: более короткий хвост, мягкий мех, густо опушенные подошвы и хорошо развитые щетинистые участки на лапах (“снегоступы”) [281,299]. В регионах с суровыми снежными зимами, таких как зона симпатрии, лесная куница может уступать соболю в прямом противостоянии, особенно мелкие самки; существуют сообщения, основанные на следах, о случаях, когда соболь преследовал и убивал куницу. Некоторые источники утверждают, что соболь более агрессивен к сородичам по сравнению с куницей: случаи, когда куница убивает другую куницу, крайне редки, тогда как соболь часто нападает и убивает других соболей (например, случаи каннибализма соболей в ловушках объясняются территориальным поведением, а не пищевой мотивацией) [470]. Таким образом, самки лесных куниц могут становиться жертвами межвидовой конкуренции в зоне симпатрии. Это также подразумевает, что митохондриальная ДНК соболя может распространяться далеко на запад от зоны симпатрии, а митохондриальная ДНК лесной куницы должна встречаться к востоку только в исключительных случаях. Несмотря на ограниченный размер нашей выборки, мы наблюдаем такую тенденцию по нашим результатам. Из всех образцов только у трёх обнаружена митохондриальная ДНК лесной куницы, тогда как у двух куниц (чистой лесной куницы S49 и гибрида S50), отобранных далеко на западе, обнаружена митохондриальная ДНК соболя, и ни у одного экземпляра, собранного восточнее, митохондриальная ДНК лесной куницы не была выявлена.

Стоит сказать, что на момент написания работы опубликовано лишь 5 полных мтДНК лесной куницы, и только 3 особи в наших выборках имеют мтДНК этого вида. Следовательно, необходимо получить больше образцов лесных куниц для более детального изучения мтДНК. Сложность заключается в том, что, как показали наши результаты, невозможно точное определение вида по фенотипическим и морфологическим признакам.

4.9 Харза как отдельный род в подсемействе *Guloninae*

На всех полученных филогенетических деревьях видно, что харзы отделились от остальных представителей рода *Martes* практически сразу после расхождения с линией росوماхи (*Gulo gulo*) и отличается от других видов *Martes* длинной ветвью, сопоставимой с ветвями росوماхой, илькой и тайрой. Узел с харзами (как и соседние узлы) поддерживается лишь примерно третью проанализированных геновых деревьев (Рисунок 28), чего впрочем достаточно для уверенной поддержки такой топологии (апостериорная вероятность 100%). Это свидетельствует об активном видообразовании вблизи общего предка *Guloninae* в позднем миоцене, вероятно сопровождающемся гибридизацией и/или неполной сортировкой линий.

Учитывая полученные результаты и беря во внимание морфологические и кариотипические ($2n=40$) отличия харзы от других видов *Martes* (см. раздел “Проблематика положения харзы (подрод *Charronia*) в роде *Martes*” Обзора литературы), мы считаем, что этот вид необходимо выделять в отдельный самостоятельный род. На данный момент существует два синонимичных варианта названия: *Charronia* (текущее название подрода) и *Lamprogale*. Мы полагаем, что более подходящим названием для рода харз будет *Charronia*, согласно правилу приоритета при выборе наименования рода. Это самое раннее наименование, предложенное Gray, J.E. в 1865 году [269] и позднее поддержанное рядом морфологов [16,18,264]. Наименование *Lamprogale* может оставаться синонимичным названием рода, поскольку синонимичные наименования среди куньих не редкость. Например, для рода американской норки альтернативами *Neogale* были наименования *Grammogale* [471] и *Neovision* [472], последнее из которых сохранилось в качестве синонима [471]. На текущий момент выделяют два вида харз: *Ch. (M.) flavigula* и *Ch. (M.) gwatkinsii*, однако наши результаты говорят, что вид *Ch. (M.) flavigula* необходимо разделить на несколько видов.

4.10 Видовая структура *M. foina* и *Ch. (M.) flavigula*

Анализ генетических расстояний мы провели в контексте проблематики видовой структуры *M. foina* и *M. flavigula*. Учитывая генетические расстояния между ядерными геномами харз, сопоставимые с межвидовыми расстояниями других видов, мы предполагаем, что исследуемые особи (Дальний Восток, Россия и п. Сычуань, Китай) относятся к разным видам. Согласно предыдущим классификациям (см. раздел “Проблематика положения харзы (подрод *Charronia*) в роде *Martes*” Обзора литературы), основанным на морфологических особенностях и местообитаниям харз, мы считаем, что наша исследуемая особь (10xmfla), предками которой были харзы с Дальнего Востока России, является видом *Ch. (M.) aterrima*, тогда как вторая южно-китайская особь (GCA_029410595.1) по всей видимости относится к *Ch. (M.) flavigula*. Отсутствие подобной картины в случае мтДНК может объясняться ее интрогрессией между этими видами, поскольку схожая мтДНК была обнаружена у других харз из Южной Кореи (MW625810.1, KM347744.1) и Северо-Восточного Китая (NC_012141.1 / FJ719367.1, NM106326.1). При этом, кластер, представленный мтДНК с большими генетическими расстояниями, не связан с происхождением особей.

Мы также предполагаем существование как минимум еще одного вида *Charronia (Martes)* в материковой части юго-восточной Азии и окружающих ее островах, например, описанного ранее *Ch. (M.) lasiotis* [17]. Однако достоверные утверждения требуют получения полногеномных данных для соответствующих особей, полностью отсутствующих на момент написания работы. Текущее выделение многочисленных таксонов харзы подвидового ранга не всегда подкреплено достаточными морфологическими доказательствами. Ярко выраженная изменчивость в размерах и окраске является основным препятствием для достоверного выделения популяций и подвидов [16]. Необходимы дальнейшие исследования на полногеномном уровне для других харз из разных частей ареала для уточнения систематики на уровне подрода.

В ходе анализа генетических расстояний на основании полных мтДНК мы выявили четкое разделение особей *M. foina* на две клады, соответствующие популяциям из Европы и Азии. Полученный результат согласуется с известной

филогеографической структурой этого вида (на основании контрольного региона мтДНК), для которой характерно разделение ареала на западную и восточную [283]. При этом, уровень различий мтДНК соответствует межвидовым, что заставляет задуматься о неизвестном ранее разнообразии *M. foina* и о возможности разделения вида (Рисунок 32В). Однако, для ядерных геномов (Рисунок 32А) расстояния между европейской и азиатской особей *M. foina* раз лишь несколько больше, чем характерные внутривидовые других представителей рода *Martes*. Таким образом, мы не можем уверенно заявить о существовании нескольких видов, но считаем это направление перспективным для дальнейших исследований. Для этого потребуется секвенирование дополнительных особей *M. foina* из различных частей ареала.

4.11 Популяционная история рода *Martes*

Датировка популяционной траектории линейно зависит от отношения времени генерации к частоте замен на поколение (g/μ) [95]. Чем больше время генерации, тем глубже сдвинуты события в прошлое, и наоборот для μ . Оба параметра трудно оценить. В данной работе в качестве опорного значения мы решили использовать $g=5$ [248] и $\mu=4,64 \times 10^{-9}$, которое было оценено в прямом замере на трио (родители и потомок) американских норок [456]. Для определения доверительного интервала датировок мы использовали границы доверительного интервала μ равные $2,94 \times 10^{-9}$ и $7,37 \times 10^{-9}$ (согласно той же статье) при фиксированном g . Учитывая сложности оценки этих параметров, важно понимать, что обсуждение датировок пиков и спадов N_e и их привязка климатическим событиям не может считаться однозначной. Тем не менее, сравнение траекторий представителей рода *Martes*, в особенности евразийских соболя и лесной куницы, все еще разумно, поскольку для этих видов должны быть схожи параметры g и μ .

Траектории лесных куниц и соболей сильно отличаются друг от друга и сходятся только в глубоком прошлом, ближе к началу реконструкции. Примечательно, что после дивергенции, траектории соболей оставались относительно стабильны, несмотря на небольшие колебания, в то время как N_e лесных куниц продолжал расти от самого начала, вплоть до резкого падения.

Вероятно, это связано с видообразованием и, предположительно, с колонизацией предками лесной куницы новых территорий в сочетании с однонаправленным непрерывным потоком генов от "материковой" популяции. Исходя из более "стабильной" траектории соболя и современного ареала обоих видов, можно предположить, что их последний общий предок населял территорию, близкую к современному ареалу соболя или его части. При этом экспансия предков лесной куницы, вероятно, происходила в западном направлении – в сторону Европы. Последующее снижение N_e у лесной куницы (в то время как у соболя оно минимально) указывает на изоляцию вида в плейстоценовую эпоху, вероятно, в ледниковых рефугиумах [473]. После пережитого «бутылочного горлышка» численность лесной куницы продолжала постепенно падать вплоть до современности (с кратковременным подъемом), что свидетельствует об исторических корнях низкой гетерозиготности этого вида.

Сопоставление траекторий популяционных историй для видов родов *Martes* и *Charonia* можно использовать для формулировки гипотезы о месте происхождения рода и расселении его видов. Прежде всего мы хотели бы подчеркнуть некоторые особенности траекторий. Траектория каменной куницы похожа на соболиную с более низким значением N_e , за исключением самого начала. В свою очередь траектория харзы похожа на траекторию каменной куницы, но с сильно сглаженным первым пиком и более высокой численностью, что можно интерпретировать как последствия влияния одних и тех же палеоклиматических событий, а значит и частично пересекающиеся ареалы всех трех видов. Важно отметить, что харза и каменная куница являются базальными видами в роде. северо-американские виды имеют весьма продолжительную общую траекторию, но в какой-то момент расходятся друг с другом. При этом они сильно отличаются от траекторий остальных видов. Траектория *M. martes* также сильно выбивается из общего паттерна.

На основании вышеизложенного и современных ареалов, мы сформулировали следующую гипотезу о расселении видов рода *Martes*. Мы предполагаем, что общий предок всех куниц обитал в Восточной Евразии. В начале произошло расселение в двух направлениях: в Европейскую часть (предки *M.*

martes) и в Северную Америку (предки *M. americana* и *M. caurina*). В дальнейшем предки Американских куниц расселялись в двух направлениях: на юг (предки *M. caurina*) и на восток (предки *M. americana*), и в какой-то момент оказались изолированы друг от друга, что и привело к формированию этих видов [474]. При этом текущие ареалы и зоны симпатрии этих видов являются следствием реколонизации оставленных ранее территорий [248]. Исходя из растущей численности в начале траектории *M. martes*, мы предполагаем, что начало расселения их предков сопровождалось односторонним потоком генов из предковой азиатской популяции. Но в дальнейшем, *M. martes* также оказалась изолирована вероятно где-то в Европе. При этом известно, что климатические изменения в плейстоцене способствовали временному расширению ареала соболя далеко на запад, включая территории Восточной Европы, вплоть до современных Литвы, Беларуси и восточной Польши [353]. Так, в холодные периоды плейстоцена соболь расширял свой ареал вглубь Европы, достигая Фенноскандии, где происходила гибридизация с лесной куницей, а затем, с потеплением климата в период межледниковий, отступал обратно в сибирские леса [475]. Для лесной куницы, напротив, происходила экспансия ареала из многочисленных рефугиумов в межледниковые периоды. Таким образом, интрогрессия мтДНК соболя в лесную куницу происходила в прошлом, приводя к вытеснению мтДНК лесной куницы, что подтверждается в том числе на основании наших результатов (см. раздел “Филогения соболя и лесной куницы на основании мтДНК”). Таким образом, современные ареалы соболя и лесной куницы, также как и американских видов, являются результатом реколонизации, связанной с климатическими колебаниями плейстоцена, следствием которых и стало образование зон симпатрии с интрогрессивной гибридизацией между ранее разошедшимися видами.

Вопрос о том, что произошло раньше, отделение и изоляция американских видов или лесной куницы, остается дискуссионным. Согласно филогении, построенной по ядерным маркерам, *M. martes* отделился раньше американских видов, но митохондриальное дерево говорит об обратном. Однако, на популяционной траектории мы видим в самом начале общий (но короткий) участок для соболя и лесной куницы, тогда как траектории американских видов отличаются

от них с самого начала. То есть сопоставление популяционных историй скорее говорит о произошедшей первой изоляции американских видов. В то же время, ареалы соболя и американских видов вероятно были связаны в прошлом за счет существования сухопутного Берингийского моста, неоднократно соединявшем Евразию и Северную Америку на протяжении тысячелетий [476]. Это говорит о возможном потоке генов между соболем и американскими видами в недавнем прошлом, что в свою очередь могло повлиять на топологию дерева. С другой стороны, общий участок популяционной истории соболя и лесной куницы является очень коротким и находится в самом начале траектории и может быть случайным, а расхождение видов могло произойти ранее, за пределами реконструкции.

Однако, полученные нами датировки траекторий противоречат более ранним работам, согласно которым дивергенция соболя и лесной куницы произошла в среднем-позднем плейстоцене $\sim 0,4 - 1,1$ млн л.н. [2,8,255,334], тогда как в нашем случае приходится на конец плиоцена (~ 3 млн л.н.) - начало плейстоцена (1,89 млн л.н.) с учетом поздней границы доверительного интервала. Таким образом, для разрешения данного вопроса необходимо: (1) увеличить размер выборки для каждого вида так, чтобы образцы равномерно покрывали ареалы, и (2) произвести совместный анализ всех видов рода *Martes*, возможно с получением и использованием межвидового пангенома всего рода.

Отметим, что траектории популяционной истории кидасов с высокой значимостью коррелируют с уровнем гибридизации: чем выше доля интрогрессии, тем сильнее траектория популяционной истории отклоняется (показывает взрывной рост) от типичной для чистого вида, и наоборот. Различия траекторий у гибридов по всей видимости связаны не только с количеством, но и со слиянием гибридных участков [96].

Заключение

В рамках данной работы мы провели тщательное исследование гибридизации соболя и лесной куницы на полногеномном уровне, однако многие выявленные закономерности требуют дальнейшего изучения. Первый блок новых вопросов напрямую связан с гибридами. Мы выявили двунаправленную интрогрессию, включая замещение мтДНК лесной куницы на мтДНК соболя в образцах, полученных в том числе за пределами зоны симпатрии. Это позволяет предположить, что текущие видовые границы, возможно, устарели и нуждаются в пересмотре. Гипотетическая датировка дивергенции соболя и лесной куницы на основании траекторий популяционных историй показала заметное расхождение с результатами опубликованных ранее исследований и требует дальнейшего изучения. Учитывая вероятную древнюю интрогрессию от близкородственной гибридной системы *M. americana* × *M. caurina*, это заслуживает отдельного масштабного исследования, которое следует проводить с использованием пангеномного подхода, а не стандартных линейных референсов. Хотя в нашей работе все анализы были выполнены дважды с использованием линейных референсов соболя и лесной куницы для исключения ошибок связанных с “эффектом референса”, именно создание и использование пангеномного графа мы рассматриваем как методологическую основу будущих исследований геномики гибридизации комплекса видов *Martes*.

Во-вторых, более низкая гетерозиготность лесной куницы в сравнении с соболем оказалась неожиданной, поскольку история обоих видов предполагала противоположную картину. Однако наши образцы охватывают лишь восточную часть ареала, так что разнообразие западных популяций этого вида остается неясным. В то же время выявленный высокий уровень гетерозиготности современных соболей позволяет впервые на полногеномном уровне заявить об успехе программ массовой реинтродукции этого вида в середине XX века. Однако до сих пор не ясно, как сокращение популяций и одна из крупнейших в истории программ реинтродукции повлияли на подвидовую и популяционную структуру, и было ли разнообразие восстановлено в действительности. Наши результаты дают основания для осторожного оптимизма, однако уровень гетерозиготности соболей

до сокращения популяций в XIX–XX веках остается неизвестным. Для решения этих вопросов требуется более широкая выборка и секвенирование сотен особей из разных частей ареала, включая музейные и древние образцы.

Наконец, в ходе работы мы сделали первый шаг в исследовании проблем отмеченной в ранних работах фертильности у гибридов соболя и лесной куницы. Мы выявили инверсию р-плеча хромосомы 11 между соболем и лесной куницей, которая была подтверждена с использованием цитогенетических данных. В пределах этой инверсии мы идентифицировали два гена-кандидата, связанных с репродуктивными процессами (*SPMIP7* и *ZPBP*). Наличие и частота этой перестройки требуют дальнейшего изучения с использованием более широкой выборки на географическом и популяционном уровне. Дальнейшие изыскания требуют детального анализа последовательностей генов, транскриптомных исследований и экспериментов по скрещиванию.

Дополнительно мы окончательно определили положение харзы в систематике подсемейства *Guloninae*, доказав необходимость ее выделения в отдельный самостоятельный род *Charronia*, включающий минимум три вида: *Ch. aterrima*, *Ch. flavigula*, и *Ch. gwatkinsii*. Возможное существование дополнительных видов, а именно *Ch. lasiotis* и потенциальных островных, требует дальнейшего изучения и дополнительных данных. Для России это имеет прямое практическое значение: присвоение видового статуса дальневосточному подвиду (*Charronia aterrima*) значительно сокращает площадь его расчетного ареала, что диктует необходимость дальнейших исследований, с возможным пересмотром действующих программ охраны и усиления мониторинга численности.

Таким образом, данная работа не только углубляет понимание эволюционных процессов в семействе куньих, но и закладывает теоретическую и методологическую основу для будущих исследований в области сравнительной геномики и природоохранной биологии.

Выводы

1. В геномах комплекса *M. zibellina* × *M. martes* выявлено мультимодальное распределение гетерозиготности, которое является следствием активной гибридизации. Присутствие протяженных участков с низким генетическим разнообразием в геномах соболя свидетельствует о недавнем инбридинге и, вероятно, являются следствием нескольких бутылочных горлышек в конце XIX – начале XX века. Высокая гетерозиготность соболей (~1,7 SNP/тыс п.н.) в настоящее время подтверждает успешность реинтродукции этого вида.
2. Относительно низкая гетерозиготность *M. martes* (~0,7 SNPs/тыс п.н.) имеет древние корни и связана с популяционной историей вида на протяжении последних одного-двух миллионов лет.
3. Интрогрессия, обнаруженная не только в геномах особей из зоны перекрывания ареалов, но и из отдаленных от зоны симпатрии регионах, подтверждает активный двунаправленный поток генов между соболем и лесной куницей, следствием которой является закрепление геномных участков одного вида в геноме другого. Широкий спектр и мозаичность хромосом гибридов подтверждают, что у них не подавлен кроссинговер.
4. Подходы, основанные на методах низкого разрешения (STR, мтДНК, единичные маркеры), непригодны для определения видовой принадлежности как исследуемых, так и других близкородственных видов, между которыми происходит активная интрогрессивная гибридизация с образованием плодовитого потомства. Выявленные противоречия видовой классификации с использованием фенотипических, митохондриальных и ядерных данных говорят о неоднозначности термина "кидас" с генетической точки зрения.
5. Проведенный филогеномный анализ подтверждает необходимость выделения харзы в отдельный род *Charronia* с как минимум тремя видами: *Ch. aterrima*, *Ch. flavigula* и *Ch. gwatkinsii*.

Список литературы

1. Wilson D. E., Reeder D. M. *Mammal Species of the World: A Taxonomic and Geographic Reference*. JHU Press, 2005. 2201 с.
2. Коефли К.-Р. и др. Multigene phylogeny of the Mustelidae: Resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation // *BMC Biol.* 2008. Т. 6, № 1. С. 10.
3. Юргенсон П. Об особенностях ареалов куниц (род *Martes*) // *Бюл МОИП Отд Биол.* 1933. Т. 42, № 1. С. 62–77.
4. Nascimento F. O. do. On the correct name for some subfamilies of Mustelidae (Mammalia, Carnivora) // *Papéis Avulsos Zool. SciELO Brasil*, 2014. Т. 54. С. 307–313.
5. Wozencraft W. C. Order carnivora // *Mammal Species World Taxon. Geogr. Ref.* Johns Hopkins Univ. Press, 2005. Т. 1. С. 532–628.
6. Yu L. и др. On the phylogeny of Mustelidae subfamilies: analysis of seventeen nuclear non-coding loci and mitochondrial complete genomes // *BMC Evol. Biol.* 2011. Т. 11, № 1. С. 92.
7. Law C. J., Slater G. J., Mehta R. S. Lineage Diversity and Size Disparity in Musteloidea: Testing Patterns of Adaptive Radiation Using Molecular and Fossil-Based Methods // *Syst. Biol.* 2018. Т. 67, № 1. С. 127–144.
8. Hassanin A. и др. Evolutionary history of Carnivora (Mammalia, Laurasiatheria) inferred from mitochondrial genomes // *PLOS ONE. Public Library of Science*, 2021. Т. 16, № 2. С. e0240770.
9. Fulton T. L., Strobeck C. Molecular phylogeny of the Arctoidea (Carnivora): effect of missing data on supertree and supermatrix analyses of multiple gene data sets // *Mol. Phylogenet. Evol. Elsevier*, 2006. Т. 41, № 1. С. 165–181.
10. Sato J. J. и др. Molecular phylogeny of arctoids (Mammalia: Carnivora) with emphasis on phylogenetic and taxonomic positions of the ferret-badgers and skunks // *Zoolog. Sci. BioOne*, 2004. Т. 21, № 1. С. 111–118.
11. Samuels J. X., Cavin J. The earliest known fisher (Mustelidae), a new species from the Rattlesnake Formation of Oregon // *J. Vertebr. Paleontol. Taylor & Francis*, 2013. Т. 33, № 2. С. 448–454.
12. Colella J. P., Johnson E. J., Cook J. A. Reconciling molecules and morphology in North American *Martes* // *J. Mammal.* 2018. Т. 99, № 6. С. 1323–1335.
13. Chutipong W. и др. *Martes flavigula*. The IUCN red list of threatened species 2016: e. T41649A45212973. 2016.
14. Abramov A. V., Rozhnov V. V., Morozov P. N. Notes on mammals of the Ngoc Linh Nature Reserve (Vietnam, Kon Tum Province) // *Russ. J. Theriol.* 2006. Т. 5, № 2. С. 85–92.
15. Юдин В. Г. О необходимости внесения харзы *Martes flavigula* Boddaert, 1785 в Красную книгу Приморского края // *БИОТА И СРЕДА ПРИРОДНЫХ ТЕРРИТОРИЙ*. 2022. Т. 10, № 2. С. 37.
16. Юдин В.Г., Юдина Е.В. Харза Дальнего Востока России: монография. Владивосток: ФНЦ Биоразнообразия ДВО РАН, 2022. 244 с.
17. Рожнов В. В. Таксономические заметки о харзе *Martes flavigula* // *Зоологический Журнал. Федеральное государственное бюджетное учреждение "Российская академия наук"*, 1995. Т. 74, № 2. С. 131–138.

18. Matyuschkin E.N. Yellow-throated marten (*Martes (Charronia) flavigula* Boddaert, 1785), (Mustelidae, Carnivora) in the Russian Far East // *Lutreola Mosc.* 1993. С. 2–9.
19. Li B. и др. Mitochondrial genomes reveal the pattern and timing of marten (*Martes*), wolverine (*Gulo*), and fisher (*Pekania*) diversification // *Mol. Phylogenet. Evol.* 2014. Т. 80. С. 156–164.
20. Hosoda T. и др. Evolutionary trends of the mitochondrial lineage differentiation in species of genera *Martes* and *Mustela* // *Genes Genet. Syst.* 2000. Т. 75, № 5. С. 259–267.
21. Kim H.-N., Jo Y.-S. Complete Mitochondrial Genome of *Martes flavigula* (Carnivora: Mustelidae) and Its Phylogenetic Status in the Genus *Martes* // *Anim. Syst. Evol. Divers.* The Korean Society of Systematic Zoology, 2024. Т. 40, № 2. С. 147–149.
22. Юргенсон П. Кидас–гибрид соболя и куницы // *Труды Печоро-Илычского Государственного Заповедника.* 1947. № 5. С. 145–178.
23. Павлинин В. Тобольский соболь: ареал, очерк морфологии, проблема межвидовой гибридизации. Институт биологии Уральского филиала АН СССР, 1963.
24. Граков Н. Н. Гибридизация соболя и лесной куницы // *Бюл МОИП Отд Биол.* 1976. № 6. С. 5–15.
25. Портнова Н. Размножение кидаса // *Кролиководство И Звероводство.* 1941. Т. 6. С. 22–23.
26. Граков Н. Н. Лесная куница. 1981.
27. Граков Н. Н. О видовой самостоятельности лесной куницы и соболя // *Сб Научно-Технической Информации ВНИИОЗ.* 1974. № 46. С. 11–14.
28. Рожнов В. В. и др. Генетический анализ популяций соболя (*Martes zibellina*) и лесной куницы (*M. martes*) в районах совместного обитания на Северном Урале // *Генетика.* Федеральное государственное унитарное предприятие Академический научно ..., 2010. Т. 46, № 4. С. 553–557.
29. Рожнов В. В. и др. О соотношении фенотипа и генотипа соболя и лесной куницы в зоне симпатрии на Северном Урале // *Вестник Московского Университета Серия 16 Биология.* Федеральное государственное бюджетное образовательное учреждение высшего ..., 2013. № 3. С. 23–26.
30. Пищулина С. Л. Взаимодействие популяций лесной куницы и соболя в зоне симпатрии: генетический аспект. Ин-т проблем экологии и эволюции им. АН Северцова РАН, 2013.
31. Насимович А., Тимофеев В. Соболя. Географические особенности питания. 1973.
32. Павлов М., Корсакова И., Лавров Н. Акклиматизация охотничье-промысловых зверей и птиц в СССР // *Киров Волго-Вятское Книжн Изд-Во.* 1973.
33. Graphodatsky A., Perelman P. L., O'Brien S. J. Atlas of mammalian chromosomes. Wiley Blackwel, 2020.
34. Nie W. и др. The genome phylogeny of domestic cat, red panda and five mustelid species revealed by comparative chromosome painting and G-banding // *Chromosome Res.* 2002. Т. 10, № 3. С. 209–222.
35. Cavagna P., Menotti A., Stanyon R. Genomic homology of the domestic ferret

- with cats and humans // *Mamm. Genome*. Springer, 2000. T. 11. C. 866–870.
36. Hameister H. и др. Zoo-FISH analysis: the American mink *Mustela vison*) closely resembles the cat karyotype // *Chromosome Res.* Springer Nature BV, 1997. T. 5, № 1. C. 5.
 37. Sanger F., Nicklen S., Coulson A. R. DNA sequencing with chain-terminating inhibitors // *Proc. Natl. Acad. Sci. U. S. A.* 1977. T. 74, № 12. C. 5463–5467.
 38. Shendure J., Ji H. Next-generation DNA sequencing: 10 // *Nat. Biotechnol.* Nature Publishing Group, 2008. T. 26, № 10. C. 1135–1145.
 39. Venter J. C. и др. The Sequence of the Human Genome // *Science*. American Association for the Advancement of Science, 2001. T. 291, № 5507. C. 1304–1351.
 40. Kchouk M., Gibrat J. F., Elloumi M. Generations of Sequencing Technologies: From First to Next Generation // *Biol. Med.* 2017. T. 09, № 03.
 41. Mardis E. R. Next-Generation DNA Sequencing Methods // *Annu. Rev. Genomics Hum. Genet.* Annual Reviews, 2008. T. 9, № 1. C. 387–402.
 42. Ronaghi M., Uhlén M., Nyrén P. A Sequencing Method Based on Real-Time Pyrophosphate // *Science*. American Association for the Advancement of Science, 1998. T. 281, № 5375. C. 363–365.
 43. Rothberg J. M. и др. An integrated semiconductor device enabling non-optical genome sequencing // *Nature*. 2011. T. 475, № 7356. C. 348–352.
 44. Bentley D. R. и др. Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry // *Nature*. 2008. T. 456, № 7218. C. 53–59.
 45. Fuentes-Pardo A. P., Ruzzante D. E. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations // *Mol. Ecol.* 2017. T. 26, № 20. C. 5369–5406.
 46. Pettersson E., Lundeberg J., Ahmadian A. Generations of sequencing technologies // *Genomics*. 2009. T. 93, № 2. C. 105–111.
 47. Goodwin S., McPherson J. D., McCombie W. R. Coming of age: ten years of next-generation sequencing technologies: 6 // *Nat. Rev. Genet.* Nature Publishing Group, 2016. T. 17, № 6. C. 333–351.
 48. Heather J. M., Chain B. The sequence of sequencers: The history of sequencing DNA // *Genomics*. 2016. T. 107, № 1. C. 1–8.
 49. Salzberg S. L., Yorke J. A. Beware of mis-assembled genomes // *Bioinforma. Oxf. Engl.* 2005. T. 21, № 24. C. 4320–4321.
 50. Eid J. и др. Real-Time DNA Sequencing from Single Polymerase Molecules // *Science*. American Association for the Advancement of Science, 2009. T. 323, № 5910. C. 133–138.
 51. Feng Y. и др. Nanopore-based Fourth-generation DNA Sequencing Technology // *Genomics Proteomics Bioinformatics*. 2015. T. 13, № 1. C. 4–16.
 52. Branton D. и др. The potential and challenges of nanopore sequencing: 10 // *Nat. Biotechnol.* Nature Publishing Group, 2008. T. 26, № 10. C. 1146–1153.
 53. Clarke J. и др. Continuous base identification for single-molecule nanopore DNA sequencing: 4 // *Nat. Nanotechnol.* Nature Publishing Group, 2009. T. 4, № 4. C. 265–270.
 54. Weisenfeld N. и др. Direct determination of diploid genome sequences // *Genome Res.* 2017. T. 27.
 55. Tjoonk N. How does 10x chromium work? [Электронный ресурс] // *Single Cell Discoveries*. 2023. URL:

- <https://www.scdiscoveries.com/blog/knowledge/how-does-10x-chromium-work/> (дата обращения: 13.11.2024).
56. de Wit E., de Laat W. A decade of 3C technologies: insights into nuclear organization // *Genes Dev.* 2012. Т. 26, № 1. С. 11–24.
 57. Lieberman-Aiden E. и др. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome // *Science*. American Association for the Advancement of Science, 2009. Т. 326, № 5950. С. 289–293.
 58. Björkegren C., Baranello L. DNA Supercoiling, Topoisomerases, and Cohesin: Partners in Regulating Chromatin Architecture? 3 // *Int. J. Mol. Sci. Multidisciplinary Digital Publishing Institute*, 2018. Т. 19, № 3. С. 884.
 59. Dekker J. и др. Capturing chromosome conformation // *science*. American Association for the Advancement of Science, 2002. Т. 295, № 5558. С. 1306–1311.
 60. Korb J. O., Lee C. Genome assembly and haplotyping with Hi-C // *Nat. Biotechnol.* Nature Publishing Group, 2013. Т. 31, № 12. С. 1099–1101.
 61. Burton J. N. и др. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions // *Nat. Biotechnol.* Nature Publishing Group, 2013. Т. 31, № 12. С. 1119–1125.
 62. Lewin H. A. и др. Precision nomenclature for the new genomics // *GigaScience*. 2019. Т. 8, № giz086.
 63. Nie W. и др. Chromosomal rearrangements and karyotype evolution in carnivores revealed by chromosome painting: 1 // *Heredity*. Nature Publishing Group, 2012. Т. 108, № 1. С. 17–27.
 64. Scherthan H. и др. Comparative chromosome painting discloses homologous segments in distantly related mammals // *Nat. Genet.* 1994. Т. 6, № 4. С. 342–347.
 65. Rens W. и др. Cross-species chromosome painting // *Nat. Protoc.* Nature Publishing Group, 2006. Т. 1, № 2. С. 783–790.
 66. Totikov A. и др. Chromosome-Level Genome Assemblies Expand Capabilities of Genomics for Conservation Biology: 9 // *Genes*. Multidisciplinary Digital Publishing Institute, 2021. Т. 12, № 9. С. 1336.
 67. Joop Ouborg N., Angeloni F., Vergeer P. An essay on the necessity and feasibility of conservation genomics // *Conserv. Genet.* Springer, 2010. Т. 11. С. 643–653.
 68. Dudchenko O. и др. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000 // *bioRxiv*. Cold Spring Harbor Laboratory, 2018. С. 254797.
 69. Shafer A. B. и др. Genomics and the challenging translation into conservation practice // *Trends Ecol. Evol.* Elsevier, 2015. Т. 30, № 2. С. 78–87.
 70. Luikart G. и др. The power and promise of population genomics: from genotyping to genome typing: 12 // *Nat. Rev. Genet.* Nature Publishing Group, 2003. Т. 4, № 12. С. 981–994.
 71. Campbell M. R. и др. Development and Application of Single-Nucleotide Polymorphism (SNP) Genetic Markers for Conservation Monitoring of Burbot Populations // *Trans. Am. Fish. Soc.* 2019. Т. 148, № 3. С. 661–670.
 72. Hoban S. и др. 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. Т. 16, № 3. С. e12953.
 73. Fernandez-Fournier P., Lewthwaite J. M. M., Mooers A. Ø. Do We Need to Identify Adaptive Genetic Variation When Prioritizing Populations for Conservation?

- // *Conserv. Genet.* 2021. T. 22, № 2. С. 205–216.
74. Bijlsma R., Loeschcke V. Genetic erosion impedes adaptive responses to stressful environments // *Evol. Appl.* 2012. T. 5, № 2. С. 117–129.
 75. Leroy G. и др. Next-generation metrics for monitoring genetic erosion within populations of conservation concern // *Evol. Appl.* 2018. T. 11, № 7. С. 1066–1083.
 76. Frankham R. и др. Predicting the probability of outbreeding depression // *Conserv. Biol. J. Soc. Conserv. Biol.* 2011. T. 25, № 3. С. 465–475.
 77. Charlesworth D., Willis J. H. The genetics of inbreeding depression: 11 // *Nat. Rev. Genet.* Nature Publishing Group, 2009. T. 10, № 11. С. 783–796.
 78. Breed M. F. и др. The potential of genomics for restoring ecosystems and biodiversity // *Nat. Rev. Genet.* Nature Publishing Group UK London, 2019. T. 20, № 10. С. 615–628.
 79. Spielman D., Brook B. W., Frankham R. Most species are not driven to extinction before genetic factors impact them // *Proc. Natl. Acad. Sci. National Acad Sciences*, 2004. T. 101, № 42. С. 15261–15264.
 80. Willoughby J. R. и др. The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings // *Biol. Conserv.* Elsevier, 2015. T. 191. С. 495–503.
 81. DeWoody J. A. и др. The long-standing significance of genetic diversity in conservation // *Mol. Ecol.* Wiley Online Library, 2021. T. 30, № 17. С. 4147–4154.
 82. Jost L. и др. Differentiation measures for conservation genetics // *Evol. Appl.* Wiley Online Library, 2018. T. 11, № 7. С. 1139–1148.
 83. Kliver S. F. Whole genome approach in conservation biology and its perspectives: 3 // *Ecol. Genet.* 2021. T. 19, № 3. С. 281–298.
 84. Mayr E. *Animal Species and Evolution* // Moscow: Mir. 1968. С. 398.
 85. Tomimatsu H., Ohara M. Genetic diversity and local population structure of fragmented populations of *Trillium camschatcense* (Trilliaceae) // *Biol. Conserv.* 2003. T. 109, № 2. С. 249–258.
 86. Hanski I. *The Shrinking World: Ecological Consequences of Habitat Loss* // *Excell. Ecol.* EE 14. 2005. T. 14.
 87. Lande R., Barrowclough G. F. Effective population size, genetic variation, and their use in population management // *Viable Popul. Conserv.* 1987. T. 87. С. 87–124.
 88. Wright S. *Random Drift and the Shifting Balance Theory of Evolution* // *Mathematical Topics in Population Genetics* / под ред. Kojima K. Berlin, Heidelberg: Springer, 1970. С. 1–31.
 89. Nevo E. Genetic variation in natural populations: Patterns and theory // *Theor. Popul. Biol.* 1978. T. 13, № 1. С. 121–177.
 90. Lewontin R. *The Genetic Basis of Evolutionary Change* // Moscow: Mir. 1978.
 91. Steiner C. C. и др. Conservation genomics of threatened animal species // *Annu Rev Anim Biosci.* Annual Reviews, 2013. T. 1, № 1. С. 261–281.
 92. Renaud G. и др. Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples // *Genetics.* Oxford University Press, 2019. T. 212, № 3. С. 587–614.
 93. Weir B. S. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data.* Oxford University Press, Incorporated, 1996. 466 с.
 94. Ritland K. Estimators for pairwise relatedness and individual inbreeding coefficients // *Genet. Res.* Cambridge University Press, 1996. T. 67, № 2. С. 175–185.

