

На правах рукописи

**Макунин Алексей Игоревич**

**Анализ генетического состава В-хромосом млекопитающих с применением  
высокопроизводительного секвенирования**

03.01.07 – Молекулярная генетика

Автореферат

диссертации на соискание ученой степени

кандидата биологических наук

**Новосибирск - 2016**

Работа выполнена в лаборатории цитогенетики животных  
Федерального государственного бюджетного учреждения науки  
Институт молекулярной и клеточной биологии СО РАН, г.  
Новосибирск

Научный руководитель:	Трифонов Владимир Александрович, кандидат биологических наук, заведующий лабораторией сравнительной геномики Федерального государственного бюджетного учреждения науки Институт молекулярной и клеточной биологии СО РАН, г. Новосибирск
Оппоненты:	Дымшиц Григорий Моисеевич, доктор биологических наук, заведующий кафедрой естественных наук Специализированного учебно-научного центра Федерального государственного автономного образовательного учреждения высшего образования "Новосибирский национальный исследовательский государственный университет", г. Новосибирск
	Кочетов Алексей Владимирович, доктор биологических наук, заместитель директора по научной работе Федерального исследовательского центра Институт цитологии и генетики СО РАН, г. Новосибирск
Ведущее учреждение:	Федеральное государственное бюджетное учреждение науки Институт цитологии РАН, г. Санкт-Петербург

Защита состоится: \_\_\_\_\_ 2016 г. в \_\_\_\_ часов на заседании диссертационного совета Д 003.074.01 по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук при ФГБУН Институт молекулярной и клеточной биологии СО РАН (630090, Новосибирск, пр. академика Лаврентьева, 8/2, тел. (383)-373-02-49, +7-952-916-7858, e-mail: [kokoza@mcb.nsc.ru](mailto:kokoza@mcb.nsc.ru) С диссертацией можно ознакомиться в библиотеке и на сайте Института молекулярной и клеточной биологии СО РАН [www.mcb.nsc.ru](http://www.mcb.nsc.ru). Автореферат разослан « \_\_\_\_ » \_\_\_\_\_ 2016 г.

Ученый секретарь  
диссертационного совета,  
кандидат биологических наук

Е. Б. Кокоза

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность.** В-хромосомы, или добавочные элементы кариотипа, обнаружены у представителей большинства основных таксонов многоклеточных эукариот. Эти хромосомы не являются необходимыми для нормальной жизнедеятельности и воспроизводства носителя, однако у некоторых организмов могут составлять значительную часть объема наследственного материала. Они характеризуются нестабильностью как в мейотических, так и в митотических делениях, и способны накапливаться от поколения к поколению с помощью различных механизмов, делающих их расхождение при клеточном делении направленно асимметричным — данное явление носит название драйва.

В-хромосомы зачастую являются гетерохроматиновыми элементами, обогащенными повторенными последовательностями. Долгое время оставались неизвестными уникальные последовательности, представленные на них. Однако развитие методов молекулярной биологии и, в особенности, высокопроизводительного секвенирования позволили выявить на добавочных хромосомах различных видов множество уникальных районов, несущих белок-кодирующие гены. Первым геном, обнаруженным на В-хромосомах млекопитающих (лисицы и енотовидной собаки), был *KIT*. В дальнейшем, список генов для этих видов был расширен, а на добавочных элементах косули были найдены три белок-кодирующих гена в составе одного протяженного района.

В отделе разнообразия и эволюции животных ИМКБ СО РАН собрана коллекция культур тканей различных видов животных, несущих В-хромосомы, а также хромосомспецифичных библиотек ДНК, в том числе и добавочных хромосом. Высокопроизводительное секвенирование данных библиотек представляется крайне интересным для понимания эволюции добавочных элементов. Однако анализ данных высокопроизводительного секвенирования ДНК отдельных хромосом не является стандартизированной процедурой и подвержен ряду ошибок, связанных с контаминацией полногеномной ДНК изучаемого вида и человека, ошибками прочтения, необходимостью работать с междвидовыми выравниваниями и т.д.

**Целью** данной работы является исследование генетического состава В-хромосом сибирской косули *Capreolus pygargus* и серого мазама *Mazama gouazoubira* с применением методов анализа данных

высокопроизводительного секвенирования (Illumina MiSeq) ДНК отдельных хромосом.

**Задачи:**

1. Разработать метод поиска уникальных геномных районов в данных высокопроизводительного секвенирования Illumina.
2. Объединить его с существующими методами поиска нуклеотидных вариантов и анализа спектров повторенных последовательностей с получением автоматизированного конвейера для воспроизводимого анализа данных.
3. Применить разработанный метод для описания состава В-хромосом сибирской косули *Capreolus pygargus* и серого мазама *Mazama gouazoubira*.
4. Проанализировать особенности эволюции добавочных хромосом этих видов.

