

*На правах рукописи*

**Добрынин Павел Владимирович**

**СБОРКА И АННОТАЦИЯ ГЕНОМА АФРИКАНСКОГО ГЕПАРДА, *ACINONUX  
JUBATUS***

специальность 1.5.7 – генетика

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени

кандидата биологических наук

Москва – 2022

Работа выполнена в лаборатории “Центр геномной биоинформатики им. Ф. Г. Добржанского” Санкт-Петербургского государственного университета, г. Санкт-Петербург

**Научный руководитель:**

**О’БРАЙЕН Стефан Джеймс**, Doctor of Philosophy (PhD), Cornell University, USA, кандидат биологических наук, Корнеллский Университет, США, профессор, иностранный член российской академии наук, член американской национальной академии наук, главный научный сотрудник научно-образовательного центра “Геномное разнообразие” федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет ИТМО», г. Санкт-Петербург

**Официальные оппоненты:**

**ХОЛОДОВА Марина Владимировна**, доктор биологических наук, главный научный сотрудник, руководитель кабинета молекулярной диагностики федерального государственного бюджетного учреждения науки Института проблем экологии и эволюции им. А.Н. Северцова Российской академии наук, г. Москва

**ТРИФОНОВ Владимир Александрович**, доктор биологических наук, заведующий лабораторией сравнительной геномики отдела разнообразия и эволюции геномов федерального государственного бюджетного учреждения науки Института молекулярной и клеточной биологии Сибирского отделения Российской академии наук, г. Новосибирск

**Ведущая организация:**

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», г. Новосибирск.

Защита состоится « » \_\_\_\_\_ в \_\_\_\_\_ часов на заседании диссертационного совета 24.1.088.01 (Д.002.214.01) в Федеральном государственном бюджетном учреждении науки Институте 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 исследования и разработанность темы исследования:**

Африканские гепарды (*Acinonyx jubatus*) являются уникальными обитателями Африки к югу от Сахары и Юго-Западной Азии. Данный вид относится к семейству Кошачьих (*Felidae*), одному из 16 семейств, входящих в отряд Хищных млекопитающих (*Carnivora*). По данным IUCN Red List за 2021 год, 39 представителей семейства *Felidae* находятся под угрозой вымирания в той или иной степени. Представители данного семейства встречаются на всех континентах, кроме Австралии и Антарктиды.

На протяжении многих лет гепард эволюционировал как превосходный, стремительный хищник, достаточно проворный, чтобы преследовать добычу и избегать конкуренции с другими хищниками. Одним из приспособлений гепарда к выживанию являются морфо-функциональные особенности строения ногтей, рассчитанные на быстрое передвижение. Крепкое сцепление с поверхностью повышает маневренность и дает возможность совершать резкие повороты на высокой скорости в случае, когда добыча меняет направление движения, чтобы избавиться от преследования.

В настоящее время гепарды встречаются только в Африке и Иране, имеется достаточно доказательств того, что представители рода возникли от одного общего предка, который жил на территории Северной Америки в Миоцене. Данная оценка основана на сравнительном анализе молекулярных маркеров митохондриальной и ядерной ДНК, ископаемых данных и сведений о современном распространении видов. Время разделения гепардов, пум и ягуарунди на отдельные виды оценивается в пределах 6,7 млн. лет назад. Палеонтологические следы существования гепарда можно встретить в Северной и Южной Америке, Европе и Азии вплоть до позднего Плейстоцена (10.000 – 12.000 лет назад). Это время примерно соответствует периоду вымирания мегафауны, которое затронуло более чем 40 видов крупных млекопитающих, включая гепардо-подобных кошек (*Miracinonyx*) и пум в Северной Америке.

Генетические исследования гепардов, начатые в 80-е годы XX века, показали, что у данного вида практически полностью отсутствует генетическое разнообразие, а точнее применяемые методы не имели достаточного разрешения, чтобы его обнаружить. С тех пор в течение 30 лет были проведены исследования с использованием различных

генетических маркеров, полностью подтвердившие первоначальное заявление о том, что гепарды имеют низкое генетическое разнообразие, а по многим локусам не имеют его вовсе. Используемые маркеры и методы включали: анализ аллозимов в фибробластах с использованием двумерного электрофореза; оценку полиморфизма длин рестрикционных фрагментов в локусах *MHC*; изучение асимметрии костей черепа, секвенирование генов *MHC-I* и *MHC-II*; секвенирование митохондриальной ДНК; изучение полиморфизма генотипов, полученных при амплификации микросателлитных локусов.

Популяционная история вида и разнообразие специализированных адапционных механизмов делают гепардов крайне интересным объектом изучения на уровне генома, в частности, в связи с возможностью обнаружения адаптаций, оставивших след в геноме гепардов. Данное исследование является важным, так как касается вопросов эволюции генома и молекулярных основ механизмов физиологической адаптации гепардов, давая возможность совершенно по-новому взглянуть на биологию этого интересного вида.

#### **Цель исследования:**

Основной целью данного исследования было полногеномное изучение генетического разнообразия, адапционных механизмов и демографической истории африканского гепарда, находящегося под угрозой исчезновения.

#### **Задачи исследования:**

1. Секвенирование и сборка высококачественного референсного генома на основе одного образца африканского гепарда;
2. Ресеквенирование шести дополнительных образцов, представляющих два подвида, и картирование ридов на референсный геном;
3. Аннотация генов, повторов и других элементов генома методами, базирующимися на поиске гомологов, и алгоритмами для *ab initio* предсказания генов;
4. Идентификация генетических основ адапционных механизмов в геноме гепарда с применением методов сравнительного геномного анализа;
5. Изучение генетического разнообразия среди семи геномов гепардов методом детекции однонуклеотидных вариантов (SNV);
6. Сборка районов *MHC* гепарда и сравнение их с гомологичными районами других представителей млекопитающих;
7. Анализ демографической истории двух субпопуляций африканских гепардов.

### **Научная новизна:**

В рамках данной работы впервые представляется первичная сборка генома *de novo* и подробная аннотация генов и генетических элементов африканского гепарда. Собранный геном гепарда представляет собой первый публично доступный геном вида, относящегося к линии *Puma*. В данном исследовании был разработан новый метод сборки с применением референсного генома, который оказался более точным по сравнению с другими методами. В геноме гепарда были описаны ранее неизвестные способы адаптации к высокоскоростному бегу и плотоядному типу питания. Нами не только была точно описана потеря генетического разнообразия по всему геному, но и реконструирована цепочка исторических событий, приведших африканского гепарда к нынешнему состоянию.

### **Теоретическая и практическая значимость работы:**

Теоретическая значимость работы определяется тем, что она посвящена одной из основных фундаментальных проблем сохранения существующих в настоящее время биологических видов – пониманию эффектов снижения генетического разнообразия на геном и фенотип у исчезающих видов в результате резкого сокращения численности популяции – феномена бутылочного горлышка и эффекта основателя.