95. Li H., Durbin R. Inference of human population history from individual whole-genome sequences // *Nature*. 2011. Т. 475, № 7357. С. 493–496.
96. Cahill J. A. и др. Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data // *Philos. Trans. R. Soc. B Biol. Sci. Royal Society*, 2016. Т. 371, № 1699. С. 20150138.
97. Hoffmann A. и др. A framework for incorporating evolutionary genomics into biodiversity conservation and management // *Clim. Change Responses*. Springer, 2015. Т. 2. С. 1–24.
98. Benestan L. M. и др. Conservation genomics of natural and managed populations: building a conceptual and practical framework // *Mol. Ecol.* 2016. Т. 25, № 13. С. 2967–2977.
99. Hoban S. и др. 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. Т. 4, № 3. С. 299–303.
100. Nazareno A. G. и др. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species // *Mol. Ecol. Resour.* 2017. Т. 17, № 6. С. 1136–1147.
101. Danecek P. и др. Twelve years of SAMtools and BCFtools // *Gigascience*. Oxford University Press, 2021. Т. 10, № 2. С. giab008.
102. McKenna A. и др. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data // *Genome Res.* Cold Spring Harbor Lab, 2010. Т. 20, № 9. С. 1297–1303.
103. Richter F. и др. Whole genome de novo variant identification with FreeBayes and neural network approaches // *BioRxiv*. Cold Spring Harbor Laboratory, 2020. С. 2020.03.24.994160.
104. Tomarovsky A. A. и др. Review of heterozygosity visualization approaches in the context of conservation research // *Ecol. Genet.* 2023.
105. Yakupova A. и др. Chromosome-Length Assembly of the Baikal Seal (*Pusa sibirica*) Genome Reveals a Historically Large Population Prior to Isolation in Lake Baikal: 3 // *Genes*. Multidisciplinary Digital Publishing Institute, 2023. Т. 14, № 3. С. 619.
106. Kliver S. и др. Chromosome-length genome assembly and karyotype of the endangered black-footed ferret (*Mustela nigripes*) // *J. Hered.* 2023. Т. 114, № 5. С. 539–548.
107. Li R. и др. The sequence and de novo assembly of the giant panda genome: 7279 // *Nature*. Nature Publishing Group, 2010. Т. 463, № 7279. С. 311–317.
108. Dobrynin P. и др. Genomic legacy of the African cheetah, *Acinonyx jubatus* // *Genome Biol.* 2015. Т. 16, № 1. С. 277.
109. Lindblad-Toh K. и др. Genome sequence, comparative analysis and haplotype structure of the domestic dog: 7069 // *Nature*. Nature Publishing Group, 2005. Т. 438, № 7069. С. 803–819.
110. Humble E. и др. 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. Т. 20, № 6. С. 1668–1681.
111. Hintze J. L., Nelson R. D. Violin plots: a box plot-density trace synergism // *Am. Stat.* Taylor & Francis, 1998. Т. 52, № 2. С. 181–184.
112. de Manuel M. и др. The evolutionary history of extinct and living lions // *Proc.*

- Natl. Acad. Sci. U. S. A. 2020. T. 117, № 20. C. 10927–10934.
113. Beichman A. C. и др. Aquatic Adaptation and Depleted Diversity: A Deep Dive into the Genomes of the Sea Otter and Giant Otter // *Mol. Biol. Evol.* 2019. T. 36, № 12. C. 2631–2655.
 114. de Ferran V. и др. Phylogenomics of the world's otters // *Curr. Biol.* 2022. T. 32, № 16. C. 3650-3658.e4.
 115. Derežanin L. и др. Multiple types of genomic variation contribute to adaptive traits in the mustelid subfamily Guloninae // *Mol. Ecol.* 2022. T. 31, № 10. C. 2898–2919.
 116. Totikov A. A. и др. Comparative genomics and phylogenomics of the Mustelinae lineage (Mustelidae, Carnivora). *bioRxiv*, 2025. C. 2025.07.01.662473.
 117. Gibson J., Morton N. E., Collins A. Extended tracts of homozygosity in outbred human populations // *Hum. Mol. Genet.* Oxford University Press, 2006. T. 15, № 5. C. 789–795.
 118. McQuillan R. и др. Runs of homozygosity in European populations // *Am. J. Hum. Genet.* Elsevier, 2008. T. 83, № 3. C. 359–372.
 119. Darwin C. The effects of cross and self fertilisation in the vegetable kingdom. *Ams PressInc*, 1877.
 120. Ceballos F. C. и др. Runs of homozygosity: windows into population history and trait architecture: 4 // *Nat. Rev. Genet.* Nature Publishing Group, 2018. T. 19, № 4. C. 220–234.
 121. Hoffman J. I. и др. High-throughput sequencing reveals inbreeding depression in a natural population // *Proc. Natl. Acad. Sci. Proceedings of the National Academy of Sciences*, 2014. T. 111, № 10. C. 3775–3780.
 122. Muir W. M. и др. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds // *Proc. Natl. Acad. Sci. National Acad Sciences*, 2008. T. 105, № 45. C. 17312–17317.
 123. Urbinati I. и др. Selection signatures in Canchim beef cattle // *J. Anim. Sci. Biotechnol.* Springer, 2016. T. 7. C. 1–9.
 124. Samuels D. C. и др. Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk // *Genetics*. 2016. T. 204, № 3. C. 893–904.
 125. Li N., Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data // *Genetics*. Oxford University Press, 2003. T. 165, № 4. C. 2213–2233.
 126. Rife D. C. Populations of hybrid origin as source material for the detection of linkage // *Am. J. Hum. Genet.* Elsevier, 1954. T. 6, № 1. C. 26.
 127. Guiblet W. M. и др. SmileFinder: a resampling-based approach to evaluate signatures of selection from genome-wide sets of matching allele frequency data in two or more diploid populations // *GigaScience*. Oxford University Press, 2015. T. 4, № 1. C. 2047-217X-4-1.
 128. Oleksyk T. K. и др. Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations // *PloS One*. Public Library of Science San Francisco, USA, 2008. T. 3, № 3. C. e1712.
 129. Allendorf F. W. *Genetics and the conservation of natural populations: allozymes to genomes.* Wiley Online Library, 2017.
 130. Edelman N. B., Mallet J. Prevalence and adaptive impact of introgression // *Annu.*

- Rev. Genet. Annual Reviews, 2021. T. 55, № 1. C. 265–283.
131. Feder J. L., Egan S. P., Nosil P. The genomics of speciation-with-gene-flow // Trends Genet. Elsevier, 2012. T. 28, № 7. C. 342–350.
 132. Harrison R. G., Larson E. L. Hybridization, Introgression, and the Nature of Species Boundaries // Journal of Heredity. 2014. T. 105, № S1. C. 795–809.
 133. Anderson E., Hubricht L. Hybridization in Tradescantia. III. The evidence for introgressive hybridization // Am. J. Bot. JSTOR, 1938. C. 396–402.
 134. Patterson N. и др. Ancient Admixture in Human History // Genetics. 2012. T. 192, № 3. C. 1065–1093.
 135. Green R. E. и др. A draft sequence of the Neandertal genome // science. American Association for the Advancement of Science, 2010. T. 328, № 5979. C. 710–722.
 136. Durand E. Y. и др. Testing for Ancient Admixture between Closely Related Populations // Mol. Biol. Evol. 2011. T. 28, № 8. C. 2239–2252.
 137. Pease J. B., Hahn M. W. Detection and polarization of introgression in a five-taxon phylogeny // Syst. Biol. Oxford University Press, 2015. T. 64, № 4. C. 651–662.
 138. Reich D. и др. Reconstructing Indian population history: 7263 // Nature. Nature Publishing Group, 2009. T. 461, № 7263. C. 489–494.
 139. Martin S. H., Davey J. W., Jiggins C. D. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci // Mol. Biol. Evol. 2015. T. 32, № 1. C. 244–257.
 140. Peter B. M. Admixture, Population Structure, and F-Statistics // Genetics. 2016. T. 202, № 4. C. 1485–1501.
 141. Yuan K. и др. Models, methods and tools for ancestry inference and admixture analysis: 3 // Quant. Biol. Higher Education Press, 2017. T. 5, № 3. C. 236–250.
 142. Davies N., Villablanca F. X., Roderick G. K. Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics // Trends Ecol. Evol. 1999. T. 14, № 1. C. 17–21.
 143. Padhukasahasram B. Inferring ancestry from population genomic data and its applications // Front. Genet. 2014. T. 5.
 144. Elgvin T. O. и др. The genomic mosaicism of hybrid speciation // Sci. Adv. American Association for the Advancement of Science, 2017. T. 3, № 6. C. e1602996.
 145. Lawson D. J., Falush D. Population Identification Using Genetic Data // Annu. Rev. Genomics Hum. Genet. 2012. T. 13, № 1. C. 337–361.
 146. Menozzi P., Piazza A., Cavalli-Sforza L. Synthetic Maps of Human Gene Frequencies in Europeans: These maps indicate that early farmers of the Near East spread to all of Europe in the Neolithic. // Science. American Association for the Advancement of Science, 1978. T. 201, № 4358. C. 786–792.
 147. Patterson N., Price A. L., Reich D. Population Structure and Eigenanalysis // PLOS Genet. Public Library of Science, 2006. T. 2, № 12. C. e190.
 148. Price A. L. и др. Principal components analysis corrects for stratification in genome-wide association studies: 8 // Nat. Genet. Nature Publishing Group, 2006. T. 38, № 8. C. 904–909.
 149. Elhaik E. Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated: 1 // Sci. Rep. Nature Publishing Group, 2022. T. 12, № 1. C. 14683.
 150. Abdellaoui A. и др. Population structure, migration, and diversifying selection in the Netherlands: 11 // Eur. J. Hum. Genet. Nature Publishing Group, 2013. T. 21, №

11. С. 1277–1285.
151. Abraham G., Qiu Y., Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets // *Bioinformatics*. 2017.
152. Purcell S. и др. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses // *Am. J. Hum. Genet.* 2007. Т. 81, № 3. С. 559–575.
153. Privé F. и др. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr // *Bioinformatics*. Oxford University Press, 2018. Т. 34, № 16. С. 2781–2787.
154. Bose A. и др. TeraPCA: a fast and scalable software package to study genetic variation in tera-scale genotypes // *Bioinformatics*. Oxford University Press, 2019. Т. 35, № 19. С. 3679–3683.
155. Agrawal A. и др. Scalable probabilistic PCA for large-scale genetic variation data // *PLoS Genet.* Public Library of Science San Francisco, CA USA, 2020. Т. 16, № 5. С. e1008773.
156. Chang C. C. и др. Second-generation PLINK: rising to the challenge of larger and richer datasets // *GigaScience*. 2015. Т. 4, № 1. С. s13742-015-0047–0048.
157. Novembre J., Stephens M. Interpreting principal component analyses of spatial population genetic variation // *Nat. Genet.* Nature Publishing Group US New York, 2008. Т. 40, № 5. С. 646–649.
158. McVean G. A genealogical interpretation of principal components analysis // *PLoS Genet.* Public Library of Science, 2009. Т. 5, № 10. С. e1000686.
159. François O. и др. Principal component analysis under population genetic models of range expansion and admixture // *Mol. Biol. Evol.* Oxford University Press, 2010. Т. 27, № 6. С. 1257–1268.
160. Pritchard J. K., Stephens M., Donnelly P. Inference of Population Structure Using Multilocus Genotype Data // *Genetics*. 2000. Т. 155, № 2. С. 945–959.
161. Alexander D. H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals // *Genome Res.* Cold Spring Harbor Lab, 2009. Т. 19, № 9. С. 1655–1664.
162. Raj A., Stephens M., Pritchard J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets // *Genetics*. Oxford University Press, 2014. Т. 197, № 2. С. 573–589.
163. Lawson D. J. и др. Inference of population structure using dense haplotype data // *PLoS Genet.* Public Library of Science San Francisco, USA, 2012. Т. 8, № 1. С. e1002453.
164. Kong S., Kubatko L. S. Comparative performance of popular methods for hybrid detection using genomic data // *Syst. Biol.* Oxford University Press, 2021. Т. 70, № 5. С. 891–907.
165. Bradburd G. S., Coop G. M., Ralph P. L. Inferring continuous and discrete population genetic structure across space // *Genetics*. Oxford University Press, 2018. Т. 210, № 1. С. 33–52.
166. Lawson D. J., van Dorp L., Falush D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat Commun* 9: 1–11. Springer US, 2018.
167. Anderson E. C., Thompson E. A model-based method for identifying species hybrids using multilocus genetic data // *Genetics*. Oxford University Press, 2002. Т. 160, № 3. С. 1217–1229.

168. Wiens B. J., DeCicco L. H., Colella J. P. *triangulaR*: an R package for identifying AIMs and building triangle plots using SNP data from hybrid zones // *Heredity*. 2025. T. 134, № 5. C. 251–262.
169. Wiens B. J., Colella J. P. That's Not a Hybrid: How to Distinguish Patterns of Admixture and Isolation By Distance // *Mol. Ecol. Resour.* T. n/a, № n/a. C. e14039.
170. Harrison R. G., Larson E. L. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones // *Mol. Ecol.* 2016. T. 25, № 11. C. 2454–2466.
171. Jin W. и др. Distribution of ancestral chromosomal segments in admixed genomes and its implications for inferring population history and admixture mapping // *Eur. J. Hum. Genet.* Nature Publishing Group, 2014. T. 22, № 7. C. 930–937.
172. Tang H. и др. Reconstructing genetic ancestry blocks in admixed individuals // *Am. J. Hum. Genet.* Elsevier, 2006. T. 79, № 1. C. 1–12.
173. Sankararaman S. и др. Estimating local ancestry in admixed populations // *Am. J. Hum. Genet.* Elsevier, 2008. T. 82, № 2. C. 290–303.
174. Falush D., Stephens M., Pritchard J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies // *Genetics*. Oxford University Press, 2003. T. 164, № 4. C. 1567–1587.
175. Montana G., Pritchard J. K. Statistical tests for admixture mapping with case-control and cases-only data // *Am. J. Hum. Genet.* Elsevier, 2004. T. 75, № 5. C. 771–789.
176. Patterson N. и др. Methods for high-density admixture mapping of disease genes // *Am. J. Hum. Genet.* Elsevier, 2004. T. 74, № 5. C. 979–1000.
177. Brisbin A. и др. PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations // *Hum. Biol. NIH Public Access*, 2012. T. 84, № 4. C. 343.
178. Maples B. K. и др. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference // *Am. J. Hum. Genet.* Elsevier, 2013. T. 93, № 2. C. 278–288.
179. Seldin M. F., Pasaniuc B., Price A. L. New approaches to disease mapping in admixed populations // *Nat. Rev. Genet.* Nature Publishing Group UK London, 2011. T. 12, № 8. C. 523–528.
180. Runemark A. и др. Variation and constraints in hybrid genome formation // *Nat. Ecol. Evol.* Nature Publishing Group UK London, 2018. T. 2, № 3. C. 549–556.
181. Kapli P., Yang Z., Telford M. J. Phylogenetic tree building in the genomic age // *Nat. Rev. Genet.* Nature Publishing Group UK London, 2020. T. 21, № 7. C. 428–444.
182. Delsuc F., Brinkmann H., Philippe H. Phylogenomics and the reconstruction of the tree of life // *Nat. Rev. Genet.* Nature Publishing Group UK London, 2005. T. 6, № 5. C. 361–375.
183. Young A. D., Gillung J. P. Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics // *Syst. Entomol.* Wiley Online Library, 2020. T. 45, № 2. C. 225–247.
184. Eisen J. A. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis // *Genome Res.* Cold Spring Harbor Lab, 1998. T. 8, № 3. C. 163–167.
185. O'Brien S. J., Stanyon R. Ancestral primate viewed // *Nature*. Nature Publishing Group UK London, 1999. T. 402, № 6760. C. 365–366.

186. Fitch W. M. Distinguishing homologous from analogous proteins // *Syst. Zool. Society of Systematic Zoology*, 1970. T. 19, № 2. C. 99–113.
187. Gabaldón T., Koonin E. V. Functional and evolutionary implications of gene orthology // *Nat. Rev. Genet. Nature Publishing Group UK London*, 2013. T. 14, № 5. C. 360–366.
188. Koonin E. V. Orthologs, paralogs, and evolutionary genomics // *Annu Rev Genet. Annual Reviews*, 2005. T. 39. C. 309–338.
189. Altenhoff A. M., Dessimoz C. Phylogenetic and functional assessment of orthologs inference projects and methods // *PLoS Comput. Biol. Public Library of Science San Francisco, USA*, 2009. T. 5, № 1. C. e1000262.
190. Kriventseva E. V. и др. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs // *Nucleic Acids Res.* 2019. T. 47, № D1. C. D807–D811.
191. Li L., Stoeckert C. J., Roos D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes // *Genome Res. Cold Spring Harbor Lab*, 2003. T. 13, № 9. C. 2178–2189.
192. Manni M. и др. BUSCO: Assessing Genomic Data Quality and Beyond // *Curr. Protoc.* 2021. T. 1, № 12. C. e323.
193. Waterhouse R. M. и др. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics // *Mol. Biol. Evol.* 2018. T. 35, № 3. C. 543–548.
194. Thompson J. D. и др. A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives // *PloS One. Public Library of Science San Francisco, USA*, 2011. T. 6, № 3. C. e18093.
195. Katoh K., Standley D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability // *Mol. Biol. Evol.* 2013. T. 30, № 4. C. 772–780.
196. Edgar R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity // *BMC Bioinformatics. BioMed Central*, 2004. T. 5, № 1. C. 1–19.
197. Lassmann T., Sonnhammer E. L. Kalign—an accurate and fast multiple sequence alignment algorithm // *BMC Bioinformatics. BioMed Central*, 2005. T. 6, № 1. C. 1–9.
198. Magis C. и др. T-Coffee: Tree-based consistency objective function for alignment evaluation // *Multiple Sequence Alignment Methods. Springer*, 2014. C. 117–129.
199. Löytynoja A. Phylogeny-aware alignment with PRANK // *Multiple Sequence Alignment Methods / под ред. Russell D. J. Totowa, NJ: Humana Press*, 2014. C. 155–170.
200. Ogden T. H., Rosenberg M. S. Multiple sequence alignment accuracy and phylogenetic inference // *Syst. Biol. Oxford University Press*, 2006. T. 55, № 2. C. 314–328.
201. Fernandes A. P., Nelson K., Beverley S. M. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. // *Proc. Natl. Acad. Sci. National Acad Sciences*, 1993. T. 90, № 24. C. 11608–11612.
202. Rodrigo A. G., Bergquist P. R., Bergquist P. L. Inadequate support for an evolutionary link between the Metazoa and the Fungi // *Syst. Biol. JSTOR*, 1994. T. 43, № 4. C. 578–584.
203. Thompson J. D. и др. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools // *Nucleic Acids Res.*

- Oxford University Press, 1997. T. 25, № 24. C. 4876–4882.
204. Herrmann G. и др. CONRAD: A method for identification of variable and conserved regions within proteins by scale-space filtering // *Bioinformatics*. Oxford University Press, 1996. T. 12, № 3. C. 197–203.
 205. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis // *Mol. Biol. Evol.* 2000. T. 17, № 4. C. 540–552.
 206. Li G. и др. Recombination-Aware Phylogenomics Reveals the Structured Genomic Landscape of Hybridizing Cat Species // *Mol. Biol. Evol.* 2019. T. 36, № 10. C. 2111–2126.
 207. Kubatko L. S., Degnan J. H. Inconsistency of phylogenetic estimates from concatenated data under coalescence // *Syst. Biol. Society of Systematic Zoology*, 2007. T. 56, № 1. C. 17–24.
 208. Di Franco A. и др. Lower statistical support with larger data sets: insights from the Ochrophyta radiation // *Mol. Biol. Evol.* Oxford University Press, 2022. T. 39, № 1. C. msab300.
 209. Whelan S., Liò P., Goldman N. Molecular phylogenetics: state-of-the-art methods for looking into the past // *TRENDS Genet. Elsevier*, 2001. T. 17, № 5. C. 262–272.
 210. Kozlov A. M. и др. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference // *Bioinformatics*. Oxford University Press, 2019. T. 35, № 21. C. 4453–4455.
 211. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood // *Comput. Appl. Biosci. Citeseer*, 1997. T. 13, № 5. C. 555–556.
 212. Guindon S. и др. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0 // *Syst. Biol.* Oxford University Press, 2010. T. 59, № 3. C. 307–321.
 213. Price M. N., Dehal P. S., Arkin A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix // *Mol. Biol. Evol.* Oxford University Press, 2009. T. 26, № 7. C. 1641–1650.
 214. Nguyen L.-T. и др. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies // *Mol. Biol. Evol.* Oxford University Press, 2015. T. 32, № 1. C. 268–274.
 215. Kalyaanamoorthy S. и др. ModelFinder: fast model selection for accurate phylogenetic estimates: 6 // *Nat. Methods. Nature Publishing Group*, 2017. T. 14, № 6. C. 587–589.
 216. Li S., Pearl D. K., Doss H. Phylogenetic tree construction using Markov chain Monte Carlo // *J. Am. Stat. Assoc. Taylor & Francis*, 2000. T. 95, № 450. C. 493–508.
 217. Huelsenbeck J. P., Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees // *Bioinformatics*. Oxford University Press, 2001. T. 17, № 8. C. 754–755.
 218. Höhna S. и др. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language // *Syst. Biol.* Oxford University Press, 2016. T. 65, № 4. C. 726–736.
 219. Suchard M. A. и др. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10 // *Virus Evol.* Oxford University Press, 2018. T. 4, № 1. C. vey016.
 220. Lartillot N., Lepage T., Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating // *Bioinformatics*. Oxford University Press, 2009. T. 25, № 17. C. 2286–2288.
 221. Yang Z. *Molecular evolution: a statistical approach*. Oxford University Press,

- 2014.
222. Chen M.-H., Kuo L., Lewis P. O. Bayesian phylogenetics: methods, algorithms, and applications. CRC Press, 2014.
223. Saitou N., Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. // *Mol. Biol. Evol.* 1987. T. 4, № 4. C. 406–425.
224. Baum B. R. PHYLIP: phylogeny inference package. Version 3.2. JSTOR, 1989.
225. Simonsen M., Mailund T., Pedersen C. N. S. Rapid Neighbour-Joining // *Algorithms in Bioinformatics* / под ред. Crandall K. A., Lagergren J. Berlin, Heidelberg: Springer, 2008. C. 113–122.
226. Howe K., Bateman A., Durbin R. QuickTree: building huge Neighbour-Joining trees of protein sequences // *Bioinformatics*. Oxford University Press, 2002. T. 18, № 11. C. 1546–1547.
227. Rannala B., Yang Z. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci // *Genetics*. Oxford University Press, 2003. T. 164, № 4. C. 1645–1656.
228. Liu L., Wu S., Yu L. Coalescent methods for estimating species trees from phylogenomic data // *J. Syst. Evol.* Wiley Online Library, 2015. T. 53, № 5. C. 380–390.
229. Liu L., Yu L. Estimating species trees from unrooted gene trees // *Syst. Biol.* Oxford University Press, 2011. T. 60, № 5. C. 661–667.
230. Vachaspati P., Warnow T. ASTRID: accurate species trees from internode distances // *BMC Genomics*. BioMed Central, 2015. T. 16, № 10. C. 1–13.
231. Sayyari E., Mirarab S. Anchoring quartet-based phylogenetic distances and applications to species tree reconstruction // *BMC Genomics*. BioMed Central, 2016. T. 17, № 10. C. 101–113.
232. Chifman J., Kubatko L. Quartet inference from SNP data under the coalescent model // *Bioinformatics*. Oxford University Press, 2014. T. 30, № 23. C. 3317–3324.
233. Richards E. J. и др. Variation across mitochondrial gene trees provides evidence for systematic error: how much gene tree variation is biological? // *Syst. Biol.* Oxford University Press, 2018. T. 67, № 5. C. 847–860.
234. Zhang C. и др. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees // *BMC Bioinformatics*. 2018. T. 19, № 6. C. 153.
235. Edwards S. V. и др. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics // *Mol. Phylogenet. Evol.* Elsevier, 2016. T. 94. C. 447–462.
236. Springer M. S., Gatesy J. The gene tree delusion // *Mol. Phylogenet. Evol.* Elsevier, 2016. T. 94. C. 1–33.
237. Gatesy J., Springer M. S. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum // *Mol. Phylogenet. Evol.* Elsevier, 2014. T. 80. C. 231–266.
238. Bravo G. A. и др. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics // *PeerJ*. PeerJ Inc., 2019. T. 7. C. e6399.
239. Shen X.-X., Steenwyk J. L., Rokas A. Dissecting incongruence between concatenation-and quartet-based approaches in phylogenomic data // *Syst. Biol.* Oxford University Press, 2021. T. 70, № 5. C. 997–1014.
240. Lozano-Fernandez J. A practical guide to design and assess a phylogenomic study // *Genome Biol. Evol.* Oxford University Press, 2022. T. 14, № 9. C. evac129.

241. Abramov A. V. и др. IUCN Red List of Threatened Species: *Martes foina* // IUCN Red List Threat. Species. 2015.
242. Chutipong W. и др. IUCN Red List of Threatened Species: *Martes flavigula* // IUCN Red List Threat. Species. 2015.
243. Mudappa D., Jathana D., Raman T. R. S. IUCN Red List of Threatened Species: *Martes gwatkinsii* // IUCN Red List Threat. Species. 2015.
244. Herrero J. и др. IUCN Red List of Threatened Species: *Martes martes* // IUCN Red List Threat. Species. 2015.
245. Abramov A. V., Kaneko Y., Masuda R. IUCN Red List of Threatened Species: *Martes melampus* // IUCN Red List Threat. Species. 2015.
246. Monakhov V. G. IUCN Red List of Threatened Species: *Martes zibellina* // IUCN Red List Threat. Species. 2015.
247. Helgen K., Reid F. IUCN Red List of Threatened Species: *Martes americana* // IUCN Red List Threat. Species. 2015.
248. Colella J. P. и др. Whole-genome resequencing reveals persistence of forest-associated mammals in Late Pleistocene refugia along North America's North Pacific Coast // *J. Biogeogr.* 2021. T. 48, № 5. С. 1153–1169.
249. Stone K. D., Cook J. A. Molecular evolution of holarctic martens (genus *Martes*, Mammalia: Carnivora: Mustelidae) // *Mol. Phylogenet. Evol.* Elsevier, 2002. T. 24, № 2. С. 169–179.
250. Simpson G. G. The principles of classification and a classification of mammals. American Museum of Natural History, 1945. Т. 85.
251. Sato J. J. и др. Deciphering and dating the red panda's ancestry and early adaptive radiation of Musteloidea // *Mol. Phylogenet. Evol.* 2009. T. 53, № 3. С. 907–922.
252. Anderson E. Quaternary evolution of the genus *Martes* (Carnivora, Mustelidae) // *Acta Zool. Fenn.* 1970. T. 130. С. 1–132.
253. Hosoda T. и др. Phylogenetic history of mustelid fauna in Taiwan inferred from mitochondrial genetic loci // *Can. J. Zool.* NRC Research Press, 2011. T. 89, № 6. С. 559–569.
254. Stone K. D., Flynn R. W., Cook J. A. Post-glacial colonization of northwestern North America by the forest-associated American marten (*Martes americana*, Mammalia: Carnivora: Mustelidae) // *Mol. Ecol.* 2002. T. 11, № 10. С. 2049–2063.
255. Sato J. J. и др. Phylogenetic relationships and divergence times among mustelids (Mammalia: Carnivora) based on nucleotide sequences of the nuclear interphotoreceptor retinoid binding protein and mitochondrial cytochrome b genes // *Zool. Sci. BioOne*, 2003. T. 20, № 2. С. 243–264.
256. Marmi J., López-Giráldez J. F., Domingo-Roura X. Phylogeny, evolutionary history and taxonomy of the Mustelidae based on sequences of the cytochrome b gene and a complex repetitive flanking region // *Zool. Scr.* Wiley Online Library, 2004. T. 33, № 6. С. 481–499.
257. Rozhnov V. V., Meshchersky I. G., Kholodova M. V. Molecular genetic study of marbled polecat (*Vormela peregusna*, Carnivora: Mustelidae) // *Doklady Biological Sciences.* Springer, 2006. T. 407. С. 169–172.
258. Agnarsson I., Kuntner M., May-Collado L. J. Dogs, cats, and kin: a molecular species-level phylogeny of Carnivora // *Mol. Phylogenet. Evol.* Elsevier, 2010. T. 54, № 3. С. 726–745.
259. Pierce A. J., Sukumal N., Khamcha D. A yellow-throated marten *Martes flavigula*

- feeding on a red muntjac *Muntiacus muntjak* carcass // *Small Carniv. Conserv.* 2014. Т. 51, № 1. С. 76–78.
260. PARR¹ J., Duckworth J. W. Notes on diet, habituation and sociality of Yellow-throated Marten *Martes flavigula* // *Small Carniv. Conserv.* 2007. Т. 36. С. 27–29.
261. Twining J. P., Mills C. Cooperative hunting in the yellow-throated marten (*Martes flavigula*): Evidence for the not-so-solitary marten? // *Ecosphere*. Wiley Online Library, 2021. Т. 12, № 3. С. e03398.
262. Grassman Jr L. I., Tewes M. E., Silvy N. J. Ranging, habitat use and activity patterns of binturong *Arctictis binturong* and yellow-throated marten *Martes flavigula* in north-central Thailand // *Wildl. Biol.* Wiley Online Library, 2005. Т. 11, № 1. С. 49–57.
263. Gong Y. и др. Prevalence of Varied Coat Coloration in a Yellow-Throated Marten (*Martes flavigula*) Population // *Anim. Open Access J.* MDPI. 2021. Т. 11, № 10. С. 2838.
264. Аристов А. А., Барышников Г. Ф. Млекопитающие фауны России и сопредельных территорий. Хищные и ластоногие. СПб, 2001.
265. Beklemisheva V. R. и др. Maps of Constitutive-Heterochromatin Distribution for Four *Martes* Species (Mustelidae, Carnivora, Mammalia) Show the Formative Role of Macrosatellite Repeats in Interspecific Variation of Chromosome Structure: 2 // *Genes*. Multidisciplinary Digital Publishing Institute, 2023. Т. 14, № 2. С. 489.
266. Yingxiang W. Chromosome study of yellow-throated marten (*Martes flavigula*). 1990.
267. Pocock R. I. The Oriental Yellow-throated Marten (*Lamprogale*). By RI РОЦОК, FRS, FZS // *Proceedings of the Zoological Society of London*. Wiley Online Library, 1936. Т. 106, № 2. С. 531–553.
268. Барышников Г. Ф., Абрамов А. В. Строение бакулюма (os penis) у куницевых, Mustelidae (Carnivora). Сообщение 1 // *Зоологический Журнал*. Федеральное государственное бюджетное учреждение "Российская академия наук", 1997. Т. 76, № 12. С. 1399–1410.
269. Gray J. E. A revision of the genera and species of amphisbaenians, with the descriptions of some new species now in the collections of the British Museum // *Proc. Zool. Soc. Lond.* 1865. Т. 1865. С. 442–455.
270. Огнёв С. И. Звери Восточной Европы и Северной Азии (Том 1. Насекомоядные и летучие мыши): *The Mammals of Eastern Europe and Northern Asia (Volume 1)*. Alexander Doweld, 1928.
271. Oleynikov A. Y. и др. Current population status of yellow-throated marten *Martes (Charronia) flavigula* in Russia // *Russ J Theriol.* 2022. Т. 21. С. 63–69.
272. Павлинов И. Я., Лисовский А. А. Млекопитающие России. Систематико-географический справочник // Сб. тр. Зоол. музея МГУ. М.: Тов-во научн. изд. КМК, 2012. С. 604.
273. Horsfield T. A catalogue of the Mammalia in the Museum of the Hon. East-India Company. J. & H. Cox, 1851.
274. Myers N. и др. Biodiversity hotspots for conservation priorities // *Nature*. Nature Publishing Group, 2000. Т. 403, № 6772. С. 853–858.
275. Anoop R. P. N. A sighting of Nilgiri Marten *Martes gwatkinsii* in Peppara Wildlife Sanctuary, southern Western Ghats, India // *Small Carniv. Conserv.* 2013. Т. 49. С.

- 51â.
276. Jathanna D. Ecology and conservation of small carnivores in the Western Ghats: final report submitted to CEPF/ATREE // Cent. Wildl. Stud. Bangalore. 2014.
277. Pocock R. I. и др. The Fauna of British India: including Ceylon and Burma. Taylor & Francis, 1912. Т. 6.
278. Corbet G. B., Hill J. E. The mammals of the Indomalayan region: a systematic review. oxford university press Oxford, 1992. Т. 488.
279. Wereszczuk A., Leblois R., Zalewski A. Genetic diversity and structure related to expansion history and habitat isolation: stone marten populating rural–urban habitats // BMC Ecol. 2017. Т. 17, № 1. С. 46.
280. Balestrieri A. и др. Spatial ecology of the stone marten in an Alpine area: combining camera-trapping and genetic surveys // Mammal Res. 2021. Т. 66, № 2. С. 267–279.
281. Гептнер В. Г., Наумов Н. П., Юргенсон П. Б. Млекопитающие Советского Союза. Морские коровы и хищные. Высшая школа, 1967.
282. Tsoupras A. и др. Phylogeography of *Martes foina* in Greece // Mamm. Biol. Springer, 2019. Т. 95. С. 59–68.
283. Ishii H. и др. Phylogeography of the stone marten (*Martes foina*: Mustelidae: Mammalia) in Eurasia, based on a mitochondrial DNA analysis // Mammal Res. 2023. Т. 68, № 3. С. 375–381.
284. Liu R., Nie W., Chen Y. Chromosome Study of Stone Marten (*Martes foina intermedia*) // Zool. Res. 1995. Т. 16, № 3. С. 275–279.
285. Beklemisheva V. R. и др. The Ancestral Carnivore Karyotype As Substantiated by Comparative Chromosome Painting of Three Pinnipeds, the Walrus, the Steller Sea Lion and the Baikal Seal (Pinnipedia, Carnivora) // PLOS ONE. Public Library of Science, 2016. Т. 11, № 1. С. e0147647.
286. Beklemisheva V. R. и др. Karyotype Evolution in 10 Pinniped Species: Variability of Heterochromatin versus High Conservatism of Euchromatin as Revealed by Comparative Molecular Cytogenetics: 12 // Genes. Multidisciplinary Digital Publishing Institute, 2020. Т. 11, № 12. С. 1485.
287. Graphodatsky A. S. и др. Comparative molecular cytogenetic studies in the order Carnivora: mapping chromosomal rearrangements onto the phylogenetic tree // Cytogenet. Cell Genet. 2002. Т. 96, № 1–4. С. 137–145.
288. Perelman P. L. и др. Chromosome painting shows that skunks (Mephitidae, Carnivora) have highly rearranged karyotypes // Chromosome Res. 2008. Т. 16, № 8. С. 1215–1231.
289. Perelman P. L. и др. Comparative Chromosome Painting in Carnivora and Pholidota // Cytogenet. Genome Res. 2012. Т. 137, № 2–4. С. 174–193.
290. Hagemeier E. M. Variation and relationships in North American marten // Can. Field-Nat. 1961. Т. 75, № 3. С. 122–138.
291. Павлюченко В. М. и др. Клеточное разведение соболей // М Колос. 1979. Т. 184. С. 49.
292. Колдаева Е. М. Пушные звери клеточного разведения // Генетика И Селекция Книга Первая М ФГУП Изд-Во «Известия. 2004.
293. Рожнов В. В. и др. Генетическая структура соболя (*Martes zibellina* L.) Евразии анализ распределения митохондриальных линий // Генетика. Федеральное государственное бюджетное учреждение" Российская академия

- наук", 2013. Т. 49, № 2. С. 251–251.
294. Тимофеев В. В., Надеев В. Н. Соболь. Alexander Doweld, 1955.
295. Filimonov P. A. и др. Genetic structure of native and naturalized populations of sable (*Martes zibellina* L.) of the Central Siberian Plateau and adjacent territories // *Biol. Invasions*. 2024. Т. 26, № 7. С. 2139–2151.
296. Haddad N. M. и др. Habitat fragmentation and its lasting impact on Earth's ecosystems // *Sci. Adv. American Association for the Advancement of Science*, 2015. Т. 1, № 2. С. e1500052.
297. Дубинин Н. П. Генетико-автоматические процессы и их значение для механизма органической эволюции // *Журн Экспер Биол*. 1931. Т. 7, № 5–6. С. 463–479.
298. Каштанов С. Н. и др. Географическая структура генофонда соболя (*Martes zibellina* L.): данные анализа микросателлитных локусов // *Генетика. Федеральное государственное бюджетное учреждение" Российская академия наук"*, 2015. Т. 51, № 1. С. 78–78.
299. Бобров В., Варшавский А., Хляп Л. Чужеродные виды млекопитающих в экосистемах России. Москва: Товарищество научных изданий КМК, 2008.
300. Monakhov V. G. Mass reintroduction of sable (*Martes zibellina*): achievements and problems over 55 years // *Martes Work. Group Newsl*. 2015. Т. 21. С. 25–33.
301. Саблина Т. Б. Восстановление промысловых запасов соболя в Якутии // *Труды Института биологии, выпуск 4, Якутского филиала Сибирского отделения АН СССР*. Москва: Издательство Академии наук СССР, 1958. Т. IV, № 136-58 В. С. Подсосенский пер., 21; 2-я типография: Шубинский пер., 10.
302. Государственный доклад «О состоянии и об охране окружающей среды Российской Федерации в 2021 году». Москва: Министерство природных ресурсов и экологии России, 2022.
303. Монахов В. Г. Географическая изменчивость и демографическая характеристика аборигенных и интродуцированных популяций соболя России // *Екатеринбург Б И–2002–361 С*. 2002.
304. Монахов В., Модоров М., Ранюк М. Генетические последствия реинтродукции соболя (*Martes zibellina* L.) в Западной Сибири // *Генетика. Федеральное государственное бюджетное учреждение" Российская академия наук"*, 2018. Т. 54, № 3. С. 361–365.
305. Monakhov V. G. The Results of Sable (*Martes zibellina*) Reintroduction Demonstrate the Founder Effect // *Dokl. Biol. Sci*. 2018. Т. 482. С. 194–197.
306. Каштанов С. Н. и др. Таксономический статус и генетическая идентификация соболя Алтая (*Martes zibellina averini* Vazhanov, 1943) // *Генетика. Федеральное государственное бюджетное учреждение" Издательство" Наука"*, 2018. Т. 54, № 11. С. 1327–1337.
307. Ranyuk M. и др. Genetic differentiation of autochthonous sable populations in Western and Eastern Siberia // *J. Zool. Syst. Evol. Res*. 2021. Т. 59, № 8. С. 2539–2552.
308. Монахов В. Г., Ранюк М. Н., Модоров М. В. Популяционная структура соболя Байкальской горной страны: анализ генетических и фенотипических характеристик // *Экология*. 2021. № 2. С. 143–152.
309. Boonman-Berson S., Turnhout E., van Tatenhove J. Invasive species: the categorization of wildlife in science, policy, and wildlife management // *Land Use*

- Policy. Elsevier, 2014. Т. 38. С. 204–212.
310. Genovesi P. и др. Alien mammals in Europe: updated numbers and trends, and assessment of the effects on biodiversity // *Integr. Zool.* Wiley Online Library, 2012. Т. 7, № 3. С. 247–253.
311. Каштанов С. Н. и др. Экспансия соболя (*Martes zibellina* L.) севера Среднесибирского плоскогорья в экосистемы тундры // *Генетика*. 2022. Т. 58, № 8. С. 933–944.
312. Монахов В. Г. Соболи (*Martes Zibellina*, Carnivora, Mustelidae) За Пределами Ареала На Среднем Урале: Факты За Период 2014-2016 // *Зоологический Журнал*. 2018. Т. 97, № 2. С. 224–229.
313. Zalewski A. и др. High mitochondrial DNA diversity of an introduced alien carnivore: comparison of feral and ranch American mink *Neovison vison* in Poland // *Divers. Distrib.* Wiley Online Library, 2011. Т. 17, № 4. С. 757–768.
314. Monakhov V. G. Tracing size distribution patterns in two *Martes* species across Eurasia: which populations should be protected? // *Biodivers. Conserv.* 2021. Т. 30, № 1. С. 37–53.
315. Van Leeuwen P. и др. Population genetic structure of the invasive American mink (*Mustela vison*) in France: evidence of a high genetic diversity and the existence of multiple genetic lineages // *Mamm. Biol.* 2024. Т. 104, № 6. С. 635–646.
316. Bonesi L., Palazon S. The American mink in Europe: status, impacts, and control // *Biol. Conserv.* Elsevier, 2007. Т. 134, № 4. С. 470–483.
317. Shimatani Y. и др. Genetic variation and population structure of the feral American mink (*Neovison vison*) in Nagano, Japan, revealed by microsatellite analysis // *Mammal Study.* BioOne, 2010. Т. 35, № 1. С. 1–7.
318. Mora M. и др. Genetic structure of introduced American mink (*Neovison vison*) in Patagonia: colonisation insights and implications for control and management strategies // *Wildl. Res.* CSIRO Publishing, 2018. Т. 45, № 4. С. 344–356.
319. Bobrov V. V., Warshavsky A. A., Khlyap L. A. Alien mammals in the ecosystems of Russia. Moscow: KMK Scientific Press Ltd, 2008.
320. Жигилева О. Н. и др. Генетическая изменчивость соболя *Martes zibellina* L., лесной куницы *M. martes* L. и их гибридов в Западной Сибири: полиморфизм белков и ДНК // *Генетика*. Федеральное государственное бюджетное учреждение "Российская академия наук", 2014. Т. 50, № 5. С. 581–581.
321. Li B. и др. Phylogeography of subspecies of the sable (*Martes zibellina* L.) based on mitochondrial genomes: implications for evolutionary history // *Mamm. Biol.* 2021. Т. 101, № 1. С. 105–120.
322. Kinoshita G. и др. Colonization history of the sable *Martes zibellina* (Mammalia, Carnivora) on the marginal peninsula and islands of northeastern Eurasia // *J. Mammal.* Oxford University Press US, 2015. Т. 96, № 1. С. 172–184.
323. Malyarchuk B., Derenko M., Denisova G. A mitogenomic phylogeny and genetic history of sable (*Martes zibellina*) // *Gene*. 2014. Т. 550, № 1. С. 56–67.
324. Fagel N., Mackay A. W. Weathering in the Lake Baikal watershed during the Kazantsevo (Eemian) interglacial: evidence from the lacustrine clay record // *Palaeogeogr. Palaeoclimatol. Palaeoecol.* Elsevier, 2008. Т. 259, № 2–3. С. 244–257.
325. Wetterich S. и др. Palaeoenvironmental dynamics inferred from late Quaternary permafrost deposits on Kurungnakh Island, Lena Delta, northeast Siberia, Russia // *Quat. Sci. Rev.* Elsevier, 2008. Т. 27, № 15–16. С. 1523–1540.