**Научная новизна и практическая ценность.** Впервые разработан воспроизводимый и автоматизированный метод поиска уникальных геномных районов в данных высокопроизводительного секвенирования отдельных хромосом. Минимальный размер обнаруживаемых уникальных районов лежит в пределах 10-20 тыс. п.н., а отличительной особенностью метода является низкий уровень ложноположительных результатов. Дополнительные данные (нуклеотидные варианты, спектры повторенной ДНК) в данном контексте получали и ранее, однако мы впервые интегрировали получение всех этих данных в один инструмент. Показана применимость метода к анализу не только добавочных хромосом, но и межхромосомных перестроек в более общем смысле, при выравнивании прочтений на референсный геном как того же, так и родственного видов. Точность метода продемонстрирована на примере определения границ межхромосомной транслокации гена *KIT* у коровы. Разработанный метод позволяет эффективно проводить анализ на наборе данных, получаемых менее чем с 1/10 запуска Illumina MiSeq. Набор получаемых характеристик (уникальные районы и замены в них, спектры повторов) являются хорошей основой для проведения целевых исследований: разработки маркеров хромосом, уточнения координат перестроек и внутренней структуры хромосом, ассоциации районов перестроек с другими геномными характеристиками, восстановления событий хромосомной эволюции.

Впервые определен набор из 26 уникальных районов, содержащих 55 генов, на добавочных хромосомах серого мазама.

Для сибирской косули подтверждено наличие крупного района, обнаруженного ранее, уточнены его границы, а также обнаружен дополнительный мелкий район. Изучены спектры повторенных последовательностей и получены обширные наборы нуклеотидных вариантов на добавочных хромосомах обоих видов. Показано наличие двух условных типов добавочных хромосом: 1) содержащих амплифицированные и дегенерированные фрагменты генома (у косули) и 2) содержащих схожие с аутосомами фрагменты генома (у мазама). Мы предполагаем, что данные типы представляют из себя различные этапы эволюции добавочных хромосом. Впервые показано независимое включение протоонкогенов (*KIT* и *RET*) в В-хромосомы у хищных и серого мазама.

**Положения, выносимые на защиту.** С помощью разработанного метода анализа высокопроизводительного секвенирования ДНК отдельных хромосом можно выявлять уникальные геномные районы, представленные на этих хромосомах. Данный метод отличается низким уровнем ложноположительных сигналов. Добавочные хромосомы сибирской косули содержат два уникальных района общим размером 2 млн п.н. с признаками дегенерации последовательности и амплификации. В-хромосомы серого мазама содержат 26 уникальных районов общим размером 9 млн п.н. Спектрами замен и повторенных элементов эти районы схожи с аутосомными. Протоонкогены *KIT* и *RET* выявляются на добавочных хромосомах как серого мазама, так и хищных (лисицы, енотовидных собак), при этом их перенос на В-хромосомы произошел независимо.

**Апробация работы.** Результаты работы были представлены на международной конференции «Высокопроизводительное секвенирование в геномике» (Новосибирск, 21-25 июля 2013), 19 международной хромосомной конференции (Болонья, Италия, 2-6 сентября 2013 года), 3 Международной конференции по В-хромосомам (Гатерслебен, Германия, 7-9 апреля 2014) и на международной конференции «Хромосома 2015» (Новосибирск, 24-28 августа 2015).

**Вклад автора.** Автором самостоятельно выполнены разработка и применение биоинформатических методов. Конвейер для последовательного запуска программ был сконструирован совместно с Кичигиным И.Г. Молекулярно-биологический и цитогенетический блок выполнены Трифоновым В.А. и Проскуряковой А.А.

Пробоподготовка и секвенирование на Illumina MiSeq выполнено Черняевой Е.Н.

**Структура диссертации.** Диссертация состоит из введения, обзора литературы, описания материалов и методов, результатов и обсуждения, а также выводов и списка цитируемой литературы, в которые входит 274 ссылок. Работа изложена на 98 страницах машинописного текста, содержит 11 таблиц, 9 рисунков и приложение на 47 страницах.

**Публикации.** По результатам и проблематике настоящего исследования опубликовано 8 работ, из них 2 статьи в рецензируемых изданиях рекомендованного перечня ВАК.

**Благодарности.** Автор выражает благодарность сотрудникам Отдела разнообразия и эволюции геномов ИМКБ СО РАН: В.А. Трифонову, И.Г. Кичигину, Н.В. Воробьевой, А.А. Проскуряковой, А.Г. Графодатскому, зав. Лаборатории иммуногенетики ИМКБ СО РАН А.В. Таранину, зав. Лаборатории клеточного деления ИМКБ СО РАН А.В. Пиндюрину, сотруднику Лаборатории молекулярной цитогенетики ИМКБ СО РАН Ф.П. Гончарову, а также сотрудникам Центра Геномной Биоинформатики им. Добржанского СПбГУ Е.Н. Черняевой и С.Дж. О'Брайену.

Работа была поддержана грантами РФФИ 12-04-00659-а, 15-29-02384, бюджетными проектами 0310-2014-0003, 0310-2014-0008 и 0310-2014-0009, а также программой РАН “Молекулярная и Клеточная Биология”.

