

*На правах рукописи*

МАНАХОВ АНДРЕЙ ДМИТРИЕВИЧ

**ГЕНОМНЫЙ АНАЛИЗ ПРЕДСТАВИТЕЛЕЙ СЕМЕЙСТВА КУНЬИХ  
(MUSTELIDAE)**

*1.5.7 – генетика*

АВТОРЕФЕРАТ

диссертации на соискание ученой степени  
кандидата биологических наук

Москва – 2022

Работа выполнена в лаборатории эволюционной геномики Федерального государственного бюджетного учреждения науки «Институт общей генетики им. Н.И. Вавилова Российской академии наук».

**Научный руководитель:** доктор биологических наук, профессор,  
член-корреспондент РАН  
**РОГАЕВ Евгений Иванович**  
заведующий лабораторией эволюционной геномики Федерального государственного бюджетного учреждения науки «Институт общей генетики им. Н.И. Вавилова Российской академии наук», г. Москва

**Официальные оппоненты:** доктор биологических наук, профессор  
**ТАРАБЫКИН Виктор Степанович**  
директор научно-исследовательского института нейронаук Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского» (ННГУ), г. Нижний Новгород

кандидат биологических наук  
**ФИШМАН Вениамин Семенович**  
ведущий научный сотрудник сектора геномных механизмов онтогенеза Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН), г. Новосибирск

**Ведущая организация:** Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет», г. Санкт-Петербург

Защита состоится «\_\_» \_\_\_\_\_ 2022 г. в \_\_\_\_ часов на заседании диссертационного совета 24.1.088.01 в Федеральном государственном бюджетном учреждении науки «Институт общей генетики им. Н.И. Вавилова Российской академии наук» по адресу: 119991, Москва, ул. Губкина, 3.

С диссертацией и авторефератом можно ознакомиться в библиотеке и на сайте Института [www.vigg.ru](http://www.vigg.ru), тел. 8-499-135-14-31.

**Автореферат разослан «\_\_» \_\_\_\_\_ 2022 года.**

Ученый секретарь диссертационного совета,  
доктор биологических наук

Горячева И. И.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Актуальность и разработанность темы исследования

Соболь (*Martes zibellina*) и лесная куница (*Martes martes*) являются типичными представителями рода куниц (*Martes*) [Гептнер и др., 1967], одного из наиболее многочисленных по числу видов среди семейства куньих (Mustelidae) [Hosoda и др., 2000]. Представители этих видов населяют в основном территорию России, куница обитает на европейской части Евразии [Herrero и др., 2016], а соболь – на территории от Уральских гор до Тихоокеанского побережья [Monakhov, 2016]. При этом на занимаемых территориях, особенно для соболя, имеет место значительная межпопуляционная дифференциация по фенотипическим признакам [Каштанов и др., 2016]. Несмотря на многолетнее изучение, систематический статус соболей из различных популяций на сегодняшний день остается не до конца выясненным.

На территории России описаны по меньшей мере две зоны симпатрии соболя и куницы, в пределах которых имеет место межвидовая гибридизация и образуются гибриды – кидусы [Кассал, Сидиров, 2013; Рожнов и др., 2010; Monakhov, 2011]. Проведенные ранее единичные эксперименты по получению межвидовых гибридов соболя и куницы в неволе показали, что возникновение таких гибридов возможно. Продемонстрирована возможность успешных возвратных скрещиваний, по меньшей мере с самцами куницы [Портнова, 1941]. При этом на территориях зон симпатрии встречаются особи, демонстрирующие всю степень переходов признака от куницы к соболу, а также особи, гибридного происхождения которых может быть установлено лишь путем анализа строения скелета и черепа [Гептнер и др., 1967]. Это позволяет предположить, что популяция кидусов состоит не только из гибридов первого поколения, но и включает в себя гибриды второго и последующих поколений, а также потомков от возвратных скрещиваний с родительскими формами [Пономарев, 1946]. Проведенные на сегодняшний день молекулярно-генетические исследования кидусов были ограничены анализом фрагмента контрольного региона мтДНК и нескольких микросателлитных локусов. Эти исследования позволили подтвердить факт гибридизации, имевшей место в прошлом, однако так и не смогли однозначно ответить на вопрос о возможной плодовитости гибридов в естественных условиях [Рожнов и др., 2010; Рожнов и др., 2013]. Кроме того, данные исследования в основном были ограничены особями из северной, более древней, зоны симпатрии. Южная зона симпатрии, расположенная на территории Омской области и существующая лишь несколько десятков лет, во многом остается недостаточно исследованной.

Как и для многих представителей семейства куньих, для соболя и лесной куницы характерен период эмбриональной диапаузы - обратимой остановки эмбрионального развития, которая происходит на стадии бластоцисты [Исакова, 2012; Fenelon, Renfree, 2018]. При этом период эмбрионального покоя у соболя и куницы является одним из наиболее продолжительных среди куньих и составляет около 245 дней [Гептнер и др., 1967]. Этот феномен с молекулярно-генетической точки зрения также остается малоизученным на сегодняшний день, и получение геномных данных может открыть новые перспективы для исследований этого явления.

Соболь и, в меньшей степени, куница являются ценными видами пушных зверей. В настоящее время на фермерских хозяйствах наблюдается «цветовой взрыв» - возникают и фиксируются новые цветовые формы, отсутствующие в природных популяциях [С.Н. Каштанов, личное сообщение]. При этом многие из окрасок демонстрируют, в соответствии с законом гомологических рядов Н.И. Вавилова, фенотипический параллелизм с ранее полученными формами окраски меха у американской норки (*Neovision vision*) [Трапезов и др., 2020]. Американская норка – образец уникальной коллекции окрасочных форм, полученных в результате искусственного отбора, описано 35 мутаций, затрагивающих окраску меха у представителей данного вида [Трапезов, Трапезова, 2009], при этом большая часть из них до сих пор не охарактеризована с молекулярно-генетической точки зрения.

Таким образом, изучение геномики представителей семейства куньих в целом, соболя и лесной куницы в частности, является актуальной и перспективной задачей с фундаментальной точки зрения, так как позволяет подробно изучить **уникальное явление межвидовой гибридизации** соболя и куницы, открывает перспективы для изучения **явления сезонной эмбриональной диапаузы** у куньих, а также является первым шагом для изучения популяционного разнообразия, **механизмов адаптации к различным экологическим условиям** этих экологически важных видов. Кроме того, изучение геномов соболя, лесной куницы и, как модельного объекта, американской норки, может быть полезным для выявления **механизмов наследования экономически ценных характеристик меха**, например окрасок, и найти применение в программах по разведению этих видов в пушном звероводстве, а также способствовать проведению **мероприятий по сохранению численности и генетического разнообразия** природных популяций соболя и куницы.

## **Цель исследования**

Осуществить сборку и аннотацию геномов представителей семейства куньих: соболя и лесной куницы, провести геномный анализ соболей, лесных куниц и их предполагаемых гибридов кидусов, определить генетические основы различных форм окраски меха у соболя и американской норки.

## **Задачи исследования**

1. Провести секвенирование геномов исследуемых представителей семейства куньих из разных географических популяций, а также характеризующихся различными типами окраски меха.
2. Провести *de novo* сборку геномов и аннотацию в них генов для соболя и лесной куницы. Провести геномный анализ соболей, лесных куниц и кидусов из различных популяций.
3. Охарактеризовать с геномной точки зрения особей, определенных по фенотипическим признакам как кидусы.
4. Разработать методологический подход анализа данных геномного секвенирования для выявления генетических факторов, обуславливающих формирование различной окраски меха у американской норки.
5. Применить разработанный подход для идентификации генетических факторов, обуславливающих формирование окраски меха пастель у соболя.

## **Научная новизна и практическая значимость исследования**

В рамках диссертационной работы впервые осуществлена *de novo* сборка геномов и аннотация в них генов для самцов соболя и лесной куницы. Полученные сборки геномов имеют практическую значимость, так как обеспечивают формирование референсных последовательностей для секвенирования и анализа индивидуальных геномов соболя и куницы.