В рамках данной работы нами были разработаны методический и биоинформатический вычислительные подходы для аннотирования генома млекопитающих и анализа генетического разнообразия. Следует отметить, что, несмотря на увеличивающиеся объемы геномных данных, проблемы при работе с таким типом данных на немодельных организмах все еще остаются.

Данные полногеномного секвенирования и первичной сборки генома африканского гепарда были размещены в базе данных NCBI GenBank, и в настоящее время находятся в открытом доступе. Вся информация об описанных вариантах, повторяющихся элементах, генах и структурных вариантах доступна в геномном браузере GARFIELD (<http://garfield.dobzhanskycenter.org/>).

Описаны вредные мутации в гене *AKAP4*, которые присутствуют у большинства гепардов. Подобраны праймеры для амплификации локуса с данными мутациями. Эффективность праймеров была подтверждена на двадцати дополнительных образцах диких африканских гепардов.

Результаты, полученные для генома гепарда, могут быть использованы в последующих работах, связанных с биологией и сохранением исчезающих видов.

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

1. Геном африканского гепарда отличается крайне низким генетическим разнообразием по сравнению с другими близкородственными видами и иными млекопитающими.
2. Геномный район *MHC* у гепарда не только содержит сравнительно небольшое число аллелей *MHC* классов I и II, но и демонстрирует потерю до 4 генов *MHC* класса I, по сравнению с геномом домашней кошки.
3. В результате демографического анализа генома африканского гепарда были найдены доказательства прохождения популяции через два бутылочных горлышка.
4. Семейства генов, связанных с размножением, у африканских гепардов аккумулируют избыточное количество несинонимичных замен, часть из которых оказалась предположительно вредной по своей природе. Кроме того, ряд серьезных мутаций в гене *AKAP4* скорее всего являются причиной формирования дефектных сперматозоидов у гепардов.
5. Признаки положительного отбора и амплификации генных семейств показали крайне специфическую адаптацию, ассоциированную с быстрым бегом и физической выносливостью гепарда.

#### **Личный вклад автора в проведенные исследования:**

Выбор темы диссертационной работы, определение цели и постановка задач исследования автор провел совместно с научным руководителем PhD, профессором, Стефаном Джеймсом О'Брайеном. Автор работы изучил литературу по теме, непосредственно участвовал в написании и подготовке материалов к публикациям, написал рукопись работы. Основная часть экспериментальной работы: сборка генома и его фрагментов, выравнивание последовательностей, поиск генов и определение генных семейств, анализ сигналов позитивного отбора, анализ демографической истории выполнены автором лично. Общий вклад автора в работу - более 80%.

### **Степень достоверности и апробация результатов:**

Работа выполнена на высоком научно-методическом уровне. Используемые методы соответствуют поставленным задачам. Доклады по теме диссертации проводились на ежегодных отчетах 2013-2018 г. Автором опубликовано 7 статей по теме диссертации в научных рецензируемых изданиях, индексируемых в базе данных Scopus и отвечающих требованиям Высшей Аттестационной Комиссии Министерства науки и высшего образования Российской Федерации

Данные результаты также были представлены на следующих конференциях: Society for Molecular Biology and Evolution meeting (Австрия, 2015), Recent Advances in Conservation Genetics Course (Венгрия, 2016) и ViATA (Санкт-Петербург, 2021).

Обсуждение диссертации проводилось на семинарах в Центре геномной биоинформатики им. Ф.Г.Добржанского, Санкт-Петербургский Государственный Университет и НОЦ Геномное разнообразие в университете ИТМО. Изложенные в диссертационном исследовании положения и выводы являются достоверными. Апробация диссертационной работы проведена на межлабораторном семинаре ИОГен РАН (протокол №15 от 15 апреля 2021 г.).

### **Структура и объем диссертационной работы:**

Диссертация включает в себя следующие главы: «Введение», «Обзор литературы», «Материалы и методы», «Результаты и обсуждение», «Заключение», «Выводы» и «Список литературы». Работа изложена на 110 страницах машинописного текста, содержит 26 рисунков и 20 таблиц. Список цитируемых литературных источников включает 135 наименований, из которых 135 – на английском языке.

## СОДЕРЖАНИЕ РАБОТЫ

### Обзор литературы

В диссертационной работе глава «Обзор литературы» состоит из 4 разделов. Первый раздел посвящен филогенетике клады Puma и описывает два главных сценария происхождения рода *Acinonyx*. Во втором разделе представлена информация о снижении численности популяций и генетического разнообразия африканских гепардов в 19-20 веках. Третий раздел описывает ряд результатов по оценке генетического разнообразия, полученных учеными за последние 50 лет. В четвертом разделе содержится информация об основных этапах геномного проекта и методах работы с геномными данными.

### Материалы и методы

#### Биоинформатические методы:

Основные результаты работы были получены с использованием баз данных GenBank и RefSeq (NCBI), Ensembl, GeneCards, MalaCards, OMIM, KEGG, Rепbase, miRbase, Rfam и TreeFam. Сборка генома осуществлялась при помощи программ SOAPdenovo и Chromosomer. Для анализа геномных данных использовались программы BLAST+, Quast, Jellyfish, bwa, bowtie2, samtools, snpEff, PolyPhen2, ANGSD, CEGMA, BUSCO, LASTZ, ProgressiveCactus, GRIMM, RepeatMasker, TRF, tRNAscan-SE, Infernal, PAML, InterPro, HMMER3, Proteinortho, Gblocks, ClustalW, PSMC, DaDi и GADMA. Реализацию авторских скриптов проводили в программной среде Python, Perl и R.

**Chromosomer:** Мы специально разработали новый метод для картирования скаффолдов генома африканского гепарда на референсный геном домашней кошки. На схеме прямоугольники соотносятся с процедурами, которые применяются к данным, соотносящимся с параллелограммами. Мы использовали результаты парного полногеномного выравнивания двух видов, полученного в программе BLAST+. Затем присвоили величину сходства для каждого картированного фрагмента. Величина сходства основывается на длине и идентичности выровненных регионов. Мы предположили, что каждый скаффолд может иметь уникальное расположение в специфическом регионе на референсном геноме домашней кошки в случае, если отношение между лучшим выравниванием и вторым по степени сходства превышает заданный порог. В случае выполнения данного условия скаффолд закрепляется на позиции, соответствующей его



выравниванию с наибольшей оценкой. Скаффолды, не имеющие локализации и конкретной позиции, были исключены из дальнейшего анализа.