## **МАТЕРИАЛЫ И МЕТОДЫ**

**Методы молекулярной биологии и цитогенетики.** Культуры фибробластов сибирской косули, серого мазама, коровы и собаки были получены из коллекций Отдела разнообразия и эволюции геномов ИМКБ СО РАН и Кэмбриджского Ресурсного Центра Сравнительной Геномики. Культивирование клеток, подготовка хромосом и хромосомный сортинг были проведены в Кэмбриджском Ресурсном Центре Сравнительной Геномики (Кэмбридж, Великобритания). Сортированные хромосомы амплифицировали с помощью DOP-ПЦР с праймером CCGACTCGAGNNNNNATGTGG. Высокопроизводительное секвенирование проводили на платформе Illumina MiSeq в Центре геномной биоинформатики им. Добржанского СПбГУ. Для проверки результатов секвенирования использовали локализацию ВАС-клонов и секвенирование по Сэнгеру библиотек кДНК, обогащенных В-хромосомными транскриптами.

**Методы биоинформатики.** Разработанный метод анализа данных секвенирования отдельных хромосом охватывает характеристику:

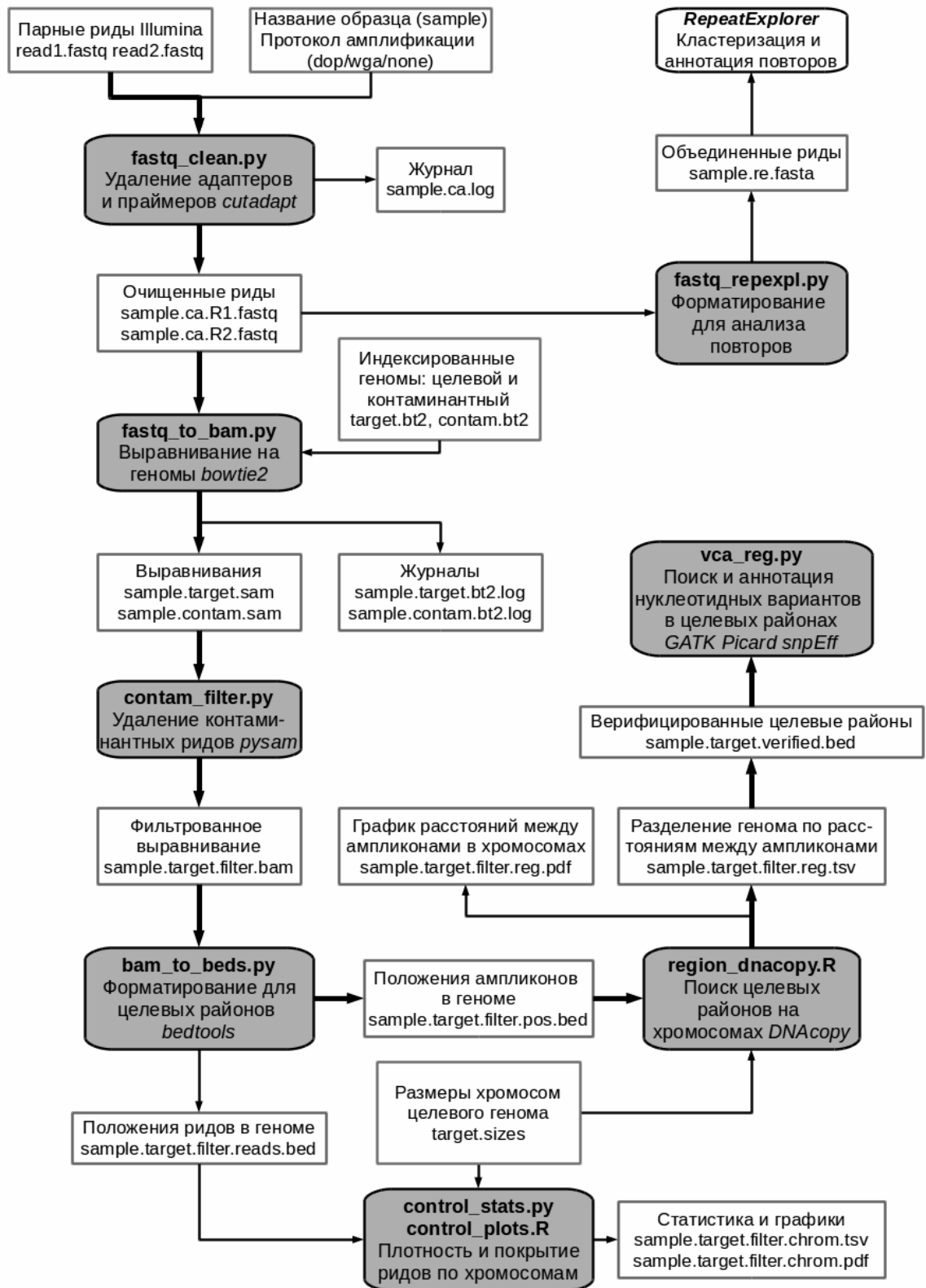
1. уникальных районов, представленных на хромосомах,
2. хромосомспецифичных нуклеотидных замен и инсерций-делеций,
3. спектров повторенных элементов.