Впервые получены и проанализированы данные геномного секвенирования соболей и куниц из разных популяций. В результате сравнительного геномного анализа выявлены четыре группы соболей, населяющих территорию России: “Итурупская”, “Дальневосточная”, “Уральская” и “Центрально-Западно Сибирская”, являющаяся результатом смешения “Дальневосточной” и “Уральской” групп. Полученные данные могут способствовать проведению мероприятий по сохранению численности и генетического разнообразия природных популяций соболя и куницы, также на их основе может быть разработана тест-система

для идентификации популяционной принадлежности соболя, что позволит эффективнее противодействовать незаконной добыче пушнины.

Впервые получены и проанализированы данные геномного секвенирования кидусов из обеих зон симпатрии на территории России. Впервые однозначно подтвержден факт происходящей в настоящее время гибридизации соболя и лесной куницы в природных условиях, а также выявлены признаки, подтверждающие возможность таких гибридов давать плодовитое потомство.

Разработан и успешно апробирован методологический подход анализа данных геномного секвенирования, позволивший впервые идентифицировать шесть мутаций, обуславливающих развитие экономически ценных форм окраски меха у американской норки. Выявлена мутация, обуславливающая формирование окраски меха пастель у соболя – первой зарегистрированной для соболя мутантной окраски мехового покрова. На основании выявленных мутаций могут быть разработаны тест-системы, позволяющие идентифицировать носителей данных мутаций. Несомненно, что подобные тест-системы найдут применение в пушном звероводстве.

#### **Положения, выносимые на защиту**

1) Лесные куницы, населяющие северную часть европейской территории России, представляют собой группу генетически близких особей. Соболя, населяющие территорию России, могут быть подразделены на три группы: “Камчатская”, “Уральская” и “Итурупская”. Популяции соболя, занимающие территории Западной и Центральной Сибири, сформированы в результате расселения соболя из рефугиумов на Северном Урале и Дальнем Востоке.

2) В зонах симпатрии лесной куниц и соболя имеет место постоянная межвидовая гибридизация, в результате которой возникают гибриды - кидусы, способные давать в природных условиях плодовитое потомство.

3) Экономически ценные серебристо-голубая, белая Хедлюнд, мойл, черный хрусталь и Шедоу окраски меха американской норки обусловлены мутациями в генах *MLPH*, *MITF*, *RAV38*, *COPA* и *KIT* соответственно.

4) Первая зарегистрированная для соболя мутантная окраска мехового покрова – пастель обусловлена мутацией в гене *TYRP1*.

## **Личный вклад автора**

Все основные результаты получены лично автором, либо при его участии в планировании и проведении экспериментов, анализе данных. Используемые в ходе исследования образцы биологического материала получены автором из коллекций к.б.н. С.Н. Каштанова (ИОГен РАН) и д.б.н. О.В. Трапезова (ИЦИГ СО РАН), часть образцов была собрана автором совместно с ними, часть - самостоятельно в зверохозяйствах. Аннотация белок-кодирующих генов в геномах соболя и лесной куницы проводилась автором совместно с к.б.н. В.В. Соловьевым (Softberry Inc).

## **Степень достоверности и апробация результатов**

Работа выполнена с использованием современных методов экспериментальной биологии, а также математических и статистических критериев для обработки данных. Доклады по теме диссертации проводились на ежегодных отчетах аспирантов ФГБУН ИОГен РАН в 2018-2021 гг. Автором опубликовано 3 статьи по теме диссертации в научных рецензируемых изданиях, индексируемых в базах данных Scopus и Web of Science, отвечающих требованиям Высшей Аттестационной Комиссии Министерства образования и науки Российской Федерации, получен 1 патент. Промежуточные и итоговые результаты диссертационной работы были представлены на российских и международных конференциях, а также на научных школах: форум QIAGEN Day (Россия, Москва, 4 апреля 2019), Российско-немецкий симпозиум «Актуальные проблемы современной биомедицины» (Россия, Сочи, 1-4 декабря 2019), международная научная школа-семинар «Genes. Brain. Behavior» (Россия, Москва, 9 декабря 2019), XII международный конгрессе «IFASA Congress» (Польша, Варшава 24-25 августа 2021), вторая студенческая научно-образовательная школа-конференция по генетике и биотехнологиям (Россия, Сочи, 13–19 декабря 2021). Изложенные в диссертационном исследовании положения и выводы являются достоверными. Апробация диссертационной работы проведена на межлабораторном семинаре ИОГен РАН (протокол №1 от 26 октября 2021 г.).

## **Структура и объем диссертационной работы**

Диссертация включает в себя следующие главы: «Введение», «Глава 1. *De novo* сборка геномов соболя и куницы», «Глава 2. Геномный анализ соболя, куницы и их гибридов», «Глава 3. Молекулярно-генетические основы многообразия форм окраски меха у соболя и американской норки», «Выводы» и «Список литературы». Работа изложена на 148 страницах машинописного текста, включая 24 страницы Приложений, содержит 23 рисунка

и 19 таблиц. Список цитируемых литературных источников включает 211 наименований, из которых 188 – на английском языке.

## СОДЕРЖАНИЕ РАБОТЫ

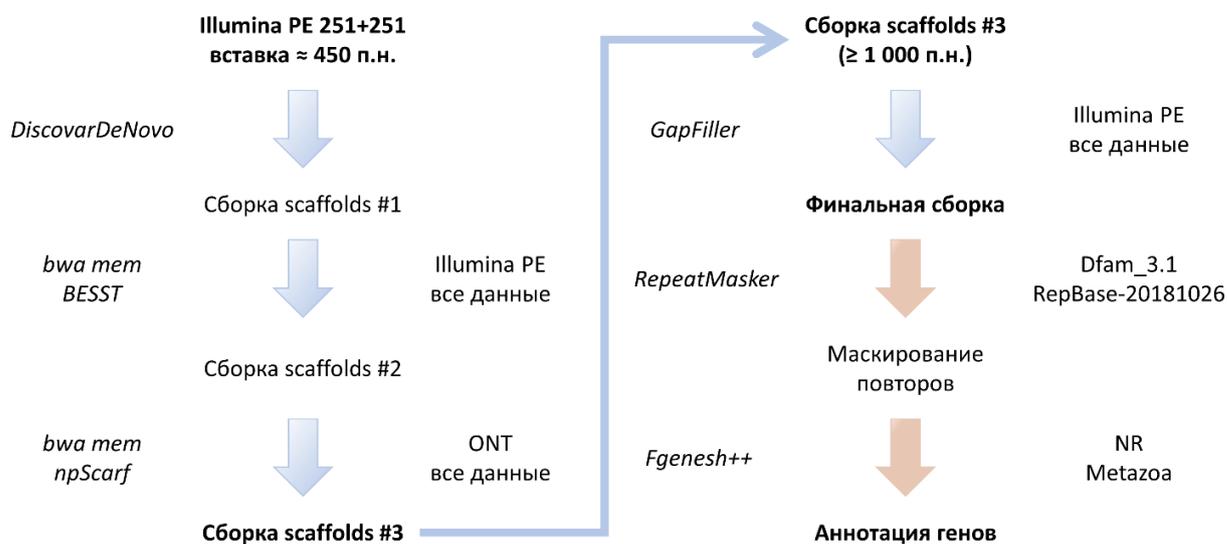
### Глава 1. *De novo* сборка геномов соболя и куницы

В диссертационной работе глава «*De novo* сборка геномов соболя и куницы» состоит из трех разделов. Раздел **Обзор литературы** посвящен биологической характеристике соболя и лесной куницы, а также описывает современное состояние геномики представителей семейства куньих.