**GADMA:** Чтобы сделать вывод о демографической истории двух популяций гепарда (восточной и южной), мы использовали инструменты GADMA и DaDi. DaDi может генерировать спектры частот аллелей в соответствии с одним или несколькими демографическими сценариями. Задача исследователя состоит в том, чтобы максимизировать сходство между предполагаемой моделью и наблюдаемыми частотами аллелей в SFS. Степень сходства оценивали путем вычисления функции правдоподобия среди разных демографических сценариев. Мы предложили метод поиска демографической модели, оптимально соответствующей наблюдаемому спектру частот аллелей, основанный на генетическом алгоритме с использованием существующих решений для симулирования спектра частот аллелей из заданной демографической модели, а именно *dad* и *moments*. Разработанный метод поддерживает до 3х популяций. Данный метод реализован в программном обеспечении GADMA (Genetic Algorithm for Demographic Model Analysis), код которого выложен в открытый доступ и может быть найден по ссылке: <https://github.com/ctlab/GADMA>. Эффективность метода была проверена на реальных данных: на геномах современных людей и бабочек *Euphydryas gillettii*.

## Результаты и обсуждение

### **Анализ данных секвенирования африканского гепарда для референсного образца и шести дополнительных особей**

Были получены и проанализированы данные секвенирования для семи образцов африканского гепарда. Для получения высокого покрытия генома для референсного образца каждая библиотека была секвенирована на нескольких дорожках, что обеспечило более 300 секвенированных млрд. п.н. первичных данных до какой-либо фильтрации и анализа. Библиотеки для шести дополнительных образцов были секвенированы совместно на одном чипе секвенатора Illumina HiSeq2000, в результате чего получены первичные данные с умеренным выходом и 5-6-кратным уровнем покрытия генома. Для получения данных высокого качества из объема первичных прочтений мы применяли ряд фильтров

для отсева дубликатов ПЦР, химерных последовательностей и прочтений с высоким уровнем ошибок.

### ***De novo* сборка генома африканского гепарда**

Для образца (Chewbacca) с высоким покрытием при сборке генома *de novo* была использована только часть полученных сырых прочтений прошедших фильтрацию по качеству и коррекцию. Откорректированные прочтения были собраны в 97.981 контигов (N50 = 28,2 т.п.н.) и 2332 скаффолда (N50 = 3,1 млн. п.н.), причем длина самого протяженного собранного скаффолда составила около 13 млн.п.н. Общий размер собранного генома составил 2,376 млрд.п.н. с менее чем 2% неизвестных пар оснований. Среднее покрытие генома, рассчитанное по картированным прочтениям, составило 73×.

### **Сборка генома африканского гепарда на основе референса домашней кошки**

Разработанная нами программа Chromosomer создает первичную сборку скаффолдов хромосомного уровня на основе карты выровненных фрагментов, которые должны быть как в целевом, так и в референсном геноме. Учитывая консервативную организацию генома у представителей *Felidae*, мы решили использовать геном домашней кошки в качестве референса для сборки хромосом гепарда. Мы успешно идентифицировали и картировали 1.519 фрагментов, входящих в состав 18 аутосом и X хромосомы. В процессе сборки с использованием референсного генома мы внесли пропуски общей длины 139 млн п.н. Размер первичной хромосомной сборки составил 2,48 млрд п.н. Общая длина всех нелокализованных скаффолдов составила 34 млн п.н., что составляет менее 1,5% длины всех собранных скаффолдов.

Кроме того, размер собранных нами хромосом гепарда оказался гораздо больше, чем было предсказано, в связи с тем, что мы добавили несколько пропусков в полученную геномную сборку для устранения перекрытий скаффолдов.

Для определения расположения скаффолдов половых хромосом мы сравнили степень покрытия и GC состав всех собранных фрагментов. Нами было замечено, что существует два распределения, одно с 20-30-кратным и второе с 40-60-кратным покрытием, со снижением степени покрытия в регионах с высоким GC составом. Снижение глубины секвенирования в GC насыщенных регионах является общей проблемой технологии секвенирования компании Illumina. Мы предположили, что

распределение с 20-30-кратной глубиной охвата соответствует X и Y хромосомам, не считая псевдоаутосомной области.

В связи с высоким содержанием повторов и небольшими размерами Y хромосомы, мы не смогли собрать её с помощью программного обеспечения Chromosomerg. Для поиска скаффолдов, которые могут быть ассоциированы с Y хромосомой гепарда, мы сравнили гены, расположенные в Y хромосоме человека, с некартированными скаффолдами гепарда. Из 54 экспрессирующих белковые продукты генов Y хромосомы человека была предсказана локализация 21 гена в скаффолдах гепарда. Таким образом, нами были найдены пять скаффолдов Y хромосомы гепарда общей длиной 1524629 п.н.

Геном африканского гепарда - это первый референсный геном представителя из филогенетической линии пумы, одной из восьми основных клад, представляющих дошедших до нас кошачьих.

### **Оценка качества геномной сборки африканского гепарда**

В случае отсутствия физической карты сцепления или референсного генома достаточно сложно определить точность сборки скаффолдов. Мы опирались на набор косвенных признаков: 1) Соответствие размера собранного генома с размером предсказанного; 2) Надежность и точность картирования исходных и дополнительных прочтений на собранный геном; 3) Представленность и целостность генов, консервативных для всех эукариот; 4) Подтверждение наблюдений о высоком уровне синтении среди кошачьих относительно других видов млекопитающих.

Для предсказания размера генома по данным первичных прочтений мы анализировали распределение 17-меров. Наши результаты показали, что размер генома африканского гепарда составляет около 2,395 млрд.п.н., что близко к размерам геномов ранее секвенированных представителей семейства кошачьих – домашней кошки (2,641 млрд.п.н.), амурского тигра (2,400 млрд.п.н.) и амурского леопарда (2,450 млрд.п.н.) [Pontius et al., 2007; Tamazian et al., 2014; Cho et al., 2013]. Предсказанный нами размер генома африканского гепарда может быть немного больше, чем фактический геном, из-за ошибок секвенирования, в результате которых могут появляться более редкие k-меры. Общая длина всех собранных скаффолдов составила 2,375 млрд п.н., что на 20 млн п.н. меньше по сравнению с предсказанием, составленным по первичным прочтениям. Данное

несоответствие может быть вызвано не только за счет переоценки размера предсказанного генома, но и вследствие недопредставленности некоторых центромерных районов. В связи с размером и сложностью нуклеотидной последовательности оказывается почти невозможным собрать перицентромерные районы генома, используя технологии, основанные на коротких прочтениях, например, Illumina [Illumina, 2014].

Картирование прочтений на первичную сборку генома гепарда показало, что около 97% всех прочтений имели уникальное положение. Значительная пропорция однозначно картированных прочтений является косвенным признаком хорошей геномной сборки.

Более 91% генов, консервативных для эукариот, присутствовали в нашей сборке генома, что говорит о высоком уровне сходства качества нашей сборки с геномами других представителей отряда хищных: собаки - 93%, домашней кошки - 95%, тигра - 93%, льва - 89%, процент обнаруженных генов предполагает аналогичную точность сборки геномов данных видов.

