

На правах рукописи



ПРОКОПОВ ДМИТРИЙ ЮРЬЕВИЧ

**ЭВОЛЮЦИЯ ПОВТОРЯЮЩИХСЯ ЭЛЕМЕНТОВ ДНК И
ПОЛОВЫХ ХРОМОСОМ ОСЕТРОВЫХ**

03.01.07 – молекулярная генетика

Автореферат

диссертации на соискание ученой степени

кандидата биологических наук

Новосибирск 2022

Работа выполнена в лаборатории сравнительной геномики Федерального государственного бюджетного учреждения науки Института молекулярной и клеточной биологии Сибирского отделения Российской академии наук, г. Новосибирск

Научный руководитель: **Трифонов Владимир Александрович**, доктор биологических наук, заведующий лабораторией сравнительной геномики Федерального государственного бюджетного учреждения науки Института молекулярной и клеточной биологии Сибирского отделения Российской академии наук, г. Новосибирск

Оппоненты: **Подгорная Ольга Игоревна**, доктор биологических наук, профессор, главный научный сотрудник Федерального государственного бюджетного учреждения науки Института цитологии Российской академии наук, г. Санкт-Петербург
Кабилев Марсель Расимович, кандидат биологических наук, старший научный сотрудник Федерального государственного бюджетного учреждения науки Института химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, г. Новосибирск

Ведущее учреждение: Федеральное государственное бюджетное учреждение науки Институт биологии развития имени Н. К. Кольцова Российской академии наук, г. Москва.

Защита состоится: 15 июня 2022 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 Институте молекулярной и клеточной биологии СО РАН (630090, г. Новосибирск, пр. академика Лаврентьева, 8/2, тел. (383)-373-02-49, e-mail: ovant@mcb.nsc.ru С диссертацией можно ознакомиться в библиотеке и на сайте Института молекулярной и клеточной биологии СО РАН https://www.mcb.nsc.ru/diss_council/autoref

Автореферат разослан «___» _____ 2022 г.

Ученый секретарь
диссертационного совета,
кандидат биологических наук



О.В. Антоненко

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность. Осетрообразные (Acipenseriformes) представляют собой один из базальных отрядов лучепёрых рыб и считаются «живыми ископаемыми». Разрушение среды обитания, строительство плотин, загрязнение рек и браконьерство ввиду высокой стоимости икры привели к сокращению численности осетровых рыб, в результате большинство видов оказались под угрозой исчезновения. Восстановлению диких популяций может способствовать развитие аквакультуры осетровых рыб, которое имеет и большое экономическое значение. Для успешного воспроизводства осетровых необходимо накопление фундаментальных знаний о молекулярно-генетических особенностях этого таксона.

Помимо двух раундов полногеномных дупликаций, которые произошли на ранних этапах эволюции позвоночных, геномы осетровых подверглись ряду дополнительных дупликаций, что было показано на основании цитогенетических и молекулярно-генетических данных и оценок размера генома. Многие детали процесса дупликации геномов осетровых остаются неясными, в частности, были ли это события авто- или аллополиплоидии.

Процессы спонтанной полиплоидизации продолжают наблюдаться в некоторых линиях осетровых, что делает представителей этого отряда перспективной моделью для исследования явлений удвоения генома, а виды, имеющие геномы с низкой ploidy (например, стерлядь, $2n=120$), могут использоваться в качестве референса в геномных исследованиях. Другие виды (например, сибирский осётр, $2n=240$) являются хорошими моделями для исследования последствий недавней полногеномной дупликации. Необходимая для отряда стандартизация кариотипов осложняется большим числом хромосом с преобладанием мелких элементов (микрохромосом).

Раннее определение пола важно для коммерчески ценных видов осетровых. Определение пола происходит генетически, однако осетровые обладают недифференцированными и не диагностируемыми микроскопически половыми хромосомами. Специфичный для пола маркер, подтверждающий гетерогаметность самок и позволяющий проводить генетические исследования, не был известен на момент начала представляемой работы.

Повторяющиеся последовательности генома относительно быстро эволюционируют, что делает их перспективным инструментом для исследования отличий в паралогичных хромосомах и поиска половых хромосом, особенно в медленно эволюционирующих геномах осетровых. Лишь малое число

повторяющихся элементов к настоящему времени было охарактеризовано в этом таксоне. Высокопроизводительное секвенирование и анализ сборки генома стерляди в качестве референсного вида является крайне перспективным подходом для описания повторяющихся элементов и поиска полового маркера.

Цель и задачи исследования. Целью данной работы является исследование распределения в геноме, эволюции и активности повторяющихся элементов осетровых, а также их связи с половыми хромосомами.