**Материалы и методы.** Для реконструкции геномов соболя и лесной куницы использовали данные секвенирования на двух платформах – Illumina и Oxford Nanopore Technologies (ONT). Для соболя были получены данные секвенирования на платформе Illumina HiSeq 2000/2500 трех библиотек с размерами вставки  $\approx 90$ , 400 и 450 п.н., а также двух библиотек со «спаренными концами» (mate pair) из фрагментов ДНК длиной  $\approx 1,5$  и 6 тыс. п.н. Для куницы были получены данные секвенирования на платформе Illumina двух библиотек с размерами вставки  $\approx 280$  и 450 п.н. Для обоих видов были получены данные нанопорового секвенирования на приборе ONT MinION. Кроме того, для соболя получены данные секвенирования на платформе Illumina трех транскриптомных библиотек, приготовленных из РНК, выделенной из мозга, сердца и печени, на основании которых с помощью программы Trinity v2.8.4 [Grabherr и др., 2011] была осуществлена *de novo* сборка транскриптома [Adam, Nathan, 2020].

*De novo* сборку генома соболя и куницы проводили в несколько этапов (**Рисунок 1**). С помощью программы DISCOVAR *de novo* и данных секвенирования библиотек с размером вставки  $\approx 450$  п.н. получена первичная сборка скаффолдов. Затем относительно нее с помощью программы BWA [Li, Durbin, 2009] картировали все имеющиеся данные секвенирования на платформе Illumina, предварительно обработанные с помощью программы AdapterRemoval v2 [Schubert, Lindgreen, Orlando, 2016]. На основании полученных выравниваний с помощью программы BESST v2.2.8 [Sahlin и др., 2014] осуществляли сборку вторичных скаффолдов. Относительно вторичной сборки картировали данные нанопорового секвенирования и с помощью программы nrScarf проводили финальное объединение скаффолдов [Cao и др., 2017]. На заключительном этапе из полученных сборок удаляли фрагменты короче 1 000 п.н. и заполняли неопределенные последовательности (gaps) с помощью программы GapFiller v1.10 [Boetzer, Pirovano, 2012] и данных секвенирования на

платформе Illumina. Оценку полноты сборок проводили с помощью программы BUSCO v3.0.1 [Simão и др., 2015]. Идентификацию и маскирование повторов проводили с помощью программы Repeat-Masker v4.1.0.



**Рисунок 1.** Стратегия *de novo* сборки геномов и аннотации генов соболя и куницы.

Для аннотации генов использовали программный конвейер Fgenesh++ [Solovyev и др., 2006]. Аннотацию проводили, используя информацию о гомологии белков для предсказания моделей генов, а также *ab initio*. Функциональную аннотацию предсказанных генов проводили путем поиска гомологичных белковых последовательностей в трех общедоступных базах данных: NR, Swiss-Prot [UniProt Consortium, 2018] и KEGG [Kanehisa и др., 2016].

**Результаты и обсуждения.** Впервые осуществлено секвенирование с глубоким покрытием генома самца соболя (**90,58x** покрытие данными секвенирования на платформе Illumina и **17,47x** покрытие данными ONT) и самца лесной куницы (**67,43x** покрытие данными секвенирования на платформе Illumina и **11,70x** покрытие данными ONT). На основании полученных данных геномного секвенирования осуществлена реконструкция полных геномов. Для соболя получена сборка генома (MarZib1) общей протяженностью 2,401 млрд. п.н., состоящая из 15 641 скаффолда. Для лесной куницы получена сборка генома (MarMaz1) общей протяженностью 2,395 млрд. п.н., состоящая из 20 239 скаффолдов (**Таблица 1**).

Полученные геномные сборки по протяженности, а также среднему GC-составу, оказались близки к ранее полученным геномам куных, а их полнота, оцененная путем

подсчета числа универсальных однокопийных генов млекопитающих в программе BUSCO, составила 96% для обоих видов.

**Таблица 1.** Описание полученных геномных сборок соболя и лесной куницы.

Показатель	MarZib1	MarMar1
Общая протяженность сборки (млрд. п.н.)	2,401	2,395
Число скаффолдов (шт.)	15 641	20 239
Число контигов (шт.)	16 848	21 496
N (%)	0,005	0,004
N50 скаффолдов (млн. п.н.)	4,315	1,090
L50 скаффолдов (шт.)	169	640
Максимальная длина скаффолда (млн. п.н.)	27,481	6,739
GC (%)	41,73	41,74

Анализ доли повторяющихся элементов в геномах соболя и лесной куницы, проведенный с помощью программы Repeat-Masker, показал, что она составляет 41% как для соболя, так и для лесной куницы, что несколько выше, чем аналогичные показатели, рассчитанные нами для опубликованных геномов хорька (*Mustela putorius furo*) – 37% и американской норки – 36%.

В результате проведенной аннотации белок-кодирующих генов в геномах соболя и куницы предсказано 19 427 и 19 388 генов, продемонстрировавших гомологию с представленными в базах данных NR (Metazoa), Swiss-Prot и KEGG.

На основании белок-кодирующих последовательностей 6 070 однокопийных генов (один к одному) у соболя, лесной куницы и девяти других млекопитающих с использованием GTR-GAMMA модели рассчитывали время дивергенции общего предка соболя и куницы с общим предком калана и хорька/норки, которое составило 17,66 млн. лет, что несколько больше, чем определялось ранее при анализе отдельных генов [Yu и др., 2011]. В целом полученные нами времена дивергенции согласуются с данными, представленными в базе данных TimeTree [Kumar и др., 2017], хотя время дивергенции соболя и лесной куницы, по нашей оценке, получилось несколько большим – 4,44 млн. лет (95% доверительный интервал 3,00-6,41 млн. лет).

## Глава 2. Геномный анализ соболя, куницы и их гибридов

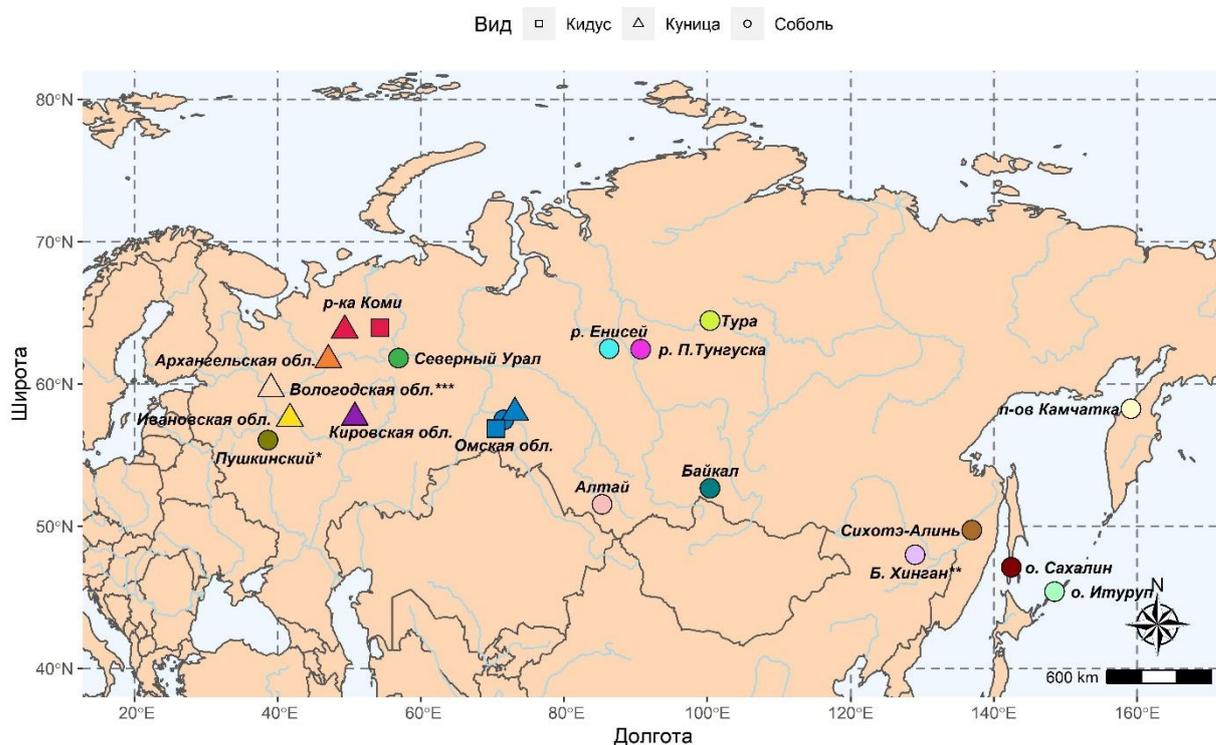
В диссертационной работе глава «Геномный анализ соболя, куницы и их гибридов» состоит из трех разделов. Раздел **Обзор литературы** посвящен описанию ареалов обитания исследуемых видов, а также проблемам систематики подвидов соболя по имеющимся на