326. Kinoshita G. и др. Ice age land bridges to continental islands: Repeated migration of the forest-dwelling sable in northeastern Asia // *J. Biogeogr.* 2024. Т. 51, № 5. С. 924–939.
327. Sato J. J. и др. Genetic diversity of the sable (*Martes zibellina*, Mustelidae) in Russian Far East and Hokkaido inferred from mitochondrial NADH dehydrogenase subunit 2 gene sequences // *Mammal Study. BioOne*, 2011. Т. 36, № 4. С. 209–222.
328. Кассал Б. Ю., Сидоров Г. Н. Расселение соболя (*Martes zibellina*) и куницы лесной (*Martes martes*) в Омской области и биогеографические последствия их гибридизации // *Российский Журнал Биологических Инвазий. Федеральное государственное бюджетное учреждение науки Институт проблем ...*, 2013. Т. 6, № 1. С. 51–65.
329. Черников Е. М. Экология соболя (*Martes zibellina* Lunneus, 1758) в Баргузинском заповеднике. Изд-во Бурятского госуниверситета, 2006.
330. Kyle C. J., Strobeck C. Genetic homogeneity of Canadian mainland marten populations underscores the distinctiveness of Newfoundland pine martens (*Martes americana atrata*) // *Can. J. Zool. NRC Research Press Ottawa, Canada*, 2003. Т. 81, № 1. С. 57–66.
331. Kinoshita E. и др. Hybridization between the European and Asian badgers (*Meles*, Carnivora) in the Volga-Kama region, revealed by analyses of maternally, paternally and biparentally inherited genes // *Mamm. Biol. Springer*, 2019. Т. 94. С. 140–148.
332. Colella J. P. и др. Whole-genome analysis of *Mustela erminea* finds that pulsed hybridization impacts evolution at high latitudes // *Commun. Biol. Nature Publishing Group UK London*, 2018. Т. 1, № 1. С. 51.
333. Etherington G. J. и др. Extensive genome introgression between domestic ferret and European polecat during population recovery in Great Britain // *J. Hered. Oxford University Press US*, 2022. Т. 113, № 5. С. 500–515.
334. Ishida K. и др. Evolutionary history of the sable (*Martes zibellina brachyura*) on Hokkaido inferred from mitochondrial *Cytb* and nuclear *Mcl1r* and *Tcf25* gene sequences // *Acta Theriol. (Warsz.)*. 2013. Т. 58, № 1. С. 13–24.
335. Терновский Д. В., Терновская Ю. Г. Экология куницеобразных. Новосибирск: ВО «Наука». 1994.
336. Павлинин В. Н. Об ареале и морфологии лесных куниц Тюменской области // *Тр Ин-Та Биол Уральск ФАН СССР Экология Позвоночных Животных Крайнего Севера*. 1965. № 38. С. 76–91.
337. Broquet T. и др. Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*) // *Landsc. Ecol. Springer*, 2006. Т. 21. С. 877–889.
338. Clevenger A. P., Chruszcz B., Gunson K. Drainage culverts as habitat linkages and factors affecting passage by mammals // *J. Appl. Ecol. Wiley Online Library*, 2001. Т. 38, № 6. С. 1340–1349.
339. Vergara M. и др. Inferring population genetic structure in widely and continuously distributed carnivores: the stone marten (*Martes foina*) as a case study // *PloS One. Public Library of Science San Francisco, CA USA*, 2015. Т. 10, № 7. С. e0134257.
340. Larroque J. и др. Divergent landscape effects on genetic differentiation in two populations of the European pine marten (*Martes martes*) // *Landsc. Ecol.* 2016. Т. 31, № 3. С. 517–531.
341. Mergey M. и др. Identifying environmental drivers of spatial genetic structure of

- the European pine marten (*Martes martes*) // *Landsc. Ecol.* Springer, 2017. Т. 32. С. 2261–2279.
342. Ruiz-Gonzalez A. и др. Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape // *Mol. Ecol.* 2015. Т. 24, № 20. С. 5110–5129.
343. Puente A. B. la, Balmori A. Electrocutions as an important cause of mortality for a mesocarnivore // *Mammalia. De Gruyter*, 2024. Т. 88, № 6. С. 504–508.
344. Sample I., editor I. S. S. Totally stuffed: Cern's electrocuted weasel to go on display // *The Guardian*. 2017.
345. Гашев С. Н., Агешин Н. Н. Видовая диагностика соболей Саранпаульского охотничьего участка Ханты-Мансийского автономного округа // *Проблемы соболиного хозяйства России: материалы Интернет-конференции.*–Киров: ВНИИОЗ. 2006. С. 25–28.
346. Сокольский С. М., Полежаев Н. М. *Martes (Martes) zibellina*, соболь // *Фауна европейского северо-востока России. Млекопитающие.* СПб.: Наука, 1998. Т. 2, № 2. С. 89–103.
347. Шварц С. С. О возрастной структуре популяций млекопитающих // *Труды Уральского отделения Московского общества испытателей природы.* 1959. С. 3–22.
348. Monakhov V. G. *Martes zibellina* (Carnivora: Mustelidae) // *Mamm. Species.* 2011. Т. 43, № 876. С. 75–86.
349. Пищулина С. Л. Взаимодействие популяций лесной куницы и соболя в зоне симпатрии: генетический аспект. М.: ИПЭЭ РАН, 2013, 2013.
350. Monakhov V. G. *Martes martes* (Carnivora: Mustelidae) // *Mamm. Species.* 2022. Т. 54, № 1022. С. seac007.
351. Монахов В. Г., Успенская О. Д. К Морфологической Определенности Гибрида Соболя И Лесной Куницы // *Доклады Академии Наук.* 2013. Т. 448, № 6. С. 732.
352. Monakhov V. G. Species-Specific Structure of the Frontal Part of the Skull in the Sable (*Martes zibellina*) and the European Pine Marten (*Martes martes*) // *Biol. Bull.* 2021. Т. 48, № 8. С. 1434–1441.
353. Davison A. и др. Mitochondrial phylogeography and population history of pine martens *Martes martes* compared with polecats *Mustela putorius* // *Mol. Ecol.* John Wiley & Sons, Ltd, 2001. Т. 10, № 10. С. 2479–2488.
354. Modorov M. и др. Microsatellite multiplex assay for sable (*Martes zibellina*) and pine marten (*Martes martes*) // *Mammal Res.* 2020. Т. 65, № 4. С. 855–862.
355. Dawson N. G. и др. Historical biogeography sets the foundation for contemporary conservation of martens (genus *Martes*) in northwestern North America // *J. Mammal.* 2017. Т. 98, № 3. С. 715–730.
356. Harper P., McCarthy L. A. *Furbearer Management Report of Survey-inventory Activities, 1 July 2009-30 June 2012.* Alaska Department of Fish and Game, Division of Wildlife Conservation, 2015.
357. Pauli J. N. и др. Examining the uncertain origin and management role of martens on Prince of Wales Island, Alaska // *Conserv. Biol.* 2015. Т. 29, № 5. С. 1257–1267.
358. Paul T. W. *Game transplants in Alaska.* Alaska Department of Fish and Game, Division of Wildlife Conservation, 2009.
359. Dawson N. G., Cook J. A. 2. Behind the Genes: Diversification of North

- American Martens (*Martes americana* and *M. caurina*) // *Biology and Conservation of Martens, Sables, and Fishers: A New Synthesis* / под ред. Aubry К. В. и др. Cornell University Press, 2012. С. 23–38.
360. Kyle C. J., Strobeck C. Connectivity of peripheral and core populations of North American wolverines // *J. Mammal. American Society of Mammalogists*, 2002. Т. 83, № 4. С. 1141–1150.
361. Bowman J. C., Robitaille J.-F. Winter habitat use of American martens *Martes americana* within second-growth forest in Ontario, Canada // *Wildl. Biol. Wiley Online Library*, 1997. Т. 3, № 2. С. 97–105.
362. Hill J. E. и др. Quantifying anthropogenic wolf mortality in relation to hunting regulations and landscape attributes across North America // *Ecol. Evol.* 2022. Т. 12, № 5. С. e8875.
363. Fortin J. K. и др. Impacts of human recreation on brown bears (*Ursus arctos*): a review and new management tool // *PLoS One. Public Library of Science San Francisco, CA USA*, 2016. Т. 11, № 1. С. e0141983.
364. Fogarty R. и др. Trapping mortality accelerates the decline of the fisher, an endangered mesocarnivore, in British Columbia, Canada // *Endanger. Species Res.* 2022.
365. Colella J. P. и др. Ecological displacement in a Rocky Mountain hybrid zone informs management of North American martens (*Martes*) // *Landsc. Ecol.* 2024. Т. 39, № 7. С. 125.
366. Thompson I. D. и др. Improved insights into use of habitat by American martens // *Biol. Conserv. Martens Sables Fish. Cornell Univ. Press Ithaca N. Y.* 2012. С. 209–230.
367. Colella J. P. и др. Implications of introgression for wildlife translocations: the case of North American martens // *Conserv. Genet. Springer*, 2019. Т. 20, № 2. С. 153–166.
368. Manlick P. J. и др. Stable isotopes reveal limited Eltonian niche conservatism across carnivore populations // *Funct. Ecol.* 2019. Т. 33, № 2. С. 335–345.
369. Merriam C. H., Stejneger L. Results of a biological survey of the San Francisco Mountain region and desert of the Little Colorado, Arizona. US Government Printing Office, 1890. № 3.
370. Wright S. Isolation by distance under diverse systems of mating // *Genetics. Oxford University Press*, 1946. Т. 31, № 1. С. 39.
371. Kyle C. J., Davis C. S., Strobeck C. Microsatellite analysis of North American pine marten (*Martes americana*) populations from the Yukon and Northwest Territories // *Can. J. Zool. NRC Research Press Ottawa, Canada*, 2000. Т. 78, № 7. С. 1150–1157.
372. Wasserman T. N. и др. Spatial scaling and multi-model inference in landscape genetics: *Martes americana* in northern Idaho // *Landsc. Ecol. Springer*, 2010. Т. 25. С. 1601–1612.
373. Powell R. A. и др. Evaluating translocations of martens, sables, and fishers // *Biol. Conserv. Martens Sables Fish. New Synth. KB Aubry WJ Zielinski MG Raphael G Proulx SW Buskirk Eds Cornell Univ. Press Ithaca N. Y.* 2012. С. 93–137.
374. Small M. P., Stone K. D., Cook J. A. American marten (*Martes americana*) in the Pacific Northwest: population differentiation across a landscape fragmented in time and space // *Mol. Ecol. Wiley Online Library*, 2003. Т. 12, № 1. С. 89–103.
375. Carr S. M., Hicks S. A. Are there two species of marten in North America? Genetic and evolutionary relationships within *Martes* // *Martes Taxon. Ecol. Tech.*

- Manag. Provincial Museum of Alberta, 1997. C. 15–28.
376. Grauer J. A. и др. Unexpected genetic composition of a reintroduced carnivore population // *Biol. Conserv.* 2017. T. 215. C. 246–253.
377. Imaizumi Y., Miyamoto T. Coloured Illustrations of the Mammals of Japan, Illus. by Takashi Miyamoto. Hoikusha, 1960.
378. Masuda R. *Martes melampus* (Wagner, 1840) // *Wild Mamm. Jpn.* SHOUKADOH Book Sellers and the Mammalogical Society of Japan, 2009. C. 250–251.
379. Proulx G. Martens and Fishers (*Martes*) in Human-Altered Environments. 2005.
380. Otani T. Seed dispersal by Japanese marten *Martes melampus* in the subalpine shrubland of northern Japan // *Ecol. Res.* Springer, 2002. T. 17, № 1. C. 29–38.
381. Wilson D. E., Wilson D. E., Mittermeier R. A. Handbook of the mammals of the world, volume 2: hoofed mammals. Lynx Ediciones Barcelona, Spain, 2011.
382. Kamada S. и др. Genetic distinctness and variation in the Tsushima Islands population of the Japanese marten, *Martes melampus* (Carnivora: Mustelidae), revealed by microsatellite analysis // *Zoolog. Sci. BioOne*, 2012. T. 29, № 12. C. 827–833.
383. Hosoda T. и др. Restriction site polymorphism in the ribosomal DNA of eight species of Canidae and Mustelidae // *Cytologia* (Tokyo). Japan Mendel Society, International Society of Cytology, 1993. T. 58, № 2. C. 223–230.
384. Ellerman J. R., Morrison-Scott T. C. S. Checklist of Palearctic and Indian mammals 1758 to 1946. 1951.
385. Sato J. J., Yasuda S. P., Hosoda T. Genetic Diversity of the Japanese Marten (*Martes melampus*) and Its Implications for the Conservation Unit // *Zoolog. Sci.* Zoological Society of Japan, 2009. T. 26, № 7. C. 457–466.
386. Kurose N. и др. Intraspecific Variation of Mitochondrial Cytochrome b Gene Sequences of the Japanese Marten *Martes melampus* and the Sable *Martes zibellina* (Mustelidae, Carnivora, Mammalia) in Japan // *Zoolog. Sci.* Zoological Society of Japan, 1999. T. 16, № 4. C. 693–700.
387. Hosoda T. и др. Genetic relationships within and between the Japanese marten *Martes melampus* and the sable *M. zibellina*, based on variation of mitochondrial DNA and nuclear ribosomal DNA // *Mammal Study.* THE MAMMAL SOCIETY OF JAPAN, 1999. T. 24, № 1. C. 25–33.
388. Hoshino L. и др. Broad-leaved forest selection of the Japanese marten (*Martes melampus*) in central Japan revealed by camera trapping // *Mammal Study.* BioOne, 2014. T. 39, № 3. C. 163–166.
389. Nakamura T. Seasonal changes in food habits of Japanese martens in Hinode-cho and Akiruno-shi, Tokyo // *Wildl. Conserv. Jpn.* 2001. T. 6. C. 15–24.
390. Abe H. A Guide to the Mammals of Japan // No Title. 2008. C. 108.
391. Anderson E. Evolution, prehistoric distribution, and systematics of *Martes* // *Martens Sables Fish.* Biol. Conserv. Cornell University Press, 1994. C. 13–25.
392. Masuda R., Yoshida M. C. A molecular phylogeny of the family Mustelidae (Mammalia, Carnivora), based on comparison of mitochondrial cytochrome b nucleotide sequences. // *Zoolog. Sci.* 1994. T. 11, № 4. C. 605–612.
393. Murakami T., Asano M., Ohtaishi N. Mitochondrial DNA variation in the Japanese marten *Martes* // *Jpn. J. Vet. Res.* Citeseer, 2004. T. 51, № 3–4. C. 135–142.
394. Inoue T. и др. Mitochondrial DNA control region variations in the sable *Martes zibellina* of Hokkaido Island and the Eurasian continent, compared with the Japanese

- marten *M. melampus* // *Mammal Study*. BioOne, 2010. T. 35, № 3. C. 145–155.
395. Liu G. и др. First Draft Genome of the Sable, *Martes zibellina* // *Genome Biol. Evol.* 2020. T. 12, № 3. C. 59–65.
396. Shajii A., Numanagić I., Berger B. Latent variable model for aligning barcoded short-reads improves downstream analyses // *bioRxiv*. Cold Spring Harbor Laboratory, 2017. C. 220236.
397. Andrews S. FastQC: a quality control tool for high throughput sequence data // *Babraham Bioinforma.* Babraham Inst. Camb. U. K. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010.
398. Kliver S. и др. KrATER: K-mer analysis tool easy to run // *Proceedings of 8-th International Moscow Conference on Computational Molecular Biology*. 2017. C. 128.
399. Marçais G., Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers // *Bioinformatics*. 2011. T. 27, № 6. C. 764–770.
400. Ranallo-Benavidez T. R., Jaron K. S., Schatz M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes: 1 // *Nat. Commun.* Nature Publishing Group, 2020. T. 11, № 1. C. 1432.
401. Starostina E. и др. Cookiecutter: a tool for kmer-based read filtering and extraction // *BioRxiv*. Cold Spring Harbor Laboratory, 2015. C. 024679.
402. Bolger A. M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data // *Bioinformatics*. 2014. T. 30, № 15. C. 2114–2120.
403. Durand N. C. и др. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments // *Cell Syst.* Elsevier, 2016. T. 3, № 1. C. 95–98.
404. Dudchenko O. и др. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds // *Science*. American Association for the Advancement of Science, 2017. T. 356, № 6333. C. 92–95.
405. Guan D. и др. Identifying and removing haplotypic duplication in primary genome assemblies // *Bioinformatics*. Oxford Academic, 2020. T. 36, № 9. C. 2896–2898.
406. Levy Karin E., Mirdita M., Söding J. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics // *Microbiome*. 2020. T. 8, № 1. C. 48.
407. Wheeler T. J., Eddy S. R. nhmmer: DNA homology search with profile HMMs // *Bioinformatics*. 2013. T. 29, № 19. C. 2487–2489.
408. Benson G. Tandem repeats finder: a program to analyze DNA sequences // *Nucleic Acids Res.* 1999. T. 27, № 2. C. 573–580.
409. Morgulis A. и др. WindowMasker: window-based masker for sequenced genomes // *Bioinforma. Oxf. Engl.* 2006. T. 22, № 2. C. 134–141.
410. Chen N. Using Repeat Masker to identify repetitive elements in genomic sequences // *Curr. Protoc. Bioinforma.* Wiley Online Library, 2004. T. 5, № 1. C. 4–10.
411. Storer J. и др. The Dfam community resource of transposable element families, sequence models, and genome annotations // *Mob. DNA*. 2021. T. 12, № 1. C. 2.
412. Brûna T., Gabriel L., Hoff K. J. Navigating Eukaryotic Genome Annotation Pipelines: A Route Map to BRAKER, Galba, and TSEBRA: arXiv:2403.19416. arXiv, 2024.
413. O’Brien D. и др. The genome sequence of the European pine marten, *Martes martes* (Linnaeus, 1758) // *Wellcome Open Res.* 2024. T. 9. C. 325.
414. Xia T. и др. Insights into cold tolerance in sable (*Martes zibellina*) from the

- adaptive evolution of lipid metabolism // *Mamm. Biol.* 2021. T. 101, № 6. С. 861–870.
415. Mei X. и др. A chromosome-level genome assembly of the yellow-throated marten (*Martes flavigula*) // *Sci. Data.* Nature Publishing Group, 2023. T. 10, № 1. С. 216.
416. Kuznetsov D. и др. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity // *Nucleic Acids Res.* 2023. T. 51, № D1. С. D445–D451.
417. Brůna T., Lomsadze A., Borodovsky M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes // *Genome Res.* 2024. T. 34, № 5. С. 757–768.
418. Stanke M. и др. AUGUSTUS: ab initio prediction of alternative transcripts // *Nucleic Acids Res.* 2006. T. 34, № suppl_2. С. W435–W439.
419. Stanke M. и др. Using native and syntenically mapped cDNA alignments to improve de novo gene finding // *Bioinformatics.* 2008. T. 24, № 5. С. 637–644.
420. Cantalapiedra C. P. и др. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale // *Mol. Biol. Evol.* 2021. T. 38, № 12. С. 5825–5829.
421. Huerta-Cepas J. и др. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses // *Nucleic Acids Res.* 2019. T. 47, № D1. С. D309–D314.
422. Indukaev F. Supervenn python package. Zenodo, 2021.
423. Armstrong J. и др. Progressive Cactus is a multiple-genome aligner for the thousand-genome era: 7833 // *Nature.* Nature Publishing Group, 2020. T. 587, № 7833. С. 246–251.
424. Krasheninnikova K. и др. halSynteny: a fast, easy-to-use conserved synteny block construction method for multiple whole-genome alignments // *GigaScience.* 2020. T. 9, № 6. С. g1aa047.
425. Romanenko S. A. и др. Integration of fluorescence in situ hybridization and chromosome-length genome assemblies revealed synteny map for guinea pig, naked mole-rat, and human // *Sci. Rep.* Nature Publishing Group, 2023. T. 13, № 1. С. 21055.
426. Quinlan A. R., Hall I. M. BEDTools: a flexible suite of utilities for comparing genomic features // *Bioinformatics.* 2010. T. 26, № 6. С. 841–842.
427. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014.
428. Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform // *Bioinformatics.* 2009. T. 25, № 14. С. 1754–1760.
429. Li H. и др. The Sequence Alignment/Map format and SAMtools // *Bioinformatics.* 2009. T. 25, № 16. С. 2078–2079.
430. Pedersen B. S., Quinlan A. R. Mosdepth: quick coverage calculation for genomes and exomes // *Bioinformatics.* 2018. T. 34, № 5. С. 867–868.
431. Macdonald P. mixdist: Finite Mixture Distribution Models. 2018.
432. Hilbe J. M. Negative binomial regression. Cambridge University Press, 2011.
433. Blischak P. D. и др. HyDe: A Python Package for Genome-Scale Hybridization Detection // *Syst. Biol.* 2018. T. 67, № 5. С. 821–829.
434. Natali C. и др. Characterization of 13 polymorphic microsatellite loci in the European pine marten *Martes martes* // *Conserv. Genet. Resour.* 2010. T. 2, № 1. С. 397–399.
435. Basto M. P. и др. Isolation and characterization of 13 tetranucleotide microsatellite loci in the Stone marten (*Martes foina*) // *Conserv. Genet. Resour.* 2010. T. 2, № 1. С.

- 317–319.
436. Davis C. S., Strobeck C. Isolation, variability, and cross-species amplification of polymorphic microsatellite loci in the family Mustelidae // *Mol. Ecol.* 1998. T. 7, № 12. С. 1776–1778.
437. Fleming M. A., Ostrander E. A., Cook J. A. Microsatellite markers for American mink (*Mustela vison*) and ermine (*Mustela erminea*) // *Mol. Ecol.* 1999. T. 8, № 8. С. 1352–1355.
438. Domingo-Roura X. Genetic distinction of marten species by fixation of a microsatellite region // *J. Mammal. American Society of Mammalogists*, 2002. T. 83, № 3. С. 907–912.
439. Vincent I. R., Farid A., Otieno C. J. Variability of thirteen microsatellite markers in American mink (*Mustela vison*) // *Can. J. Anim. Sci.* 2003. T. 83, № 3. С. 597–599.
440. Gardner S. N., Slezak T. Simulate_PCR for amplicon prediction and annotation from multiplex, degenerate primers and probes // *BMC Bioinformatics.* 2014. T. 15, № 1. С. 237.
441. Sadedin S. P., Oshlack A. Bazam: a rapid method for read extraction and realignment of high-throughput sequencing data // *Genome Biol.* 2019. T. 20, № 1. С. 78.
442. Willems T. и др. Genome-wide profiling of heritable and de novo STR variations // *Nat. Methods.* Nature Publishing Group, 2017. T. 14, № 6. С. 590–592.
443. Jakobsson M., Rosenberg N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure // *Bioinformatics.* 2007. T. 23, № 14. С. 1801–1806.
444. Clavijo B. J. и др. W2RAP: a pipeline for high quality, robust assemblies of large complex genomes from short read data // *BioRxiv.* Cold Spring Harbor Laboratory, 2017. С. 110999.
445. Alonge M. и др. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing // *Genome Biol.* 2022. T. 23, № 1. С. 258.
446. Huerta-Cepas J., Serra F., Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data // *Mol. Biol. Evol.* 2016. T. 33, № 6. С. 1635–1638.
447. Meng G. и др. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization // *Nucleic Acids Res.* 2019. T. 47, № 11. С. e63–e63.
448. Donath A. и др. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes // *Nucleic Acids Res.* 2019. T. 47, № 20. С. 10543–10552.
449. Manakhov A. D. и др. Genome Analysis of Sable Fur Color Links a Lightened Pigmentation Phenotype to a Frameshift Variant in the Tyrosinase-Related Protein 1 Gene: 2 // *Genes.* Multidisciplinary Digital Publishing Institute, 2021. T. 12, № 2. С. 157.
450. Li B., Wu D., Malyarchuk B. Complete mitochondrial genome of European pine marten, *Martes martes* // *Mitochondrial DNA.* 2014. T. 25, № 5. С. 372–373.
451. Hua Y. и др. Complete mitochondrial genome reveals the phylogenetic relationship of sable *Martes zibellina linkouensis* // *Mitochondrial DNA Part A.* Taylor & Francis, 2017. T. 28, № 2. С. 263–264.
452. Xu C. и др. The complete mitochondrial genome of sable, *Martes zibellina* //

- Mitochondrial DNA. Taylor & Francis, 2012. Т. 23, № 3. С. 167–169.
453. Capella-Gutiérrez S., Silla-Martínez J. M., Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses // *Bioinforma. Oxf. Engl.* 2009. Т. 25, № 15. С. 1972–1973.
454. Fu L. и др. CD-HIT: accelerated for clustering the next-generation sequencing data // *Bioinformatics.* 2012. Т. 28, № 23. С. 3150–3152.
455. Leigh J. W., Bryant D. POPART: full-feature software for haplotype network construction // *Methods Ecol. Evol.* 2015. Т. 6, № 9. С. 1110–1116.
456. Bergeron L. A. и др. Evolution of the germline mutation rate across vertebrates // *Nature.* Nature Publishing Group, 2023. Т. 615, № 7951. С. 285–291.
457. Tamazian G. и др. Draft de novo Genome Assembly of the Elusive Jaguarundi, *Puma yagouaroundi* // *J. Hered.* 2021. Т. 112, № 6. С. 540–548.
458. Tomarovsky A. и др. Chromosome-length genome assembly of the stone marten (*Martes foina*, Mustelidae): A new view on one of the cornerstones in carnivore cytogenetics // *J. Hered. Oxford Academic*, 2025. Т. 116, № 4. С. 548–557.
459. Partanen J. и др. Escaped From Fur Farms-Now Not Only The American Mink And Blue Artic Fox, But Also The Russian Sable // *Vestn. Okhotovedeniya.* 2020. Т. 17, № 4. С. 293–299.
460. Wilson D. E., Reeder D. M. Mammal species of the world: a taxonomic and geographic reference. JHU press, 2005. Т. 1.
461. СССР У. делами С. О натуральном налоге на пушнину // *Собрание Узаконений И Распоряжений Правительства За 1921 Г Москва.* 1944. Т. 61. С. 430–431.
462. Twining J. P. и др. All forests are not equal: population demographics and denning behaviour of a recovering small carnivore in human modified landscapes // *Wildl. Biol. Wiley Online Library*, 2020. Т. 2020, № 4. С. 1–10.
463. Kretschmer E. J., Olsen J. B., Wenburg J. K. Characterization of eight microsatellite loci in Sea Otter, *Enhydra lutris*, and cross-species amplification in other Mustelidae // *Conserv. Genet.* 2009. Т. 10, № 3. С. 775–777.
464. Larson S. и др. Loss of genetic diversity in sea otters (*Enhydra lutris*) associated with the fur trade of the 18th and 19th centuries // *Mol. Ecol. Wiley Online Library*, 2002. Т. 11, № 10. С. 1899–1903.
465. Secomandi S. и др. Pangenome graphs and their applications in biodiversity genomics // *Nat. Genet. Nature Publishing Group*, 2025. Т. 57, № 1. С. 13–26.
466. Андрианов Б. В. и др. Изменчивость митохондриального генома domestцированного соболя (*Martes zibellina*) // *Генетика. Федеральное государственное бюджетное учреждение" Российская академия наук"*, 2012. Т. 48, № 4. С. 529–529.
467. Gooley R. M. и др. Comparison of genomic diversity and structure of sable antelope (*Hippotragus niger*) in zoos, conservation centers, and private ranches in North America // *Evol. Appl.* 2020. Т. 13, № 8. С. 2143–2154.
468. Abramov A. V., Tumanov I. L. Sexual dimorphism in the skull of the European mink *Mustela lutreola* from NW part of Russia // *Acta Theriol. (Warsz.)*. 2003. Т. 48, № 2. С. 239–246.
469. Абрамов А. В., Пузаченко А. Ю. О вероятной гибридизации между *Meles meles* и *Meles leucurus* (Carnivora, Mustelidae) на границе видовых ареалов в Западном Тянь-Шане. Млекопитающие горных территорий. Материалы

- международной конференции. Нальчик, 13–18 августа 2007 г. Товарищество научных изданий КМК // Москва. 2007. С. 4–7.
470. McErlain-Ward P. и др. Martens, Sables and Fishers' Biology and Conservation // *J. Anim. Ecol.* 1996. Т. 65. С. 130.
471. Patterson B. D. и др. On the nomenclature of the American clade of weasels (Carnivora: Mustelidae) // *J. Anim. Divers.* Ilostan university, 2021. Т. 3, № 2. С. 1–8.
472. Abramov A. V. A taxonomic review of the genus *Mustela* (Mammalia, Carnivora) // *Zoosystematica Ross.* 2000. Т. 8, № 2. С. 357–364.
473. Ruiz-González A. и др. Phylogeography of the forest-dwelling European pine marten (*Martes martes*): new insights into cryptic northern glacial refugia // *Biol. J. Linn. Soc.* 2013. Т. 109, № 1. С. 1–18.
474. Lynch L. M. Fossil Calibration of Mitochondrial Phylogenetic Relationships of North American Pine Martens, *Martes*, Suggests an Older Divergence of *M. americana* and *M. caurina* than Previously Hypothesized // *J. Mamm. Evol.* 2020. Т. 27, № 3. С. 535–548.
475. Stojak J., Jędrzejewska B. Extinction and replacement events shaped the historical biogeography of Arctic mammals in Europe: new models of species response // *Mammal Rev.* Wiley Online Library, 2022. Т. 52, № 4. С. 507–518.
476. Hu A. и др. Influence of Bering Strait flow and North Atlantic circulation on glacial sea-level changes // *Nat. Geosci.* Nature Publishing Group UK London, 2010. Т. 3, № 2. С. 118–121.

Приложения

Приложение 1. Количество прочтений до и после фильтрации для каждого образца, оцененные на основании 23-меров покрытия и размеры генома и доля, используемая для случайной подвыборки прочтений (даунсемплинга) для достижения 22-х кратного покрытия.

ID	NCBI ID	Пол	Длина прочтений	Кол-во прочтений до фильтрации	Кол-во прочтений после фильтрации	%GC до фильтрации	%GC после фильтрации	Доля сохраненных прочтений, %	Покрытие генома (Genomescope2)	Размер генома (Genomescope2), млрд п.н.	Доля даунсемплинга *
10xmmar	SRR22412800	F	128+150	633.01	578.26	41	41	91.35	49.62	2.45	0.44
S44	SRR28749753	F	150+150	380.48	369.50	43	43	97.11	37.32	2.49	0.59
S46	SRR28749752	F	150+150	427.28	417.27	43	43	97.66	41.6	2.53	0.53
S49	SRR28749751	M	150+150	291.26	284.51	41	41	97.68	28.92	2.48	0.76
S50	SRR28749750	F	150+150	250.61	244.71	42	42	97.65	24.36	2.54	0.9
T149	SRR28749749	M	150+150	351.28	343.66	42	42	97.83	35.18	2.47	0.63
T151	SRR28749748	M	150+150	377.11	367.49	42	42	97.45	37.36	2.48	0.59
10xmzib	SRR22412799	F	128+150	619.13	566.65	42	42	91.52	49.7	2.4	0.44
china	PRJNA495455	F	120+120	977.18	955.63	42	42	97.79	76.78	2.43	0.29
S26	SRR28749757	F	150+150	269.04	261.32	42	42	97.13	25.32	2.53	0.87
T104	SRR28749730	M	150+150	275.03	266.11	41	41	96.76	27.74	2.41	0.79
T118	SRR28749729	M	150+150	352.81	342.09	42	42	96.96	35.64	2.41	0.62
T148	SRR28749728	F	150+150	294.84	286.42	42	42	97.15	29.68	2.42	0.74
T150	SRR28749755	F	150+150	341.00	331.36	43	43	97.17	33.76	2.47	0.65
T18	SRR28749745	M	150+150	356.65	345.64	43	43	96.91	36.14	2.4	0.61
T194	SRR28749754	F	150+150	214.47	207.90	42	42	96.93	21.24	2.45	NA

T26	SRR28749734	F	150+150	319.08	309.51	43	43	97	31.82	2.44	0.69
T50	SRR28749733	M	150+150	437.93	423.14	43	43	96.62	44.02	2.43	0.5
T72	SRR28749732	M	150+150	361.13	349.85	43	43	96.88	36.68	2.39	0.6
T8	SRR28749756	M	150+150	366.90	357.24	43	43	97.37	36.88	2.44	0.6
T90	SRR28749731	F	150+150	232.63	225.76	43	43	97.05	22.98	2.47	0.96
T24	SRR28749747	M	150+150	376.90	365.39	43	43	96.95	36.48	2.48	0.6
T76	SRR28749746	F	150+150	307.97	302.50	43	43	98.22	30.14	2.5	0.73
T77	SRR28749744	M	150+150	295.04	289.54	43	43	98.14	29.06	2.48	0.76
T78	SRR28749743	F	150+150	312.95	306.79	43	43	98.03	30.8	2.48	0.71
T79	SRR28749742	F	150+150	253.90	245.87	43	43	96.84	24.28	2.49	0.91
T81	SRR28749741	M	150+150	305.17	299.30	43	43	98.08	30.12	2.48	0.73
T82	SRR28749740	F	150+150	370.46	363.71	43	43	98.18	36.58	2.48	0.6
T83	SRR28749739	M	150+150	317.82	311.74	43	43	98.08	31.62	2.46	0.7
T84	SRR28749738	M	150+150	292.12	287.13	43	43	98.29	29.26	2.45	0.75
T85	SRR28749737	F	150+150	315.36	309.73	43	43	98.21	31.26	2.47	0.7
T86	SRR28749736	M	150+150	251.00	246.61	43	43	98.25	25.16	2.45	0.87
T87	SRR28749735	M	150+150	471.25	462.92	43	43	98.23	47.18	2.45	0.47
10xmfoi	SRR22412409	M	128+150	564.76	522.94	42	41	92.59	46.8	2.37	0.47
10xmfla	SRR22412516	M	128+150	637.84	572.26	43	43	89.72	48.44	2.46	0.45
SRR11575343	SRR11575343	M	150+150	368.09	363.26	46	46	98.69	36.34	2.43	0.61
SRR11575346	SRR11575346	M	150+150	368.86	362.97	45	45	98.40	31.3	2.83	0.70
SRR11575348	SRR11575348	M	150+150	267.84	260.25	42	42	97.17	26.16	2.36	0.84
SRR11575351	SRR11575351	F	150+150	366.12	345.32	43	42	94.32	32.16	2.47	0.68
SRR11575352	SRR11575352	M	150+150	261.15	249.92	44	44	95.70	22.4	2.61	0.98

* *NA* – незатронутые образцы.

Приложение 2. Доля основных категорий повторяющихся элементов в геномных сборках *M. zibellina*, *M. martes*, *M. foina* и *M. flavigula*.