Для достижения данной цели были поставлены следующие задачи:

1. Идентифицировать тандемноорганизованные повторяющиеся последовательности (ТПП) и проанализировать их распределение в геноме стерляди.

2. Локализовать ТПП стерляди на метафазных хромосомах сибирского осетра.

3. Разработать молекулярно-цитогенетические маркеры для идентификации отдельных хромосом и их паралогов в кариотипах осетровых с разным уровнем полиплоидизации геномов — стерляди и сибирского осетра.

4. Идентифицировать рассеянные повторяющиеся элементы в сборке генома стерляди, проанализировать их распределение, экспрессию и проследить историю экспансии на паралогичных хромосомах.

5. Выявить полоспецифичный маркер, используя данные секвенирования самцов и самок стерляди, и проследить его консерватизм в геномах других видов осетровых.

Научная новизна, теоретическая и практическая ценность.

Результаты данной работы способствуют расширению фундаментальных знаний о структуре геномов осетровых, в т.ч. о составе и эволюции повторяющихся элементов ДНК. Впервые подробно охарактеризована быстро эволюционирующая фракция консервативного генома стерляди. С помощью методов высокопроизводительного секвенирования обнаружены наиболее представленные тандемноорганизованные последовательности в геноме стерляди. Картирование методом FISH показало, что они неравномерно распределены в геноме с тенденцией к кластеризации на определенных хромосомах.

Создана физическая карта хромосом сибирского осетра с молекулярными маркерами. С помощью выявленных в геноме стерляди тандемноорганизованных последовательностей было обнаружено, что последнее событие дупликации генома у сибирского

осетра сопровождалось экспансией и редукцией некоторых семейств повторяющейся ДНК.

Показано, что геном стерляди содержит 40,3% повторяющихся элементов, по большей части рассеянных. Предположено автополиплоидное происхождение генома стерляди. Обнаружена экспрессия некоторых мобильных генетических элементов (МГЭ) и показан разный уровень их экспрессии в разных тканях.

Подтверждена ZW-система определения пола у осетровых и выявлена специфичная для самок последовательность длиной 16 т.п.н. (тысяч пар нуклеотидов) Разработан тест для ПЦР-генотипирования, дающий специфические для самок продукты у шести видов осетровых.

Положения, выносимые на защиту. Тандемноорганизованные последовательности неравномерно распределены в геноме стерляди с тенденцией к кластеризации в определенных областях. Некоторые из сателлитных ДНК позволяют отличать отдельные хромосомы от их паралогов и дают возможность однозначно идентифицировать 20 пар хромосом в кариотипе стерляди.

Последнее событие дупликации генома сибирского осетра сопровождалось экспансией и редукцией некоторых повторяющихся элементов.

Содержание МГЭ и анализ их экспансии на паралогичных хромосомах указывает на автополиплоидное происхождение генома стерляди. Некоторые МГЭ активно экспрессируются в разных тканях, а экспрессия других подавлена.

Геномы самок содержат специфичный маркер пола, а разработанный тест ПЦР-генотипирования, позволяет амплифицировать специфические для самок продукты у шести видов осетровых.

Степень достоверности и апробация результатов. Научные положения и выводы полностью обоснованы. Результаты работы были представлены на международной мини-конференции «Chromosomes and Mitosis» (5 Ноября, 2016, Новосибирск), 21-ой международной конференции «Chromosome» (10-13 июля 2016, Фоз ди Игуасу, Бразилия), всероссийской конференции с международным участием «Водные экосистемы Сибири и перспективы их использования» (22–24 ноября 2016, Томск), VII международной школе молодых учёных по молекулярной генетике «Геномика и биология живых систем» (14-18 ноября 2016, Звенигород), 54-й Международной научной студенческой конференции (6-20 апреля 2016, Новосибирск), XXIII Международной конференции студентов и молодых учёных «Ломоносов» (11-15 апреля 2016, Москва), пятой

научно-практической конференции молодых ученых ФГУП «ВНИРО» с международным участием «Современные проблемы и перспективы рыбохозяйственного комплекса» (17-18 апреля, Москва, 2017), 11-ой Европейской Цитогенетической Конференции (1-4 июля 2017, Флоренция, Италия), Второй всероссийской конференции с международным участием «Высокопроизводительное секвенирование в геномике» (18-23 июня 2017, Новосибирск), международной конференции «Bioinformatics: from Algorithms to Applications» (16-19 июля 2018, Санкт-Петербург), международной конференции «Хромосома 2018» (20-24 августа 2018, Новосибирск), VIII международной школы молодых учёных по молекулярной генетике «Геномика и биология живых систем» (19-23 ноября 2018, Звенигород), XXVI Международной конференции студентов и молодых учёных «Ломоносов» (8-12 апреля 2019, Москва), международной конференции «Smart Bio» (02-04 мая 2019, Каунас, Литва), международной мини-конференции «Chromosomes and Mitosis» (21 ноября, 2019, Новосибирск), молодежном научном форуме «ЛОМОНОСОВ-2020» (10-27 ноября 2020, Москва).