сегодняшний день ограниченным молекулярно-генетическим данным (отдельные участки мтДНК, несколько десятков полных митохондриальных геномов и отдельные ядерные микросателлитные локусы). Начиная с XVIII–XIX вв. в литературе встречаются упоминания о *кидусах* – гибридах соболя и куницы, которые возникают в зонах симпатрии этих видов [Павлинин, 1963]. В XX в. в результате единичных экспериментов кидусы были получены и в неволе, в одном случае была продемонстрирована возможность получить потомство от кидусов при возвратном скрещивании с самцом куницы [Граков, 1976; Портнова, 1941]. Проведенные на сегодняшний день молекулярно-генетические исследования кидусов из природных популяций были ограничены исследованием отдельных участков мтДНК и нескольких микросателлитных локусов. Кроме того, преимущественно исследовались особи, происходящие из северной, более древней, зоны симпатрии. Эти работы позволили подтвердить факт гибридизации между соболем и лесной куницей, а также выявили среди исследуемых особей таких, которые не могут быть однозначно отнесены к одному из видов на основании анализа микросателлитных локусов [Пищулина, 2013; Рожнов и др., 2010; Рожнов и др., 2013]. Тем не менее они не смогли однозначно ответить на вопрос о возможной плодовитости гибридов в естественных условиях.

**Материалы и методы.** Для исследования использовали коллекцию биологического материала соболей и куниц, как из аллопатрических, так и симпатрических популяций, а также кидусов из обеих зон симпатрии (северной – Республика Коми и Северный Урал и южной – Омская обл.), предоставленные к.б.н. С.Н. Каштановым (ИОГен РАН). В общей сложности на платформе Illumina осуществлено секвенирование полных геномов 42 соболей, 16 куниц и 10 кидусов из 15 природных популяций России и 1 фермерской популяции (**Рисунок 2**). Нашей целью было получать в среднем  $\times 10$  покрытие для геномов особей из симпатрических и  $\times 5$  для особей из аллопатрических популяций. Кроме того, в работе были использованы археологические образцы (800–1 000 лет) из средневекового памятника Минино I (Вологодская обл.), предоставленные д.б.н. А.Б. Савинецким (ИПЭЭ РАН).

Данные глубокого секвенирования картировали относительно реконструированных нами геномов соболя (MarZib1) и лесной куницы (MarMar1) с помощью функции *bwa mem* из пакета программ BWA [Li, Durbin, 2009] и обрабатывали с помощью функции *MarkDuplicates* из пакета программ picard-tools v2.22.2 [Picard toolkit, 2019]. В картированных относительно референсных геномов данных глубокого секвенирования вели поиск генетических вариантов (SNPs, InDels, MNPs) с помощью функций *HaplotypeCaller*, *CombineGVCFs* и *GenotypeGVCFs* из пакета программ GATK v4 [McKenna и др., 2010]. Данные секвенирования археологических образцов лесной куницы обрабатывали программой

AdapterRemoval v2 и картировали относительно мтДНК (KC660129.1) [Li, Wu, Malyarchuk, 2014] и реконструированного нами генома лесной куницы (MarMar1) с помощью функций *bwa aln* и *samse* с параметрами, адаптированными для древней ДНК [Schubert и др., 2012].



**Рисунок 2.** Географическая карта районов сбора образцов. \* - фермерская популяция соболя, \*\* - образец соболя из статьи [Liu и др., 2020], \*\*\* - археологические образцы лесной куницы.

Для анализа использовали весь спектр доступных из данных геномного секвенирования маркеров, как аутосомных, так и маркеров мтДНК и Y-хромосомы. Полные последовательности мтДНК реконструировали из данных глубокого геномного секвенирования с использованием программы MitoZ v2.4 [Meng и др., 2019]. На основе выявленных 48 гаплотипов мтДНК (длиной 16 276–16 278 п.н., после удаления 5′–AC–(AC)<sub>n</sub>–AC–3′ повторов из контрольного региона) с помощью программы BEAST v1.10.4 [Suchard и др., 2018] строили филогенетическое дерево, используя последовательность мтДНК американкой норки в качестве внешней группы. Реконструкцию последовательностей мтДНК археологических образцов лесной куницы, ввиду небольшой глубины, проводили вручную с помощью визуализации результатов картирования данных секвенирования в программе IGV v2.4.8 [Robinson и др., 2018].

Для идентификации скаффолдов, относящихся к Y-хромосоме, использовали два подхода. Первый был основан на поиске геномной синтении с хромосомами собаки, в

качестве Y-хромосомы использовали ее ранее реконструированный, специфичный для самцов регион (MSY) [Li и др., 2013]). Выравнивание геномов соболя и собаки проводили с помощью протокола LASTZ-ChainNet [Harris, 2007; Kent и др., 2003], реализованного в программе CNEr [Tan, Polychronopoulos, Lenhard, 2019]. Второй подход был основан на анализе глубины покрытия скаффолдов у особей известного пола. С помощью функции DepthOfCoverage из пакета программ GATK v3.8 рассчитали среднюю глубину покрытия каждого скаффолда. Скаффолды, средняя глубина покрытия которых составляла менее 0,2 от средней глубины покрытия всего генома для самок и от 0,35 до 0,65 для самцов, считали потенциально относящимися к Y-хромосоме. Скаффолды, отобранные обоими методами, верифицировали методом ПЦР-амплификации с использованием ДНК по крайней мере 1 самца и 1 самки каждого вида.

Для анализа генетической структуры популяции соболей, куниц и их гибридов - кидусов на основании данных геномного секвенирования использовали несколько подходов. Анализ методом главных компонент (PCA) проводили, используя программу PLINK v1.9 [Picard toolkit, 2019]. Филогенетическое дерево на основе однонуклеотидных полиморфизмов (ОНП) по всему геному строили с помощью метода максимального правдоподобия в программе RAxML v8.2.12, в качестве внешней группы использовали американскую норку. Генетическую структуру популяции реконструировали с использованием программы ADMIXTURE v1.3 [Alexander, Novembre, Lange, 2009]. Предварительно для анализа с помощью программы PLINK отобрали несцепленные ОНП. Анализ ADMIXTURE проводили несколько раз, изменяя число предполагаемых генетических кластеров, задавая значения параметра  $K$  от 2 до 10. Для каждого значения  $K$  проводили 3 независимых расчета. Оптимальное значение параметра  $K$  было выбрано путем сравнения выдаваемого программой *cv-error* значения, а также с помощью  $\Delta K$  метода [Evanno, Regnaut, Goudet, 2005].

Помимо этого, использовали собственный алгоритм, основанный на выявлении специфичных для куницы полиморфизмов в данных полногеномного секвенирования. Аллель признавали специфичным для куницы, если все куницы из аллопатрических популяций являются гомозиготами по одному аллелю, и данный аллель не встречается, ни в гомо-, ни в гетерозиготном состоянии, в группе соболей из аллопатрических и фермерской популяций. Для каждого животного индивидуально рассчитывали долю специфичных для куницы аллелей по всему его геному. Дополнительно для скаффолдов длиннее 10 млн. п.н. (23 шт., суммарная длина 309 млн. п.н., т.е. 12,89% генома) выполняли аналогичный анализ, рассчитывая долю специфичных для куницы аллелей в скользящем окне длиной 100 тыс. п.н. с шагом 50 тыс. п.н.