Основные процедуры были автоматизированы с помощью конвейера `DOPseq_analyzer`, свободно доступного по интернет-адресу [https://github.com/ilyakichigin/DOPseq\\_analyzer](https://github.com/ilyakichigin/DOPseq_analyzer). Схема анализа приведена на Рисунке 1.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

**Характеристика хромосомных библиотек.** В настоящей работе были использованы две контрольные библиотеки: хромосомы 12 собаки *Canis familiaris* (CFA12) и смеси мелких хромосом коровы *Bos taurus* (BTAMix), а также три экспериментальных библиотеки: две — В-хромосом сибирской косули *Capreolus pygargus*, полученные независимыми отборами из одного эксперимента по сортированию (CPYB1, CPYB2), и библиотека В-хромосом серого мазама *Mazama gouazoubira* (MGOB). У сибирской косули и серого мазама В-хромосомы отличаются малыми размерами и высоким процентом ГЦ-пар, благодаря чему они хорошо отделялись от других хромосом при проточной цитометрии. Специфичность библиотек подтверждена методом FISH.

**Разработка метода анализа данных высокопроизводительного секвенирования хромосомспецифичных библиотек.** Ключевые требования к биоинформатическому анализу — воспроизводимость, простота использования и возможность контроля на различных этапах. Для того чтобы удовлетворить их, мы сконструировали программу-конвейер `DOPseq_analyzer` (схема представлена на Рисунке 1). С точки зрения архитектуры она является набором скриптов, написанных на языках программирования Python и R, автоматизирующих работу сторонних программ и проводящих дополнительные вычисления. Также написан конвейер, последовательно запускающий скрипты, для получения максимально полного объема данных в одну команду. Для управления конвейером используется конфигурационный файл с указанием всех необходимых параметров.



**Рисунок 1. Схема анализа данных секвенирования хромосомспецифичных библиотек.** Прямоугольники - входные и выходные файлы. Скругленные прямоугольники - процедуры: серый фон - скрипты, написанные в данной работе, курсив - используемые внешние программы и пакеты.



Для удаления последовательностей праймеров и адаптеров Illumina использовали программу cutadapt, работающую с вырожденными последовательностями ДНК. Выравнивание парных прочтений на референсный геном производили с помощью программы bowtie2, эффективно работающей со сниженной из-за дивергенции между видами и псевдогенизации гомологией. Для удобства поиска целевых районов повторно картированные прочтения отбрасывались. Для оценки и удаления загрязнения человеческой ДНК одновременно с выравниванием на целевой геном (собаки для CFA12, коровы для остальных библиотек) проводили выравнивание на геном человека, отбрасывая прочтения с более высоким качеством картирования на геном человека и с низким качеством выравнивания на целевой геном. В прочтениях сортированных хромосом содержатся несколько процентов ДНК человека, а процент целевых выровненных прочтений колеблется от 12,7 (CPYB2) до 49,3 (BTAMix) %.

Для поиска целевых районов, представленных на хромосомах, перекрывающиеся позиции картирования прочтений интерпретировали как положения ампликонов. Идентификация целевых хромосом для контрольных образцов CFA12 и BTAMix стала возможна уже после изучения распределения по хромосомам референсных геномов расстояний между ампликонами (средние значения на целевых хромосомах заметно ниже, чем на нецелевых) и их покрытия ридами (на целевых хромосомах повышается разброс значений). Таким образом, уже на этом этапе мы смогли определить состав BTAMix: туда входили хромосомы 23, 26, 28 и 29 коровы.

Для автоматизации поиска целевых районов мы использовали пакет DNAsoru для языка программирования R, созданный для поиска дубликаций и делеций в данных сравнительной геномной гибридизации на микрочипах. На его основе был написан скрипт region\_dnascoru.R, разделяющий заданные хромосомы целевого генома по средним значениям расстояний между ампликонами: низкие значения характерны для целевых районов (присутствующих на хромосоме из образца), высокие — для нецелевых (контаминации другими хромосомами).

Целевые и нецелевые районы легко разделялись для контрольных образцов: средние расстояния между ампликонами составляли 3,5-6,7 тыс. п.н. против 11,7 тыс. п.н. для CFA12, 3,6-22 тыс. п.н. против 80 тыс. п.н. для BTAMix. После определения целевых районов необходимо проверять и корректировать точные границы перестроек.

В ходе анализа было выявлено несколько ограничений нашего метода поиска районов, представленных на хромосомах: точность определения границ перестроек ограничена средним расстоянием между ампликонами DOP-ПЦР (до 5-10 тыс. п.н.); ампликоны внутри целевых районов распределены неравномерно; нет возможности определить взаимное расположение, ориентацию и копияность целевых районов на хромосоме. Эти проблемы отчасти могут быть решены полногеномной амплификацией, основанной на использовании ф29 ДНК-полимеразы и случайного гексапраймера, обеспечивающей более равномерное покрытие генома.