Тип повторов	<i>M. zibellina</i>		<i>M. martes</i>		<i>M. foina</i>		<i>M. flavigula</i>	
	Длина, п.о.	Длина, %	Длина, п.о.	Длина, %	Длина, п.о.	Длина, %	Длина, п.о.	Длина, %
Ретроэлементы	870871585	36.37	878539291	36.54	875539264	36.19	912279015	36.86
SINEs	237444556	9.92	238992690	9.94	238158404	9.84	246594632	9.96
Penelope	57707	0.00	56129	0.00	56066	0.00	55804	0.00
LINEs	518969807	21.67	524763476	21.83	521971290	21.57	546947504	22.10
CRE/SLACS	0	0.00	0	0.00	0	0.00	0	0.00
L2/CR1/Rex	88603222	3.70	88635894	3.69	88385088	3.65	88662032	3.58
R1/LOA/Jockey	0	0.00	0	0.00	0	0.00	0	0.00
R2/R4/NeSL	83901	0.00	82868	0.00	85171	0.00	82828	0.00
RTE/Bov-B	2875393	0.12	2879106	0.12	2859653	0.12	2876245	0.12
L1/CIN4	427361651	17.85	433119245	18.01	430594586	17.80	455280984	18.39
LTR-элементы	114457222	4.78	114783125	4.77	115409570	4.77	118736879	4.80
BEL/Pao	0	0.00	0	0.00	0	0.00	0	0.00
Ty1/Copia	0	0.00	0	0.00	0	0.00	0	0.00
Gypsy/DIRS1	3450189	0.14	3459917	0.14	3425117	0.14	3443291	0.14
Retroviral	108815544	4.54	109130344	4.54	109791419	4.54	113099281	4.57
ДНК-транспозоны	69076893	2.88	69092488	2.87	69012541	2.85	69741546	2.82
hobo-Activator	48344438	2.02	48360389	2.01	48291179	2.00	48800534	1.97
Tc1-IS630-Pogo	19325274	0.81	19326908	0.80	19322784	0.80	19532221	0.79
En-Spm	0	0.00	0	0.00	0	0.00	0	0.00
MULE-MuDR	65089	0.00	61858	0.00	58406	0.00	62517	0.00
PiggyBac	198772	0.01	200754	0.01	199207	0.01	203349	0.01
Tourist/Harbinger	62378	0.00	63052	0.00	61753	0.00	61914	0.00
Другие (Mirage, P-element, Transib)	0	0.00	0	0.00	0	0.00	0	0.00
Rolling-circles	342852	0.01	340629	0.01	334299	0.01	357197	0.01
Неклассифицированные диспергированные повторы	544264	0.02	544192	0.02	538075	0.02	551664	0.02
Общее кол-во повторов	940550449	39.28	948232100	39.44	945145946	39.07	982628029	39.70
Малые РНК	175718980	7.34	177343940	7.38	176650893	7.30	184536005	7.46
Сателлиты	0	0.00	0	0.00	0	0.00	0	0.00

Простые повторы	30352239	1.27	30480662	1.27	31282371	1.29	30857877	1.25
Последовательности низкой сложности	5544570	0.23	5614311	0.23	5632322	0.23	5517380	0.22

Приложение 3. Количество замаскированных оснований для исследуемых образцов соболя, лесной куницы и их гибридов из зоны симпатрии для геномных сборок соболя (*M. zibellina*) и лесной куницы (*M. martes*).

ID	<i>M. martes</i>			<i>M. zibellina</i>		
	Общее кол-во	>2.5 *	<0.33 **	Общее кол-во	>2.5 *	<0.33 **
10xmzib	100019837	5946170	94073667	70512427	5936822	64575605
S26	86692485	6645479	80047006	63937509	6594431	57343078
T8	101837884	5980124	95857760	79208271	5831246	73377025
T26	88052151	5761283	82290868	65370993	5727879	59643114
T50	102323097	5886465	96436632	79189229	5822456	73366773
T72	101444464	5403974	96040490	79133305	5354574	73778731
T90	85484068	5872449	79611619	62786786	5938766	56848020
T104	96212242	6146189	90066053	74343770	6060894	68282876
T118	102631121	5851772	96779349	80263274	5768144	74495130
T148	93117777	5465243	87652534	70092723	5510204	64582519
T150	86308923	5496513	80812410	63912325	5515943	58396382
T194	86096441	5340166	80756275	63437198	5336734	58100464
china	142339413	9961222	132378191	119773932	9872441	109901491
10xmmar	77690575	7467627	70222948	87535669	7305340	80230329
S44	69089660	6343178	62746482	75107591	6273326	68834265
S46	69657638	6987587	62670051	76100527	6653635	69446892
S49	81211898	7019223	74192675	86805493	6606650	80198843
T149	79969683	7324894	72644789	85938085	6879475	79058610
T24	81822420	6482883	75339537	88133795	6325230	81808565
T76	69756766	7024754	62732012	76533331	7211061	69322270
T77	82307204	7410651	74896553	88194290	7116188	81078102
T82	68684974	6571699	62113275	75503772	6794480	68709292
T18	100666826	5918099	94748727	82704420	5779188	76925232
S50	69897077	7064285	62832792	72228119	6799836	65428283
T151	87339644	6795080	80544564	88869345	6386397	82482948
T78	71096727	7165845	63930882	74136549	6640974	67495575
T79	69094119	6453865	62640254	71360191	6233641	65126550

T81	84016376	7218598	76797778	89078733	6727928	82350805
T83	82673798	7271922	75401876	87357337	6978491	80378846
T84	89544850	6625818	82919032	77274482	6373443	70901039
T85	71181116	6487501	64693615	69953242	6229687	63723555
T86	89144285	7444519	81699766	87406188	7218063	80188125
T87	85447452	6411189	79036263	75624899	6212422	69412477

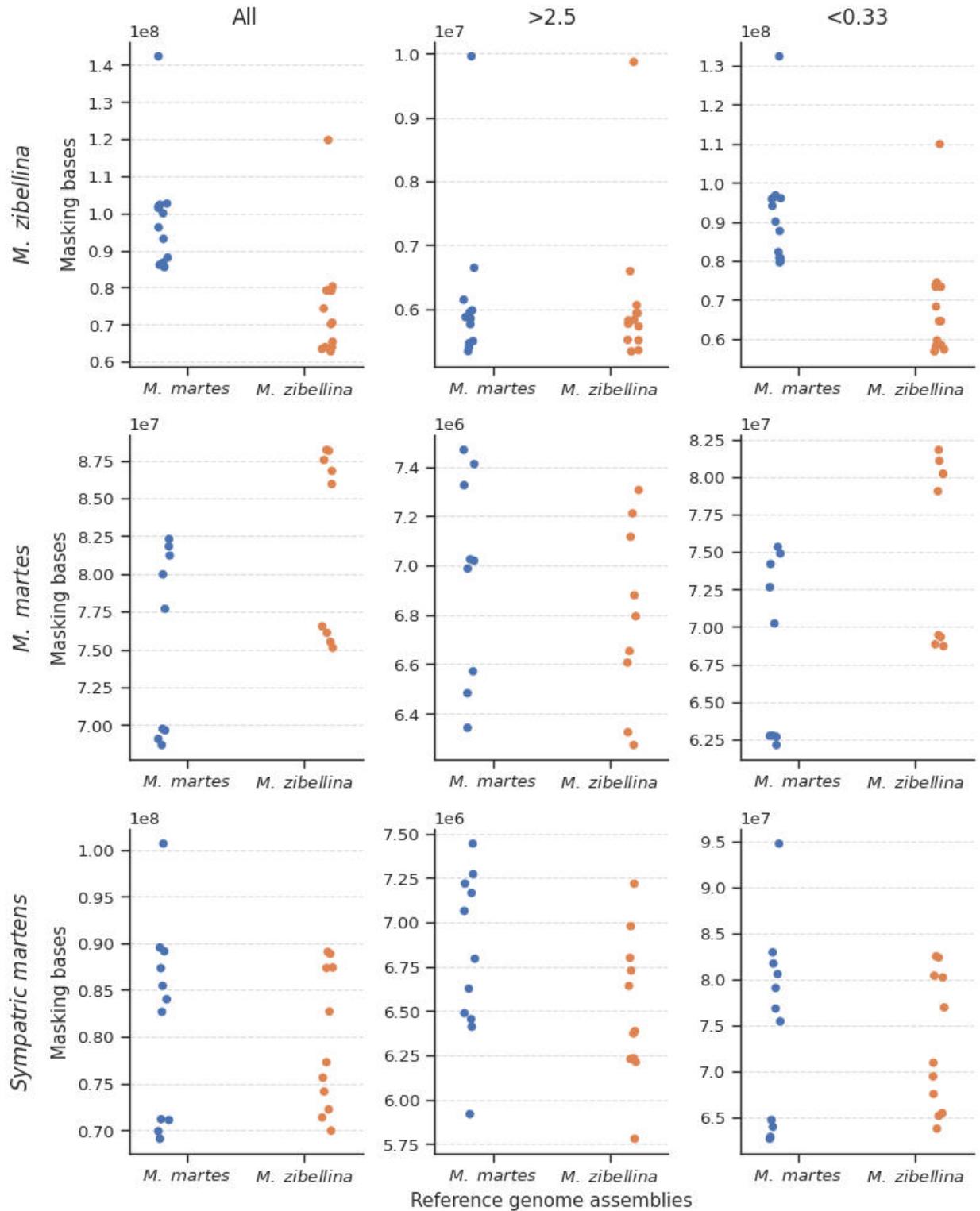
* >2.5 – количество оснований, покрытие которых в 2.5 раз больше, чем медианное всего генома.

** <0.33 – количество оснований, покрытие которых в 0.33 раз меньше, чем медианное всего генома.

Приложение 4. Количество замаскированных оснований для исследуемых образцов харзы, каменной куницы, американской куницы и тихоокеанских куниц.

ID	Референс	Общее кол-во	>2.5 *	<0.33 **
10xmfla	M. flavigula	144379492	10149636	134229856
10xmfoi	M. foina	76833566	9285007	67548559
SRR11575343	M. zibellina	273123120	80494075	192629045
SRR11575346	M. zibellina	171531361	62924104	108607257
SRR11575348	M. zibellina	82433964	10580443	71853521
SRR11575351	M. zibellina	64112186	11144856	52967330
SRR11575352	M. zibellina	93866576	18062654	75803922

Приложение 5. Сравнение количества замаскированных оснований на основании покрытия сборок геномов соболя и лесной куницы для исследуемых образцов. Образцы разделены на группы *M. zibellina*, *M. martes* и *Sympatric martens* согласно классификации. Значения в Приложении 4.



All – суммарное количество замаскированных оснований.

>2.5 – количество оснований, покрытие которых в 2.5 раз больше, чем медианное всего генома.

<0.33 – количество оснований, покрытие которых в 0.33 раз меньше, чем медианное всего генома.

Приложение 6. Статистики сравнения замаскированных оснований на основании покрытия сборок геномов соболя и лесной куницы для исследуемых образцов. Образцы разделены на группы *M. zibellina*, *M. martes* и *Sympatric martens* согласно классификации.

Comparison	Statistic	Type *	Masked bases **		
			All	>2.5	<0.33
M. zibellina - M. zibellina	Two-sided U-test	P-value	0.00022	0.71961	0.00022
		P-value (corrected)	0.00200	6.47653	0.00200
M. martes - M. martes	Two-sided U-test	P-value	0.07739	0.37722	0.07739
		P-value (corrected)	0.69650	3.39502	0.69650
Kidases - Kidases	Two-sided U-test	P-value	0.94764	0.10067	1.00000
		P-value (corrected)	8.52880	0.90601	9.00000
M. zibellina - M. zibellina	One-sided U-test	P-value	0.00011	0.35981	0.00011
		P-value (corrected)	0.00100	3.23826	0.00100
M. martes - M. martes	One-sided U-test	P-value	0.03869	0.18861	0.03869
		P-value (corrected)	0.34825	1.69751	0.34825
kidases - kidases	One-sided U-test	P-value	0.47382	0.05033	0.50000
		P-value (corrected)	4.26440	0.45300	4.50000

Приложение 7. Медианное, среднее, максимальное и минимальное значение покрытия всего генома для образцов соболя, лесной куницы и их гибридов на основании обоих референсных геномных сборок.

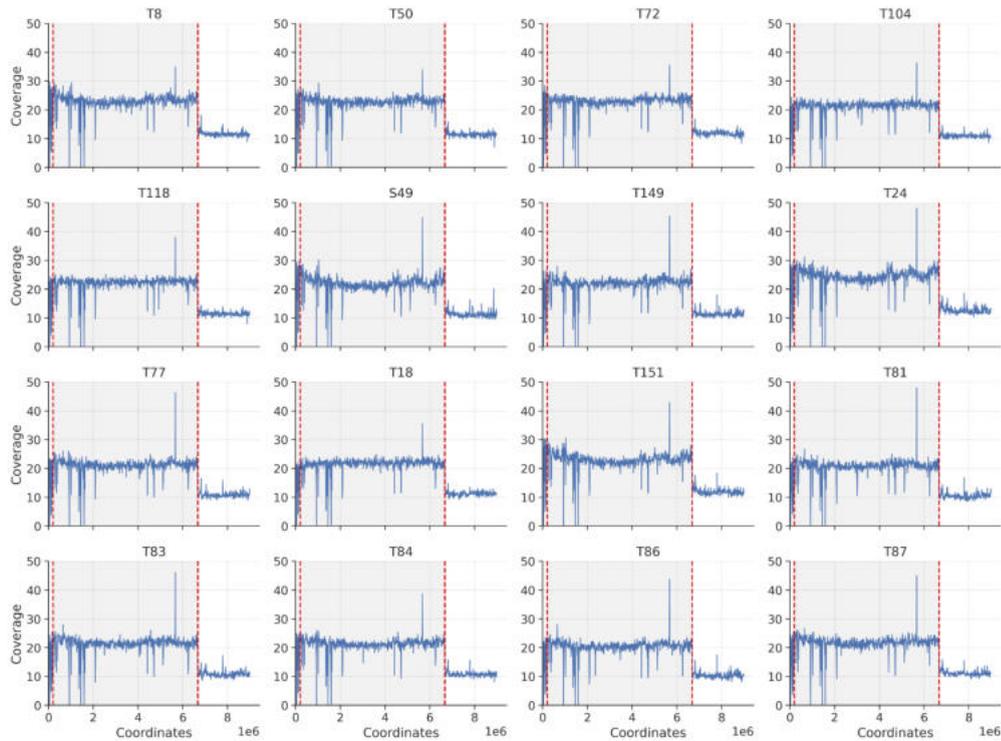
ID	<i>M. martes</i>				<i>M. zibellina</i>			
	Медиана	Среднее	Макс.	Мин.	Медиана	Среднее	Макс.	Мин.
10xmmar	23	24	38987	0	23	23.91	78676	0
S44	22	22.45	59070	0	22	22.43	60435	0
S46	22	22.54	68078	0	22	22.51	64371	0
S49	21	21.23	54065	0	21	21.19	73018	0
S50	21	21.39	54873	0	21	21.42	59636	0
T149	21	21.6	58932	0	21	21.54	62018	0
T151	22	22.6	78958	0	22	22.59	68279	0
10xmzib	23	22.99	36551	0	23	23.34	70122	0
china	20	20.84	39693	0	20	21.07	41207	0
S26	23	23.46	78413	0	23	23.71	59272	0
T104	21	21.13	54566	0	21	21.34	55593	0

T118	22	21.99	58505	0	22	22.22	49561	0
T148	22	21.81	54344	0	22	22.05	46180	0
T150	22	22.33	64501	0	22	22.56	49432	0
T18	22	21.69	52075	0	22	21.89	43685	0
T194	21	20.8	66051	0	21	21.02	64791	0
T26	22	22.34	58942	0	22	22.57	37577	0
T50	22	22.08	65570	0	22	22.32	55577	0
T72	23	22.46	53159	0	23	22.68	44591	0
T8	22	22.21	64849	0	22	22.44	59392	0
T90	21	21.36	55955	0	21	21.57	49280	0
T24	23	23.57	86673	0	23	23.51	50622	0
T76	21	21.04	69063	0	20	21.03	47816	0
T77	20	20.71	50848	0	20	20.68	53836	0
T78	20	20.4	48726	0	20	20.41	50296	0
T79	24	24.24	66740	0	24	24.25	60921	0
T81	20	20.32	57566	0	20	20.29	59904	0
T82	21	20.87	53307	0	20	20.85	61273	0
T83	20	20.57	59175	0	20	20.54	71353	0
T84	20	20.56	57272	0	20	20.7	67450	0
T85	20	20.24	45092	0	20	20.29	47703	0
T86	19	19.78	52040	0	19	19.79	55570	0
T87	21	21.06	50136	0	21	21.19	52188	0

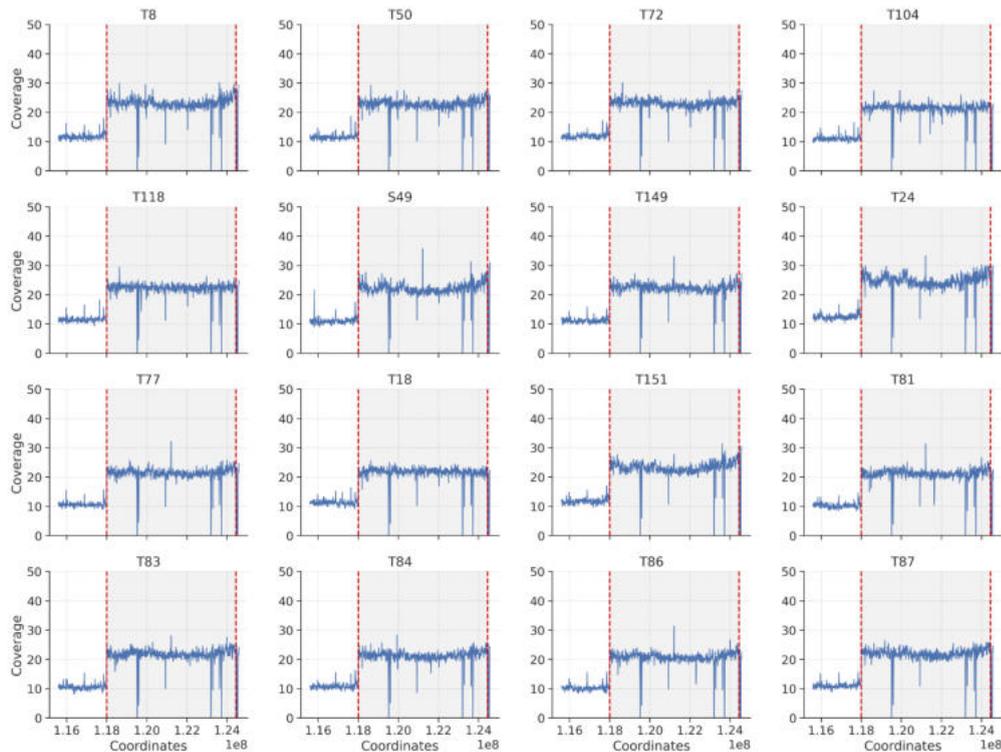
Приложение 8. Медианное, среднее, максимальное и минимальное значение покрытия всего генома для образцов харзы, каменной куницы, американской куницы и тихоокеанских куниц.

ID	Референс	Медиана	Среднее	Макс.	Мин.
10xmfla	<i>M. flavigula</i>	22	22.18	37900	0
10xmfoi	<i>M. foina</i>	18	18.57	522298	0
SRR11575343	<i>M. zibellina</i>	16	18.32	65258	0
SRR11575346	<i>M. zibellina</i>	20	22.9	65461	0
SRR11575348	<i>M. zibellina</i>	20	20.55	146163	0
SRR11575351	<i>M. zibellina</i>	24	24.71	66260	0
SRR11575352	<i>M. zibellina</i>	23	24.97	157996	0

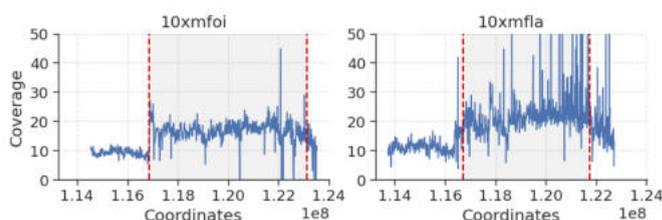
Приложение 9. Медианное покрытие в неперекрывающихся окнах по 10 тыс п.н. для самцов в регионе PAR в геномной сборке *M. zibellina*.



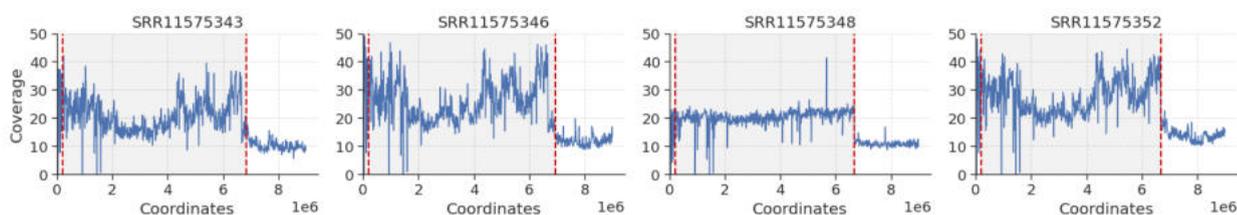
Приложение 10. Медианное покрытие в неперекрывающихся окнах по 10 тыс п.н. для самцов в регионе PAR в геномной сборке *M. martes*.



Приложение 11. Медианное покрытие в неперекрывающихся окнах по 10 тыс п.н. для самцов в регионе PAR в геномной сборке *M. flavigula* и *M. foina*.



Приложение 12. Медианное покрытие в неперекрывающихся окнах по 10 тыс п.н. для самцов *M. saurina* (SRR11575343, SRR11575346, SRR11575348) и *M. americana* (SRR11575352) в регионе PAR (Референс: *M. zibellina*).



Приложение 13. Сопоставление результатов анализа синтении на основании множественных полногеномных выравниваний и опубликованных результатов хромосомного пейнтинга для геномной сборки каменной куницы (*Martes foina*, MFOI) против сборки домашней собаки (*Canis familiaris*, CFAM). Зеленым обозначены синтенные блоки, присутствующие на опубликованной в работе Nie et al. 2002 карте хромосомного пейнтинга; желтым – отсутствующие короткие синтенные блоки (≤ 5 млн п.н.); красным – отсутствующие длинные синтенные блоки (> 5 млн п.н.).

Скаффолд целевой сборки	Скаффолд запрашиваемой сборки	Начало	Конец	Длина	Цвет (HEX)
MFOI 1	CFAM 37'	485147	1567994	1082847	#007FFF
MFOI 1	CFAM 16	1560624	39992354	38431730	#B6B915
MFOI 1	CFAM 25	40172165	56847978	16675813	#88E1FE
MFOI 1	CFAM 15	57295605	65956568	8660963	#06BFBD
MFOI 1	CFAM 25	65993679	67550248	1556569	#88E1FE
MFOI 1	CFAM 15	67607324	81621334	14014010	#06BFBD
MFOI 1	CFAM 32	81748596	114177818	32429222	#04685D
MFOI 1	CFAM 19	114111359	134919956	20808597	#A401B5
MFOI 1	CFAM 3	135027058	169480485	34453427	#5F04F8
MFOI 1	CFAM 13'	169480814	194777853	25297039	#4A1753
MFOI 1	CFAM 32	195373099	202687427	7314328	#04685D

MFOI 2	CFAM 31'	0	40109089	40109089	#FA81C8
MFOI 2	CFAM 33	40109811	66384352	26274541	#BC052A
MFOI 2	CFAM 23'	66454784	82440445	15985661	#7FFF00
MFOI 2	CFAM 34	82440453	113356870	30916417	#3600AC
MFOI 2	CFAM 33	113965625	120888071	6922446	#BC052A
MFOI 2	CFAM 23'	120944456	146168239	25223783	#7FFF00
MFOI 2	CFAM 20	146181169	146270445	89276	#643D02
MFOI 2	CFAM 23'	146270727	149242026	2971299	#7FFF00
MFOI 2	CFAM 20	149210534	149287262	76728	#643D02
MFOI 2	CFAM 23'	149325314	156725239	7399925	#7FFF00
MFOI 2	CFAM 20	157042225	216435893	59393668	#643D02
MFOI 3	CFAM 22'	17830	62255308	62237478	#FC307A
MFOI 3	CFAM 25	62277294	81671651	19394357	#88E1FE
MFOI 3	CFAM 19	81770819	114419870	32649051	#A401B5
MFOI 3	CFAM 36'	114516946	144407875	29890929	#FF7F00
MFOI 3	CFAM 19	144543885	147427246	2883361	#A401B5
MFOI 3	CFAM 25	148093850	161778361	13684511	#88E1FE
MFOI 3	CFAM 37'	161710458	192717086	31006628	#007FFF
MFOI 3	CFAM 36'	192797871	195015576	2217705	#FF7F00
MFOI 4	CFAM 13'	5634	38921489	38915855	#4A1753
MFOI 4	CFAM 29'	38911830	82606634	43694804	#FFFF00
MFOI 4	CFAM 17'	83751577	96958237	13206660	#0EB94A
MFOI 4	CFAM 6'	96898587	135847099	38948512	#D048C3
MFOI 4	CFAM 5	135826312	148894951	13068639	#3FF73D
MFOI 4	CFAM 15	148895546	165655121	16759575	#06BFBD
MFOI 4	CFAM 2	165736500	183084882	17348382	#FF0000
MFOI 4	CFAM 5	183587720	190994362	7406642	#3FF73D
MFOI 5	CFAM 14	38189	2640191	2602002	#002CFA
MFOI 5	CFAM 14	4164536	4278926	114390	#002CFA
MFOI 5	CFAM 14	5379113	6199388	820275	#002CFA
MFOI 5	CFAM 11'	6997011	21767510	14770499	#0A6503
MFOI 5	CFAM 3	22412654	54766176	32353522	#5F04F8
MFOI 5	CFAM 2	54850601	78364180	23513579	#FF0000
MFOI 5	CFAM 11'	78363743	89450774	11087031	#0A6503
MFOI 5	CFAM 4	89660644	142738591	53077947	#D77F69
MFOI 5	CFAM 34	142740496	154445468	11704972	#3600AC
MFOI 6	CFAM 1	677994	17025510	16347516	#4C8D61
MFOI 6	CFAM 12'	17025510	52977160	35951650	#3E59AB
MFOI 6	CFAM 1	53201050	82660587	29459537	#4C8D61
MFOI 6	CFAM 14	82682461	82906598	224137	#002CFA
MFOI 6	CFAM 12'	82921076	121345663	38424587	#3E59AB
MFOI 6	CFAM 35	121387979	147786526	26398547	#7FBF7F

MFOI 7	CFAM 8'	247079	73900677	73653598	#4CFBC4
MFOI 7	CFAM 30'	73999137	93421997	19422860	#00FFFF
MFOI 7	CFAM 8'	93408840	93496328	87488	#4CFBC4
MFOI 7	CFAM 30'	93583541	93647173	63632	#00FFFF
MFOI 7	CFAM 8'	93633823	93693210	59387	#4CFBC4
MFOI 7	CFAM 30'	93723330	93882858	159528	#00FFFF
MFOI 7	CFAM 15	93915600	95691056	1775456	#06BFBD
MFOI 7	CFAM 30'	95755584	117195651	21440067	#00FFFF
MFOI 7	CFAM 3	117277534	143815711	26538177	#5F04F8
MFOI 8	CFAM 24'	7841	47454684	47446843	#8F5B6C
MFOI 8	CFAM 23'	47469215	49988648	2519433	#7FFF00
MFOI 8	CFAM 10	50120437	59471525	9351088	#B9FE7E
MFOI 8	CFAM 14	59471560	59839314	367754	#002CFA
MFOI 8	CFAM 10	59843784	86608235	26764451	#B9FE7E
MFOI 8	CFAM 17'	86608240	138172325	51564085	#0EB94A
MFOI 9	CFAM 2	5557	32573283	32567726	#FF0000
MFOI 9	CFAM 27	32713417	41224355	8510938	#4FA101
MFOI 9	CFAM 10	41350281	42247143	896862	#B9FE7E
MFOI 9	CFAM 1	42250848	42547832	296984	#4C8D61
MFOI 9	CFAM 10	42649904	49391208	6741304	#B9FE7E
MFOI 9	CFAM 3	49334422	49550903	216481	#5F04F8
MFOI 9	CFAM 27	49551102	49794477	243375	#4FA101
MFOI 9	CFAM 10	49708544	49781265	72721	#B9FE7E
MFOI 9	CFAM 27	49950112	87645130	37695018	#4FA101
MFOI 9	CFAM 10	87839776	97116410	9276634	#B9FE7E
MFOI 9	CFAM 15	97094969	121976061	24881092	#06BFBD
MFOI 9	CFAM 10	121979048	139574765	17595717	#B9FE7E
MFOI 10	CFAM 5	52484	29571371	29518887	#3FF73D
MFOI 10	CFAM 21	29990482	80683960	50693478	#F8C66F
MFOI 10	CFAM 18	80684526	100620492	19935966	#6297D7
MFOI 10	CFAM 21	100620994	101214265	593271	#F8C66F
MFOI 10	CFAM 18	101246542	111960802	10714260	#6297D7
MFOI 11	CFAM 16	1504	18152951	18151447	#B6B915
MFOI 11	CFAM 14	18156545	76704404	58547859	#002CFA
MFOI 11	CFAM 18	76704929	83158502	6453573	#6297D7
MFOI 11	CFAM 18	83203461	83406428	202967	#6297D7
MFOI 11	CFAM 18	83461877	90816255	7354378	#6297D7
MFOI 11	CFAM 18	90349035	102466261	12117226	#6297D7
MFOI 11	CFAM 16	102466519	103737546	1271027	#B6B915
MFOI 11	CFAM 1	104843079	105045426	202347	#4C8D61
MFOI 12	CFAM 4	1	10046372	10046371	#D77F69
MFOI 12	CFAM 26	10046725	18108099	8061374	#9143FA

MFOI 12	CFAM 4	18573920	43889494	25315574	#D77F69
MFOI 12	CFAM 28'	43956690	85117410	41160720	#00FF7F
MFOI 13	CFAM 1	14142	23912806	23898664	#4C8D61
MFOI 13	CFAM 7'	23867004	61154676	37287672	#C1B3CA
MFOI 13	CFAM 26	62121821	74246343	12124522	#9143FA
MFOI 13	CFAM 19	74265707	74385079	119372	#A401B5
MFOI 13	CFAM 26	74367642	76097212	1729570	#9143FA
MFOI 13	CFAM 19	76098725	76165964	67239	#A401B5
MFOI 13	CFAM 26	76174906	92486881	16311975	#9143FA
MFOI 14	CFAM 9	32559	13448630	13416071	#C14D19
MFOI 14	CFAM 11'	13439868	61298539	47858671	#0A6503
MFOI 14	CFAM 1	61559875	90093638	28533763	#4C8D61
MFOI 15	CFAM 5	13	12319602	12319589	#3FF73D
MFOI 15	CFAM 9	12350840	58969525	46618685	#C14D19
MFOI 16	CFAM 7'	10528	40921365	40910837	#C1B3CA
MFOI 16	CFAM 38	40930132	64768901	23838769	#FF00FF
MFOI 16	CFAM 7'	64768928	67759797	2990869	#C1B3CA
MFOI 17	CFAM 1	81168	21601114	21519946	#4C8D61
MFOI 17	CFAM 2	22162449	32032174	9869725	#FF0000
MFOI 17	CFAM 5	32071005	57585412	25514407	#3FF73D
MFOI 18	CFAM 6'	264492	39190701	38926209	#D048C3
MFOI X	CFAM X'	64820	123511915	123447095	#00FF00

Приложение 14. Сопоставление результатов анализа синтении на основании множественных полногеномных выравниваний и опубликованных результатов хромосомного пейнтинга для геномной сборки каменной куницы (*Martes foina*, MFOI) против сборки человека (*Homo sapiens*, HSAP). Зеленым обозначены синтенные блоки, присутствующие на опубликованной в работе Nie et al. 2002 карте хромосомного пейнтинга; желтым – отсутствующие короткие синтенные блоки (≤ 5 млн п.н.); красным – отсутствующие длинные синтенные блоки (> 5 млн п.н.).

Скаффолд целевой сборки	Скаффолд запрашиваемой сборки	Начало	Конец	Длина	Цвет (HEX)
MFOI 1	HSAP 8	262433	6726532	6464099	#D048C3
MFOI 1	HSAP 4	6916093	16215617	9299524	#F8C66F
MFOI 1	HSAP 8	16237388	23884129	7646741	#D048C3
MFOI 1	HSAP 4	23888428	27221608	3333180	#F8C66F
MFOI 1	HSAP 8	27290532	46301410	19010878	#D048C3

MFOI 1	HSAP 4	46581537	51641896	5060359	#F8C66F
MFOI 1	HSAP 8	51652801	56834031	5181230	#D048C3
MFOI 1	HSAP 4	56957120	134919956	77962836	#F8C66F
MFOI 1	HSAP 4	135030352	202687427	67657075	#F8C66F
MFOI 2	HSAP 21'	864	28657482	28656618	#FF0000
MFOI 2	HSAP 3	28660980	201326646	172665666	#FC307A
MFOI 2	HSAP 19'	201357640	216431510	15073870	#5F04F8
MFOI 3	HSAP 13	41381	22907705	22866324	#3FF73D
MFOI 3	HSAP 13	22917813	81499606	58581793	#3FF73D
MFOI 3	HSAP 2	81603223	114419877	32816654	#7FFF00
MFOI 3	HSAP 2	114521142	195015576	80494434	#7FFF00
MFOI 4	HSAP 8	9112	83095363	83086251	#D048C3
MFOI 4	HSAP 1	83729293	190971406	107242113	#8F5B6C
MFOI 5	HSAP 1	42286	1092131	1049845	#8F5B6C
MFOI 5	HSAP 1	4149487	6976532	2827045	#8F5B6C
MFOI 5	HSAP 5'	7037071	118530286	111493215	#88E1FE
MFOI 5	HSAP 5'	118572101	154445468	35873367	#88E1FE
MFOI 6	HSAP 6'	41058	82660258	82619200	#9143FA
MFOI 6	HSAP 7	82682522	82905682	223160	#FA81C8
MFOI 6	HSAP 6'	82921076	102652360	19731284	#9143FA
MFOI 6	HSAP 6'	102707936	147790778	45082842	#9143FA
MFOI 7	HSAP 14'	271879	69169423	68897544	#4FA101
MFOI 7	HSAP 14'	69170523	73944536	4774013	#4FA101
MFOI 7	HSAP 15'	74000380	93075066	19074686	#D77F69
MFOI 7	HSAP 14'	93345179	94040754	695575	#4FA101
MFOI 7	HSAP 15'	93915291	94109088	193797	#D77F69
MFOI 7	HSAP 14'	94211592	95678603	1467011	#4FA101
MFOI 7	HSAP 15'	95755584	144049927	48294343	#D77F69
MFOI 8	HSAP 20'	10122	48806630	48796508	#00FF7F
MFOI 8	HSAP 2	49017332	49123311	105979	#7FFF00
MFOI 8	HSAP 20'	49390860	49984217	593357	#00FF7F
MFOI 8	HSAP 2	50126024	59465581	9339557	#7FFF00
MFOI 8	HSAP 7	59473020	59839254	366234	#FA81C8
MFOI 8	HSAP 2	59843788	136188681	76344893	#7FFF00
MFOI 8	HSAP 8	131774962	132070355	295393	#D048C3
MFOI 8	HSAP 2	136234252	138170743	1936491	#7FFF00
MFOI 9	HSAP 10	6948	32675713	32668765	#04685D
MFOI 9	HSAP 22	32714287	33776313	1062026	#4C8D61
MFOI 9	HSAP 12	33779133	124947253	91168120	#FFFF00
MFOI 9	HSAP 22	124953419	139573147	14619728	#4C8D61
MFOI 10	HSAP 11	46519	111959358	111912839	#00FFFF
MFOI 11	HSAP 7	15113	66100111	66084998	#FA81C8

MFOI 11	HSAP 8	66091273	66167022	75749	#D048C3
MFOI 11	HSAP 7	66100148	72769185	6669037	#FA81C8
MFOI 11	HSAP 7	72842058	104550758	31708700	#FA81C8
MFOI 12	HSAP 1	405	9724309	9723904	#8F5B6C
MFOI 12	HSAP 10	9730947	85008084	75277137	#04685D
MFOI 13	HSAP 18'	17332	61990591	61973259	#BC052A
MFOI 13	HSAP 22	62129490	74367334	12237844	#4C8D61
MFOI 13	HSAP 12	74408453	92456856	18048403	#FFFF00
MFOI 14	HSAP 9'	32559	89983935	89951376	#3600AC
MFOI 15	HSAP 17	175	59722643	59722468	#7FBF7F
MFOI 16	HSAP 1	10990	67759797	67748807	#8F5B6C
MFOI 17	HSAP 19'	85421	21843441	21758020	#5F04F8
MFOI 17	HSAP 16	21762706	56236624	34473918	#007FFF
MFOI 17	HSAP 16	56048285	57600736	1552451	#007FFF
MFOI 18	HSAP 7	68265	16191412	16123147	#FA81C8
MFOI 18	HSAP 16	16256249	36802964	20546715	#007FFF
MFOI 18	HSAP 16	36802970	39131678	2328708	#007FFF
MFOI X	HSAP X'	74825	74673306	74598481	#00FF00
MFOI X	HSAP X'	76422267	123403585	46981318	#00FF00

Приложение 15. Сопоставление результатов анализа синтении на основании множественных полногеномных выравниваний и опубликованных результатов хромосомного пейнтинга для геномной сборки каменной куницы (*Martes foina*, MFOI) против сборки домашней кошки (*Felis catus*, FCAM). Зеленым обозначены синтенные блоки, присутствующие на опубликованной в работе Nie et al. 2002 карте хромосомного пейнтинга; желтым – отсутствующие короткие синтенные блоки (≤ 5 млн п.н.); красным – отсутствующие длинные синтенные блоки (> 5 млн п.н.).