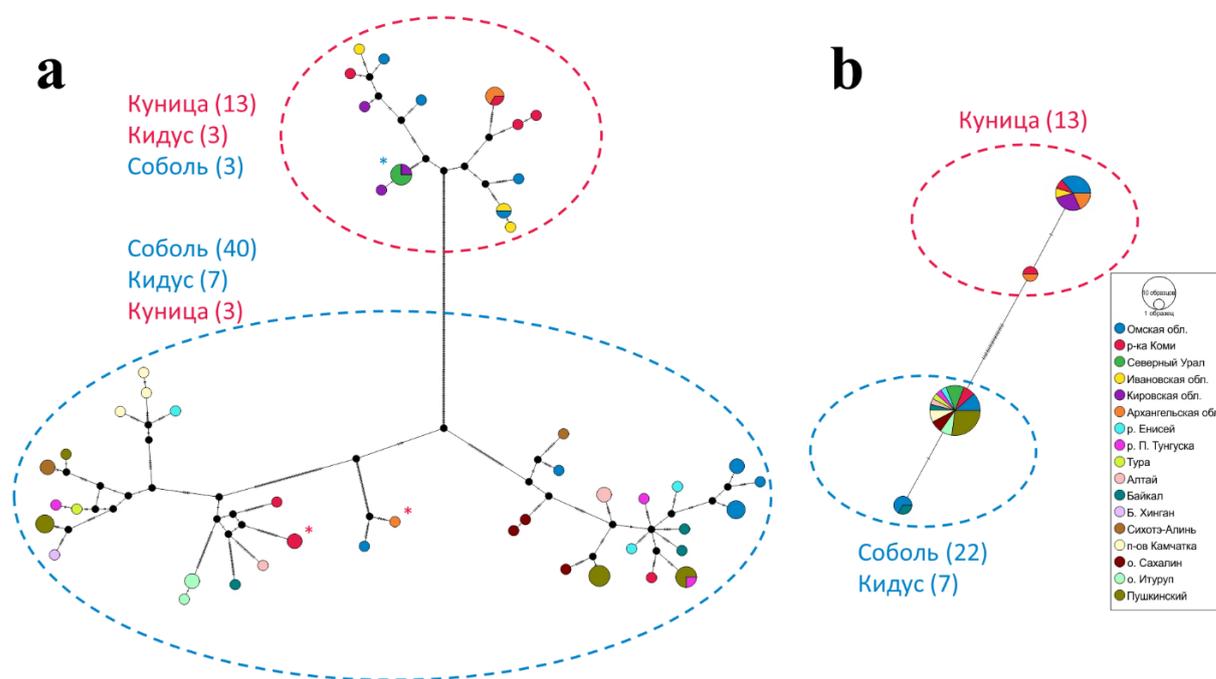
**Результаты и обсуждения.** В ходе проведенного исследования впервые было осуществлено секвенирование геномов 31 соболя и 16 лесных куниц из природных популяций, 11 фермерских соболей, а также 10 особей, фенотипически идентифицированных как кидусы, из двух районов совместного обитания соболя и лесной куницы. В общей сложности получены геномы для соболей, лесных куниц и кидусов из 15 природных и 1 фермерской популяции (**Рисунок 2**). Также в ходе анализа использовали ранее полученные данные геномного секвенирования соболя из популяции Большого Хингана (Китай) [Liu и др., 2020].

*Анализ гаплотипов митохондриальной ДНК.* Из данных глубоко геномного секвенирования реконструировали 69 последовательностей мтДНК. Всего было выявлено 48 уникальных гаплотипов мтДНК длиной 16 276-16 278 п.н. Анализ отношений между которыми показал, что они образуют 2 специфичных кластера с высоким уровнем поддержки для митотипов «*M. zibellina*» и «*M. martes*» (**Рисунок 3**). В большинстве случаев наблюдается соответствие между видовой принадлежностью, определенной по фенотипическим признакам и типом мтДНК. Однако выявлено 3 соболя из уральской популяции, вошедших в митогруппу «*M. martes*», и 3 куницы из популяций Коми и Архангельска, вошедших в митогруппу «*M. zibellina*». Среди кидусов 7 особей вошли в митогруппу «*M. zibellina*» и 3 - в митогруппу «*M. martes*». При этом все 5 кидусов, происходящих из Омской симпатрической популяции, имели митотип «*M. zibellina*».

Проведенный анализ гаплотипом мтДНК двух археологических образцов лесной куницы гаплотипов показал, что они относятся кластеру «*M. martes*», и при этом демонстрируют наибольшее сходство с гаплотипом образца *marten\_1732*, происходящим с территории Омской области.

*Анализ вариантов Y-хромосомы.* Впервые выявлены и подтверждены экспериментально геномные регионы, относящиеся к MSY району Y-хромосомы соболя (3 скаффолда, суммарной длиной 483 тыс. п.н.). Для самцов соболя (22 шт.), куниц (13 шт.) и кидусов (7 шт.) реконструированы 4 уникальных гаплотипа, длиной 23 п.н. Анализ отношений между ними показал, что они образуют 2 специфичных кластера (**Рисунок 3**). Два гаплотипа оказались специфичными для «*M. zibellina*» и два – для «*M. martes*». На анализируемой выборке не выявлено ни одного случая несовпадения между видовой принадлежностью, определенной по фенотипическим признакам и типам Y-хромосомы. При этом все самцы кидусы имеют гаплотип Y-хромосомы типа «*M. zibellina*». Интересно, что в ранее проведенных экспериментах по получению кидусов в неволе успешными оказались лишь скрещивания самок соболя, а также самок кидуса (F<sub>1</sub>) с самцами лесной куницы [Граков, 1976;

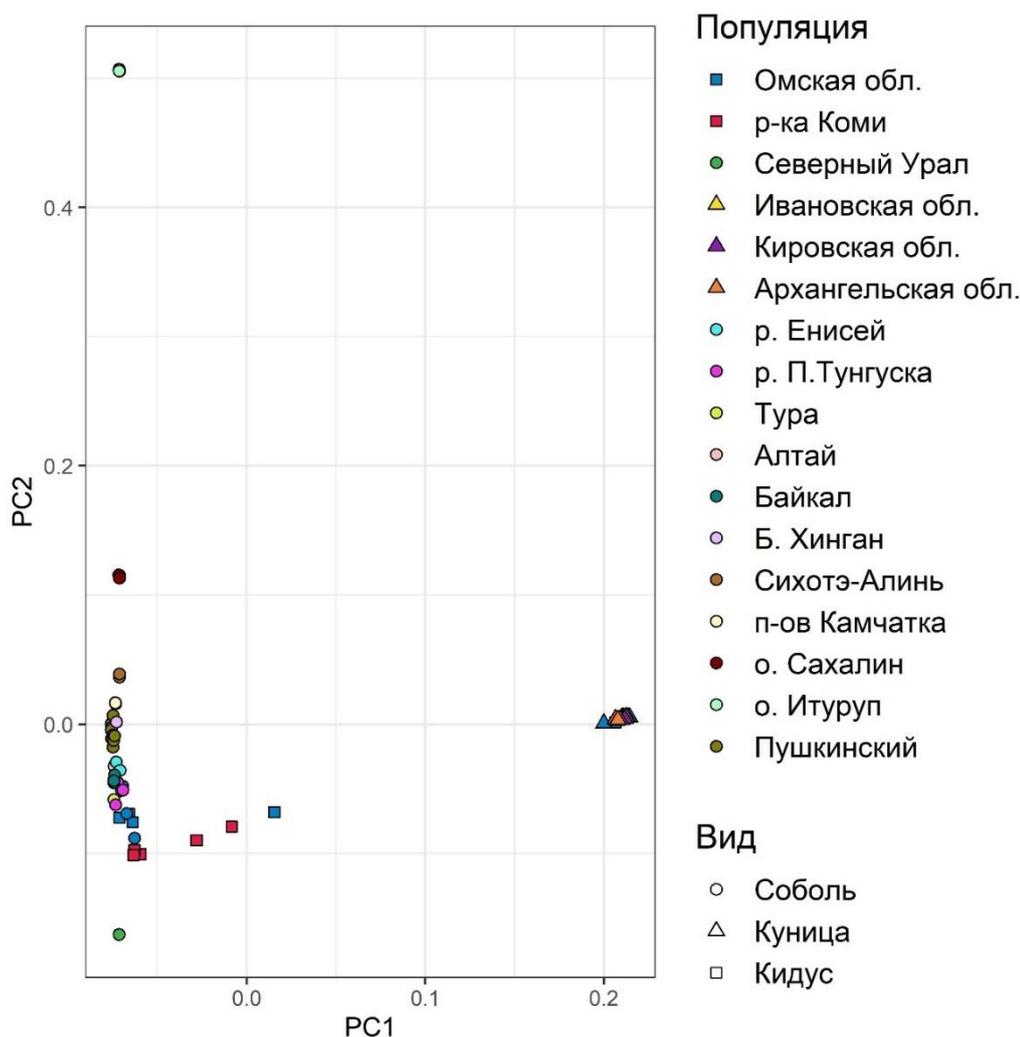
Портнова, 1941]. Полученные нами данные позволяют предположить, что в естественных условиях более распространены скрещивания самок лесной куницы и кидусов с самцами соболя.