Мы произвели выравнивание геномов африканского гепарда, домашней кошки, тигра, льва и домашней собаки для подсчета количества и размера блоков синтении. Мы ожидали, что количество и размер блоков синтении будет больше между близкими видами, чем между далекими. Кроме того, мы искали крупные геномные перестройки в геномах домашней кошки и африканского гепарда, поскольку наличие большого количества таких перестроек будет говорить о возможных ошибках, полученных в процессе сборки скаффолдов. Процент совпадения генома африканского гепарда по сравнению с геномами других животных составил: 93,6 % для тигра (*Panthera tigris*), 91,6 % для льва (*Panthera lion*), 91,1 % для домашней кошки (*Felis catus*) и 74,1 % для собаки (*Canis familiaris*). Как и было предсказано, наибольшая разница была найдена между наиболее отдаленными видами (гепардом и собакой).

Принимая во внимание полученные результаты, мы уверены в высоком качестве *de novo* геномной сборки африканского гепарда.

### **Поиск генов и повторов в геноме африканского гепарда, *Acinonyx jubatus***

Всего в геноме гепарда было идентифицировано 20343 белок-кодирующих генов и 110431 генов некодирующих РНК-элементов. Доля обнаруженных повторов составила

39,48%. Данные показатели являются консервативными как для кошачьих, так и для более широкой групп хищников.

### **Поиск сигналов положительного отбора**

Мы искали следы недавнего естественного отбора по всем генам гепарда, оценивая соотношения  $kA/kS$  в соответствии с ортологами из геномов льва, тигра, кошки, человека и мыши. В частности, мы использовали PAML для тестирования сигналов позитивного отбора вдоль филогенетической линии гепарда и обнаружили 947 генов со значимыми сигналами ( $p < 0,05$ , скорректированных для множественного тестирования), десять из которых показали обогащение в конкретных GO-терминах. Пять генов со следами отбора были связаны с регуляцией сокращения сердечной и поперечнополосатой мускулатуры (*ADORA1*, *RGS2*, *SCN5A*, *ADRA1* и *CACNA1C*); два гена (*TAOK2* и *ADORA1*) были связаны с активностью MAPK, важной для ответа на стрессовые воздействия, включая тепловой стресс, и четыре гена (*APOC3*, *DDIT4*, *SUFU* и *PPARA*) были связаны с негативной регуляцией катаболических процессов.

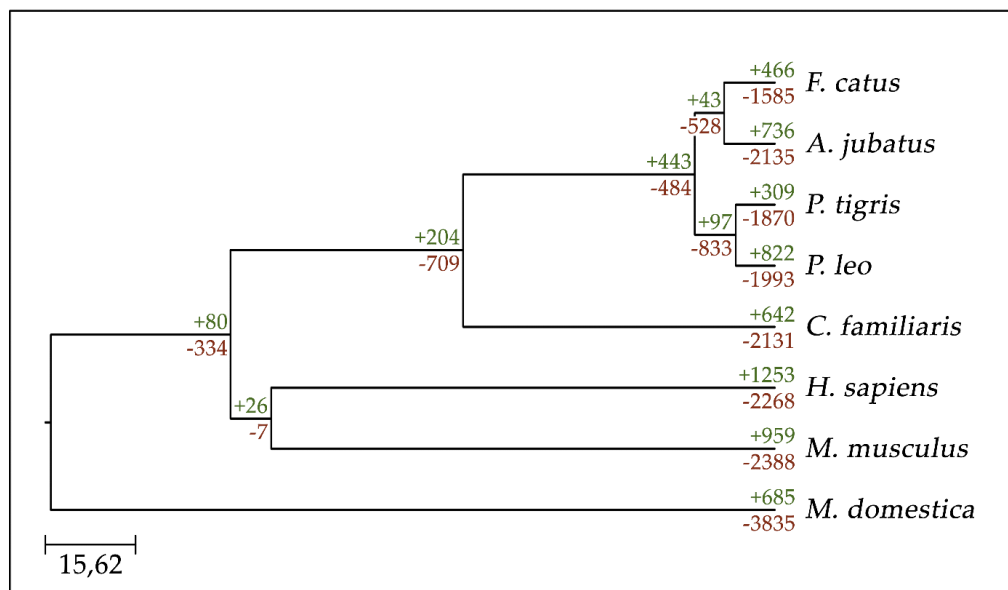
### **Анализ экспансии и сокращения генных семейств**

Геном гепарда содержит 17863 семейства ортологичных генов. Среди них 10983 семейства ортологичных генов были общими для всех девяти анализируемых геномов, и 12114 семейства были общими для группы *Felidae*. В то же время 112 кластеров были общими исключительно для генома гепарда и домашней кошки. Было проаннотировано 1335 генов, уникальных для гепардов; 812 из них содержали 2293 белковых домена, определенных при помощи InterPro. Анализ, основанный на сравнении ортологичных генных семейств среди восьми видов млекопитающих (африканский гепард, амурский тигр, африканский лев, домашняя кошка, человек, собака, мышь и опоссум), показал, что геном гепарда имеет 814 расширенных и 2169 сокращенных генных семейства по сравнению с общим предком кошачьих (Рис. 1).

Расширенные генные семейства включали множество GO-терминов, таких как обонятельные и G-связанные белковые рецепторы (также подвергшиеся экспансии у других представителей *Felidae*), которые могут быть связаны с физиологией гепарда. Семейство генов *LDH-A* и *LDH-B*, кодирующих ферменты лактатдегидрогеназы, являются

важным элементом анаэробного метаболического пути. Показано двукратное увеличение числа генов у некоторых видов *Felidae* (кошка, гепард и лев) по сравнению с другими млекопитающими, что потенциально является характерной адаптацией для хищнического образа жизни, типичного для большинства видов кошачьих.

Анализ динамики экспансии и сокращения генных семейств может дать ключ к пониманию эволюционных событий, которые сформировали геномы млекопитающих. Предполагается, что дупликация гена является мощным двигателем эволюционных изменений и инноваций, в то время как потеря гена считается распространенной реакцией на изменения в селективном давлении в процессе эволюции, например, в происхождении человека.



**Рисунок 1. Экспансия и сокращение генных семейств в геноме гепарда.** Кладограмма, отражающая филогенетические отношения между изучаемыми видами млекопитающих. Нанесенные на нее цифры, отмеченные зеленым и красным цветом, отражают количество генных семейств, подвергнувшихся экспансии и сокращению соответственно.