Скаффолд целевой сборки	Скаффолд запрашиваемой сборки	Начало	Конец	Длина	Цвет (HEX)
MFOI 1	FCAT B2'	4912	202402	197490	#9143FA
MFOI 1	FCAT B1'	276282	202687427	202411145	#F8C66F
MFOI 2	FCAT C2	0	156745565	156745565	#FC307A
MFOI 2	FCAT A2'	156754851	156817878	63027	#06BFBD
MFOI 2	FCAT C2	156950349	157027612	77263	#FC307A
MFOI 2	FCAT A2'	157155775	216435893	59280118	#06BFBD
MFOI 3	FCAT A1	20056	81660417	81640361	#3E59AB
MFOI 3	FCAT C1	81770819	183165551	101394732	#4A1753

MFOI 3	FCAT C1	184891376	195015576	10124200	#4A1753
MFOI 4	FCAT F2'	5781	83104300	83098519	#002CFA
MFOI 4	FCAT C1	83730843	190985427	107254584	#4A1753
MFOI 5	FCAT A1	31877	153919645	153887768	#3E59AB
MFOI 6	FCAT B2'	41058	82660587	82619529	#9143FA
MFOI 6	FCAT A2'	82682461	82906598	224137	#06BFBD
MFOI 6	FCAT B2'	82921076	147790778	64869702	#9143FA
MFOI 7	FCAT B3'	229686	144048594	143818908	#4FA101
MFOI 8	FCAT A3	7833	59471525	59463692	#00FF7F
MFOI 8	FCAT A2'	59471926	59839314	367388	#06BFBD
MFOI 8	FCAT A3	59843788	138171024	78327236	#00FF7F
MFOI 9	FCAT B4	6701	42247143	42240442	#FFFF00
MFOI 9	FCAT D4'	42250848	42547832	296984	#3600AC
MFOI 9	FCAT B4	42649904	139577817	96927913	#FFFF00
MFOI 10	FCAT D1	46425	111960802	111914377	#00FFFF
MFOI 11	FCAT A2'	1439	104555085	104553646	#06BFBD
MFOI 12	FCAT D2	1	85117557	85117556	#04685D
MFOI 13	FCAT D3'	14142	74272935	74258793	#BC052A
MFOI 13	FCAT D3'	74222829	74300073	77244	#BC052A
MFOI 13	FCAT D3'	74347314	92489707	18142393	#BC052A
MFOI 14	FCAT D4'	32559	90093638	90061079	#3600AC
MFOI 15	FCAT E1	256	59722643	59722387	#7FBF7F
MFOI 16	FCAT F1	10989	67755343	67744354	#FF7F00
MFOI 17	FCAT E2	38623	57600736	57562113	#007FFF
MFOI 18	FCAT E3	69502	39190701	39121199	#FF00FF
MFOI X	FCAT X'	71144	123546935	123475791	#00FF00

Приложение 16. Координаты центромер в геномных сборках *M. zibellina*, *M. martes*, *M. foia* и *M. flavigula*.

Хромосома	<i>M. zibellina</i>		<i>M. martes</i>		<i>M. foia</i>		<i>M. flavigula</i>	
	Начало	Конец	Начало	Конец	Начало	Конец	Начало	Конец
Chr1	162388431	162550792	39953621	40106605	39992354	40172165	164060472	164308841
Chr2	59588730	59934821	59288566	59673842	156725239	157042225	57912802	58089253
Chr3	113319631	113418987	81435655	81486707	81671651	81770819	112165928	112504580
Chr4	106687407	107702027	106953259	107931513	82606634	83751577	103985091	105083946
Chr5	147229757	148006540	6179540	6674003	4818601	4918601	4215062	4315062
Chr6	26318136	26354437	120931299	120966934	121345663	121387979	25633597	25672911
Chr7	26605235	26678365	117787880	117871564	117195651	117277534	32424678	32535455
Chr8	87734355	87931460	49940042	50248109	47454684	47469215	47843302	48539558

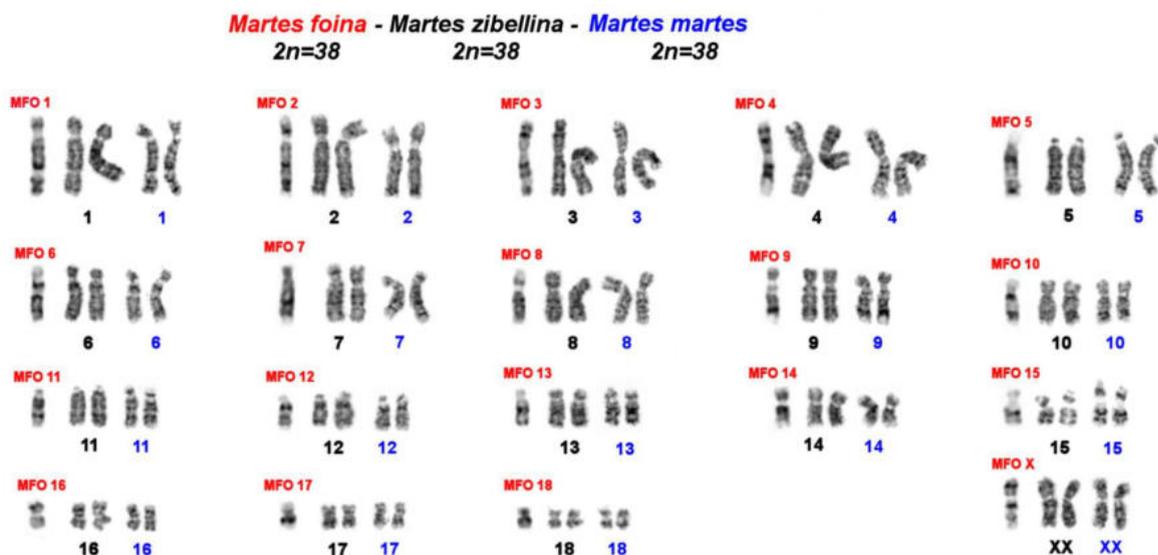
Chr9	106342467	106477938	32490535	32628034	32573283	32713417	0	100000
Chr10	81282155	81690829	81877421	82293558	29571371	29990482	29120863	29880058
Chr11	10553932	10653932	11104254	11204254	93879150	93979150	93173131	93273131
Chr12	66553920	67017660	66509297	66970460	18108099	18573920	12437577	12537577
Chr13	15232218	15332218	76997299	77097299	77304351	77404351	18348423	18823395
Chr14	61135396	61442819	28445474	28746857	61298539	61559875	60337577	60641920
Chr15	12185599	12199109	12359108	12373676	12319602	12350840	16879435	16979435
Chr16	26868126	26876880	26821390	26830144	40921365	40930132	40281067	40289852
Chr17	22060369	22603058	35446870	35890592	21601114	22162449	33914132	34376502
Chr18	26058618	26068618	13383322	13483322	13063567	13163567	12563403	12663403
Chr19	-	-	-	-	-	-	13201904	13301904
ChrX	48984497	49212382	75341268	75555172	74680060	75366282	74234304	74727652

Приложение 17. Соответствие и ориентация номеров хромосом и скаффолдов в геномных сборках.

Хромосома	<i>M. zibellina</i>	<i>M. martes</i>	<i>M. foina</i>	<i>M. flavigula</i>
Chr1	HiC_scaffold_2*	HiC_scaffold_2	HiC_scaffold_2	HiC_scaffold_2*
Chr2	HiC_scaffold_1*	HiC_scaffold_1*	HiC_scaffold_1	HiC_scaffold_1*
Chr3	HiC_scaffold_3*	HiC_scaffold_3	HiC_scaffold_3	HiC_scaffold_3*
Chr4	HiC_scaffold_4*	HiC_scaffold_4*	HiC_scaffold_4	HiC_scaffold_4*
Chr5	HiC_scaffold_5*	HiC_scaffold_5	HiC_scaffold_5	HiC_scaffold_5
Chr6	HiC_scaffold_6*	HiC_scaffold_6	HiC_scaffold_6	HiC_scaffold_6*
Chr7	HiC_scaffold_7*	HiC_scaffold_7	HiC_scaffold_7	HiC_scaffold_7
Chr8	HiC_scaffold_9*	HiC_scaffold_9	HiC_scaffold_9	HiC_scaffold_8
Chr9	HiC_scaffold_8*	HiC_scaffold_8	HiC_scaffold_8	HiC_scaffold_9*
Chr10	HiC_scaffold_10*	HiC_scaffold_10*	HiC_scaffold_10	HiC_scaffold_10
Chr11	HiC_scaffold_11*	HiC_scaffold_11*	HiC_scaffold_11	HiC_scaffold_11
Chr12	HiC_scaffold_14*	HiC_scaffold_14*	HiC_scaffold_14	HiC_scaffold_16
Chr13	HiC_scaffold_12*	HiC_scaffold_12	HiC_scaffold_12	HiC_scaffold_14
Chr14	HiC_scaffold_13	HiC_scaffold_13*	HiC_scaffold_13	HiC_scaffold_13
Chr15	HiC_scaffold_16	HiC_scaffold_16	HiC_scaffold_16	HiC_scaffold_12*
Chr16	HiC_scaffold_15*	HiC_scaffold_15*	HiC_scaffold_15	HiC_scaffold_15
Chr17	HiC_scaffold_17	HiC_scaffold_17*	HiC_scaffold_17	HiC_scaffold_17*
Chr18	HiC_scaffold_18*	HiC_scaffold_18	HiC_scaffold_18	HiC_scaffold_18
Chr19	-	-	-	HiC_scaffold_19*
ChrX	HiC_scaffold_19*	HiC_scaffold_19	HiC_scaffold_19	HiC_scaffold_20

* – разнонаправленные скаффолды относительно геномной сборки *M. foina*.

Приложение 18. Сравнение G-бэндинга кариотипов соболя, лесной куницы и каменной куницы, выполнено к.б.н. Виолеттой Робертовной Беклемишевой (ИМКБ СО РАН).

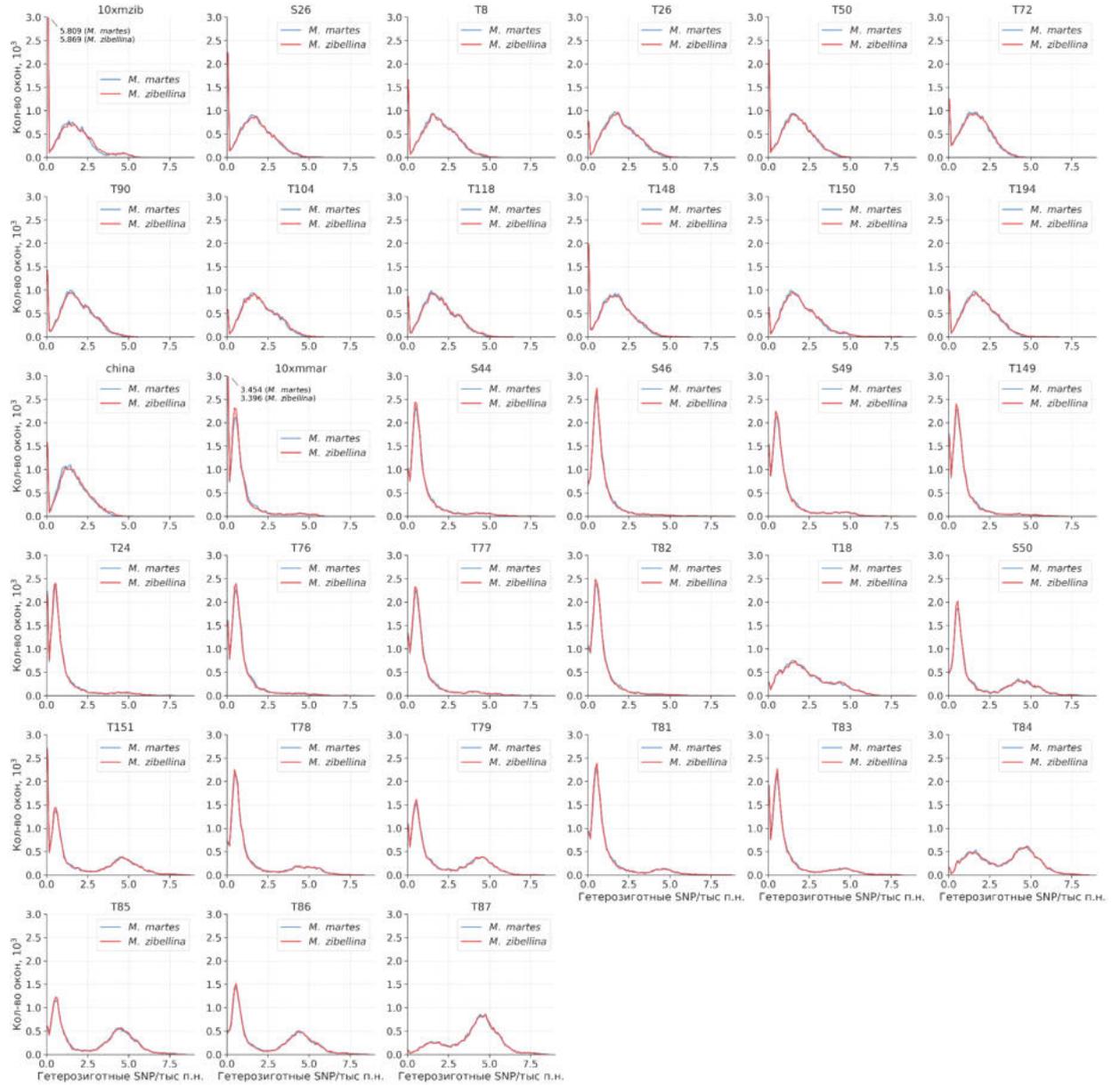


Приложение 19. Средние и медианные значения гетерозиготности (SNP/тыс п.н.) для образцов соболя, лесной куницы и их гибридов из зоны симпатрии на основании геномных сборок соболя (*M. zibellina*) и лесной куницы (*M. martes*). Хромосомы X исключены.

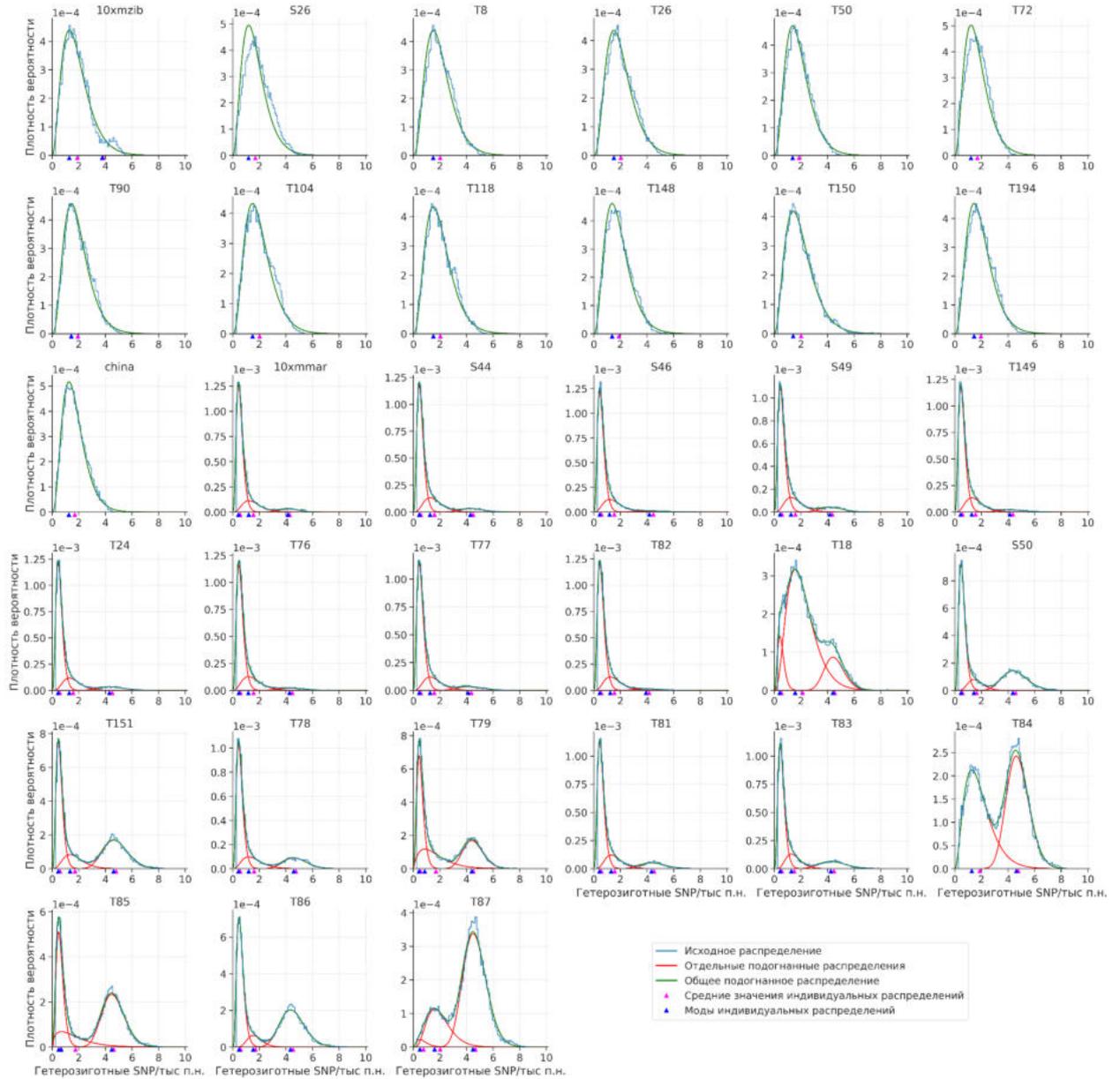
Образец	<i>M. zibellina</i>			<i>M. martes</i>		
	Кол-во гетероSNP, млн	Среднее (SNP/тыс п.н.)	Медиана (SNP/1 тыс п.н.)	Кол-во гетероSNP, млн	Среднее (SNP/тыс п.н.)	Медиана (SNP/1 тыс п.н.)
10xmzib	3.12	1.42	1.34	3.24	1.37	1.28
S26	3.89	1.74	1.69	3.99	1.7	1.65
T8	4.23	1.9	1.85	4.35	1.85	1.8
T26	4.44	2	1.9	4.58	1.95	1.85
T50	3.81	1.71	1.68	3.91	1.67	1.63
T72	3.67	1.65	1.62	3.77	1.6	1.58
T90	4.09	1.85	1.76	4.21	1.79	1.71
T104	4.47	2.02	1.9	4.60	1.96	1.85
T118	4.42	1.99	1.88	4.55	1.94	1.83
T148	3.88	1.75	1.7	3.99	1.7	1.66

T150	4.45	2	1.82	4.57	1.95	1.78
T194	4.23	1.91	1.83	4.36	1.85	1.78
china	3.53	1.63	1.55	3.72	1.54	1.46
l0xmmar	2.01	0.84	0.56	1.93	0.86	0.59
S44	2.42	1.03	0.66	2.37	1.05	0.68
S46	2.14	0.9	0.64	2.09	0.92	0.66
S49	2.51	1.07	0.67	2.46	1.09	0.68
T149	2.17	0.92	0.64	2.13	0.94	0.65
T24	2.33	0.99	0.62	2.29	1.01	0.63
T76	2.24	0.95	0.65	2.19	0.97	0.67
T77	2.45	1.04	0.66	2.40	1.06	0.68
T82	2.09	0.88	0.64	2.03	0.9	0.66
T18	5.31	2.38	2.08	5.41	2.34	2.03
S50	4.60	2	0.89	4.57	2.01	0.93
T151	4.90	2.14	0.94	4.88	2.14	0.97
T78	3.58	1.55	0.73	3.54	1.57	0.75
T79	4.64	2.03	1.02	4.62	2.04	1.05
T81	2.81	1.2	0.68	2.76	1.22	0.7
T83	2.92	1.25	0.67	2.88	1.27	0.69
T84	7.62	3.39	3.69	7.67	3.36	3.65
T85	6.23	2.73	2.86	6.21	2.73	2.86
T86	5.66	2.47	1.54	5.63	2.48	1.6
T87	9.01	3.99	4.31	9.04	3.97	4.29

Приложение 20. Распределения гетерозиготных SNP, подсчитанные в скользящих окнах размером 1 млн п.н. с шагом 100 тыс п.н. на основании геномных сборок *M. zibellina* (красный) и *M. martes* (синий).



Приложение 21. Визуализация лучших моделей и их компонент для распределений гетерозиготности на основе геномной сборки *M. zibellina*.



Приложение 23. Моды и средние значения для каждой компоненты распределений гетерозиготности (только SNP) соболей, лесных куниц и их гибридов на основании геномной сборки *M. zibellina*.

Виды	Образец	Модель	QDE	Комбинированное аппроксимированное распределение		Индивидуальные аппроксимированные распределения					
				Y (SNP / 1 kbp)		Y1 (SNP/1 kbp)		Y2 (SNP/1 kbp)		Y3 (SNP/1 kbp)	
				Мода 1	Мода 2	Среднее	Мода	Среднее	Мода	Среднее	Мода
<i>M. zibellina</i>	10xmzib	Соболь с интрогрессией	3.07E-08	1.283	-	-	-	1.879	1.283	3.833	3.756
	S26	"Чистый" соболь	1.73E-07	1.195	-	-	-	1.700	1.195	-	-
	T8	"Чистый" соболь	4.34E-08	1.496	-	-	-	2.015	1.496	-	-
	T26	"Чистый" соболь	3.52E-08	1.506	-	-	-	2.032	1.506	-	-
	T50	"Чистый" соболь	3.76E-08	1.373	-	-	-	1.867	1.373	-	-
	T72	"Чистый" соболь	1.03E-07	1.215	-	-	-	1.700	1.215	-	-
	T90	"Чистый" соболь	2.90E-08	1.435	-	-	-	1.936	1.435	-	-
	T104	"Чистый" соболь	3.34E-08	1.487	-	-	-	2.024	1.487	-	-
	T118	"Чистый" соболь	2.82E-08	1.503	-	-	-	2.031	1.503	-	-
	T148	"Чистый" соболь	3.17E-08	1.372	-	-	-	1.883	1.372	-	-
	T150	"Чистый" соболь	1.23E-08	1.413	-	-	-	2.012	1.413	-	-
	T194	"Чистый" соболь	3.45E-08	1.446	-	-	-	1.954	1.446	-	-
china	"Чистый" соболь	1.73E-08	1.253	-	-	-	1.700	1.253	-	-	
<i>M. martes</i>	10xmmar	Полная модель	1.95E-08	0.442	4.089	0.559	0.435	1.550	1.207	4.276	4.150
	S44	Полная модель	6.74E-09	0.450	4.215	0.566	0.442	1.588	1.259	4.490	4.307
	S46	Полная модель	1.60E-08	0.452	-	0.573	0.443	1.534	1.181	4.488	4.186
	S49	Полная модель	8.71E-09	0.452	4.105	0.582	0.443	1.578	1.254	4.362	4.179

	T149	Полная модель	6.68E-09	0.448	3.953	0.573	0.441	1.566	1.268	4.362	4.144
	T24	Полная модель	7.88E-09	0.446	4.222	0.564	0.440	1.575	1.277	4.514	4.289
	T76	Полная модель	6.55E-09	0.464	4.003	0.584	0.454	1.560	1.177	4.482	4.266
	T77	Полная модель	7.88E-09	0.449	4.012	0.576	0.441	1.581	1.256	4.369	4.140
	T82	Полная модель	8.09E-09	0.453	-	0.578	0.444	1.551	1.235	4.165	3.931
Гибриды	T18	Полная модель	7.70E-09	1.557	4.019	0.536	0.413	2.118	1.575	4.538	4.415
	S50	Полная модель	1.08E-08	0.464	4.361	0.579	0.462	1.641	1.438	4.567	4.363
	T151	Полная модель	1.29E-08	0.461	4.607	0.573	0.454	0.573	1.339	4.808	4.618
	T78	Полная модель	1.58E-08	0.441	4.564	0.559	0.433	1.606	1.198	4.755	4.584
	T79	Полная модель	8.19E-09	0.446	4.364	0.561	0.437	1.702	0.857	4.521	4.394
	T81	Полная модель	8.62E-09	0.443	4.328	0.567	0.437	1.569	1.304	4.541	4.349
	T83	Полная модель	1.33E-08	0.444	4.242	0.562	0.439	1.506	1.277	4.485	4.254
	T84	Соболь с интрогрессией	1.16E-08	1.274	4.575	-	-	1.873	1.274	4.774	4.620
	T85	Полная модель	1.16E-08	0.480	4.456	0.613	0.477	1.752	0.683	4.633	4.473
	T86	Полная модель	7.54E-09	0.469	4.341	0.597	0.468	1.715	1.530	4.535	4.342
T87	Полная модель	1.73E-08	1.560	4.479	0.763	0.497	2.022	1.610	4.669	4.499	

Приложение 24. Моды и средние значения для каждой компоненты распределений гетерозиготности (только SNP) соболей, лесных куниц и их гибридов на основании геномной сборки *M. martes*.

Виды	Образец	Модель	QDE	Комбинированное аппроксимированное распределение		Индивидуальные аппроксимированные распределения					
				Y (SNP / 1 kbp)		Y1 (SNP/1 kbp)		Y2 (SNP/1 kbp)		Y3 (SNP/1 kbp)	
				Мода 1	Мода 2	Среднее	Мода	Среднее	Мода	Среднее	Мода

<i>M. zibellina</i>	10xmzib	Соболь с интрогрессией	4.55E-08	1.222	4.578	-	-	1.698	1.222	4,649.260	4.626
	S26	"Чистый" соболь	1.31E-07	1.210	-	-	-	1.700	1.207	-	-
	T8	"Чистый" соболь	4.53E-08	1.460	-	-	-	1.956	1.458	-	-
	T26	"Чистый" соболь	3.78E-08	1.470	-	-	-	1.971	1.470	-	-
	T50	"Чистый" соболь	3.74E-08	1.350	-	-	-	1.816	1.347	-	-
	T72	"Чистый" соболь	7.93E-08	1.220	-	-	-	1.700	1.218	-	-
	T90	"Чистый" соболь	3.66E-08	1.410	-	-	-	1.879	1.405	-	-
	T104	"Чистый" соболь	3.88E-08	1.450	-	-	-	1.963	1.449	-	-
	T118	"Чистый" соболь	2.98E-08	1.470	-	-	-	1.970	1.466	-	-
	T148	"Чистый" соболь	3.51E-08	1.330	-	-	-	1.825	1.332	-	-
	T150	"Чистый" соболь	1.34E-08	1.380	-	-	-	1.958	1.380	-	-
	T194	"Чистый" соболь	3.62E-08	1.410	-	-	-	1.893	1.407	-	-
	china	"Чистый" соболь	1.77E-08	1.190	-	-	-	1.616	1.188	-	-
<i>M. martes</i>	10xmmar	Полная модель	1.41E-08	0.451	4.315	0.573	0.438	1.497	1.058	4.464	4.371
	S44	Полная модель	5.14E-09	0.456	4.186	0.578	0.447	1.586	1.257	4.487	4.294
	S46	Полная модель	8.81E-09	0.457	-	0.582	0.447	1.545	1.196	4.490	4.147
	S49	Полная модель	5.78E-09	0.457	4.086	0.590	0.447	1.580	1.285	4.333	4.152
	T149	Полная модель	6.65E-09	0.452	3.916	0.582	0.443	1.572	1.288	4.324	4.111
	T24	Полная модель	1.19E-08	0.449	4.225	0.571	0.442	1.580	1.291	4.510	4.290
	T76	Полная модель	1.05E-08	0.469	3.886	0.594	0.457	1.565	1.197	4.483	4.251
	T77	Полная модель	5.43E-09	0.453	3.779	0.577	0.444	1.491	1.241	4.169	3.892
	T82	Полная модель	4.80E-09	0.460	-	0.589	0.451	1.560	1.250	4.163	3.920
Гибриды	T18	Полная модель	7.59E-09	1.512	3.933	0.548	0.417	2.065	1.538	4.522	4.393
	S50	Полная модель	1.04E-08	0.473	4.308	0.594	0.472	1.634	1.477	4.528	4.309

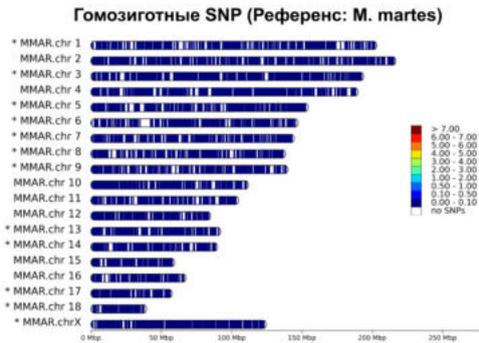
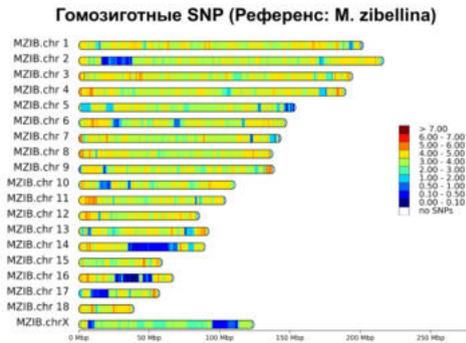
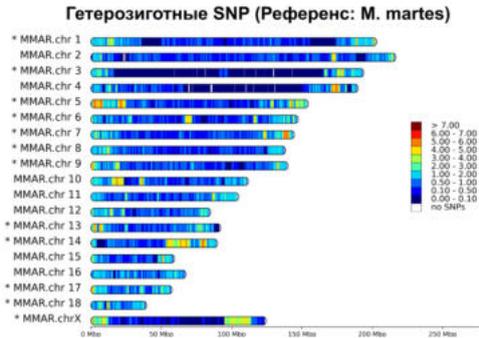
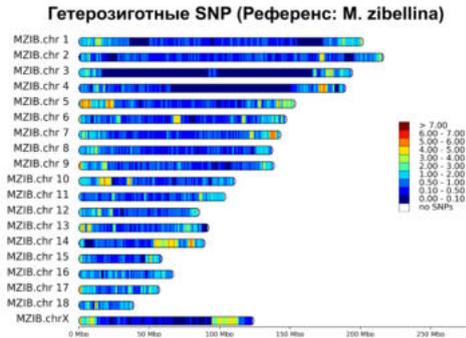
	T151	Полная модель	1.22E-08	0.468	4.543	0.588	0.463	1.727	1.450	4.753	4.549
	T78	Полная модель	9.27E-09	0.448	4.541	0.568	0.440	1.602	1.258	4.739	4.558
	T79	Полная модель	7.95E-09	0.456	4.341	0.573	0.449	1.705	0.720	4.498	4.375
	T81	Полная модель	5.84E-09	0.451	4.312	0.577	0.445	1.575	1.327	4.530	4.332
	T83	Полная модель	8.08E-09	0.451	4.188	0.572	0.445	1.501	1.287	4.446	4.200
	T84	Полная модель	1.13E-08	1.425	4.515	0.750	0.588	1.853	1.538	4.711	4.536
	T85	Полная модель	1.34E-08	0.486	4.422	0.628	0.478	1.752	0.874	4.611	4.440
	T86	Полная модель	1.02E-08	0.479	4.316	0.613	0.478	1.716	1.551	4.516	4.317
	T87	Полная модель	1.95E-08	1.528	4.470	0.765	0.494	2.006	1.581	4.666	4.492

Приложение 25. Количество SNP, средние и медианные значения гетерозиготности (hetSNP/1 kbp) для образцов каменной куницы (10xmfoi) и харзы (10xmfla), американской куницы (SRR11575352) и тихоокеанских куниц: материковых (SRR11575346, SRR11575351) и островных (SRR11575343, SRR11575348). В качестве референса для 10xmfla и 10xmfoi использованы полученные геномные сборки соответствующих видов. Для образцов *M. caurina* и *M. americana* в качестве референса использована геномная сборка соболя. Хромосомы X исключены.

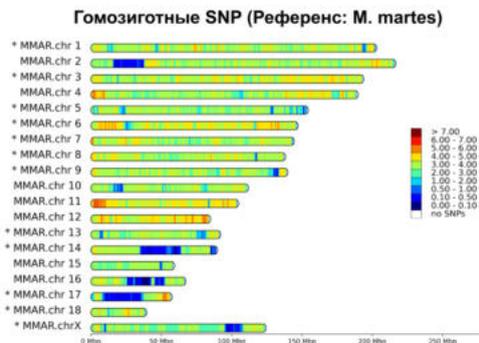
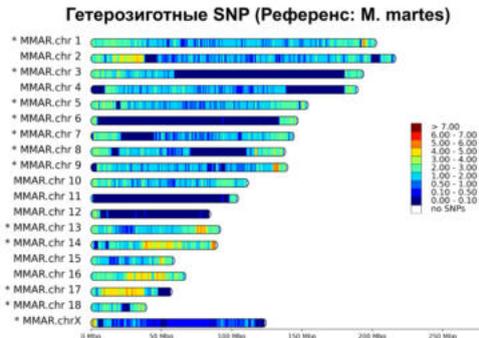
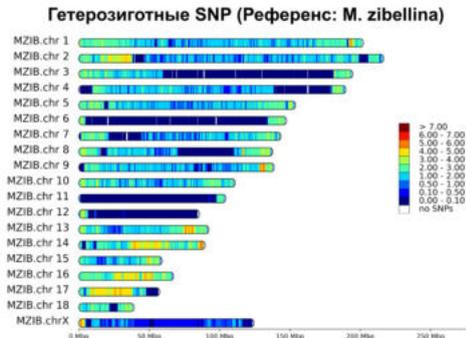
ID	Кол-во гетероSNP, млн	Среднее (SNP/1 тыс п.н.)	Медиана (SNP/1 тыс п.н.)
10xmfla	1.91	0.82	0.78
10xmfoi	0.98	0.42	0.38
SRR11575343	0.27	0.11	0.06
SRR11575346	1.18	0.49	0.47
SRR11575348	0.06	0.02	0.01
SRR11575351	1.35	0.58	0.52
SRR11575352	1.91	0.83	0.82

Приложение 26. Распределения гетерозиготных и гомозиготных SNP (SNPs/kbp) для исследуемых образцов. SNP были подсчитаны в скользящих окнах размером 1 Mbp с шагом 100 kbp и пересчитаны в SNP/kbp. Инвертированные скаффолды в геномной сборке лесной куницы отмечены “*”.