**Рисунок 3.** Генетическая структура популяций, построенная на основании гаплотипов мтДНК (а) и вариантов Y-хромосомы (б) соболей, куниц и кидусов. В скобках указано число особей. Синими и красными звездочками отмечены соболя, попадающие в кластер «*M. Martes*», и куницы, попадающие в кластер «*M. zibellina*» соответственно.

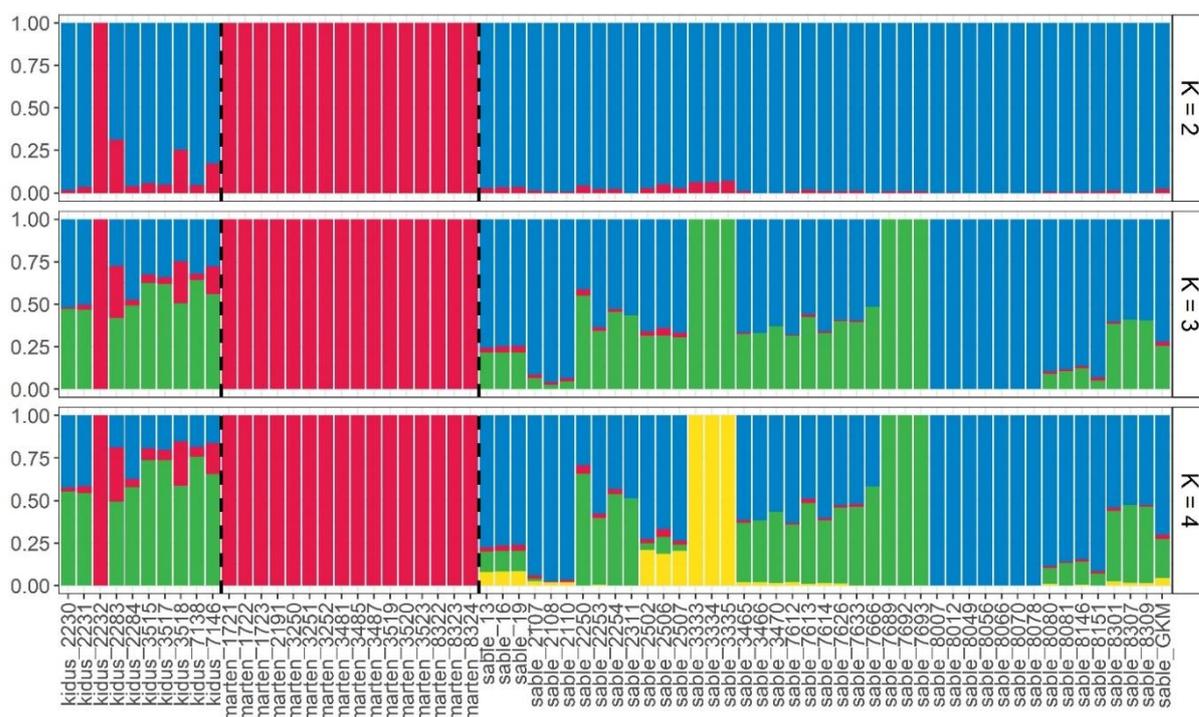
*Анализ генетических вариантов по всему геному.* В результате анализа методом главных компонент 69 образцов и генотипов по 19 млн. ОНП, а также в результате реконструкции филогенетического дерева по 14 млн. ОНП для этих же образцов установлено, что образцы соболей и куницы образуют два очень четко отделимых друг от друга кластера (**Рисунок 4**). При этом кластер куницы более плотный, чем кластер соболей, который может быть разделен на несколько частей. Обособлено от всех остальных группируются соболя с о. Итуруп, что, возможно, обусловлено длительным изолированным существованием этой популяции, ее низкой численностью, эффектами дрейфа генов и бутылочного горлышка. Еще 2 группы соболей располагаются несколько независимо от образцов соболя других популяций, однако это выражено в меньшей степени, чем для соболей о. Итуруп. Это соболя с о. Сахалин и Северного Урала, что также объясняется длительной изолированностью популяции о. Сахалин и высказываемыми ранее предположениями о существовании обособленного рефугиума соболей на Северном Урале [Каштанов и др., 2015]. Оставшиеся образцы соболя образуют плотный кластер, который можно подразделить на две части. В

первую часть могут быть отнесены образцы соболя из дальневосточных популяций (п-ов Камчатка, Сихотэ-Алинь и Большой Хинган), с ними же группируются фермерские образцы соболей из зверохозяйства «Пушкинский». Во вторую часть могут быть отнесены материалы, отобранные из популяций Западной и Центральной Сибири, Байкала и Алтая, что свидетельствует об их генетической близости. Среди кидусов один образец кластеризуется с образцами куниц, шесть – с образцами соболей Северного Урала и Западной Сибири, а оставшиеся три (kidus\_2283, kidus\_3518 и kidus\_7146) занимают промежуточное между кластерами соболей и куницы положение, со смещением в сторону первого. Подобное промежуточное положение свидетельствует о присутствии в их геномах как компонента от генома лесной куницы, так и соболя при преобладании последнего, что подтверждается результатами исследований другими методами, которые, в отличие от методов филогенетического анализа и анализа методами главных компонент, позволяют количественно оценить соотношение долей предковых компонентов в исследуемых геномах.



**Рисунок 4.** Результаты анализа методом главных компонент на основании 19 млн. ОНП по всему геному.

Результаты проведенного по полногеномным данным анализа ADMIXTURE также показали, что популяции соболей и куниц однозначно отделяются друг от друга (**Рисунок 5**). Установлено, что оптимальным числом предковых популяций (значение параметра  $K$ ) для анализируемых данных является  $K=2$  и  $4$ . Согласно результатам ADMIXTURE, исследуемые образцы лесных куниц являются генетически однородными. В исследуемых образцах соболя может быть выделено 3 предковых компонента: “Камчатский”, “Уральский” и “Итурупский”. При этом соболя Западной и Центральной Сибири, Байкала и Алтая имеют в геноме компоненты как “Камчатской”, так и “Уральской” предковых популяций в соотношении, близком к 1:1.