Полученные данные высокопроизводительного секвенирования позволяют провести поиск и аннотацию вариантов (замен, коротких инсерций и делеций), специфичных для хромосом. Идентификацию нуклеотидных вариантов проводили с помощью программы GATK HaplotypeCaller, после чего с помощью программы GATK SelectVariants выбирали варианты, принадлежащие целевым районам. Полученные наборы вариантов аннотировали по генам из базы данных snpEff (основана на аннотациях Ensembl) либо по генам RefSeq с помощью инструмента Variant Annotation Integrator (для генома коровы). Мы протестировали фильтр по глубине покрытия, позволяющий отсеять ошибки секвенирования, интерпретируемые как варианты при низком покрытии, на данных библиотеки BTAMix. Доля известных полиморфизмов, внесенных в базу данных dbSNP 138, составляла ~65% вне зависимости от выбранного значения минимального покрытия (3, 4 или 6). Таким образом, данный фильтр оказался неэффективным.

Были исследованы общие тенденции распределения нуклеотидных вариантов: уровни гетерозиготности, частоты замен в кодирующих участках, соотношения частот синонимичных и несинонимичных замен. Индивидуальные замены после дополнительной верификации можно использовать как доказательства псевдогенизации на В-хромосомах (например, нонсенс-мутации), а также как потенциальные маркеры В-хромосом (замены, характерные для В-хромосом, но не для основного генома).

Анализ состава повторенной ДНК проводили непосредственно на ридах с помощью инструмента RepeatExplorer. Данная программа производит кластеризацию прочтений по перекрытию, аннотацию по базе данных повторов RepeatMasker для млекопитающих, а также сборку консенсусных последовательностей кластеров. Для библиотеки

ВТАМix было проведено сравнение результатов этого метода с аннотацией повторов в сборке генома коровы. Были выявлены значительные различия: в геномной сборке отсутствует центромерный сателлит, а в данных секвенирования многие повторы оказываются неаннотированными. Кроме того, меняются соотношения различных типов повторов. Характерным для данных секвенирования является присутствие широко представленного в геноме человека семейства Alu. Таким образом, корректное сравнение возможно только между результатами RepeatExplorer для образцов, полученных с помощью одного и того же типа амплификации.

**Обнаружение межхромосомной перестройки KIT у коровы.** При анализе смешанных хромосом 23, 26, 28 и 29 коровы был выявлен дополнительный район референсного генома ВТА6:72,525,912-73,007,603, содержащий полный ген *KIT*. В литературе описана перестройка с практически идентичными координатами и, вероятно, кольцевым интермедиатом, ассоциированная со специфической окраской — белой полосой вдоль хребта (color sidedness).

**Уникальные районы В-хромосом.** Поиск уникальных геномных районов с помощью DOPseq\_analyzer выявил на добавочных хромосомах косули два участка с тремя генами общей протяженностью 1,42-1,98 млн п.н., а на добавочных хромосомах мазама — 26 участков с 55 генами общей протяженностью 8,28-9,31 млн п.н. (Таблица 1). Разброс в оценках размеров связан с вероятными делециями внутри целевых районов.

**Таблица 1. Уникальные районы и гены RefSeq генома коровы (bosTau7), обнаруженные на добавочных хромосомах сибирской косули и серого мазама. (f) — фрагменты генов на границе районов, (f?) — гены, не покрытые прочтениями (вероятные делеции), кДНК — подтверждение прочтением клонов В-хромосомспецифичной кДНК, ВАС — подтверждение локализацией ВАС-клонов. Жирным выделены общие В-хромосомные районы серого мазама и хищных: лисицы и/или енотовидной собаки.**

<b>Сибирская косуля (<i>Capreolus pygargus</i>)</b>			
Хромосома: млн п.н	Гены	кДНК	ВАС
ВТА3:74,548-76,487	<i>LRRIQ3, FPGT, TNNT3K</i> (f)	+	+
ВТА28:11,360-11,401	-		

Таблица 1. Продолжение

Серый мазама( <i>Mazama gouazoubira</i> )			
Хромосома: млн п.н	Гены	кДНК	ВАС
ВТА1:80,558-80,816	<i>LPP</i> (f)		
ВТА1:132,196-132,249	<i>PIK3CB</i> (f)		
ВТА3:53,940-54,036	<i>GFII</i>		
ВТА3:97,346-97,530	<i>ACOT11</i> (f), <i>SSBP3</i> (f)		
ВТА5:112,038-112,093	<i>CCND2</i>		
ВТА6:67,475-69,303	<i>COX7B2</i> , <i>GABRA4</i> , <i>GABRB1</i> (f?), <i>LOC536190</i> (f?), <i>ATP10D</i> , <i>NFXL1</i> , <i>CNGA1</i> , <i>NIPAL1</i> , <i>TXK</i>	+	
<b>ВТА6:72,726-72,802</b>	<b><i>KIT</i> (f)</b>		+
ВТА6:101,763-101,983	<i>COQ2</i> , <i>HPSE</i> , <i>MIR2446</i> , <i>MRPS18C</i> , <i>FAM175A</i> (f)		
ВТА6:115,148-115,688	<i>MIR2448</i> , <i>FBXL5</i> , <i>CD38</i> (f?), <i>FGFBP1</i>		
ВТА7:11,113-11,188	<i>MAN2B1</i> (f)		
ВТА7:21,377-21,646	-	+	
ВТА8:0,078-0,782	<i>ANXA10</i> , <i>MIR2466</i>		
ВТА10:28,364-29,133	<i>TMCO5B</i>	+	+
ВТА14:5,824-6,377	<i>KHDRBS3</i> (f)		
ВТА17:46,546-46,775	<i>MBD3L1</i> , <i>CHFR</i> , <i>ZNF268</i> (f)		
ВТА19:59,130-59,367	<i>CDC42EP4</i> , <i>FAM104A</i> , <i>COG1</i> (f)		
ВТА22:1,766-1,921	<i>EOMES</i> (f)		
ВТА22:11,626-11,985	<i>ACAA1</i> , <i>MYD88</i> , <i>OXSRI</i> , <i>XYLB</i> , <i>ACVR2B</i> , <i>DLEC1</i> (f)		
ВТА23:8,994-9,084	<i>C23H6orf106</i> (f), <i>SNRPC</i> (f)	+	
ВТА23:50,678-50,701	<i>SERPINB9</i>		
ВТА25:37,085-37,276	-		
ВТА25:41,480-41,788	-		