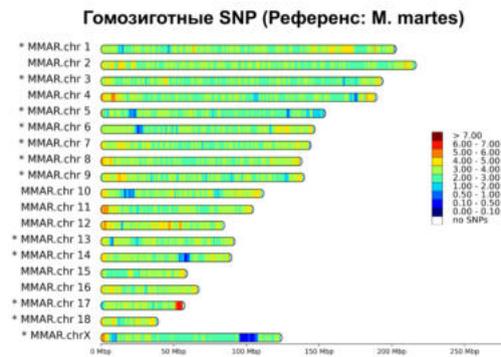
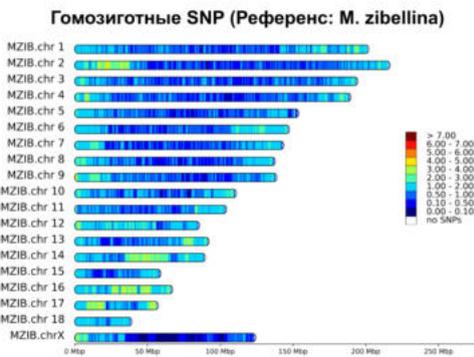
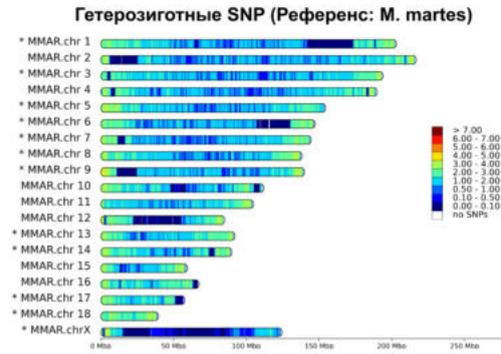
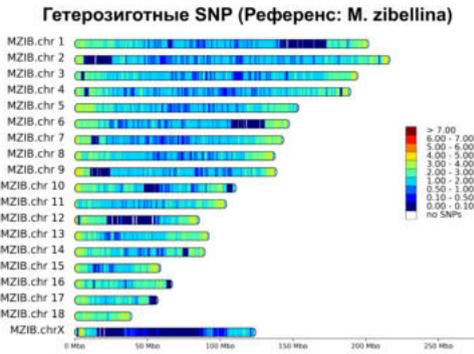
10xmmar



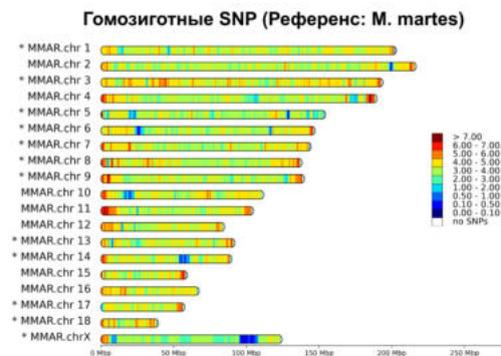
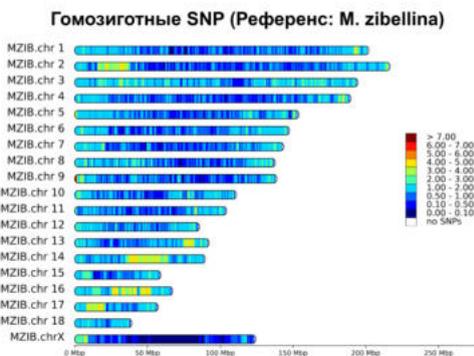
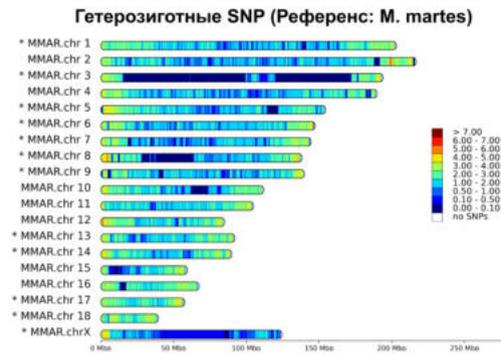
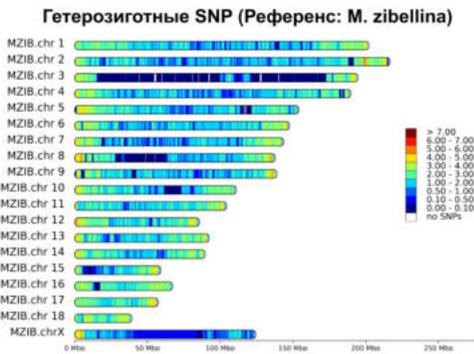
10xmzib



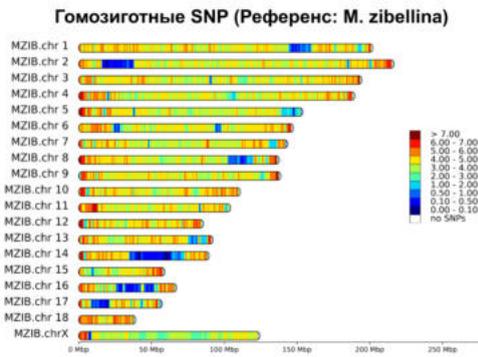
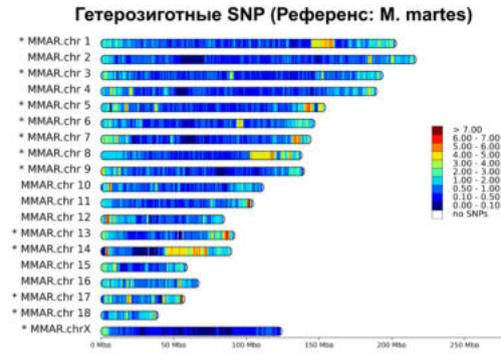
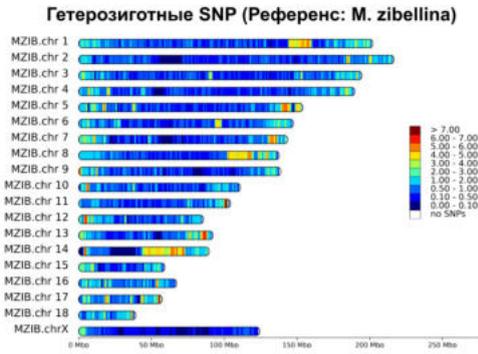
china



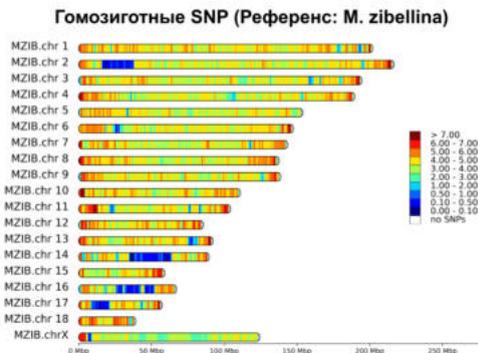
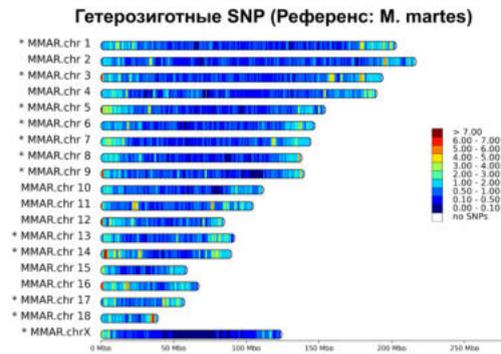
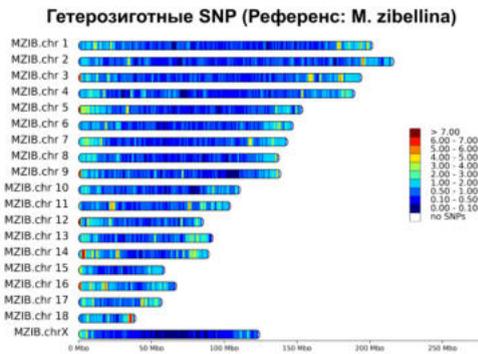
S26



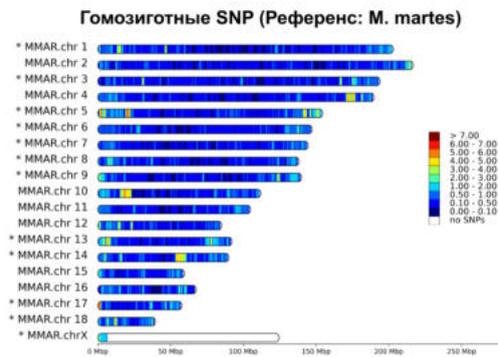
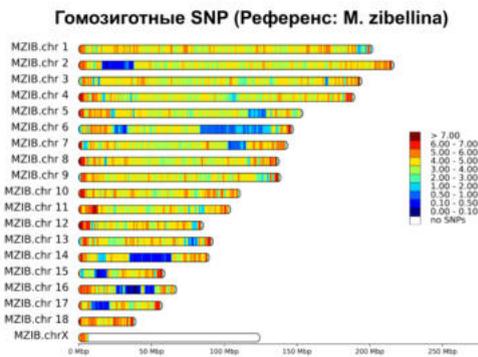
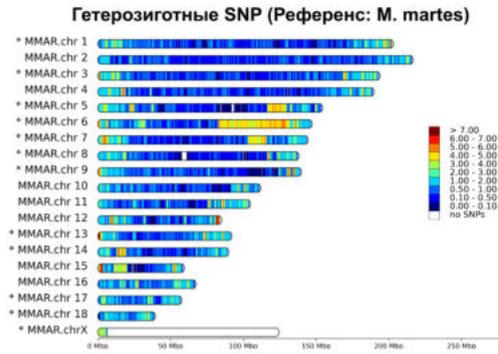
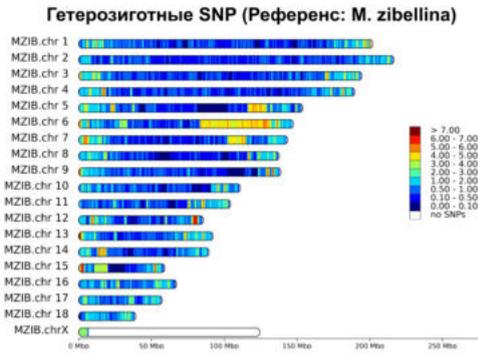
S44



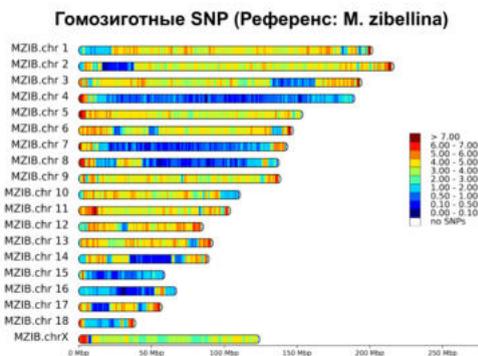
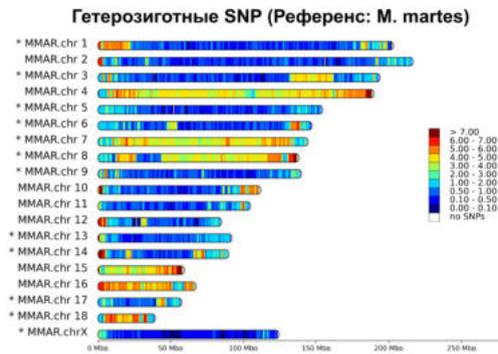
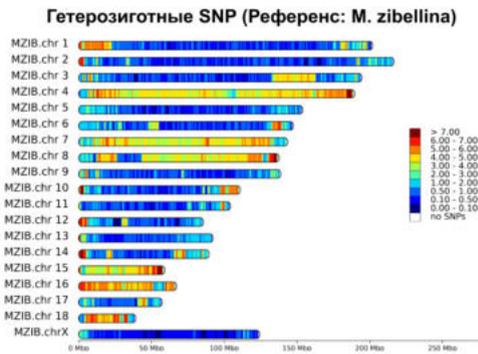
S46



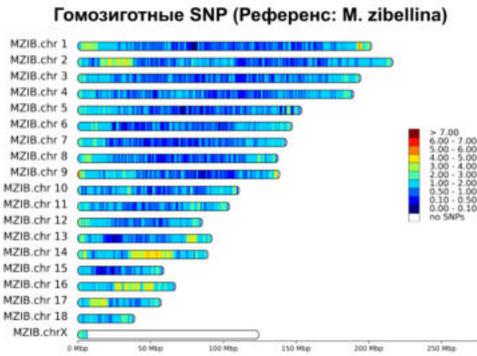
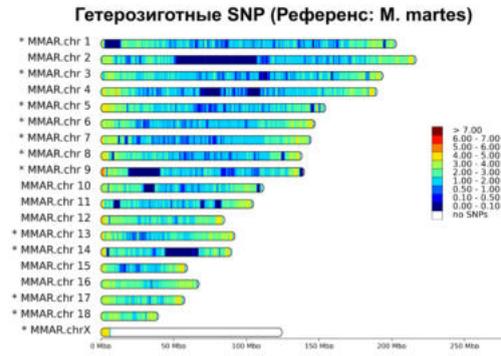
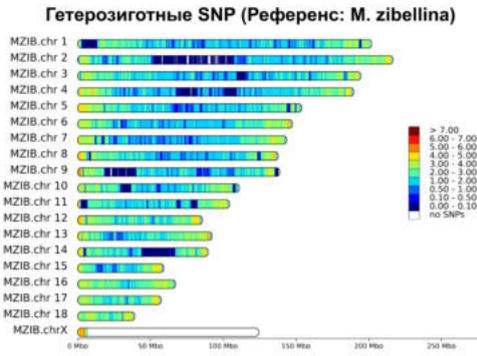
S49



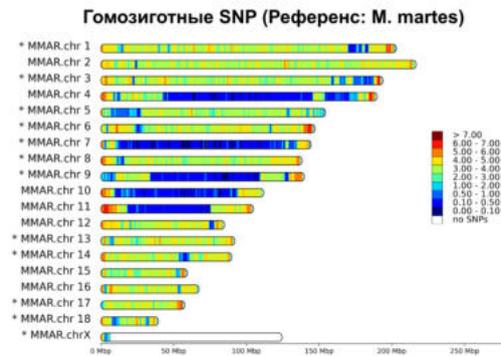
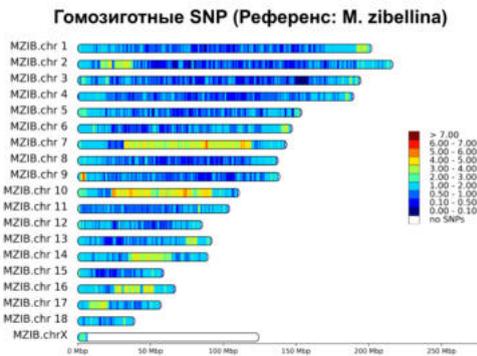
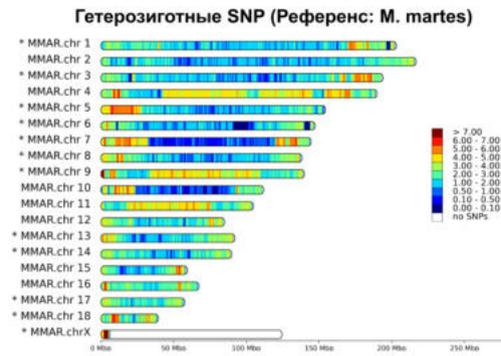
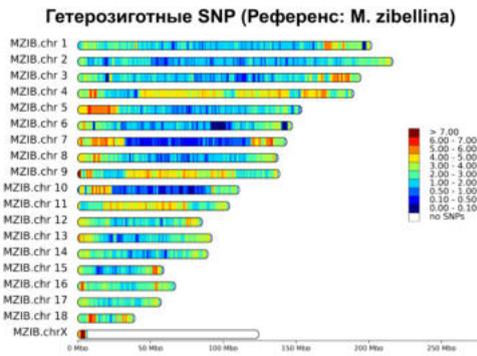
S50



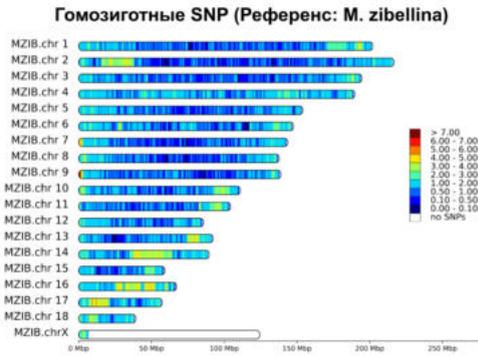
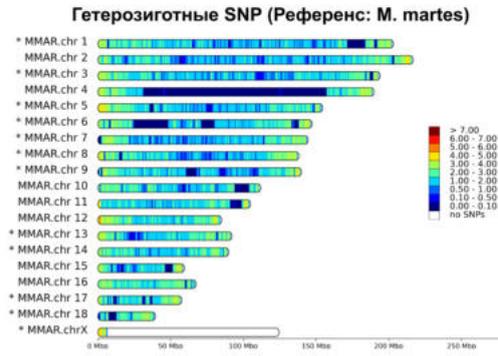
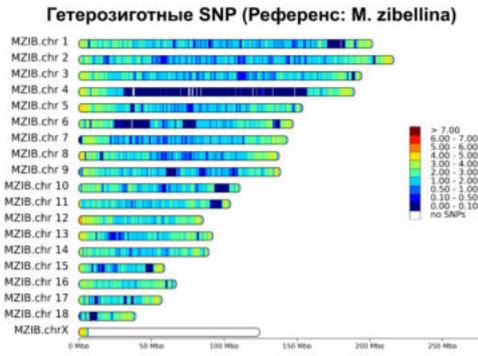
T8



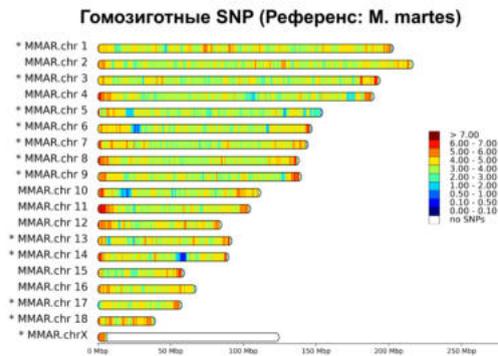
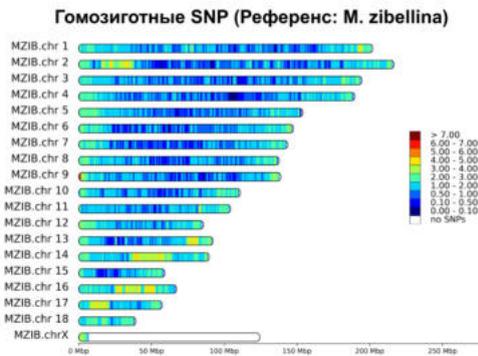
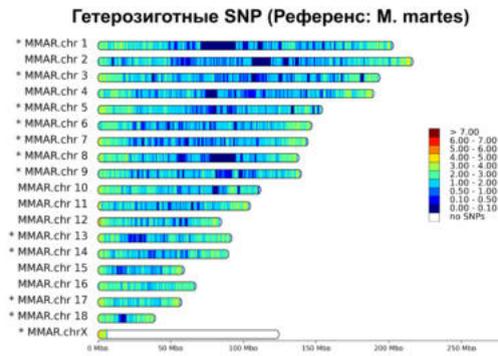
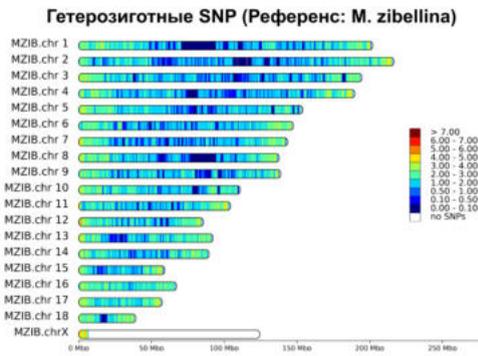
T18



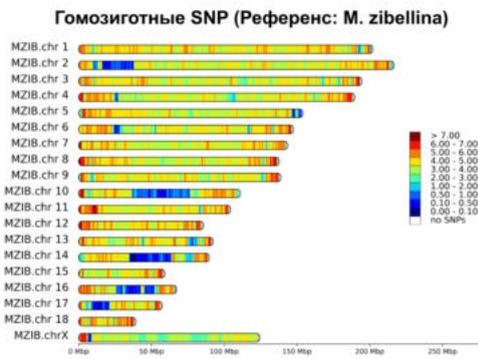
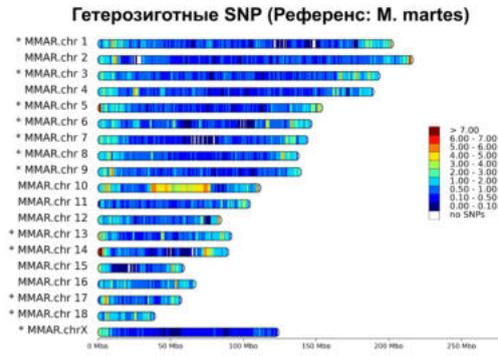
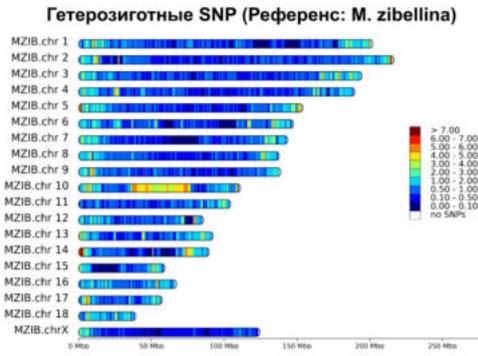
T50



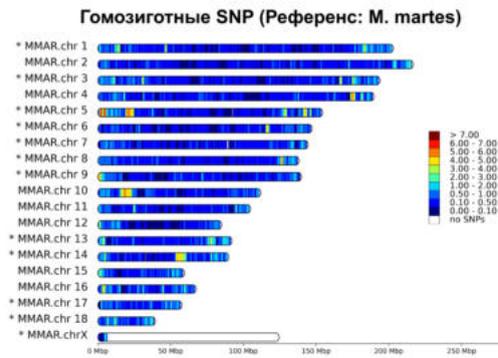
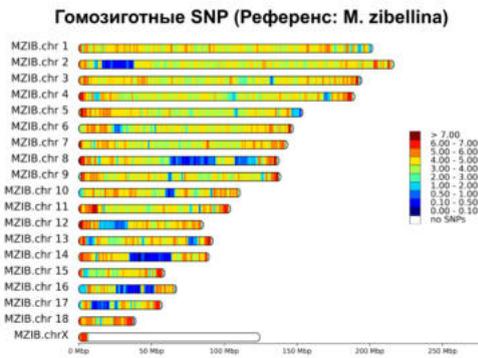
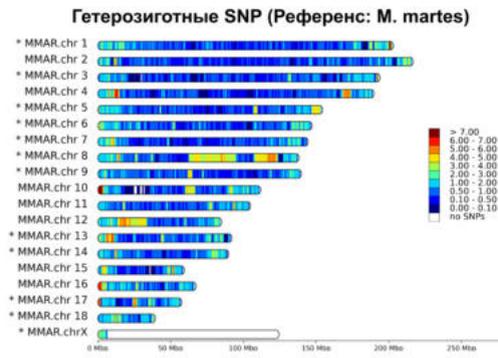
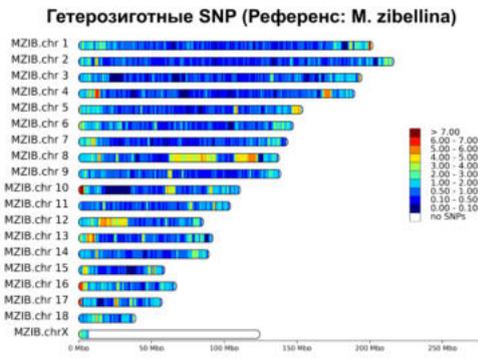
T72



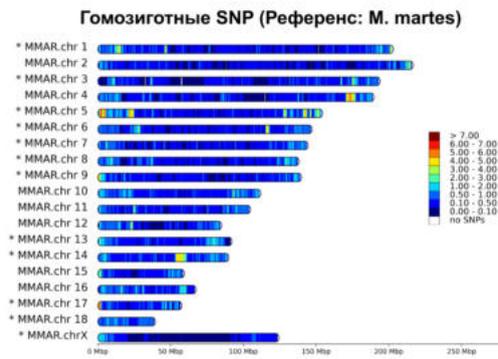
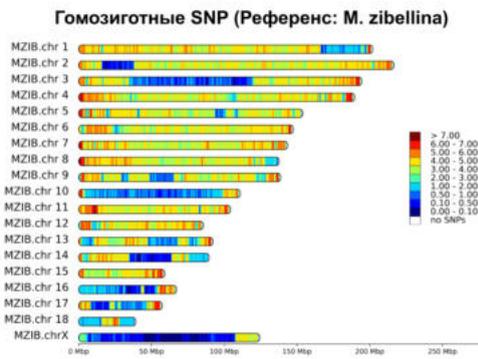
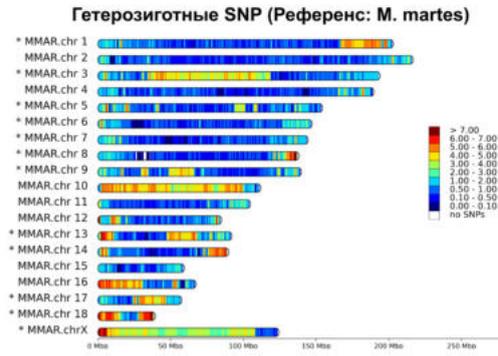
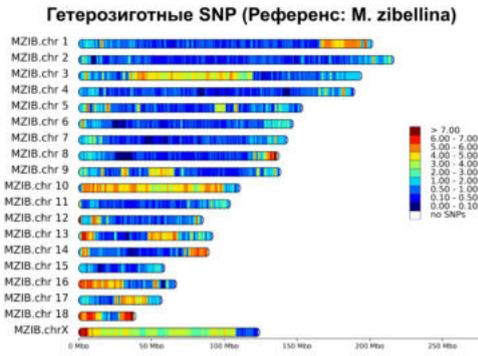
T76



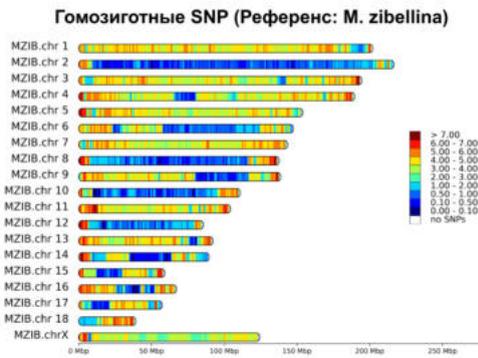
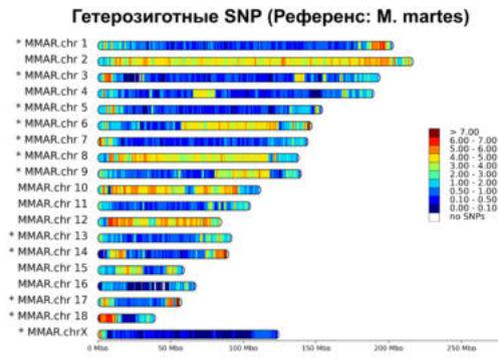
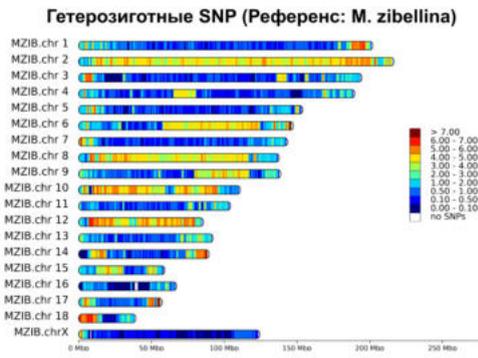
T77



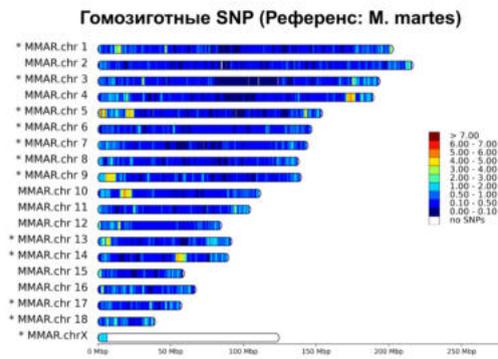
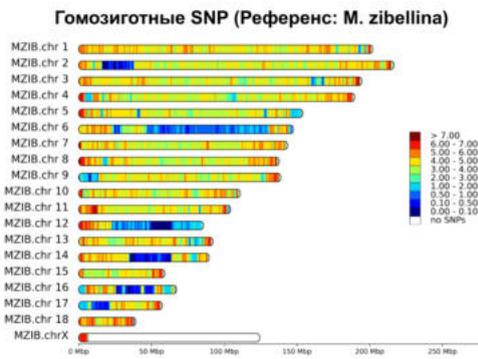
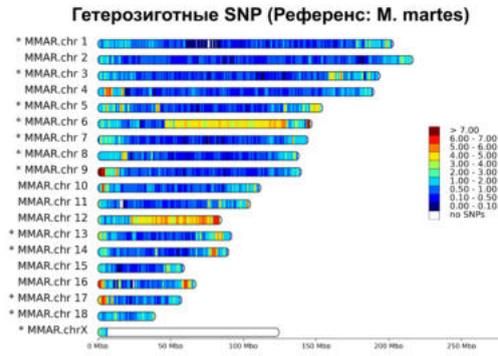
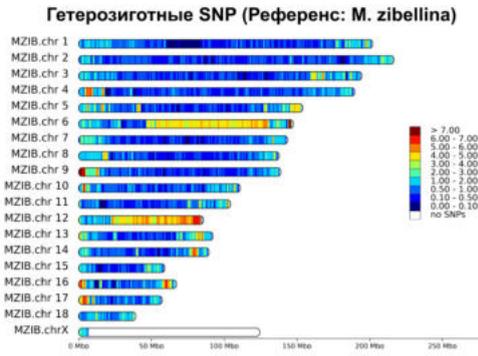
T78



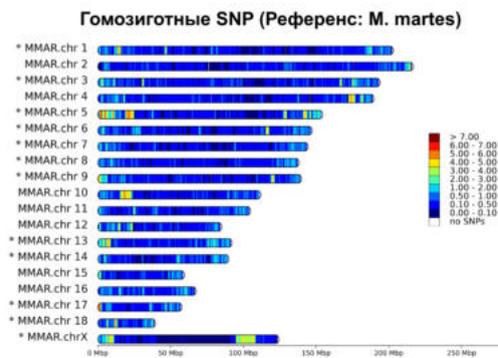
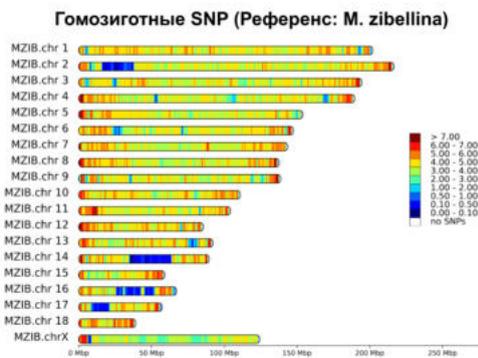
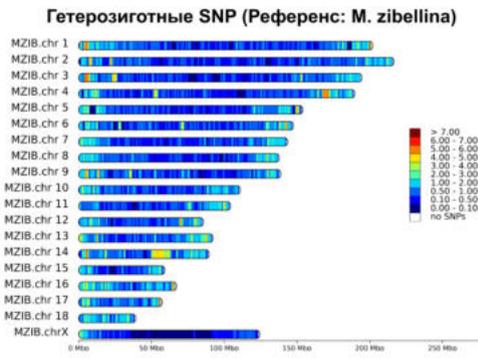
T79



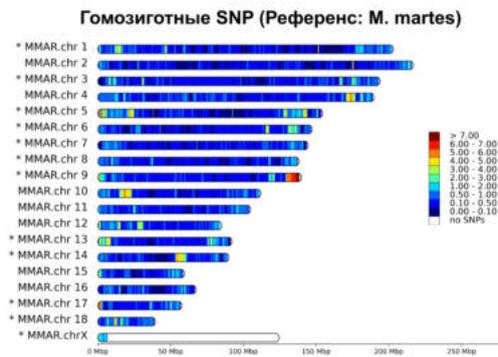
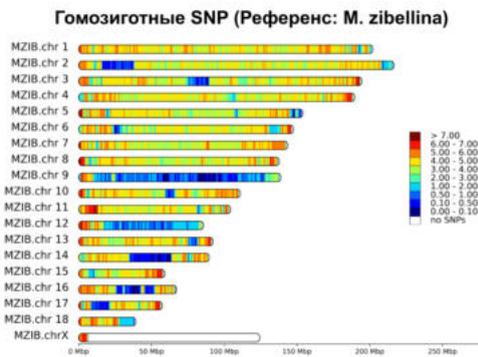
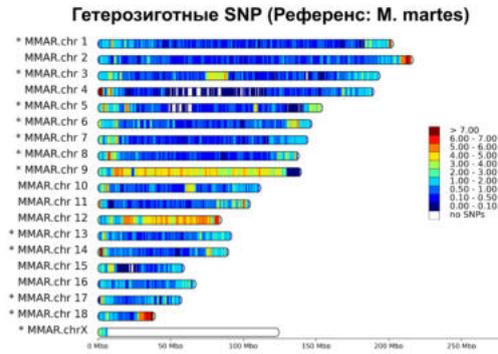
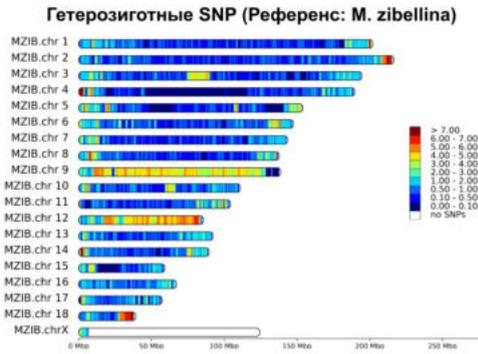
T81



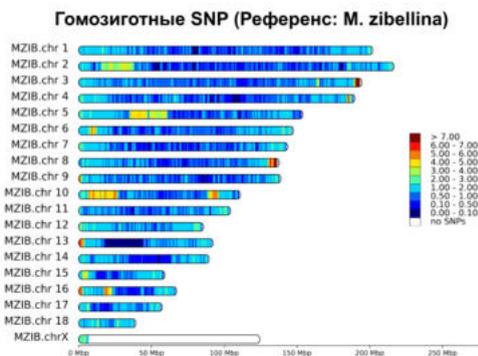
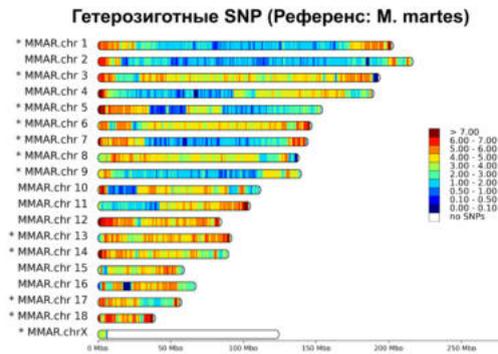
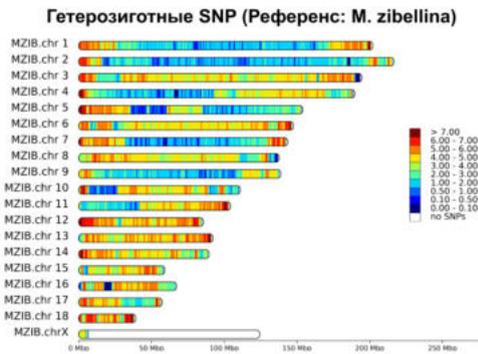
T82



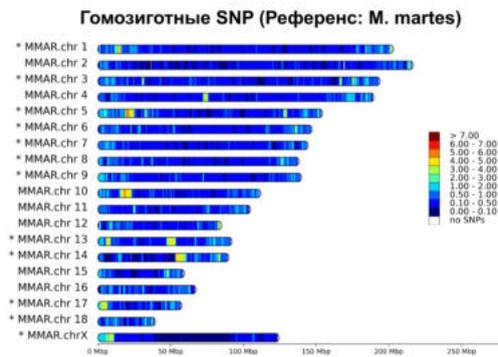
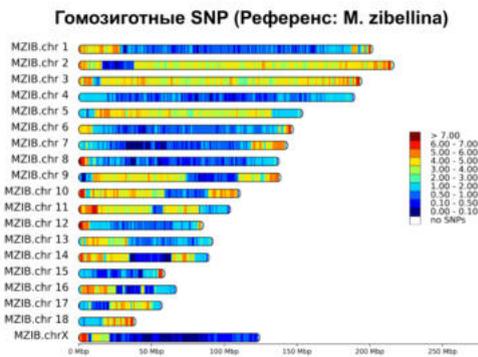
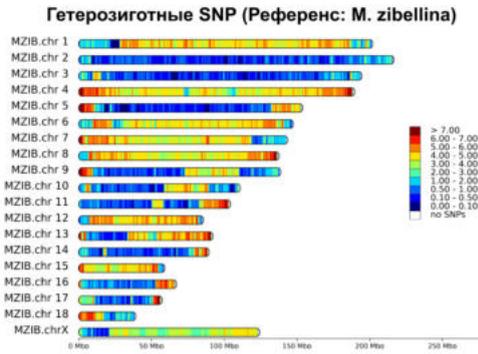
T83



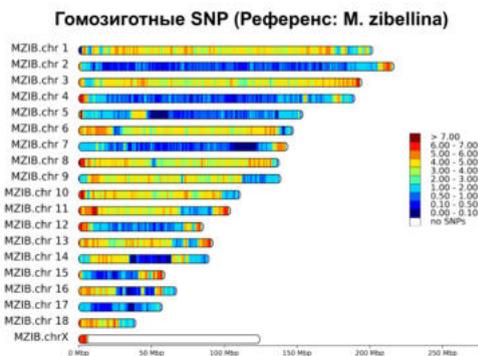
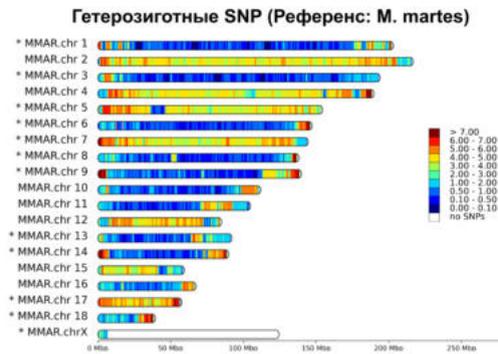
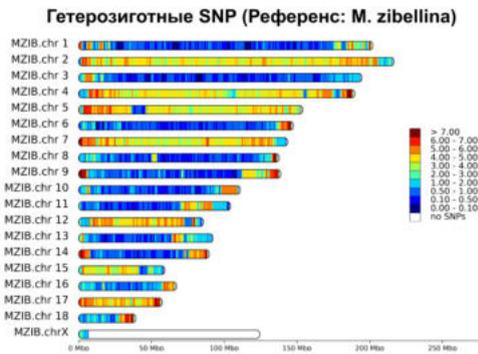
T84



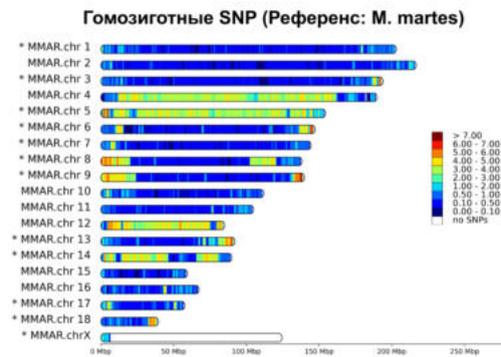
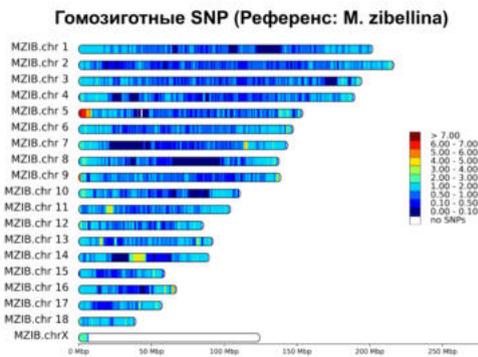
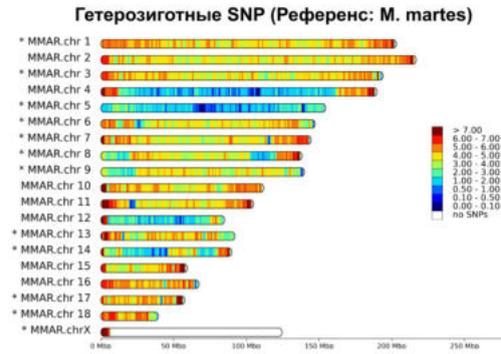
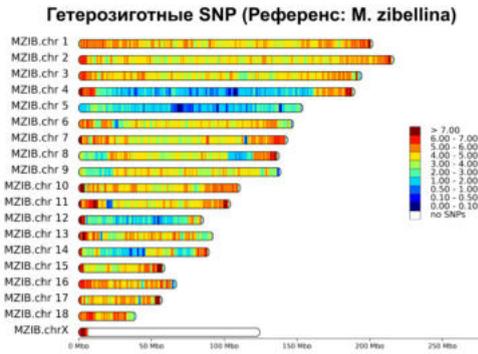
T85



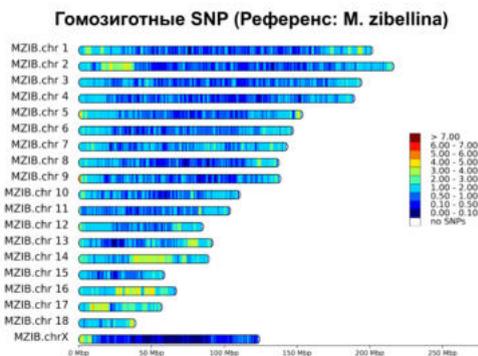
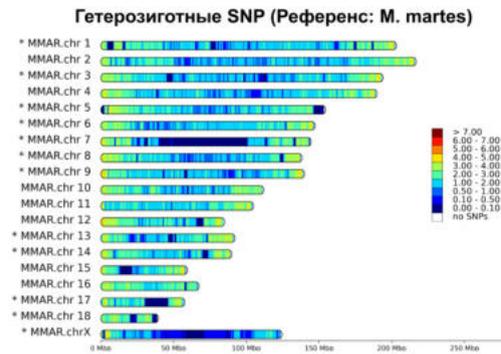
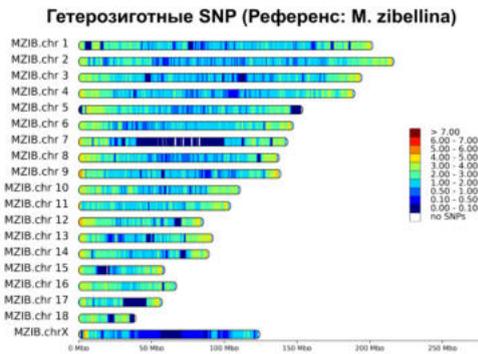
T86