**Рисунок 5.** Результаты ADMIXTURE анализа на основании 615 тыс. несцепленных ОНП по всему геному. Представлены результаты анализа с предположением от 2 до 4 предковых популяций (значение параметра  $K$ ). Цветами отмечены разные предковые популяции, выявленные в ходе анализа.

ADMIXTURE анализ особей, фенотипически определенных как кидусы, подтвердил и значительно дополнил результаты PCA анализа. Один образец показал наличие 99,99% предкового компонента, характерного для куниц, для шести образцов доля этого компонента составила от 2,18% до 6,83%, а для трех (kidus\_2283, kidus\_3518 и kidus\_7146) составила 31,75%, 26,18% и 18,23% соответственно. Тот факт, что соотношения компонентов соболя и лесной куницы в геномах отклоняются от ожидаемых для гибридов  $F_1$  50% и смещены в сторону соболя, позволяет утверждать, что данные особи являются гибридами

следующих поколений. Наиболее вероятно они являются результатом 1–2 возвратных скрещиваний с соболем, но можно предположить и более сложные сценарии.

Аналогичные результаты получены и при применении разработанного нами метода выявления и подсчета доли специфичных для куницы аллелей. В отличие от ADMIXTURE анализа, предполагающего использование несцепленных ОНП, наш подход позволил провести анализ распределения доли специфичных для куницы аллелей вдоль наиболее крупных скаффолдов (более 10 млн. п.н.) в реконструированном нами геноме соболя. Для особей из симпатрических популяций обнаружено присутствие компонента генома другого вида в диапазоне от 1% до 5%. При этом в геномах кидусов с долей специфичных для куницы аллелей до 5% обнаружены протяженные участки длиной до 18 млн. п.н., в которых эти аллели находятся в гетерозиготном состоянии. Эти протяженные участки представляют собой еще не разрушенные рекомбинацией остатки гаплотипов соболя и лесной куницы, унаследованные от F<sub>1</sub> гибридов.

### **Глава 3. Молекулярно-генетические основы многообразия форм окраски меха у соболя и американской норки**

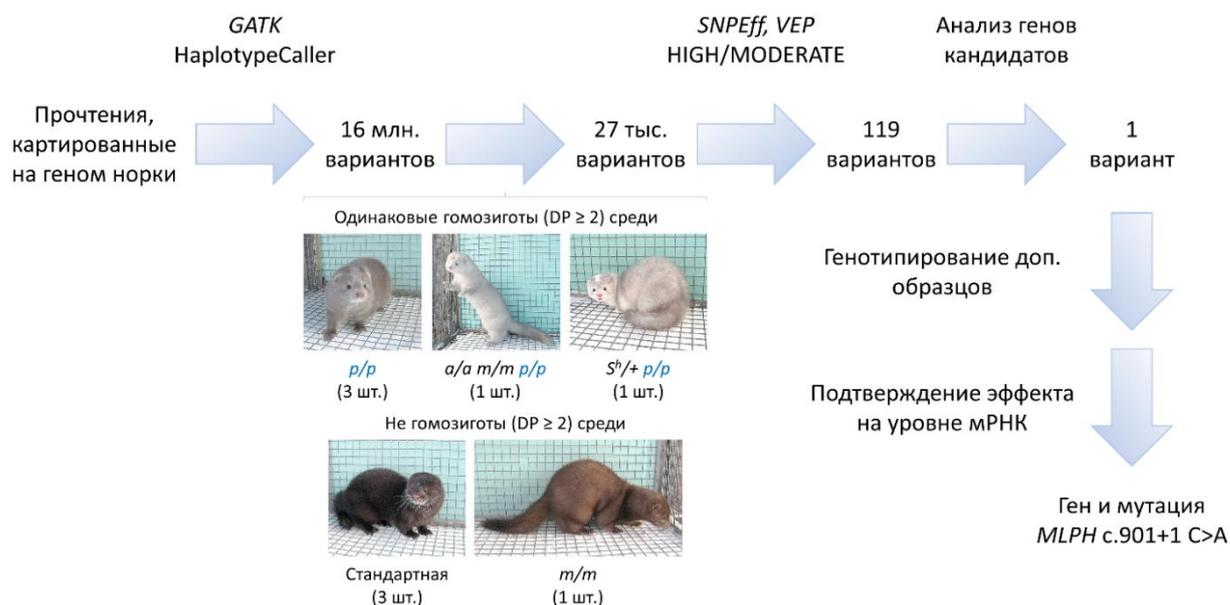
В диссертационной работе данная глава также состоит из трех разделов. Раздел **Обзор литературы** описывает наблюдаемый в настоящее время в фермерских популяциях соболя «цветовой взрыв» - возникновение и фиксацию новых цветовых форм, отсутствующих в природных популяциях. Опираясь на проведенные ранее исследования фенотипического параллелизма в окрасках меха у соболя и американской норки, предлагаем использовать обширную коллекцию линий американской норки с различными окрасками меха (описано по меньшей мере 35 различных мутантных форм [Трапезов, Трапезова, 2009]) в качестве модельного объекта для разработки стратегии анализа данных широкомасштабного параллельного секвенирования с целью идентификации генов и мутаций, лежащих в основе многообразия форм окраски меха у этих видов. С этой целью приводится обзор современного состояния проблемы генетики окраски у американской норки.

**Материалы и методы.** В общей сложности в ходе исследования были приготовлены геномные библиотеки и проведено полногеномное секвенирование на платформе Illumina 13-ти особей американской норки с семью типами окраски меха: стандартной темно-коричневой, серебристо-голубой (*p/p*), белой Хедлюнд (*h/h*), мойл (*m/m*), фиолет (*a/a k/k p/p*), черный хрусталь (*C'/C'*), Шедоу серебристо-голубой (*S<sup>h</sup>+ p/p*), а также двух особей соболя с окраской меха пастель (*b/b*). Ранее для всех изучаемых окрасок меха американской норки классическими методами генетического анализа установлен тип наследования [Трапезов,

Трапезова, 2009; Robinson, 1975]. Тип наследования окраски меха пастель у соболя установлен нами в результате генеалогического и гибридологического анализов.

Данные глубокого секвенирования картировали относительно геномов американской норки (*Neovison vison*, Ensembl NNQGG.v01) или соболя (MarZib1) с помощью программы BWA [Li, Durbin, 2009]. Затем в картированных относительно референсных геномов данных выявляли генетические варианты с помощью пакета программ GATK v4.

Мутации, обуславливающие формирование различных окрасок меха, идентифицировали по общей схеме (**Рисунок 6**). Например, для идентификации мутации, обуславливающей формирование серебристо-голубой окраски меха у американской норки, которая наследуется как моногенный аутосомно-рецессивный признак, сравнивали геномы животных, характеризующихся серебристо-голубой (*p/p*) окраской меха, а также потенциально включающими в себя аллель *p* окрасками: виолет (*a/a m/m p/p*) и Шедоу серебристо-голубой (*Sh/+ p/p*), с геномами животных стандартной темно-коричневой и мойл (*m/m*) окрасками мехового покрова. Из данных глубокого секвенирования отобрали такие генетические варианты, по которым исследуемые животные являются гомозиготами (с глубиной покрытия  $\geq 2$ ) по одному и тому же аллелю, а среди животных со сравниваемыми окрасками гомозиготные особи по данному аллелю отсутствуют.



**Рисунок 6.** Стратегия поиска мутаций, обуславливающих формирование различных окрасок меха, на примере анализа серебристо-голубой окраски меха у американской норки.

Аналогично, но с поправкой на тип наследования идентифицировали мутации, обуславливающие формирование белой Хедлунд (моногенная, кодоминантная,

эпистатирующая над другими типами окрасок), мойл (моногенная, рецессивная, описано по меньшей мере 2 аллеля), черный хрусталь (моногенная, кодоминантная, эпистатирующая над другими типами окрасок) и Шедоу (моногенная, доминантная, с рецессивным летальным действием) окрасок меха у американской норки.