### Анализ генов, связанных с размножением

Современные представители вида *Acinonyx jubatus* демонстрируют множественные физиологические корреляты инбридинга как в неволе, так и в диких популяциях. Гепарды проявляют конститутивные нарушения в размножении, включая низкую плодовитость в неволе, в среднем 80% сперматозоидов с аномальной морфологией на эякулят и

повышенную частоту акросомных дефектов, как это наблюдалось у других инбредных природных популяций

Мы обнаружили в геноме гепарда 92 гена, связанных с размножением (например, оогенез и сперматогенез), с повышенными значениями  $kA/kS$ . В дальнейшем эти гены были вручную проверены с использованием общедоступных баз данных. Был сформирован финальный список из 18 генов для поиска в базах данных генетических заболеваний, включая OMIM, KEGG и MalaCards, а также для скрининга всех несинонимичных замен и делеций. Патогенность мутаций оценивали с использованием PolyPhen2 с белками человека в качестве модели. Среди 18 генов мы обнаружили один ген, который показал избыток возможных повреждающих миссенс-мутаций, и был связан с важными функциями сперматогенеза: *AKAP4*. Мы использовали секвенирование по Sanger для подтверждения мутаций *AKAP4* у 10 дикоживущих гепардов из Намибии. Четыре из пяти несинонимичных замен были подтверждены в 9 образцах и оказались гомозиготными. Пятая мутация пока не была изучена, поскольку она была расположена в одной из праймерных последовательностей, которые мы сконструировали для амплификации *AKAP4*.

Мутации в гене *AKAP4*, обнаруженные в геноме гепарда, не наблюдались у тигра, домашней и дикой кошки (*Felis silvestris*), а также у азиатского льва, чья единственная популяция обитает в национальном парке Гир-Форест, Индия, и демонстрирует крайнее генетическое истощение и обширные репродуктивные дефекты. Обнаруженные нами нарушения репродуктивных генов являются важными фактами для описания особенностей репродуктивной физиологии, которая наблюдается у всех гепардов.

Геномный анализ выявил убедительные данные для групп генов, участвующих в размножении, которые накапливали избыток аминокислотных изменений у гепардов по сравнению с другими кошачьими. Мы идентифицировали восемь фиксированных мутаций в локусе гена *AKAP4*, который экспрессируется исключительно в семенниках, чьи гомологи играют критическую роль в развитии и подвижности сперматозоидов и возникновении аномалий у сперматозоидов для некоторых видов млекопитающих. Пять гомозиготных мутаций в гене *AKAP4*, весьма вероятно, объясняют высокую долю дефектных сперматозоидов (в среднем 81,6% поврежденных сперматозоидов), обнаруженную у гепарда.

## Изучение генетического разнообразия в свободноживущих популяциях африканского гепарда, *Acinonyx jubatus*

Для всесторонней оценки степени генетического разнообразия мы использовали шесть различных способов: 1) Встречаемость однонуклеотидных замен на всю протяженность генома; 2) Плотность однонуклеотидных замен в окнах генома протяженностью в 50 т.п.н.; 3) Количество вариативных позиций в кодирующих областях; 4) Протяженность участков гомозиготности; 5) Количество достоверных гетерозигот на геном; 6) Генетическое разнообразие с использованием митохондриальных последовательностей.

Гепарды демонстрируют самую низкую встречаемость SNV в геноме среди 11 видов, включая человека, домашнюю кошку, гориллу (*Gorilla gorilla*), льва (*Panthera leo*) и тасманского дьявола (*Sarcophilus harrisii*), частота встречаемости SNV на 90% меньше, чем, у домашнего кота (Boris), рис. 2.

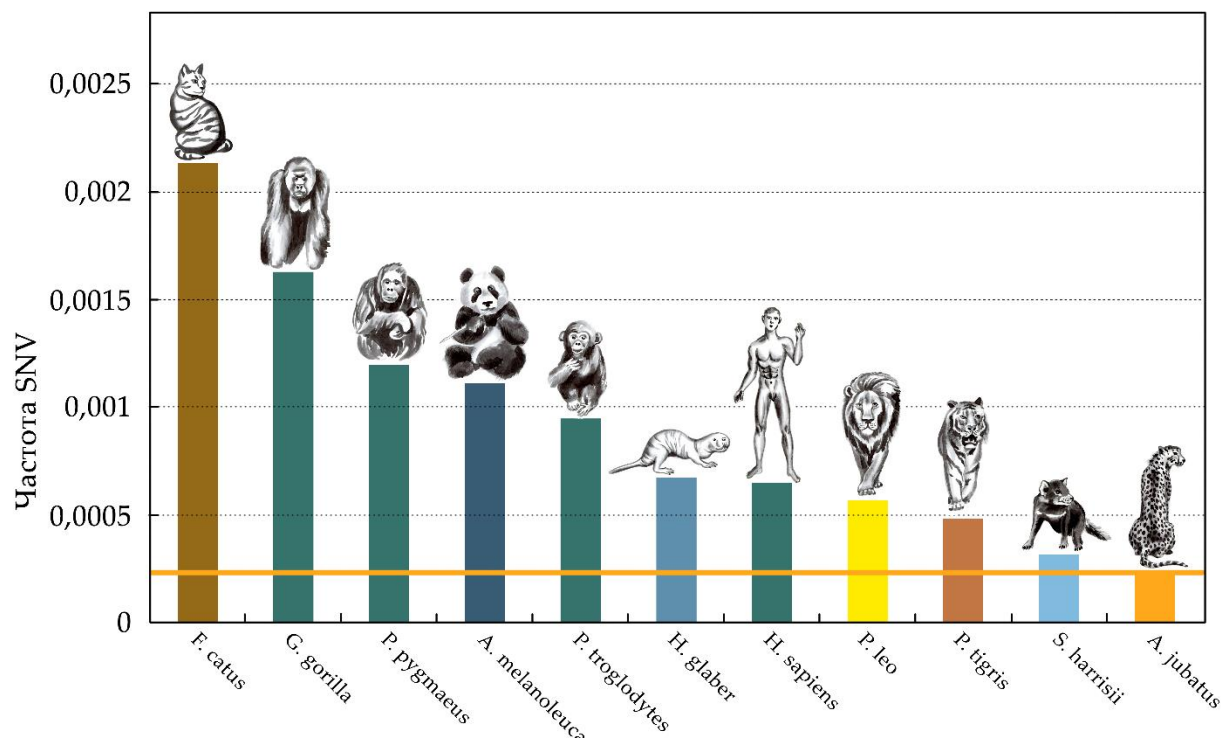
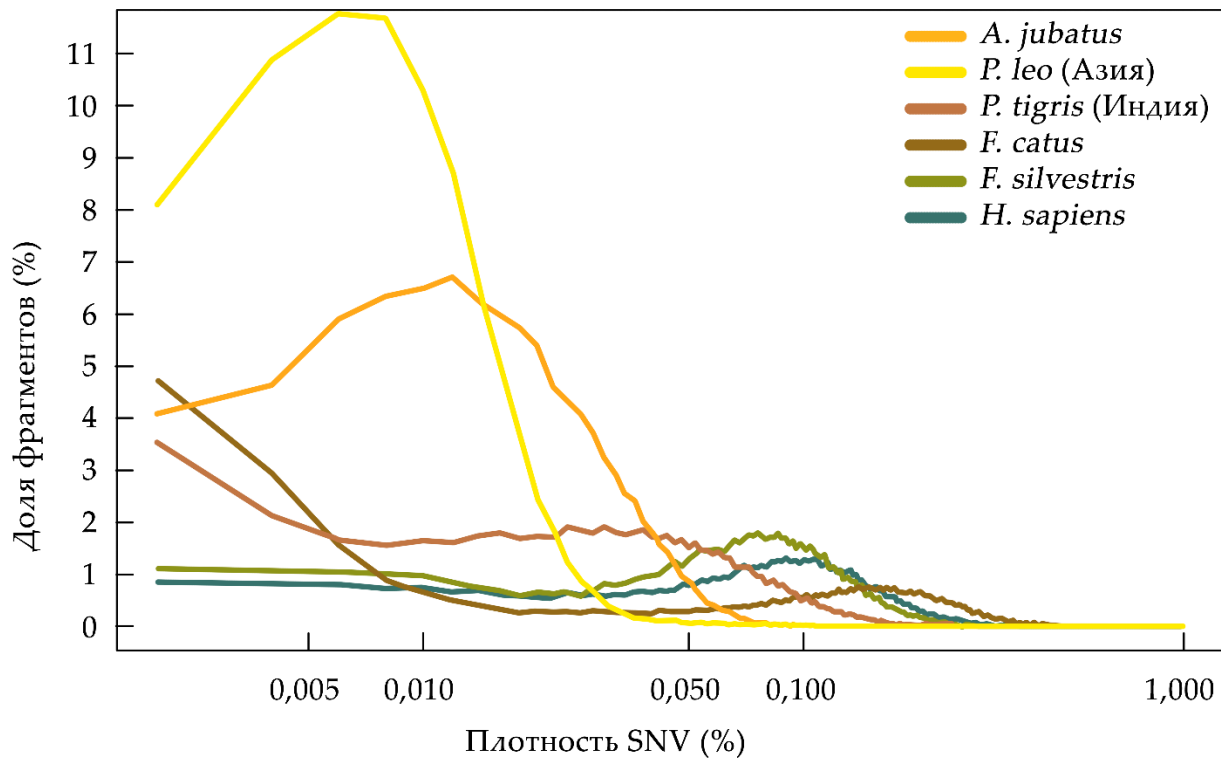