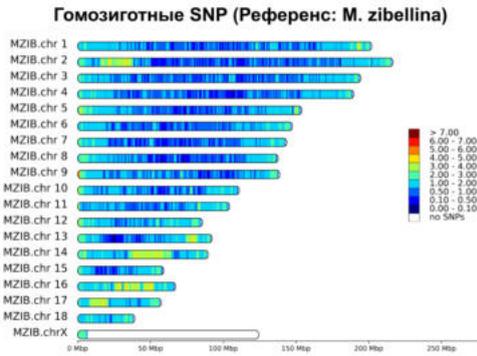
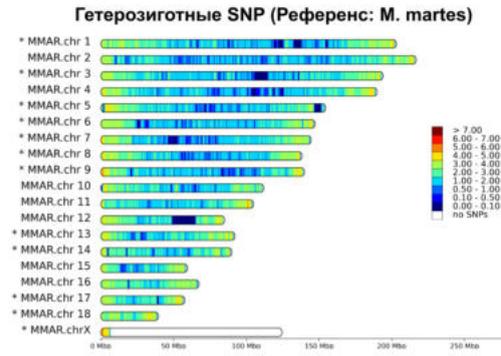
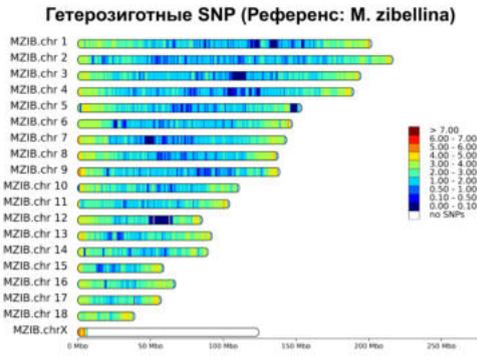
T87



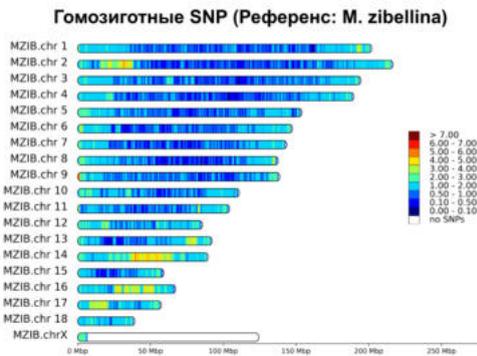
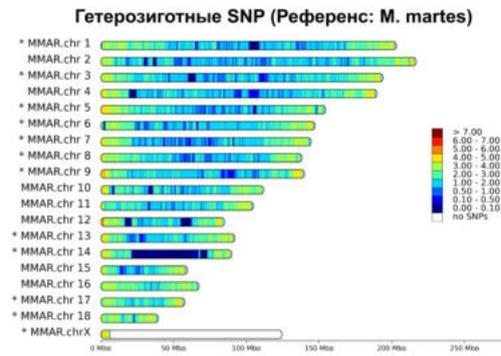
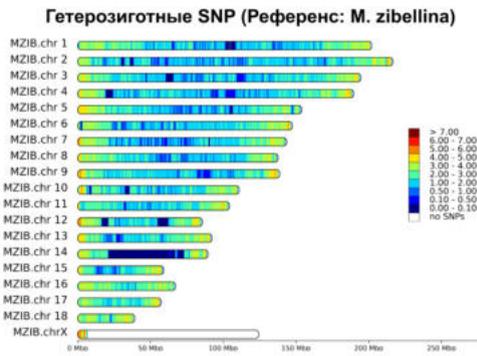
T90



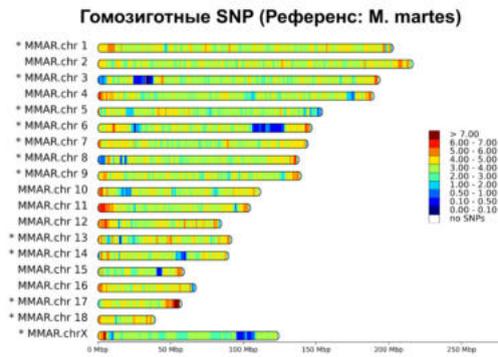
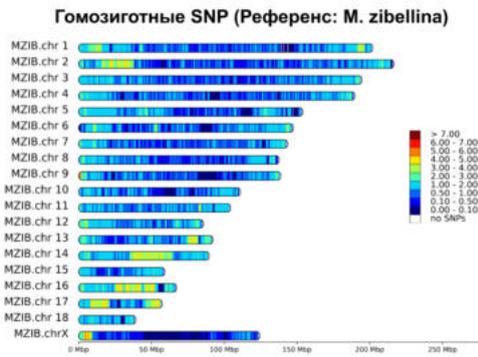
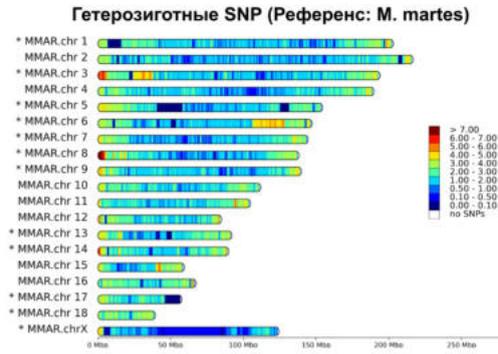
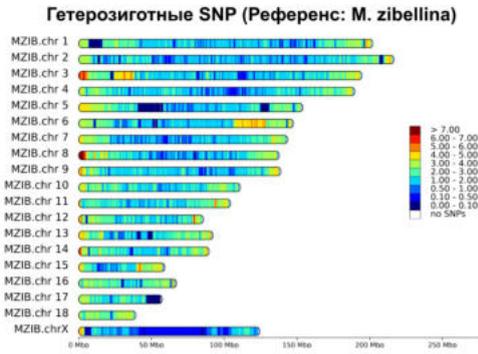
T104



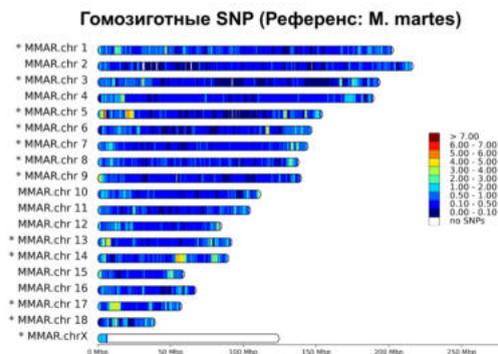
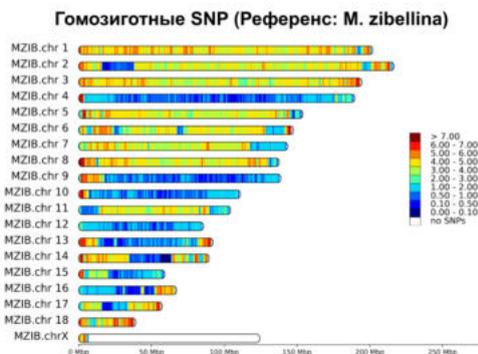
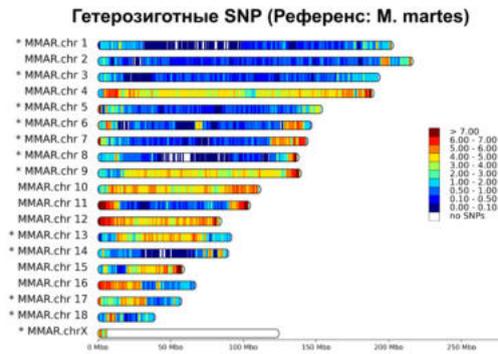
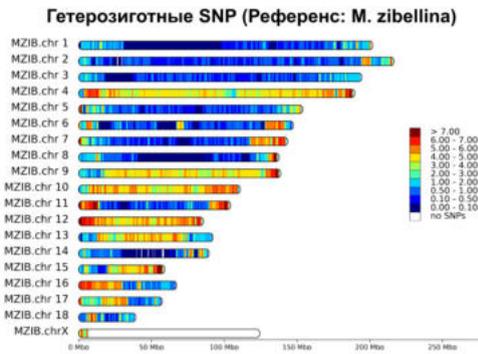
T118



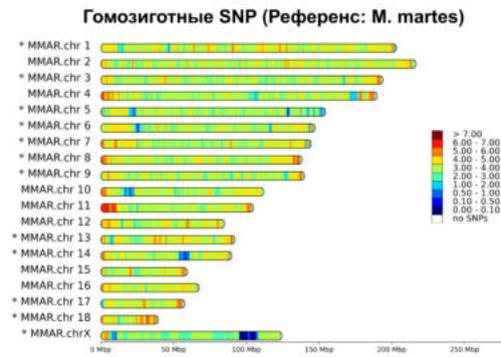
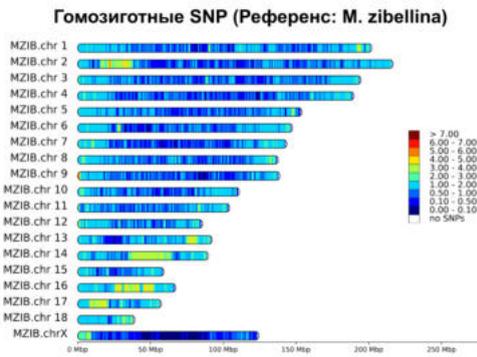
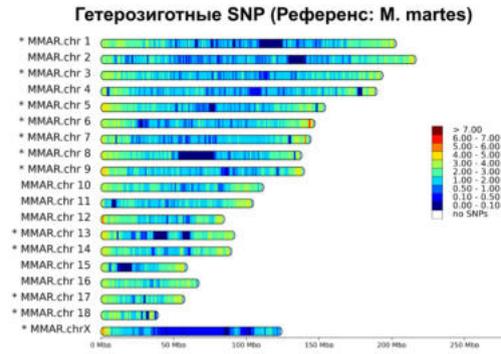
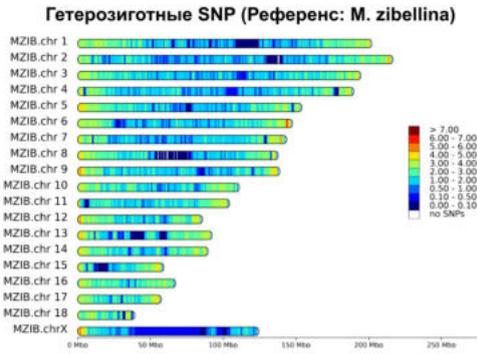
T150



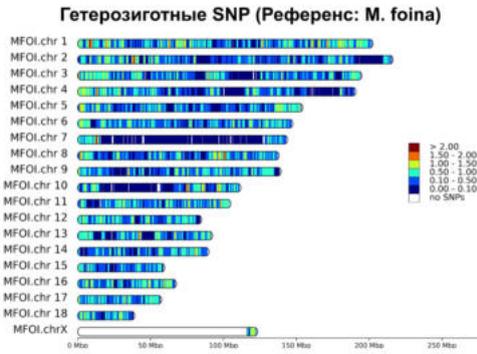
T151



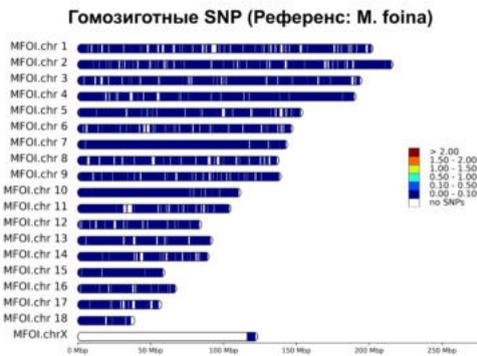
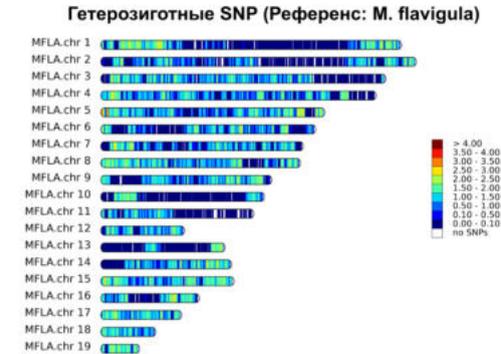
T194



10xmfoi

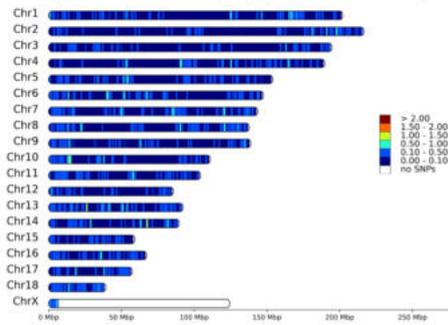


10xmfla



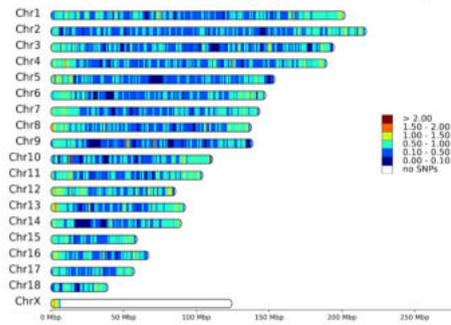
SRR11575343

Гетерозиготные SNP (Референс: *M. zibellina*)



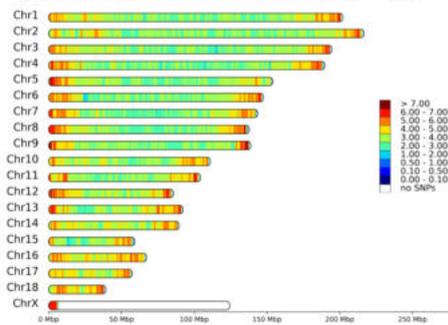
SRR11575346

Гетерозиготные SNP (Референс: *M. zibellina*)



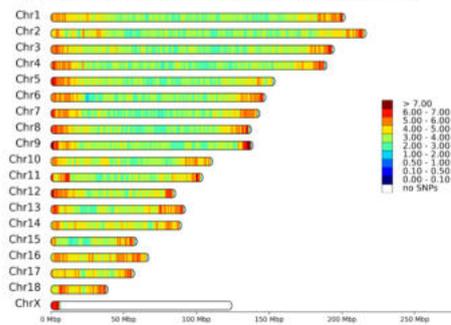
SRR11575343

Гомозиготные SNP (Референс: *M. zibellina*)



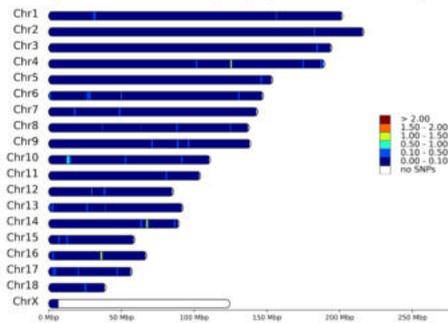
SRR11575346

Гомозиготные SNP (Референс: *M. zibellina*)



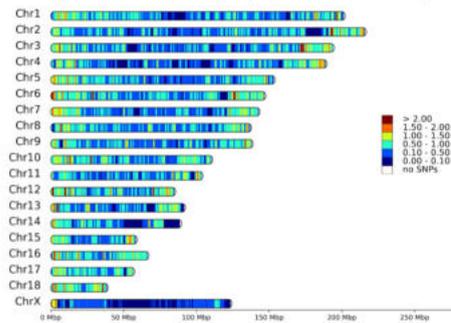
SRR11575348

Гетерозиготные SNP (Референс: *M. zibellina*)



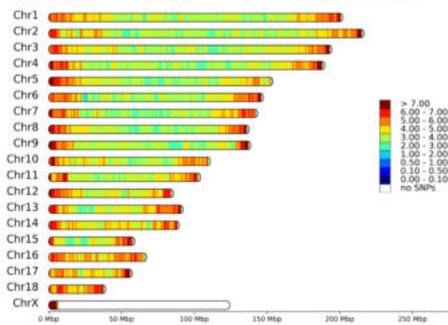
SRR11575351

Гетерозиготные SNP (Референс: *M. zibellina*)



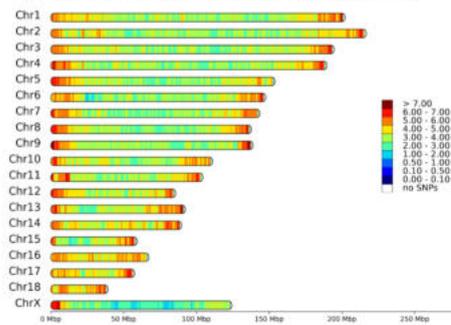
SRR11575348

Гомозиготные SNP (Референс: *M. zibellina*)

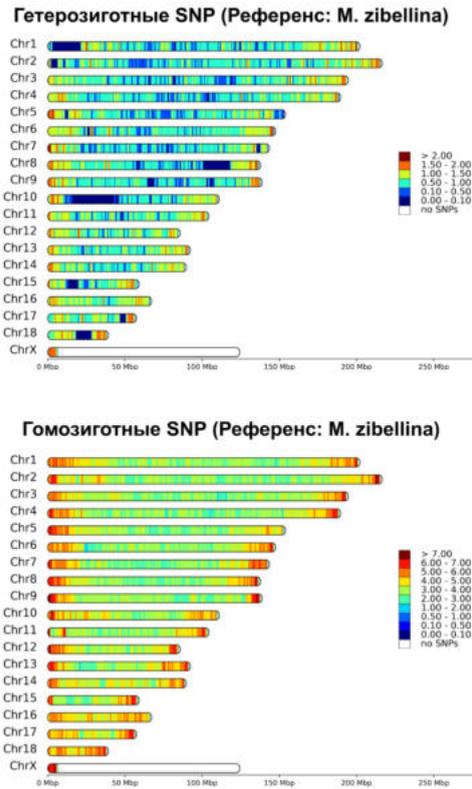


SRR11575351

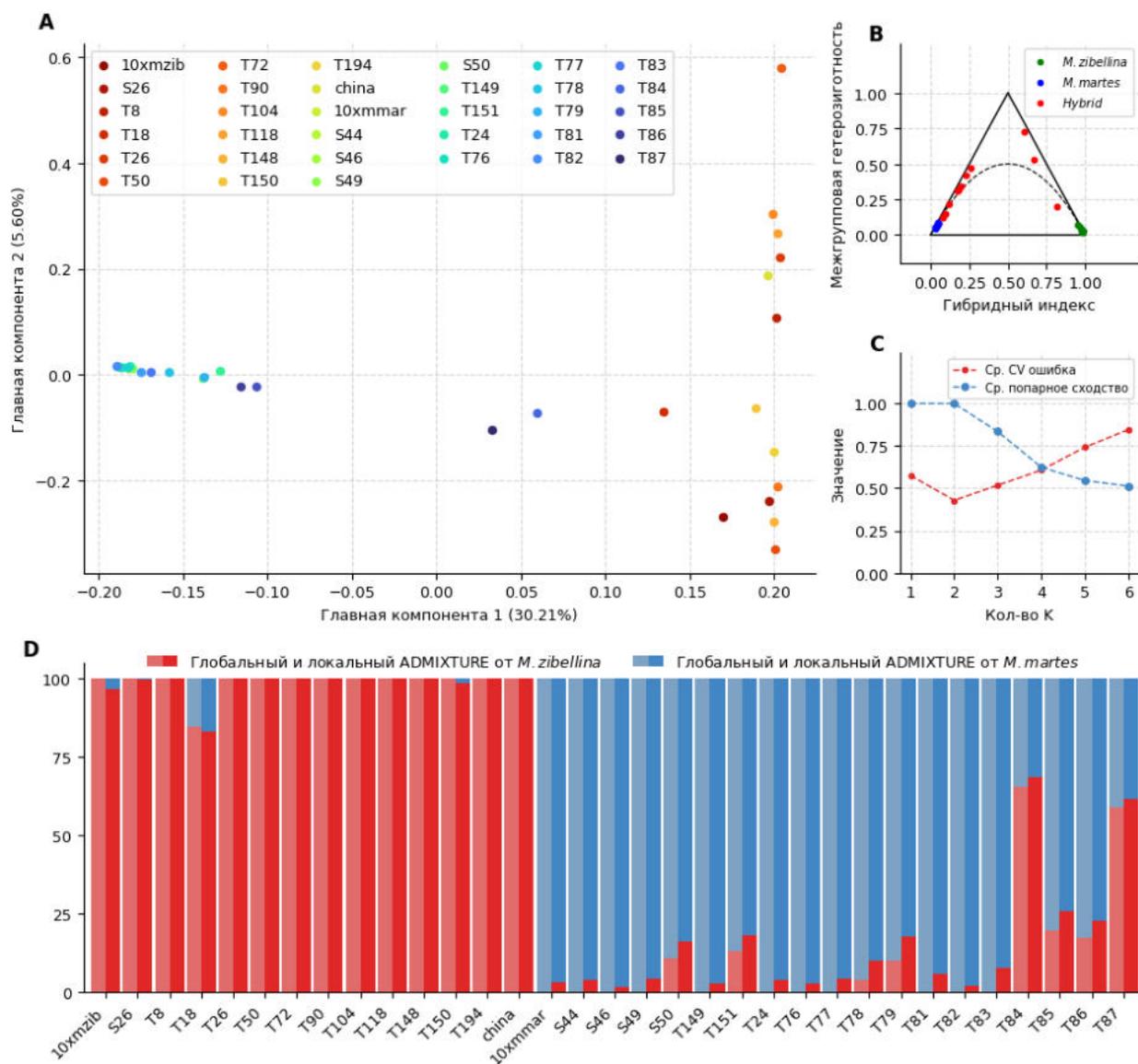
Гомозиготные SNP (Референс: *M. zibellina*)



SRR11575352



Приложение 27. Анализ происхождения на основании геномной сборки *M. martes*: А – Анализ главных компонент (PCA) на основе гетерозиготных SNPs из аутосом и псевдоаутосомного региона (PAR) сборки *M. martes*. Пунктирной линией обозначены кластеры лесных куниц (слева), соболей (справа) и их гибридов (посередине). Образец T18 находится вне кластеров. Первая и вторая главные компоненты представлены на осях X и Y соответственно. В скобках указан процент дисперсии, объясняемой соответствующей компонентой; В – Треугольная диаграмма зависимости индекса гибридизации (ось X) и гетерозиготности (ось Y). Гибридные образцы отмечены красным, лесные куницы – синим, а соболя – зеленым (Референс: *M. martes*); С – Средние значения ошибки перекрестной валидации (синий) и среднего парного сходства (красный) на основе глобальной ADMIXTURE (Референс: *M. martes*); D – Глобальный ADMIXTURE (левая половина) и локальный ADMIXTURE (правая половина), определенный в скользящих окнах размером 1 млн п.н. с шагом 100 тыс п.н. (Референс: *M. martes*). Образцы обозначены по оси X, а по оси Y – средняя вероятность происхождения.



Приложение 28. Результаты HyDe-анализа, рассчитанные для каждой гибридной особи относительно выявленных групп соболей и лесных куниц.

A (Source 1)	B (Source 2)	Гибрид	γ	Z-score
Соболя	Лесные куницы	T18	0.74	196.79
		S50	0.09	142.40
		T151	0.12	168.55
		T78	0.04	72.38
		T79	0.10	152.80
		T81	0.02	44.18
		T83	0.04	67.63

		T84	0.53	289.06
		T85	0.16	207.59
		T86	0.14	186.33
		T87	0.47	304.55

Приложение 29. Значения F3-статистики, рассчитанные для каждой гибридной особи относительно выявленных групп соболей и лесных куниц.

A (Source 1)	B (Source 2)	C (Target)	F3	Стандартная ошибка	Z-score
Соболя	Лесные куницы	T18	-0.056455	0.00701	-8.053
		S50	-0.134154	0.007635	-17.572
		T151	-0.136879	0.008127	-16.843
		T78	-0.094713	0.008567	-11.055
		T79	-0.126304	0.008348	-15.129
		T81	-0.037167	0.008966	-4.145
		T83	-0.035522	0.010772	-3.298
		T84	-0.172081	0.005606	-30.696
		T85	-0.183036	0.005761	-31.773
		T86	-0.169621	0.006112	-27.75
		T87	-0.211518	0.004697	-45.032

Приложение 30. Значения D- и F4-статистик, рассчитанные между группами соболей, лесных куниц и их гибридов. В качестве аутгруппы использована каменная куница.

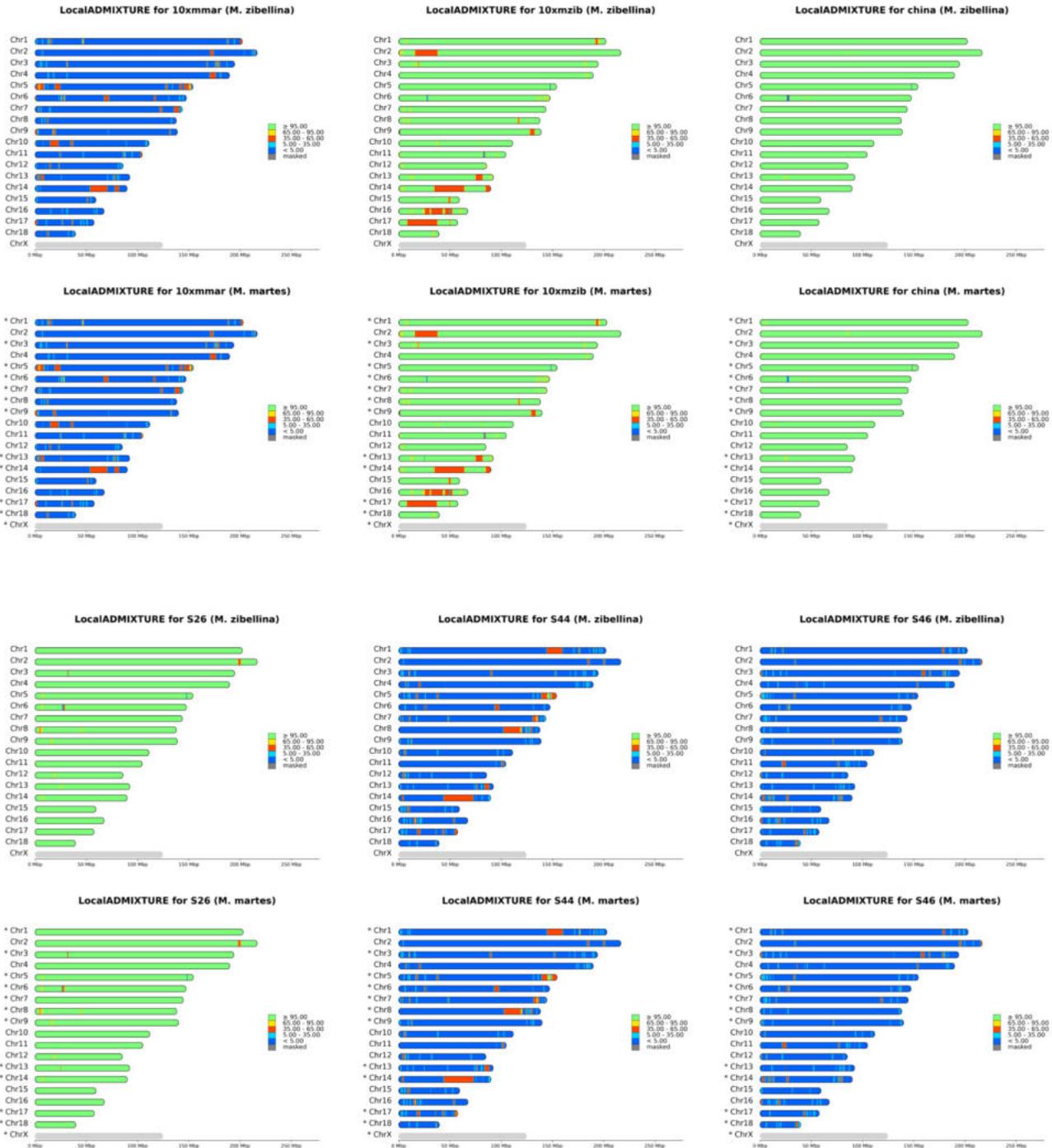
W (Target)	X (Source 1)	Y (Source 2)	Z (Outgroup)	Значение статистик	Стандартная ошибка	Z-score
Кидасы	Лесные куницы	Соболя	<i>M. foina</i>	D = 0.4642	0.010433	44.49
				F4 = 0.019508	0.000572	34.106
Кидасы	Соболя	Лесные куницы	<i>M. foina</i>	D = 0.7847	0.006701	100
				F4 = 0.082066	0.001425	57.585

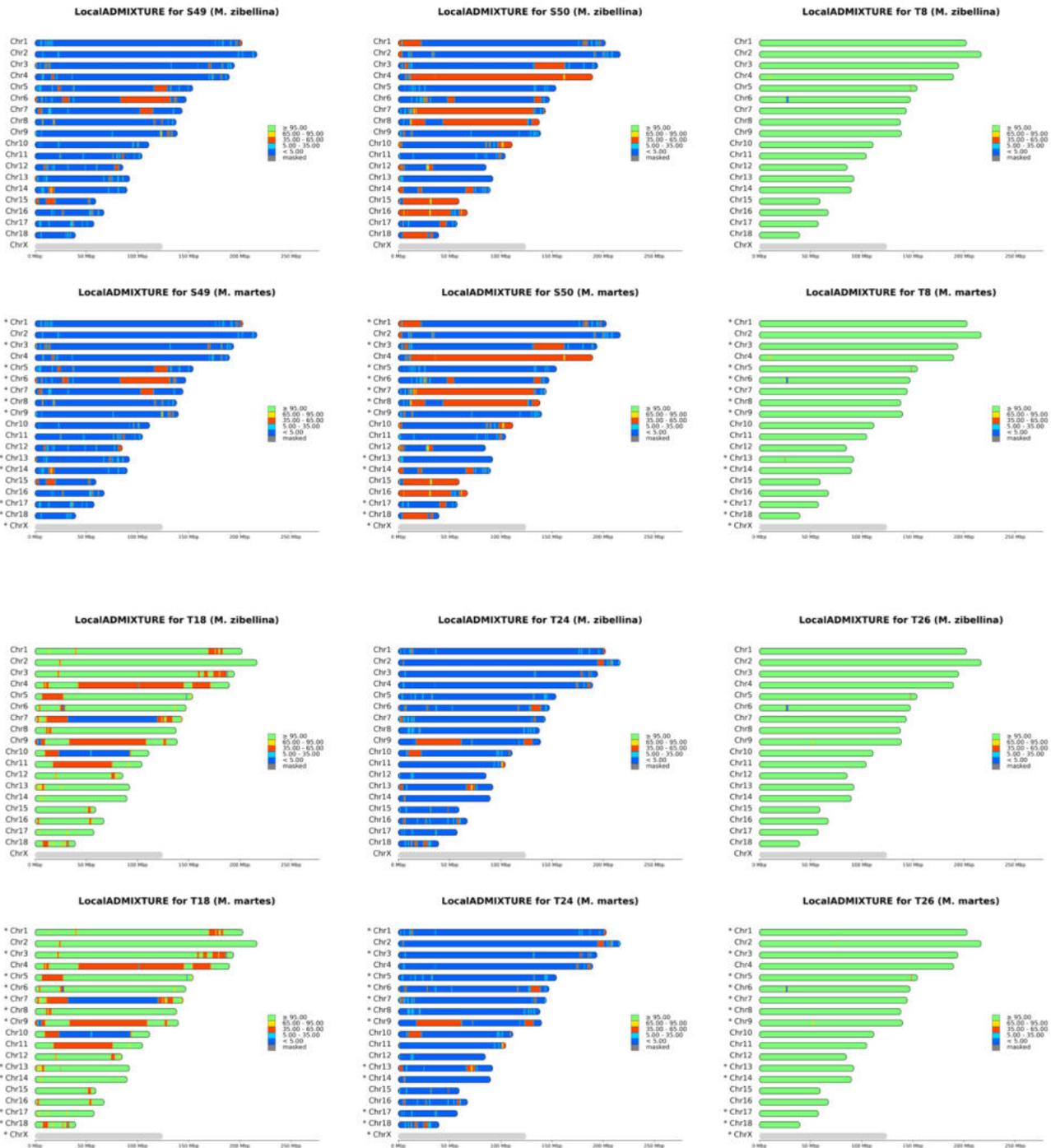
Приложение 31. Значения D- и F4-статистик, рассчитанные для каждой гибридной особи относительно выявленных групп соболей и лесных куниц. В качестве аутгруппы использована каменная куница.