Таблица 1. Продолжение

<b>Серый мазама (<i>Mazama gouazoubira</i>)</b>			
Хромосома: млн п.н	Гены	кДНК	ВАС
ВТА28:0,628-0,915	<i>TRIM67, FAM89A, ARV1, TTC13</i> (f)		
<b>ВТА28:12,111-12,175</b>	<b><i>RET</i></b>		
ВТА29:34,729-35,503	<i>OPCML</i> (f)		
ВТА29:49,339-50,253	<i>DHCR7</i> (f?), <i>NADSYN1</i> , <i>CARS</i> (f)	+	+

В настоящем исследовании мы впервые обнаружили 26 уникальных районов, представленных на В-хромосомах серого мазама. Их размеры лежат в диапазоне от 23 до 1827 тыс. п.н. Полученные ранее данные секвенирования кДНК, обогащенной В-хромосомными транскриптами, хорошо согласуются с данными высокопроизводительного секвенирования. Из 17 секвенированных клонов, 5 оказались повторенными, 3 происходили из района ВТА29:49,339-50,253 млн п.н., по 2 — из ВТА7:21,377-21,646 млн п.н. и ВТА10:28,364-29,133 млн п.н., 1 клон из ВТА23:8,994-9,084 млн п.н., 1 — из неспецифического района, а еще два не обнаружили гомологии в геноме коровы. Оставшийся клон происходил из внехромосомного скэффолда Un\_JH125058 размером 65 тыс. п.н., по данным гомологии с геномом человека принадлежащего уже известному на В-хромосомах мазама району ВТА6:115,148-115,688 млн п.н. Данные по локализации ВАС-клонов также не вступают в противоречие с результатами секвенирования (Таблица 1).

На В-хромосомах косули было обнаружено всего два района общей протяженностью не более 2 млн п. н. Большой из них (ВТА3:74,55-76,49 млн п.н.) был ранее идентифицирован при секвенировании библиотек кДНК, обогащенных В-специфичными последовательностями и подтвержден ПЦР-картированием и локализацией ВАС-клонов. Полученные нами данные совпадают с полученными ранее, однако наблюдаются два участка, не покрытых ридами (ВТА3:74,88-75,30 млн п.н. и 75,68-75,82 млн п.н.). При этом один из пробелов имеет подтверждение локализацией ВАС-клона и может быть интерпретирован как ложноотрицательный результат метода, а второй по большей части занят пропуском в сборке генома коровы.

**Гены В-хромосом и обнаружение протоонкогенов *KIT* и *RET* у серого мазама и хищных.** Состав генов на В-хромосомах сибирской косули не изменился относительно предыдущего исследования — все данные указывают на то, что присутствуют гены *LRRIQ3*, *FPGT*, а также начало гена *TNNI3K*. Второй В-хромосомный район размером 42 тыс. п.н. расположен в богатом генами районе между генами *CHRM3* и *ZNF33B*, однако не включает их.

Генный состав добавочных хромосом мазама гораздо более разнообразен и включает 55 генов, из которых 18 попадают на границы целевых районов и входят в них лишь частично, а еще 4 попадают в вероятные делеции.

Наиболее примечательным оказалось присутствие на В-хромосомах серого мазама нескольких протоонкогенов, обнаруженных ранее на добавочных элементах хищных: лисицы и двух подвидов енотовидной собаки. Первый из них, *KIT*, состоит из 21 экзона и кодирует рецепторную тирозин-киназу, активную в дифференциации меланоцитов, кроветворных и предшественников зародышевых клеток. Мутации в этом гене активируют канцерогенез. Среди млекопитающих распространены специфические формы окраски (зачастую — белая пятнистость), ассоциированные с мутациями в этом гене. В описанном выше случае межхромосомной транслокации у коровы размер переносимого фрагмента отличается от размеров участков на В-хромосомах в пределах одного порядка — 580 тыс. п.н. по сравнению 76 тыс. п.н. на В-хромосомах мазама и 202 тыс. п.н. на В-хромосомах лисицы. Однако если у коровы ген переносится на другую хромосому целиком, то на В-хромосомах серого мазама наблюдаются только экзоны 1-20 (также обнаруживается тринуклеотидная вставка в 10 экзоне), а на В-хромосомах лисицы — экзоны 2-21. Кроме того, у енотовидных собак на В-хромосомах представлен фрагмент соседнего протоонкогена *KDR*, не обнаруживаемого у серого мазама и лисицы. Локализация ВАС-клона собаки, ранее использованного для исследования В-хромосом хищных, подтвердила наличие гена *KIT* на В-хромосомах мазама.