Рисунок 2. Частота SNV у одиннадцати видов млекопитающих



Референсный геном африканского гепарда разбивали окнами по 50 тысяч п.н., которые использовали для оценки плотности SNV. В общей сложности 46 787 окон составили 2,337 Гб или 99,12% от общей длины генома. Анализ большинства окон показал, что гепарды в 8-15 раз менее генетически разнообразны, чем человек, домашний кот (Boris) или Европейский дикий кот (*Felis silvestris*) (Рис. 3). Единственным видом или популяцией с сопоставимой, либо более низкой плотностью SNV, чем гепард, был азиатский лев из Национального парка Гир-Форест в Индии.

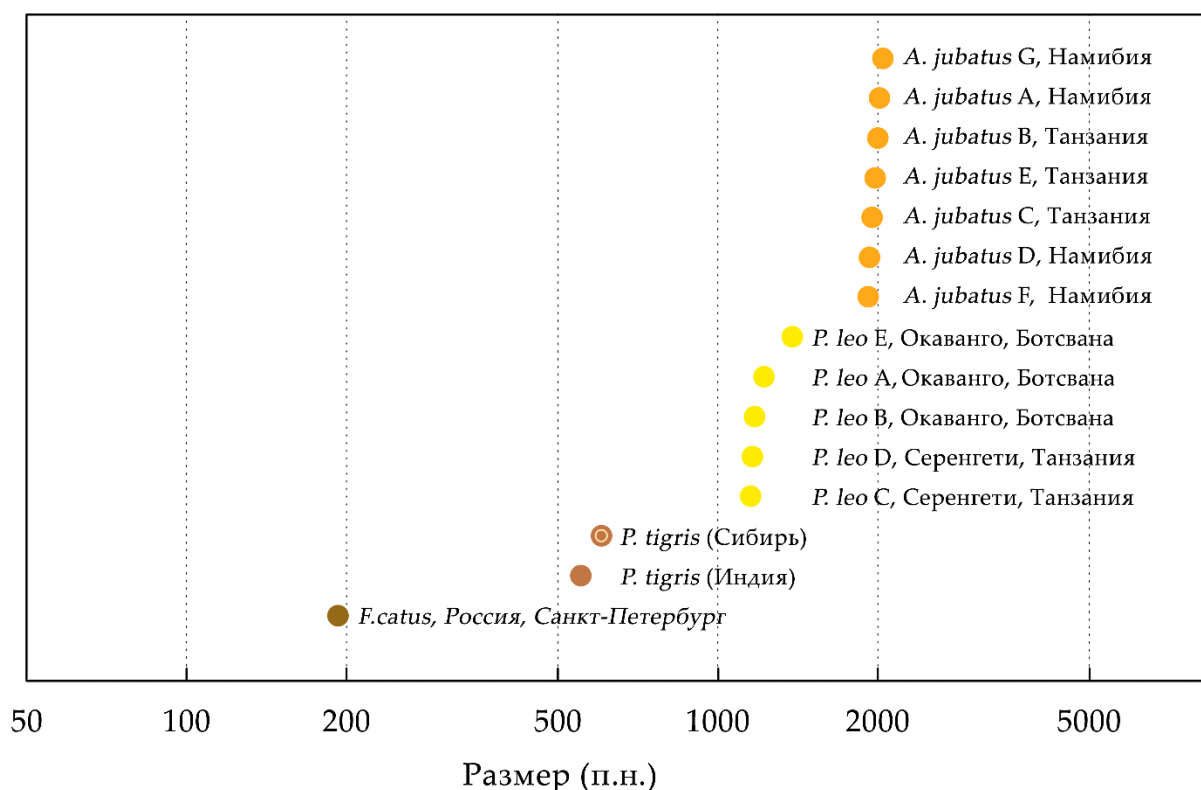


**Рисунок 3.** Плотность SNV в скользящих окнах по 50 тысяч п.н. у гепарда, тигра, дикой и домашней кошки (Boris), и человека.

Кодирующие гены гепарда обнаружили резкое 50-кратное (~98%) сокращение генетической вариации к уровню вариации в геноме у домашней или дикой кошки. Экстремальное сокращение вариативных позиций в кодирующих генах объясняет первоначально открытую низкую генетическую изменчивость гепарда, обнаруженную три десятилетия назад в исследованиях с использованием аллозимов, электрофоретических

вариантов клеточного белка и полиморфизма длины рестрикционных фрагментов на основе генов (RFLPs).

Гепарды демонстрируют в среднем в 10 раз более протяженные участки гомозиготности по сравнению с геномом дикой и домашней кошки. В среднем 93% генома каждого гепарда состоит из областей гомозиготности (ROH - то есть областей генома, в которых не было обнаружено SNV) с медианной длиной около 2000 п.н. Медиана длины регионов гомозиготности (ROH) у гепардов (семь особей), африканских львов (пять особей), сибирских и бенгальских тигров и домашней кошки представлены на рис. 4.