Аннотацию отобранных генетических вариантов проводили с использованием программ VEP [McLaren и др., 2016] и/или SnpEff [Cingolani и др., 2012]. Отобранные мутации в генах кандидатах проверяли на расширенной выборке животных с помощью секвенирования по Сенгеру. Для мутаций, потенциально оказывающих влияние на сплайсинг, проводили анализ на уровне мРНК методом ОТ-ПЦР и последующего секвенирования по Сенгеру.

**Результаты и обсуждения.** В рамках данной работы впервые осуществили секвенирование геномов американских норок из линий, характеризующихся различными типами окраски мехового покрова (**Рисунок 6**), что позволило выявить гены и мутации, обуславливающие развитие шести исследованных мутантных форм окраски меха у американской норки (**Таблица 2**).

Разработанный и апробированный на данных геномного секвенирования американской норки подход применяли для поиска молекулярно-генетических факторов, обуславливающих развитие окраски меха пастель у соболя – первой зарегистрированной для соболя мутантной формы окраски мехового покрова. В результате генеалогического и гибридологического анализов результатов скрещиваний соболей черной и пастелевой окрасок меха (73 скрещивания, 237 потомка) установлено, что данный тип окраски наследуется как моногенный, аутосомно-рецессивный признак (**Таблица 3**).

В ходе сравнительного геномного анализа выявлена однонуклеотидная инсерция в седьмом экзоне гена *TYRP1* (с.1503\_1504insT), приводящая к сдвигу рамки считывания и возникновению преждевременного стоп-кодона на месте 509 аминокислотного остатка. Ген *TYRP1* кодирует мембрансвязанный тирозиназоподобный фермент, которой вместе с продуктом гена *DCT* участвует в поддержании стабильности фермента тирозиназы, модулирует его каталитическую активность в синтезе эумеланина, а также способен гидроксिलировать тирозин и продуцировать меланин, но в меньшей степени чем тирозиназа [Bennett, Lamoreux, 2003; Kamaraj, Purohit, 2013; Kobayashi и др., 1998].

**Таблица 2.** Идентифицированные и подтвержденные на расширенной выборке животных мутации, обуславливающие развитие различных форм окраски меха у американской норки.

Символ	Название	Ген	Мутация	Эффект мутации	Функции гена
<i>p/p</i>	Серебристо-голубая	<i>MLPH</i>	с.901+1 G>A	Нарушение сплайсинга	Транспорт зрелых меланосом
<i>h/h</i>	Белая Хедлюнд	<i>MITF</i>	с.33+1 G>A	Нарушение сплайсинга изоформы М	Выживание и дифференциация меланобластов
<i>m/m</i>	Мойл	<i>RAB38</i>	с.20-21dup	Сдвиг рамки считывания	Транспорт TYR и TYRP1 в созревающие меланосомы
<i>m<sup>c</sup>/m<sup>c</sup></i>	Камео	<i>RAB38</i>	с.574-589del	Сдвиг рамки считывания	
<i>C<sup>r</sup>/+</i>	Черный хрусталь	<i>COPA</i>	с.478 C>T	р.160 R>C в WD40 домене	Одна из субъединиц белкового комплекса COP1, участвующего в транспорте и сортировке трансмембранных белков из аппарата Гольджи в эндоплазматический ретикулум
<i>S<sup>h</sup>/+</i>	Шедоу	<i>KIT</i>	с.2374 G>T	р.792 D>Y активный сайт белка	Выживание и дифференциация меланобластов

**Таблица 3.** Гибридологический анализ скрещиваний соболей черной и пастелевой окрасок меха.

Скрещивание	Потомство					
	Число пар	Число потомков	Черный	Пастель	$\chi^2$	<i>p</i>
Пастель × Пастель	7	21	0	21	-	-
Черный × Пастель	22	71	71	0	-	-
	5	22	11	11	0,00	1,00
Черный × Черный	37	117	117	0	-	-
	2	6	4	2	0,22	0,64
Всего	73	237	203	34	-	-

## ВЫВОДЫ

1) Получены и охарактеризованы методом глубокого секвенирования геномы 42 соболей, 16 лесных куниц и 10 кидусов из разных географических популяций, а также 13 американских норок, характеризующихся различными типами окраски мехового покрова.

2) Впервые проведено секвенирование и *de novo* сборка геномов самцов соболя и лесной куницы, а также аннотация в них генов. В результате сравнительного геномного анализа особей лесной куницы выявлена генетическая близость особей, происходящих из северной части европейской территории России. Для соболя выявлены 3 основные предковые компоненты: “Камчатская”, “Уральская” и “Итурупская”. Выделение последней является результатом ее длительной изолированности от материковых популяций, эффекта бутылочного горлышка и дрейфа генов. Выделение “Камчатской” и “Уральской” предковых компонент подтверждает ранее предположенное расположение в этих районах ледниковых рефугиумов и заселение из них территорий Западной и Центральной Сибири.

3) Впервые методами геномного секвенирования охарактеризованы особи, по фенотипическим признакам определяемые как кидусы. Показано присутствие в исследованных образцах компонентов как соболиного, так и куньего генома, в том числе в отличных от ожидаемых для гибридов F<sub>1</sub> соотношениях. Полученные данные впервые демонстрируют плодовитость гибридов соболя и куницы в природных популяциях.

4) Разработан и успешно апробирован методологический подход анализа данных геномного секвенирования для выявления генетических факторов, обуславливающих формирование различной окраски меха у американской норки. Впервые выявлены мутации в генах *MLPH*, *MITF*, *RAB38*, *COPA* и *KIT*, обуславливающие развитие серебристо-голубой, белой Хедлюнд, мойл, черный хрусталь и Шедоу окрасок меха у американской норки.

5) С помощью разработанного методологического подхода анализа данных геномного секвенирования впервые выявлена мутация в гене *TYRP1*, обуславливающая развитие первой зарегистрированной для соболя мутантной окраски мехового покрова – па-стель.

## СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

### Публикации в журналах, соответствующих Перечню ВАК

1. **Manakhov A.D.**, Andreeva T.V., Trapezov O.V., Kolchanov N.A., Rogaev E.I. Genome analysis identifies the mutant genes for common industrial Silverblue and Hedlund white coat colours in American mink. // Sci. Rep. 2019. Т. 9. № 1. С. 4581.
2. **Manakhov A.D.**, Mintseva M.Yu., Andreev I.A., Uralsky L.I., Andreeva T.V., Trapezov O.V., Rogaev E.I. Genome analysis of American minks reveals link of mutations in Ras-related protein-38 gene to Moyle brown coat phenotype. // Sci. Rep. 2020. Т. 10. № 1. С. 15876.
3. **Manakhov A.D.**, Mintseva M.Y., Andreeva T.V., Filimonov P.A., Onokhov A.A., Chernova I.E., Kashtanov S.N., Rogaev E.I. Genome Analysis of Sable Fur Color Links a Lightened Pigmentation Phenotype to a Frameshift Variant in the Tyrosinase-Related Protein 1 Gene. // Genes (Basel). 2021. Т. 12. № 2. С. 157.

### Патенты, тезисы докладов в материалах конференций и симпозиумов

1. Рogaев Е.И., **Манахов А.Д.**, Андреева Т.В., Трапезов О.В., Колчанов Н.А. Молекулярно-генетические маркеры цветовых вариаций американской норки и способ выявления особей, являющихся носителем аллелей, обуславливающих формирование желаемой цветовой вариации. №2722564. Дата регистрации 01.06.2020
2. **Manakhov A.**, Andreeva T., Mintseva M., Andreev I., Uralsky L., Trapezov O., Rogaev E. Identification of molecular genetic factors linked to phenotypic diversity fur colours in American mink (*Neovison vison*) by whole genome sequencing analysis. The XII International Scientific Congress in Fur Animal Production (the IFASA Congress), August 24-25, 2021, Warsaw, Poland. *Онлайн доклад.*