Другой протоонкоген, независимо обнаруживаемый у мазама и хищного (китайской енотовидной собаки), — *RET*. Обогащение ассоциированными с раком генами является характерной особенностью В-хромосом хищных — ранее обнаружены гены *LRIG1*, *LRP1B* и *CTNND2*. На В-хромосомах мазама также обнаруживаются

связанные с онкогенезом гены *PIK3CB*, *CCND2* и *OXSRI*. Анализ функционального обогащения по базе данных Gene Ontology с помощью инструмента DAVID выявляет лишь незначительное обогащение функциями эмбрионального развития, генами киназ, и позитивных регуляторов киназной активности. Связь между функцией генов и их переносом на В-хромосомы работает далеко не для всех случаев: так, лишь 5 из 26 районов несут протоонкогены и супрессоры опухолей, 5 — гены киназ и их регуляторов, а для остальных функциональных групп количество районов оказывается еще меньше.

**Нуклеотидные варианты В-хромосом.** Спектр нуклеотидных вариантов (замен и коротких инсерций/делеций), характерных для В-хромосом сибирской косули и не встречающихся в геноме европейской косули заметно отличается от такового для аутосом коровы. Так, плотность вариантов в три раза выше, чем у коровы (1 вариант на 96 п.н. против 1 на 316 п.н.), при этом большинство вариантов в кодирующих областях являются несинонимичными, что свидетельствует о сниженном давлении отбора. Уровень гетерозиготности крайне высок. Неоднородность может существовать как между хромосомами (в культуре клеток, использованной для сортировки, было 8 В-хромосом), так и между копиями внутри отдельной хромосомы (наличие нескольких копий фрагментов уникального района на В-хромосоме было показано ранее методом ПЦР в реальном времени).

Для В-хромосом серого мазама соотношение между синонимичными и несинонимичными заменами было близко к значениям для аутосом коровы, а плотность замен в кодирующих областях — в 1,65 раз ниже, чем у В-хромосом сибирской косули, что свидетельствует о меньшей степени псевдогенизации В-хромосом серого мазама по сравнению с В-хромосомами сибирской косули. Плотность гетерозиготных вариантов на В-хромосомах мазама была в 6,5 раз ниже, чем на В-хромосомах косули (768 п.н. на гетерозиготную позицию против 118 п.н.), и всего лишь незначительно выше, чем на аутосомах коровы (952 п.н.). Это может свидетельствовать о низком разнообразии добавочных хромосом у мазама или о незначительной степени амплификации уникальных последовательностей на В-хромосомах. В пользу второго объяснения говорит точечный характер локализации ВАС-клона, содержащего ген *KIT*, а также общий размер уникальных последовательностей на В-хромосомах — около 9 млн п.н. у мазама и менее 2 млн п.н. у косули. Впрочем, локализация

других ВАС-клонов на хромосомах мазама демонстрирует наличие повторенных сигналов. Кроме того, различие в уровне гетерозиготности может быть связано и с количеством В-хромосом в исследованных культурах клеток (3 у мазама и 8 у косули).

**Повторенная ДНК В-хромосом.** Суммарная доля повторенной ДНК ниже всего в аутосомах коровы (42,73%), в полтора раза выше в В-хромосомах мазама (63,23%) и максимальна в В-хромосомах косули (77,76 и 79,45% в СРУВ1 и СРУВ2). Число кластеров, включающих более 1% прочтений, выше для добавочных хромосом. У косули наиболее представленные кластеры содержат в основном неаннотированные повторы (повторы с низкой сложностью, накопившие большое количество замен). Далее следуют кластеры, содержащие сателлитную, в том числе и центромерную, ДНК. По сравнению с аутосомами коровы резко снижена доля SINE. У серого мазама процент неаннотированных повторов заметно повышен относительно коровы, однако полностью неаннотированные кластеры не являются наиболее крупными. Высокое разнообразие LINE и SINE на В-хромосомах мазама сближает их с аутосомами коровы, но с некоторыми особенностями. Обнаруживается двукратное обогащение LINE L1, повышение доли SINE BovA и снижение доли SINE RTE-BovB. Особенно выделяется кластер, полностью состоящий из эндоретровируса ERVK с консервативной последовательностью — вероятный след недавней ретровирусной инфекции.

**Сравнение В-хромосом сибирской косули и серого мазама.** В-хромосомы мазама находятся в более ранней эволюционной стадии — из фрагментов образовалась В-хромосома, но процессы, связанные с отсутствием отбора еще не продвинулись. В-хромосомы косули демонстрируют уже протекшие процессы: накопление замен, повторов, амплификация и, вероятно, утрата исходных фрагментов.