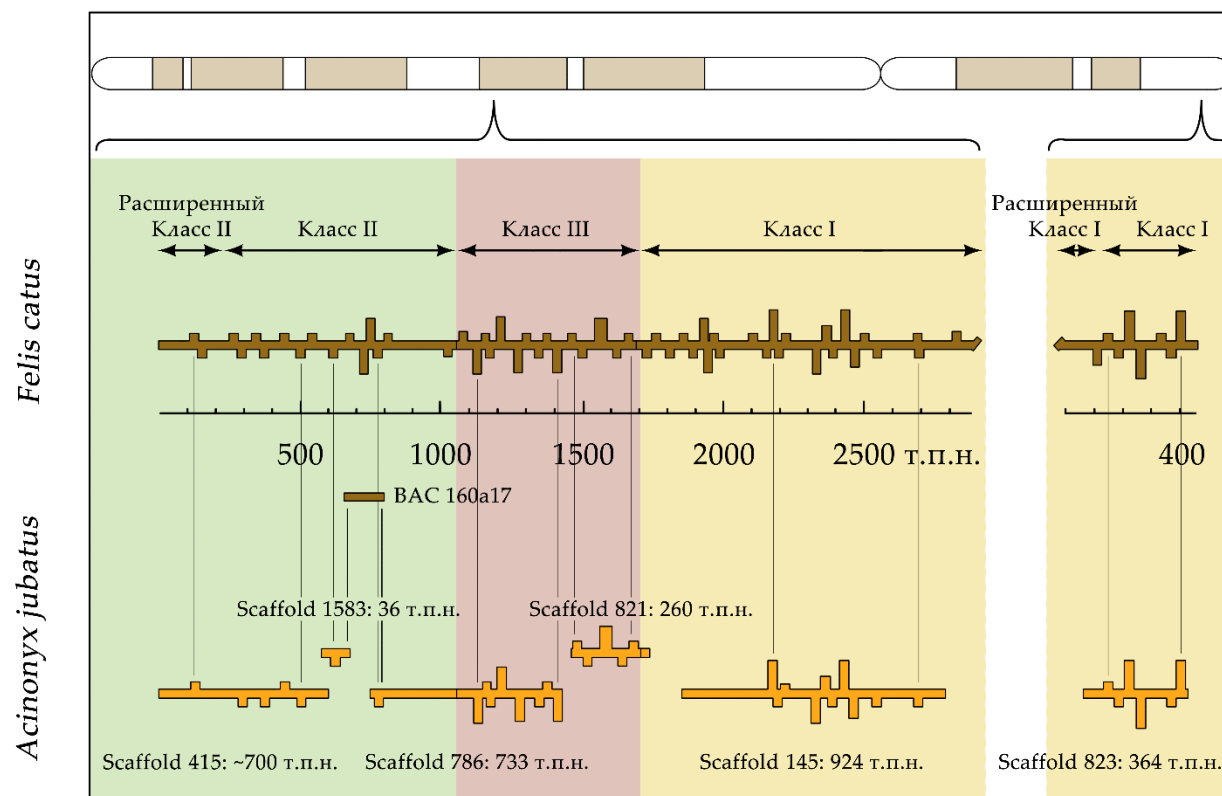
**Рисунок 4. Размер участков гомозиготности в геномах.** Для каждого образца отмечена медиана длин всех регионов гомозиготности.

В среднем в геноме гепарда обнаружено очень мало гетерозиготных SNV, 0,019-0,021%, что составляет до 50-61% гетерозигот у тигров, 30% у людей и 15% у домашних

кошек. Полные митохондриальные геномы гепарда аналогичным образом показывают снижение количества SNV по отношению к другим видам в среднем на 90%.

### Сборка района МНС при помощи референсной последовательности

Мы идентифицировали скаффолды, предположительно принадлежащие к области МНС, на основе гомологии с FLA-последовательностью домашней кошки. Чтобы восстановить область МНС класса I, мы картировали скаффолды гепарда на исходную последовательность FLA в качестве референса. Последовательности ВАС от домашней кошки использовали для дополнительной поддержки ориентации и порядка идентифицированных фрагментов (Рис. 5).



**Рисунок 5. Часть региона МНС гепарда, собранного на основе референса.** Сверху – референсный район МНС из домашней кошки (*Felis catus* FLA). Снизу – Скаффолды из de novo сборки генома гепарда, ориентированные по порядку генов и ВАС домашней кошки.

Из сборки гепарда удалось выделить 13 скаффолдов и 7 контигов (общий размер 8,3 млн п.н.), которые содержали 278 генов, относящихся к области FLA, причем относительный порядок генов у гепарда был аналогичен таковому в геноме *Felis catus*. Генетическое разнообразие в МНС области генома гепарда соответствовало уровню разнообразия по всему геному, и мы смогли найти лишь несколько полиморфных позиций.

Мы сравнили количество обнаруженных SNV (синонимичных и несинонимичных) в генах МНС гепарда (из Намибии и Танзании), домашней и дикой кошки, человека и собаки. Мы обнаружили 95-98% -ное снижение SNV разнообразия в обеих популяциях гепардов, а также у Cinnamon (высокоинбредного абиссинского кота) по сравнению с обилием SNV у бродячего домашнего кота (Boris), человека и собаки.

Сокращение количества SNV в областях МНС инбредной кошки и гепарда включает как синонимичные, так и несинонимичные аминокислотные замены. Подобные функциональные варианты обычно являются целью балансирующего отбора и отражают историю взаимодействия вида с различными патогенами.

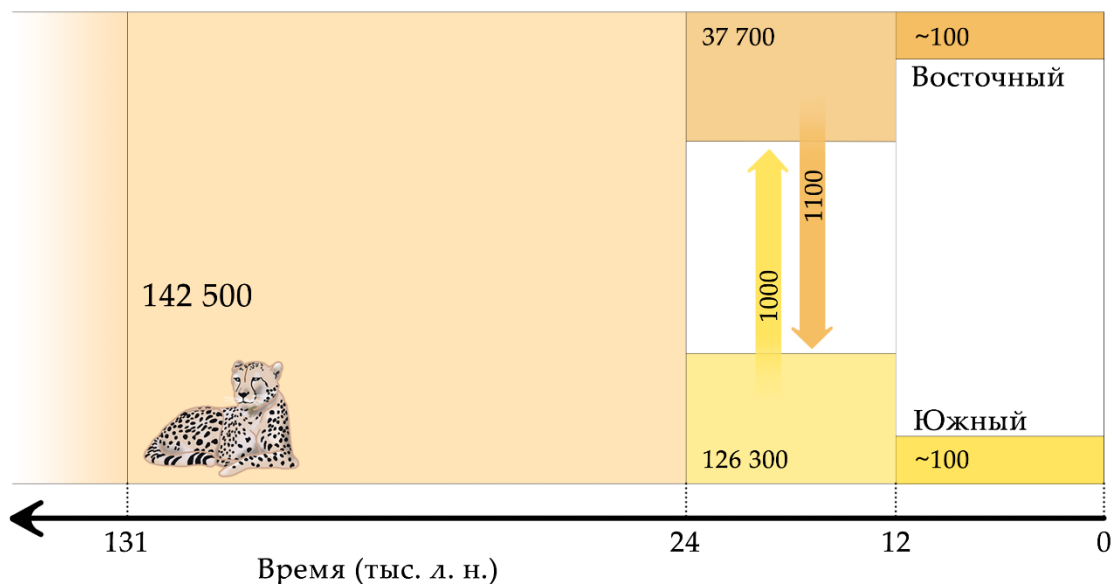
### **Изучение популяционной истории африканского гепарда на основе данных полногеномного секвенирования**