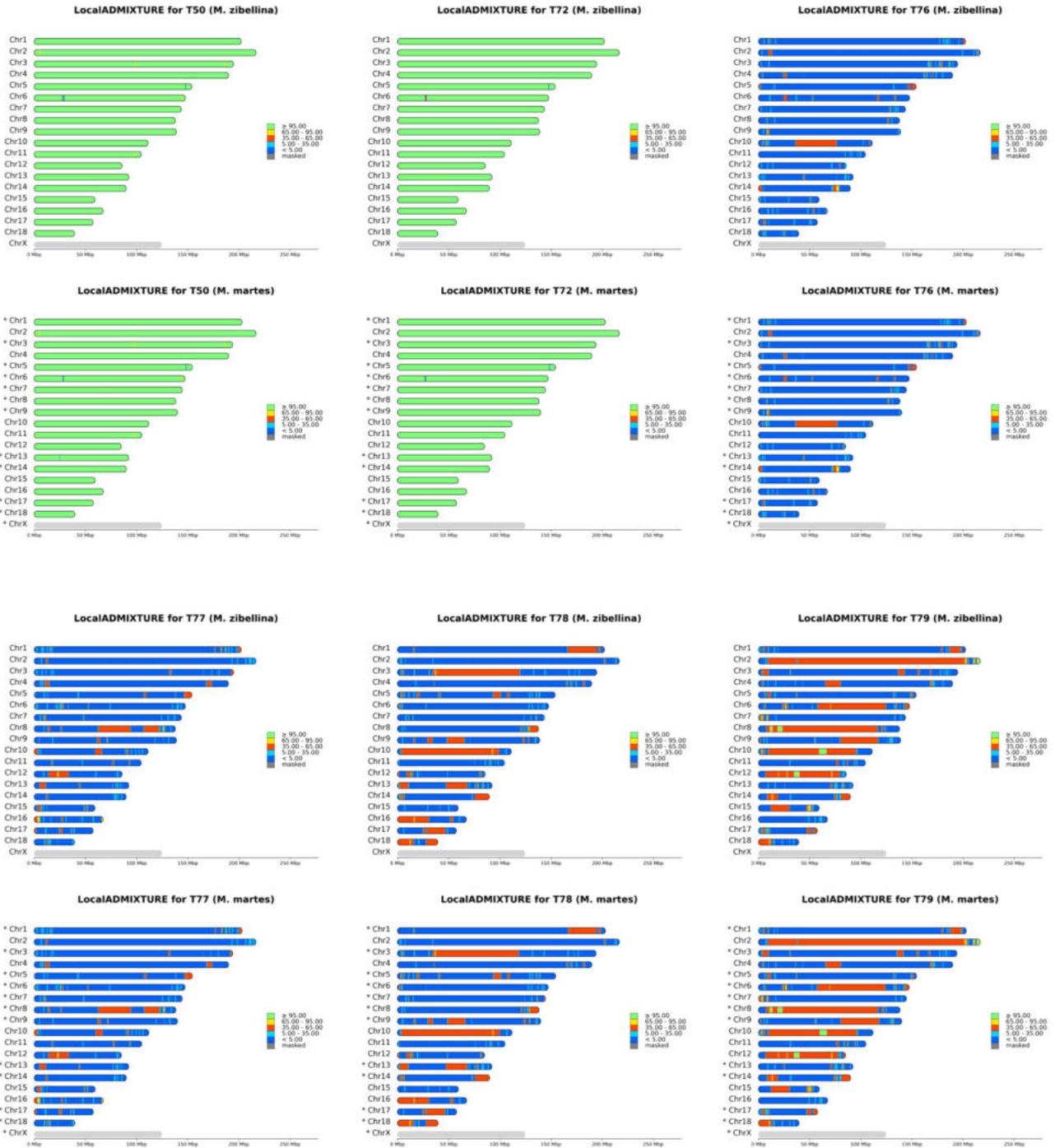
W (Target)	X (Source 1)	Y (Source 2)	Z (Outgroup)	Значение статистик	Стандартная ошибка	Z-score
------------	--------------	--------------	--------------	--------------------	--------------------	---------

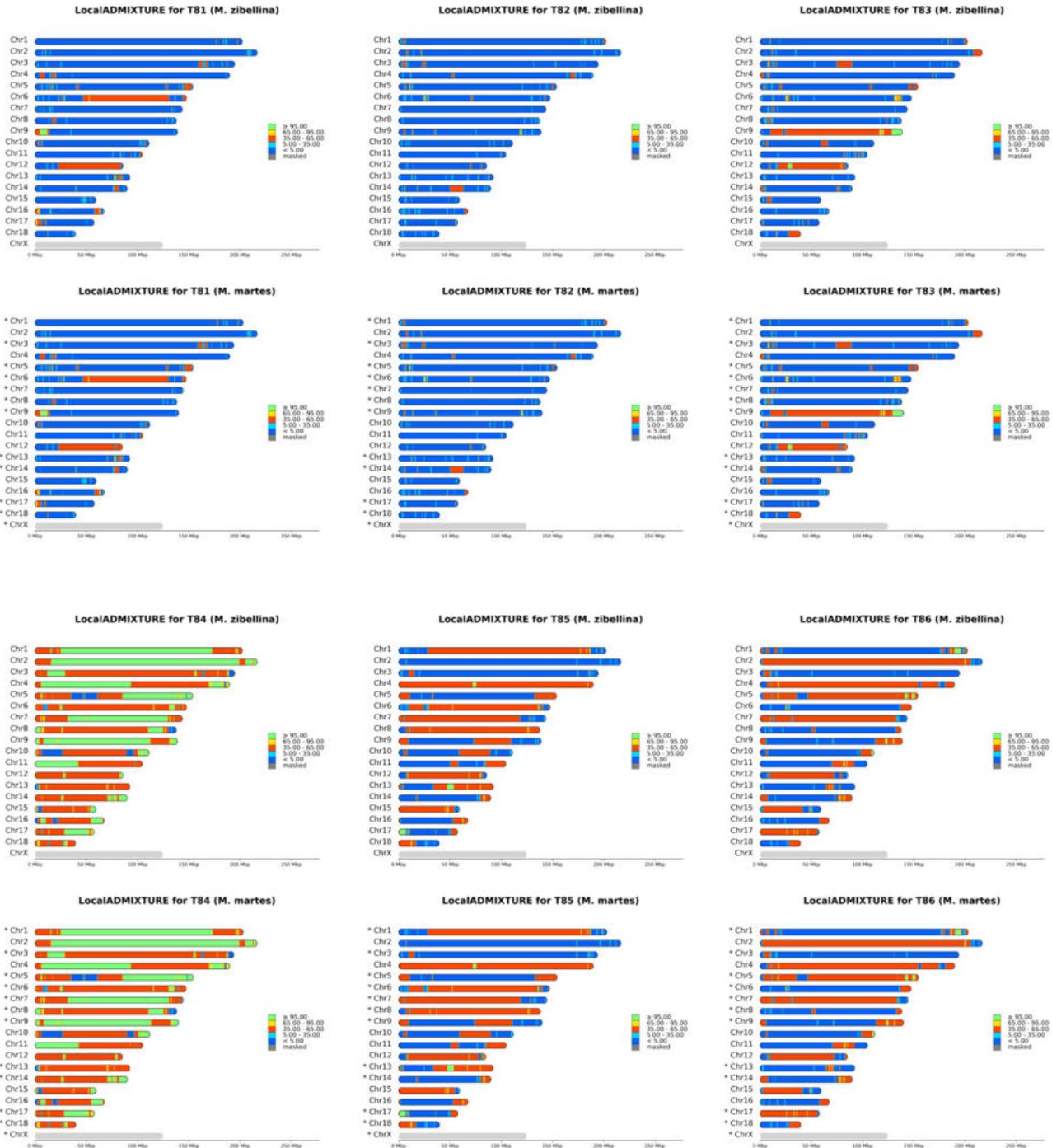
T18	Лесные куницы	Соболя	<i>M. foina</i>	D = 0.6211	0.007773	79.898
				F4 = 0.053196	0.001246	42.691
S50				D = 0.3386	0.020577	16.454
				F4 = 0.010265	0.000918	11.177
T151				D = 0.3829	0.017885	21.412
				F4 = 0.012809	0.000963	13.307
T78				D = 0.1938	0.019473	9.953
				F4 = 0.004603	0.000586	7.856
T79				D = 0.3603	0.020848	17.28
				F4 = 0.011151	0.00096	11.614
T81				D = 0.1266	0.025182	5.028
				F4 = 0.002652	0.000609	4.352
T83				D = 0.1856	0.025343	7.325
				F4 = 0.004219	0.000719	5.869
T84	D = 0.6014	0.009368	64.195			
	F4 = 0.043018	0.001266	33.988			
T85	D = 0.4456	0.01571	28.361			
	F4 = 0.017083	0.001021	16.735			
T86	D = 0.41	0.016547	24.777			
	F4 = 0.014739	0.000963	15.299			
T87	D = 0.5991	0.009369	63.944			
	F4 = 0.040847	0.00125	32.685			

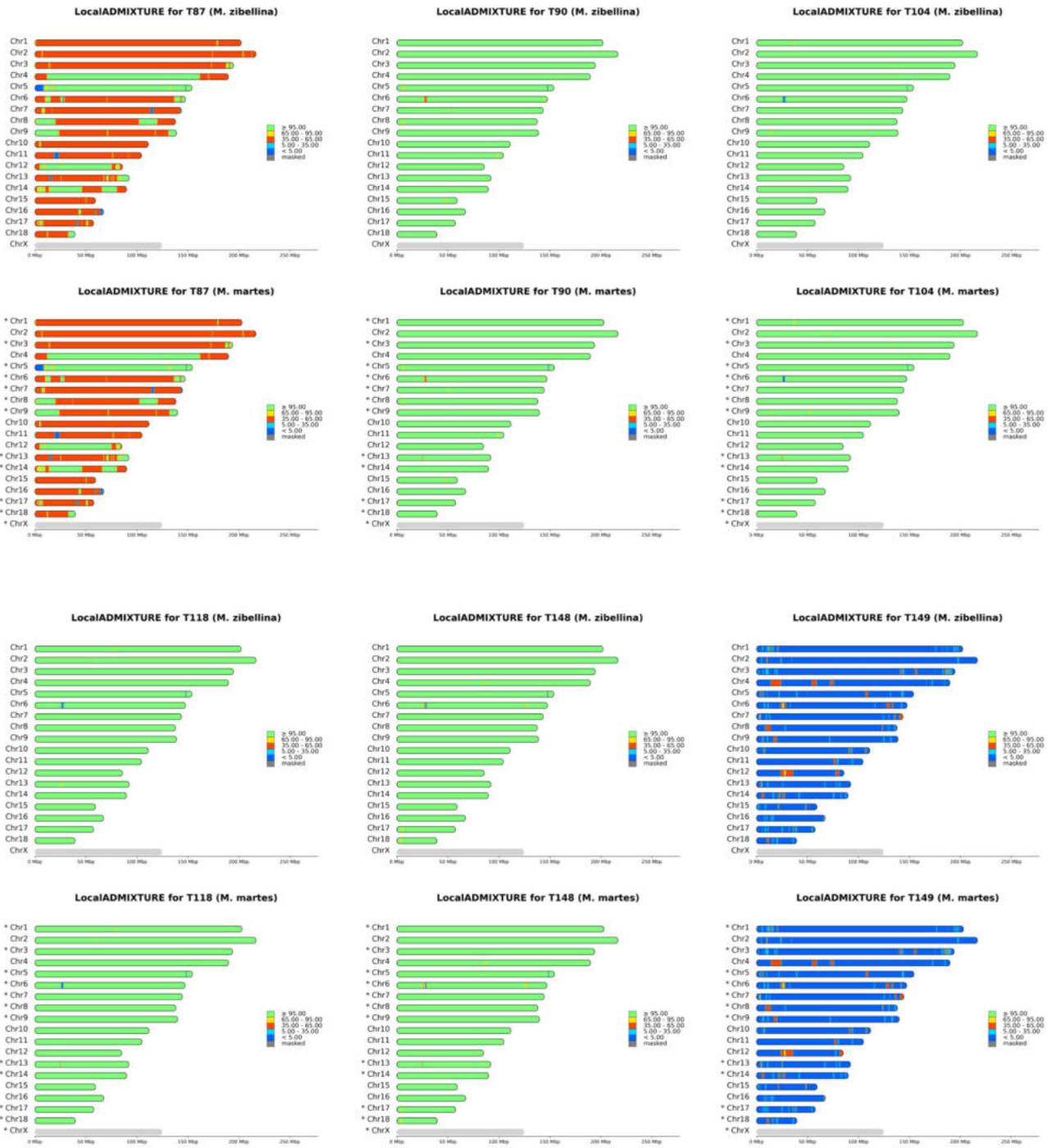
Приложение 32. Результаты локального ADMIXTURE-анализа для образцов соболя, лесной куницы и их гибридов на основе геномных сборок обоих видов.

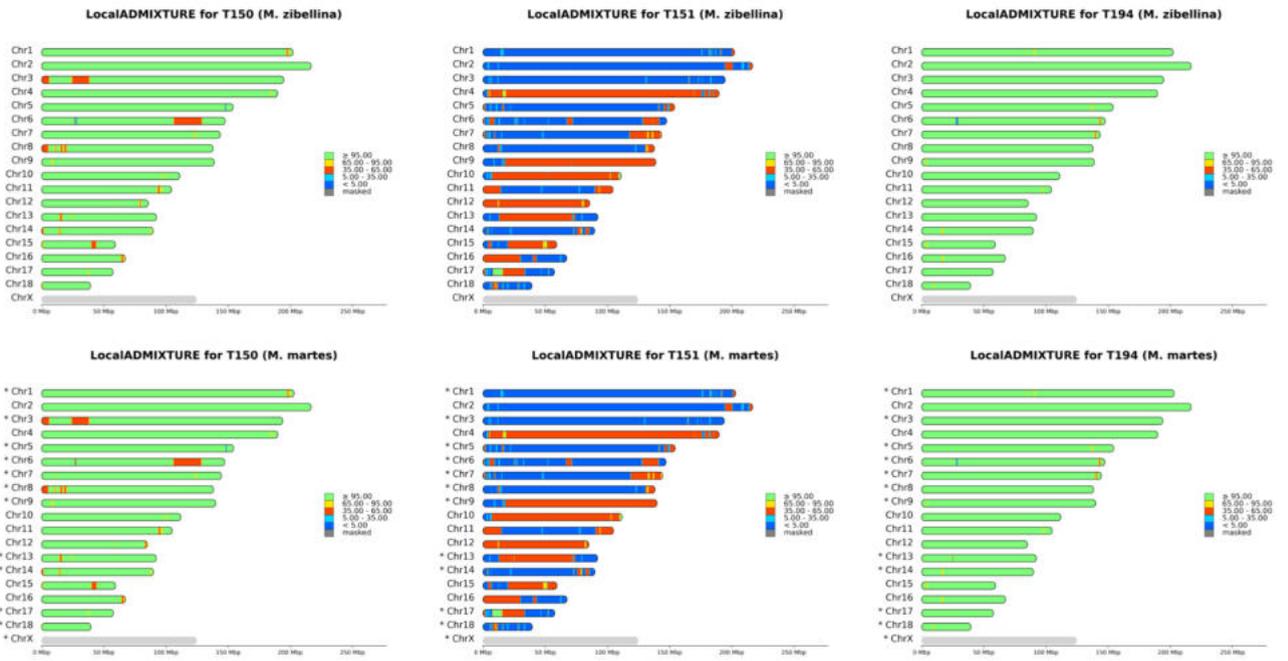










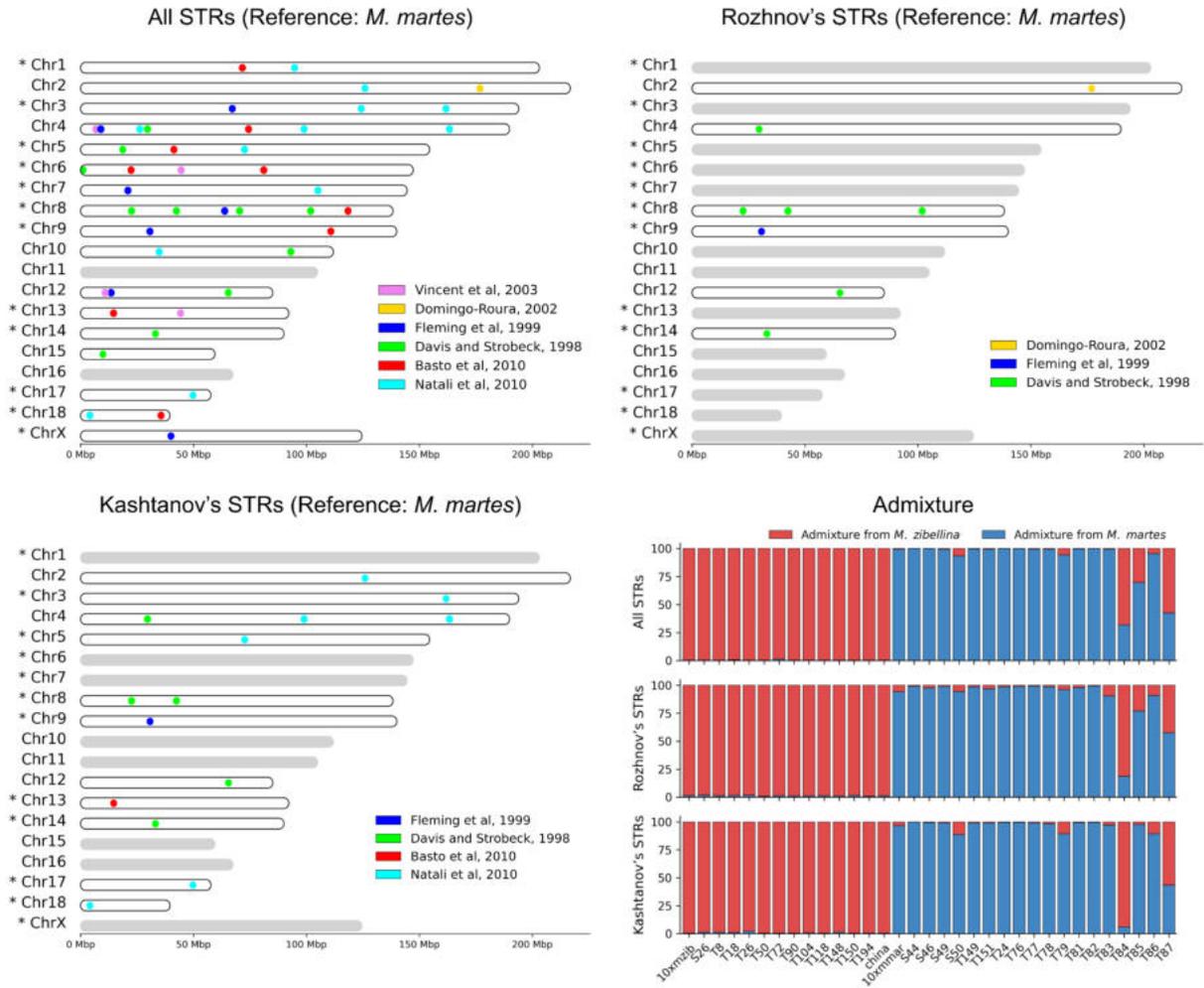


Приложение 33. Описание STR-локусов *M. martes* и *M. zibellina*.

ID	Исходная публикация	Скаффолд	Мотив	Начало повтора	Конец повтора	Мотив из оригинальной статьи
Ma-1	Davis and Strobeck, 1998	HiC_scaffold_4	(CA)1(AA)1(CA)19(TA)1(CA)4(TA)1(CA)3	29396510	29396570	(TG)4TA(TG)19
Ma-11	Davis and Strobeck, 1998	HiC_scaffold_5	(AA)1(CA)8	18721409	18721427	(TG)10
Ma-15	Davis and Strobeck, 1998	HiC_scaffold_13	(CA)6(AC)2(TC)1(GG)1(GC)1(AA)1(CA)10	32985386	32985430	(TG)9
Ma-18	Davis and Strobeck, 1998	HiC_scaffold_9	(TG)2(TA)1(TG)5(TA)5(TG)3(TA)17	22554859	22554925	(TG)4(TA)12
Ma-19	Davis and Strobeck, 1998	HiC_scaffold_14	(TA)1(AT)2(TT)1(AT)3(TA)6(TG)1(AT)1(GT)1(CA)1(TA)1(CA)14(TA)1	65484498	65484564	(TG)16
Ma-2	Davis and Strobeck, 1998	HiC_scaffold_9	(AC)1(AT)1(AC)17(AG)2	70041016	70041058	(TG)17
Ma-5	Davis and Strobeck, 1998	HiC_scaffold_16	(CA)1(TA)1(CA)1(TA)1(CA)5(GA)1(CA)9(GA)1	9759141	9759181	(TG)17
Ma-8	Davis and Strobeck, 1998	HiC_scaffold_9	(AA)2(CA)18(CG)1	42161287	42161329	(TG)21
Ma-9	Davis and Strobeck, 1998	HiC_scaffold_10	(TAAA)4(ATCT)1(A)17	92435521	92435558	(T)14(TG)4
Mar08	Natali et al, 2010	HiC_scaffold_17	(AC)13	50003347	50003373	(CA)22
Mar14	Natali et al, 2010	HiC_scaffold_10	(C)5(AC)1(CA)17(TA)1(CT)11(CC)2(CT)2	34621411	34621484	(TC)11AT(CA)15
Mar15	Natali et al, 2010	HiC_scaffold_3	(CT)1(TT)1(CT)13(CA)9(CC)1(CA)3	125075184	125075240	(CT)12(CA)11
Mar19	Natali et al, 2010	HiC_scaffold_7	(GT)14(GA)12(CA)1	104269245	104269299	(CT)15(CA)8
Mar21	Natali et al, 2010	HiC_scaffold_1	(TT)1(GT)11	125963317	125963341	(CA)24
Mar36	Natali et al, 2010	HiC_scaffold_3	(GC)1(GT)5(GC)1(GT)12(TG)1(GT)1	162533105	162533147	(CA)24
Mar43	Natali et al, 2010	HiC_scaffold_5	(GG)1(GT)18(AT)1(GT)1(AT)1(GT)3(TT)1	72528867	72528919	(CA)24
Mar56	Natali et al, 2010	HiC_scaffold_4	(CA)22(GA)1	26000630	26000676	(CA)21
Mar58	Natali et al, 2010	HiC_scaffold_18	(GG)2(GT)18(AT)1	4232518	4232560	(CA)12G(CA)7
Mar64	Natali et al, 2010	HiC_scaffold_4	(GT)23	98840520	98840566	(CA)21
Mel10	Domingo-Roura, 2002	HiC_scaffold_1	(CT)1(GT)13(GA)1(GT)11(TT)1	176999894	176999948	(TG)3(C)1(TG)3(C)1(T)3(C)1(TA)9(AC)14

Mer005	Fleming et al, 1999	HiC_scaffold_14	(CA)1(GT)1(CA)20	13562510	13562554	(CA)16
Mer009	Fleming et al, 1999	HiC_scaffold_7	(GT)21(GA)9(AA)1	20900210	20900272	(CA)18
Mf1.1	Basto et al, 2010	HiC_scaffold_6	(AAAT)1(GAAT)6	81538536	81538564	(ATTC)11
Mf1.11	Basto et al, 2010	HiC_scaffold_9	(TATC)6(TATT)1(TTTC)2(TATC)1(ATC)1(TATC)3(TATT)1	117922565	117922624	(TATC)8
Mf1.18	Basto et al, 2010	HiC_scaffold_18	(GCAT)1(AGAT)1(AAAT)1(AGAT)5(GAT)1(AGAT)1	35371034	35371073	(ATCT)8
Mf2.13	Basto et al, 2010	HiC_scaffold_2	(TATC)9(ATC)1(TATC)1(ACC)1(TATC)1(TAAC)1	71450511	71450565	(TATC)9
Mf3.2	Basto et al, 2010	HiC_scaffold_4	(TTTC)6(TCTC)1(TTTC)3(TTTT)1	74274785	74274829	(TTTC)12
Mf3.7	Basto et al, 2010	HiC_scaffold_12	(ATAG)4(ATGG)1(GTGG)1(ATAG)4(ATTG)1	14705291	14705335	(TAGA)4(TGGG)(TGGA)(TAGA)9
Mf8.7	Basto et al, 2010	HiC_scaffold_5	(AGAG)1(TAGA)10(CAGA)1	41353930	41353978	(TCTA)11
Mvi_1271	Vincent et al, 2003	HiC_scaffold_12	(AC)17	44349170	44349204	(CA)15
Mvi_1341	Vincent et al, 2003	HiC_scaffold_6	(CA)23	44981624	44981670	(CA)17
Mvi_1342	Vincent et al, 2003	HiC_scaffold_4	(GG)1(GC)1(GT)8(TG)1(GG)2(GT)1(GC)1(GG)2(GC)1(GG)1	6611614	6611652	(AC)14
Mvis002	Fleming et al, 1999	HiC_scaffold_3	(AT)4(GT)5(AT)1(GT)4(AT)1(CT)1(ATGT)3(ATCT)2	67784124	67784176	(CA)11
Mvis027	Fleming et al, 1999	HiC_scaffold_9	(AA)1(CA)10	63469073	63469095	(CA)10
Mvis072	Fleming et al, 1999	HiC_scaffold_8	(TA)1(CA)19	30259005	30259045	(CA)15
Mvis075	Fleming et al, 1999	HiC_scaffold_4	(GA)1(GT)4(TT)1(GC)1(GT)20	8763234	8763288	(CA)12

Приложение 34. Кластеризация с использованием 3 наборов STR-маркеров: А – полный набор, включающий все STR-локусы, В – набор из исследования Каштанова и др. 2022 и С – набор Рожнова и др. 2013 (Референс: *M. martes*).

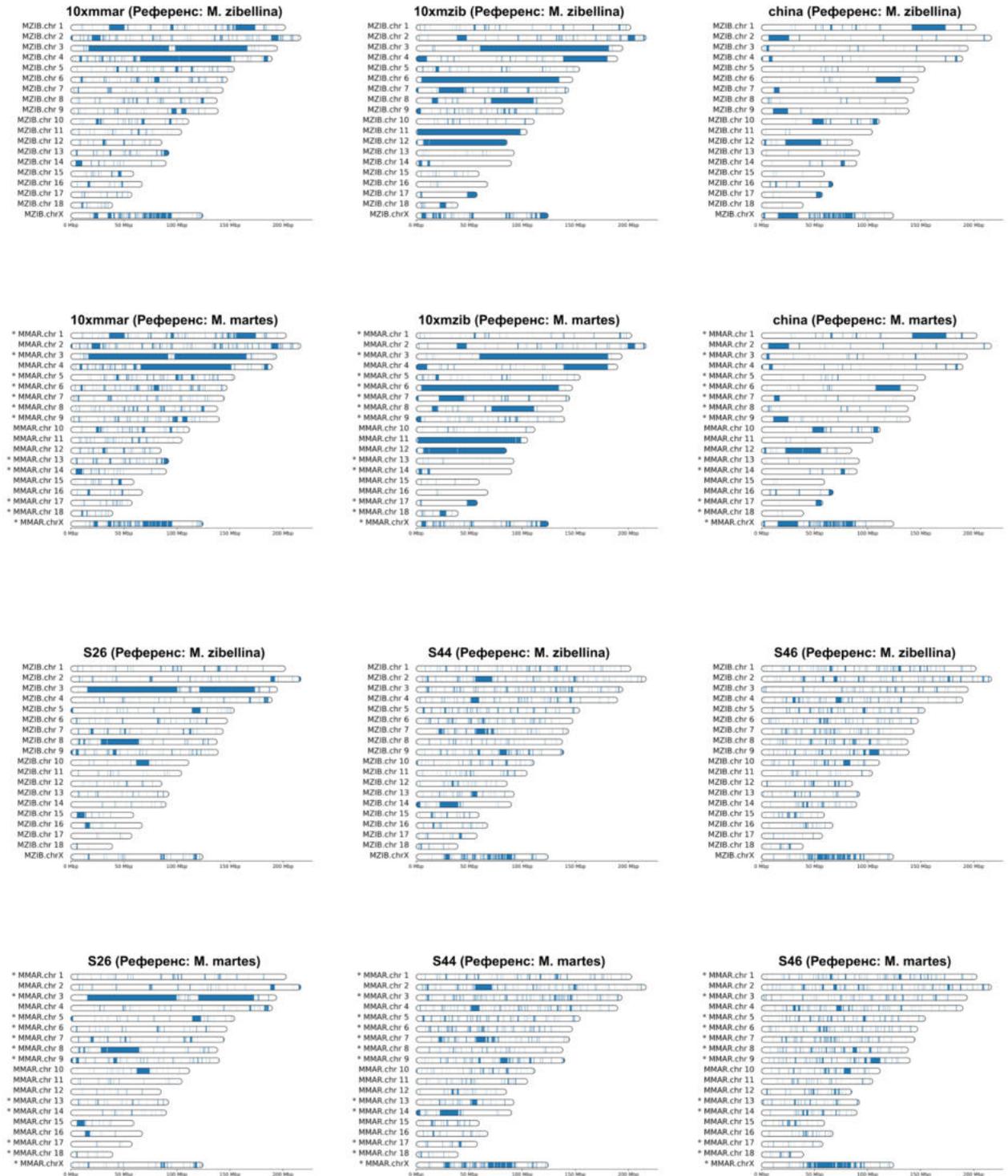


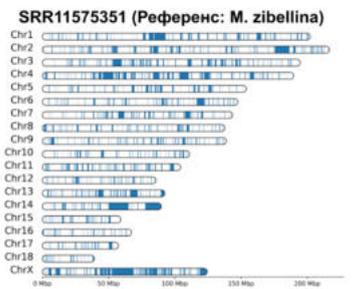
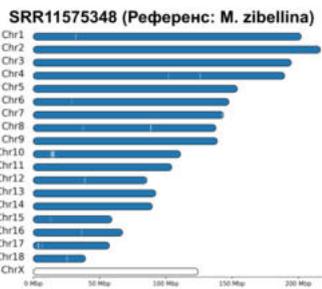
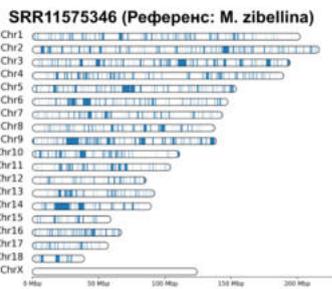
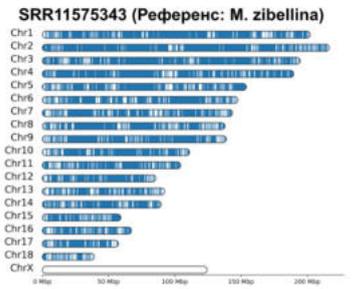
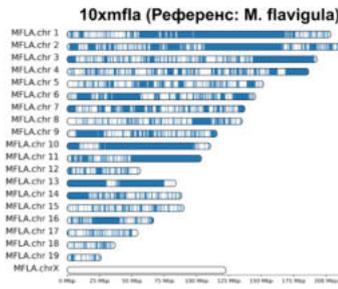
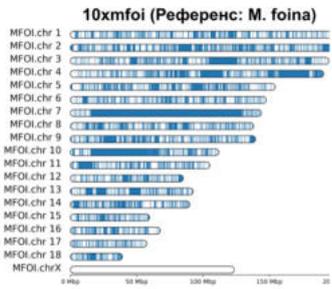
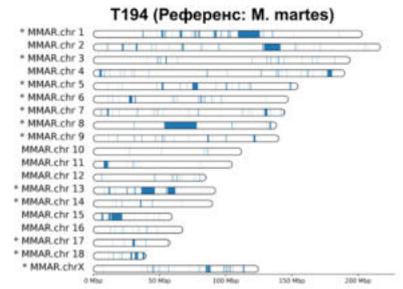
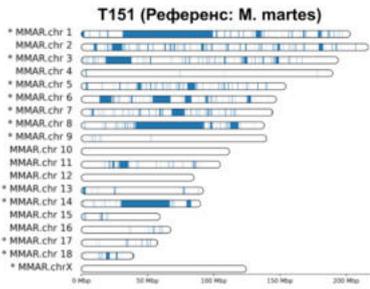
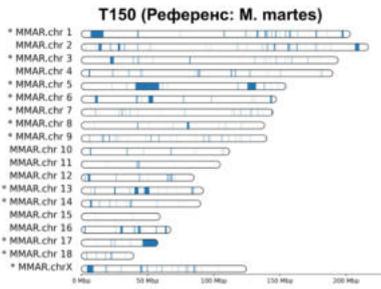
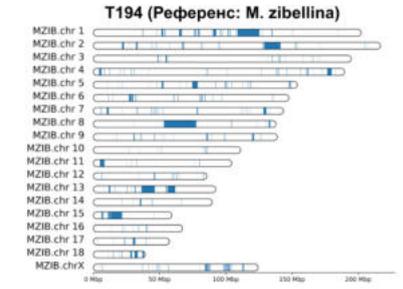
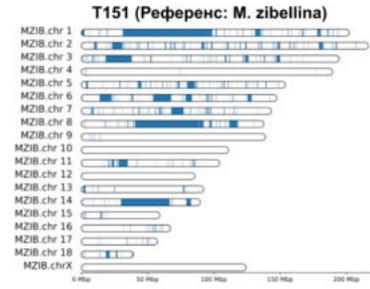
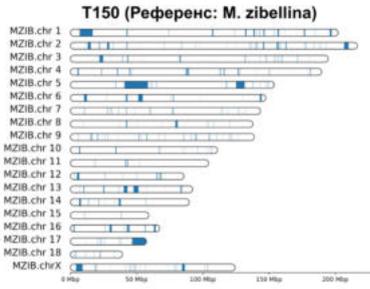
Приложение 35. Количество, суммарная длина (млн п.н.), доля генома (без учета хромосомы X) RoH относительно геномной сборки соболя и лесной куницы.

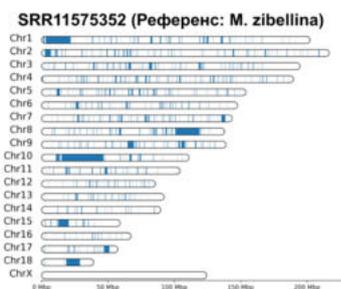
Образец	<i>M. zibellina</i>							<i>M. martes</i>						
	Кол-во RoH	Длина (млн п.н.)	% генома	He RoH, %	Короткие RoH <1 млн п.н., %	Длинные RoH >=1 млн п.н., %	Ультрадлинные RoH >=10 млн п.н., %	Кол-во RoH	Длина (млн п.н.)	% генома	He RoH, %	Короткие RoH <1 млн п.н., %	Длинные RoH >=1 млн п.н., %	Ультрадлинные RoH >=10 млн п.н., %
10xmmar	513	476.1	21.3	78.7	5	4.5	11.8	603	515.8	23.1	76.9	6.3	5	11.7
S44	625	264.7	11.9	88.1	6.5	3.9	1.5	668	279.1	12.5	87.5	6.8	4.2	1.5
S46	667	238.9	10.7	89.3	6.7	4	0	700	256.3	11.5	88.5	7.1	4.3	0
S49	600	307.3	13.8	86.2	6.1	5.6	2.1	634	324.3	14.5	85.5	6.6	5.8	2.1
T149	644	347.2	15.5	84.5	6.6	6.3	2.6	673	365.5	16.4	83.6	7.1	6.7	2.7
T24	598	385.8	17.3	82.7	5.8	6.7	4.7	641	403.8	18.1	81.9	6.5	6.8	4.7
T76	652	332.2	14.9	85.1	6.9	4.6	3.4	670	344.5	15.4	84.6	7.1	4.9	3.5
T77	632	298.9	13.4	86.6	6.8	5.7	0.8	651	317.3	14.2	85.8	7.3	5.6	1.3
T82	647	277.7	12.4	87.6	6.4	6	0	684	298.0	13.3	86.7	6.8	6.5	0
10xmzib	294	681.3	30.5	69.5	3.3	3.1	24.1	255	659.7	29.5	70.5	2.4	3.3	23.8
S26	345	322.7	14.5	85.5	3.5	3	8	329	316.1	14.2	85.8	3.3	2.9	8
T8	196	232.0	10.4	89.6	2.1	2.6	5.7	192	228.2	10.2	89.8	2	2.6	5.7
T26	140	115.5	5.2	94.8	1.5	1.3	2.4	142	114.4	5.1	94.9	1.5	1.2	2.4
T50	351	346.7	15.5	84.5	3.6	4.3	7.6	335	342.0	15.3	84.7	3.5	4	7.8
T72	769	392.1	17.6	82.4	9.3	6.1	2.2	776	382.5	17.1	82.9	9.1	5.8	2.2
T90	257	224.2	10.0	90	2.5	4.1	3.5	253	219.6	9.8	90.2	2.4	4.3	3.1

T104	162	114.7	5.1	94.9	1.8	2.2	1.2	169	113.1	5.1	94.9	1.8	2.5	0.7
T118	125	129.3	5.8	94.2	1.2	2.5	2.1	130	130.2	5.8	94.2	1.3	2.5	2.1
T148	267	296.7	13.3	86.7	2.5	5.3	5.5	266	294.2	13.2	86.8	2.6	5.1	5.5
T150	231	137.6	6.2	93.8	2.6	2.3	1.3	233	136.0	6.1	93.9	2.5	2.3	1.3
T194	227	169.1	7.6	92.4	2.2	3	2.4	247	168.9	7.6	92.4	2.2	2.5	2.8
china	187	219.0	9.8	90.2	2	1.9	6	187	220.7	9.9	90.1	2	1.9	6
T18	256	95.0	4.3	95.7	2.6	1.1	0.5	257	96.1	4.3	95.7	2.7	1.1	0.5
S50	436	160.7	7.2	92.8	4.3	2.9	0	452	171.5	7.7	92.3	4.6	3	0
T151	400	376.9	16.9	83.1	4	4.5	8.3	418	388.0	17.4	82.6	4.3	4.6	8.5
T78	524	207.8	9.3	90.7	5.3	3.4	0.5	565	221.0	9.9	90.1	5.8	3.6	0.5
T79	463	239.8	10.7	89.3	5	3.9	1.8	491	254.3	11.4	88.6	5.3	4.2	1.8
T81	629	255.1	11.4	88.6	6.7	3.7	1.1	678	273.6	12.3	87.7	7.2	3.9	1.1
T83	582	355.2	15.9	84.1	6.2	4.1	5.6	612	367.6	16.5	83.5	6.4	5	5.1
T84	155	59.5	2.7	97.3	1.7	1	0	161	58.2	2.6	97.4	1.6	1	0
T85	371	155.7	7.0	93	3.6	3.3	0	379	162.2	7.3	92.7	3.9	3.4	0
T86	366	145.0	6.5	93.5	3.6	2.9	0	396	155.2	7.0	93	4	3	0
T87	125	38.0	1.7	98.3	1.2	0.5	0	124	36.6	1.6	98.4	1.3	0.4	0

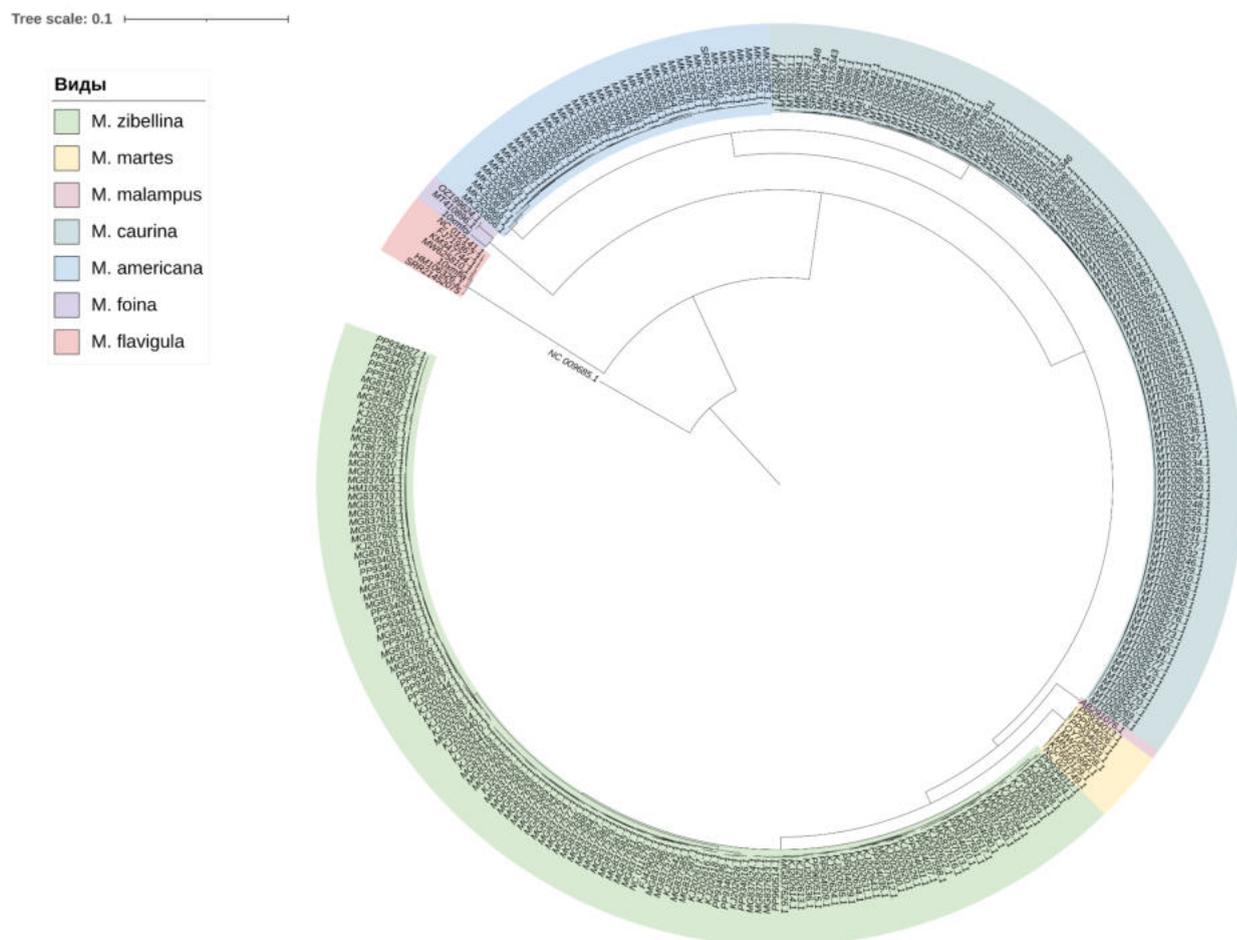
Приложение 36. Участки гомозиготности (RoHs) для исследуемых образцов. RoH выявлены в скользящих окнах размером 100 кbp с шагом 10 кbp. Инвертированные скаффолды в геномной сборке лесной куницы отмечены “*”.







Приложение 37. Филогенетическое дерево рода *Martes* на основании мтДНК. В качестве аутгруппы использован митохондриальный геном росомahi *Gulo gulo* (NC_009685.1).



Приложение 38. Корреляция между уровнем интрогрессии и высотой пика "популяционного взрыва" у гибридов. Частота замен на поколение (μ) = $4,64 \times 10^{-9}$.