Поскольку среди косуль В-хромосомы обнаружены только у сибирского, но не у европейского вида, наиболее вероятным представляется возникновение добавочных элементов после расхождения видов 2-3 млн лет назад. В роде *Mazama* В-хромосомы представлены в большинстве видов и, вполне возможно, имеют общее происхождение. При этом наиболее базальная дивергенция видов с В-хромосомами произошла ~5 млн лет назад. Таким образом, В-хромосомы серого мазама являются более древними, но консервативными, а В-хромосомы сибирской косули — более молодыми и измененными, что свидетельствует о неоднородности



скорости и путей эволюции В-хромосом даже внутри отдельного подсемейства млекопитающих.

### ВЫВОДЫ

1. Разработан метод биоинформатического анализа данных высокопроизводительного секвенирования отдельных хромосом, позволяющий, основываясь на выравнивании на референсный геном того же или родственного вида, идентифицировать уникальные районы размером от 10-20 тыс. п.н. в составе хромосом. Данный метод характеризуется низким уровнем ложноположительных результатов. Впервые сконструирован автоматизированный конвейер, объединяющий поиск уникальных участков, характеристику замен в них и описание спектров повторенных последовательностей.
2. Применение данного конвейера к данным Illumina MiSeq секвенирования библиотек сортированных В-хромосом оленей показало, что добавочные хромосомы сибирской косули (*Capreolus pygargus*) содержат два уникальных района общим размером 2 млн п.н. с 3 генами и характеризуются вырожденностью последовательности вкуче с амплификацией.
3. Впервые показано, что добавочные хромосомы серого мазама (*Mazama gouazoubira*) содержат 26 уникальных районов (9 млн п.н.) с 55 полными и фрагментированными генами, а по характеру замен и спектру повторенных последовательностей близки к аутосомам.
4. Отсутствие общих уникальных районов на В-хромосомах косули и мазама свидетельствует о независимом происхождении В-хромосом этих двух видов. Различия в характеристиках последовательностей позволяют заключить, что В-хромосомы мазама находятся на более ранней стадии эволюции, а В-хромосомы косули — на более поздней.
5. Повторное обнаружение протоонкогенов *KIT* и *RET* на В-хромосомах у хищных и серого мазама иллюстрирует тенденцию к предпочтительному включению некоторых геномных районов в состав добавочных хромосом.

### ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Макунин А.И., Трифонов В.А. Современные методы анализа добавочных хромосом // Цитология. 2013. Т. 55. № 3. С. 148–152.

2. **Makunin A.I.**, Dementyeva P.V., Graphodatsky A.S., Volobouev V.T., Kukekova A.V., Trifonov V.A. Genes on B chromosomes of vertebrates // *Molecular Cytogenetics*. 2014. V. 7. N. 1. P. 99.

3. **Makunin A.I.**, Kichigin I.G., Larkin D.M., O'Brien P.M., Ferguson-Smith M.A., Yang F., Proskuryakova A.A., Vorobieva N.V., Chernyaeva E.N., O'Brien S.J., Graphodatsky A.S., Trifonov V.A. Contrasting origin of B chromosomes in two cervids (Siberian roe deer and grey brocket deer) unraveled by chromosome-specific DNA sequencing // *BMC Genomics* (submitted)

4. Кичигин И.Г., Джиованотти М., Кабилов М.Р., Тупикин А.Е., Фергюсон-Смит М.А., **Макунин А.И.**, Графодатский А.С., Трифонов В.А. Эволюция половых хромосом в роде анолисов (*Anolis*) // *Материалы международной конференции Хромосома 2015*. Новосибирск, 2015. С. 103–104.

5. **Макунин А.И.**, Кичигин И.Г., Ларкин Д.М., О'Брайен П.С.М., Фергюсон-Смит М.А., Янг Ф., Проскурякова А.А., Воробьева Н.В., Черняева Е.Н., О'Брайен С.Д., и др. Секвенирование библиотек сортированных В хромосом для исследования их происхождения и эволюции // *Материалы международной конференции Хромосома 2015*. Новосибирск, 2015. С. 122–123.

6. **Макунин А.И.**, Черняева Е.Н., О'Брайен С.Д., Трифонов В.А. Высокопроизводительное секвенирование добавочных хромосом млекопитающих // *Тезисы международной конференции «Высокопроизводительное секвенирование в геномике»*. Новосибирск, 2013. С. 19.

7. Трифонов В.А., Тупикин А.Е., Кичигин И.Г., **Макунин А.И.**, Джиованотти М., Кабилов М.Р. Изучение эволюции половых хромосом рептилий с помощью высокопроизводительного секвенирования хромосомспецифичных библиотек // *Тезисы международной конференции «Высокопроизводительное секвенирование в геномике»*. Новосибирск, 2013. С. 25.

8. **Makunin A.**, Chernyaeva E., Trifonov V. Discovery of unique regions on B chromosomes in mammals // *Program and abstract of the 3rd B-Chromosome Conference Gatersleben (Germany)*. Gatersleben, Germany, 2014. P. 28.

9. Trifonov V., Dementyeva P., **Makunin A.**, Larkin D., O'Brien P.M., Perelman P., Yang F., Ferguson-Smith M., Graphodatsky A. Amplification of genes on mammalian B chromosomes // *The 19th International Chromosome Conference*. Bologna, Italy, 2013. P. 167.