Результаты PSMC показали постепенное сокращение эффективной численности популяции в течение времени без каких-либо свидетельств о наличии эффекта бутылочного горлышка в прошлом. Подобный результат может быть обусловлен спецификой анализа PSMC, имеющего более низкую чувствительность к событиям в более недавнем прошлом (поскольку меньше коалесцентных событий оставили след в геноме) и/или, в результате того, что событие, приведшее к сокращению численности, было настолько быстрым и тяжелым, что привело к потере всех более ранних следов.

Мы использовали паттерны изменений геномных вариаций для моделирования и определения истории популяции гепардов из восточной и южной части Африки (из Танзании и Намибии, соответственно) при помощи модели диффузной аппроксимации и спектров частот аллелей (SFS). Подход, реализованный в программном обеспечении DaDi, сравнивает ожидаемую и наблюдаемую частоту аллеля в SFS, вычисляя совокупный показатель правдоподобия для наилучшего подбора нескольких различных, но

правдоподобных эволюционных сценариев. Сценарии были смоделированы с использованием данных SFS, результаты были использованы для определения лучшей модели, которая наиболее правдоподобна с точки зрения наблюдаемых данных.

Модель, предполагающая расширяющуюся предковую популяцию, которая разделяется на две субпопуляции, показала наибольшую вероятность соответствия, основываясь на низкой дисперсии и высоком значении максимального правдоподобия. Результаты моделирования предполагают наличие эффекта основателя у гепардов около 100 000 лет назад, возможно, это следствие их долгой миграционной истории, когда в позднем плейстоцене они перешли из Северной Америки через Берингов пролив в Азию, а затем на юг в Африку, и в связи с регулярным постоянным сокращением численности, а также с ограничением потока генов (Рис. 6).



**Рисунок 6. Анализ демографической истории африканского гепарда.** Демографическая история двух популяций гепарда (юга Намибии и восточной части Танзании)

В качестве альтернативного сценария был предложен иной вариант трактовки наблюдаемых результатов. Основываясь на данных изучения древней ДНК *Miracinonyx trumani* (американских гепардов), можно предположить, что современные африканские гепарды происходят из Азии, что указывает на то, что эффект основателя 10.000 лет назад совпадает с миграцией гепардов из Азии в Африку.

## ВЫВОДЫ

1) За последние сто тысяч лет в результате по крайней мере двух резких сокращений эффективной численности популяции *Acinonyx jubatus* была потеряна существенная часть генетического разнообразия, которое является характерной чертой для дико живущих представителей семейства *Felidae* и других млекопитающих.

2) Высокое генетическое сходство и низкое разнообразие аллелей генов МНС классов I и II, а также потеря нескольких генов МНС класса I дают наиболее вероятное объяснение двум ранее наблюдаемым фенотипическим свойствам у *Acinonyx jubatus*: а) Приживаемость кожных трансплантатов между неродственными особями и б) Высокая смертность от коронавирусной инфекции.

3) Данные анализа сайт-частотных спектров и гаплотипов между двумя популяциями Африканских гепардов позволили с высокой точностью реконструировать демографическую историю вида вплоть до событий в районе 100 тысяч лет назад. Были получены данные, подтверждающие и уточняющие ранее высказанные гипотезы о прохождении предковых популяций через бутылочное горлышко в районе 12 тысяч лет назад.

4) Анализ мутационного груза в группе генов, связанных с размножением, обнаружил избыточное количество потенциально вредных мутаций относительно генома домашней кошки и других млекопитающих, а также выявил вероятную причину высокой доли аномальных сперматозоидов у Африканских гепардов в виде гена *AKAP4*, накопившем ряд серьезных мутаций.

5) Следы положительного отбора, экспансия некоторых генных семейств, а также структурные вариации в генах, отвечающих за регуляцию сокращения сердечной и поперечнополосатой мускулатуры, тепловой стресс и регуляцию катаболических процессов, демонстрируют генетические основы уникальных адаптаций *Acinonyx jubatus* к хищническому образу жизни и стилю охоты, который выделяет их на фоне не только группы *Felidae*, но и *Carnivora* в целом.

## СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

### Публикации в журналах, соответствующих Перечню ВАК

1. Tamazian G. Annotated features of domestic cat – *Felis catus* genome / G. Tamazian, S. Simonov, P. Dobrynin et al. // GigaScience. – 2014. – Т.3. – №.1. – С.13.
2. Dobrynin, P. Genomic legacy of the African cheetah, *Acinonyx jubatus* / P. Dobrynin, S. Liu, G. Tamazian et al. // Genome biology. – 2015. – Т.16. - №.1. – С.277.
3. O'Brien, S.J. Response to Comment by Faurby, Werdelin and Svenning / S.J. O'Brien, K.P. Koepfli, E. Eizirik, W. Johnson, C. Driscoll, A. Antunes, A. Schmidt-Kuntzel, L. Marker, P. Dobrynin // Genome biology. – 2016. – Т.17. - №.1. – С.90.
4. Tamazian G. Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences // G. Tamazian, P. Dobrynin, K. Krasheninnikova et al. // GigaScience. – 2016. – Т.5. - №.1. – С.38.
5. Noskova E. GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data / E. Noskova, V. Ulyantsev, K.P. Koepfli, S.J. O'Brien, P. Dobrynin // GigaScience. – 2020. – Т.9. - №.3. – С.giaa005.
6. Tamazian G. Draft de novo Genome Assembly of the Elusive Jaguarundi, *Puma yagouaroundi* / G. Tamazian, P. Dobrynin, A. Zhuk et al. // Journal of Heredity. – 2021. – Т.112. - №.6. – С.540-548.
7. Fyodorova A. Genome-wide detection of positive selection in new African cheetah assembly / A. Fyodorova, A. Zhuk, P. Dobrynin // BMC Bioinformatics. – 2021. – Т.22. - №.S16.