Вклад автора. Автором выполнена вся экспериментальная работа, включающая приготовление зондов и локализацию сателлитной ДНК и хромосомоспецифичных библиотек с помощью FISH. Автором также проведён биоинформатический поиск повторяющихся последовательностей в сборке генома стерляди и изучение их экспрессии. Поиск tandemноорганизованных последовательностей на сырых прочтениях был проведён совместно с к.б.н. А.С. Комиссаровым (Национальный исследовательский университет ИТМО, Санкт-Петербург). Стандартизация кариотипа стерляди и сибирского осетра с помощью tandemноорганизованных маркёров и хромосомного пэинтинга была проведена совместно с к.б.н. Л.С. Билтуевой (ИМКБ СО РАН, Новосибирск). Анализ полосспецифичного маркёра осетровых был проведён совместно с Х. Кулем (Институт пресноводной экологии и внутреннего рыболовства Ассоциации Лейбница, Берлин). Обработка, анализ, интерпретация данных и подготовка публикаций были выполнены при участии автора.

Структура диссертации. Диссертация состоит из введения, обзора литературы, описания материалов и методов, результатов, обсуждения, а также выводов и списка цитируемой литературы, в который входит 250 ссылок. Работа изложена на 124 страницах машинописного текста, содержит 8 таблиц и 31 рисунок.

Публикации. По результатам и проблематике настоящего исследования опубликовано 4 работы в рецензируемых изданиях рекомендованного перечня ВАК.

СОДЕРЖАНИЕ РАБОТЫ ОБЗОР ЛИТЕРАТУРЫ

Обзор литературы включает пять разделов и заключение. Подробно описаны подходы, которые применяются для секвенирования геномов рыб. Охарактеризованы полногеномные дубликации в геномах лучепёрых рыб. Рассмотрены системы определения пола позвоночных, в т.ч. рыб. Большая часть обзора литературы посвящена повторяющимся последовательностям генома с акцентом на описание организации таких элементов в геномах лучепёрых рыб. Изложены современные представления об эволюции и структуре геномов осетрообразных.

МАТЕРИАЛЫ И МЕТОДЫ

Культуры клеток, использованные в исследовании, и методы классической и молекулярной цитогенетики. В работе были использованы фиксированные суспензии метафазных хромосом стерляди и сибирского осетра. Первичные культуры фибробластов были получены из коллекций Отдела разнообразия и эволюции геномов ИМКБ СО РАН. Использованы следующие методы классической и молекулярной цитогенетики: GTG-окрашивание, выявление гетерохроматиновых AT- и GC-богатых блоков после G-окрашивания, микродиссекция хромосом, флуоресцентная гибридизация *in situ*, сравнительная геномная гибридизация.

Методы молекулярной биологии. В работе использованы стандартные методы молекулярной биологии: выделение ДНК фенол-хлороформным методом и ПЦР. Подготовку библиотеки для секвенирования на платформе Illumina проводили согласно протоколу производителя. Амплификацию и мечение микродиссекционных проб проводили с использованием наборов Whole Genome Amplification (Sigma) в соответствии с протоколом производителя.

Методы биоинформатического анализа. Для идентификации повторяющихся последовательностей использовали инструменты для поиска *de novo*, такие как: Lyrebird, RepeatExplorer2, TAREAN, RepeatModeler. Для поиска *ab initio* использовали инструменты LTR_FINDER, LTRharvest и LTR-retriever. Для поиска по гомологии был использован программный пакет RepeatMasker. Для аннотации повторяющихся последовательностей использовали следующие базы данных: RepeatPeps, RepBase, FishTEDB, SWISS-PROT, REXdb. Для кластеризации последовательностей использовали CD-HIT-est. Для

попарного выравнивания использовали blastn и blastx. Подсчёт k-меров проводили с помощью Jellyfish. Среднюю длину теломер подсчитывали инструментом Telomerecat. Прочтения обрезали с использованием fastp. Прочтения выравнивали на геномную сборку с использованием bowtie2 и Minimap2. Данные выравнивания преобразовывали с использованием SAMtools. Поиск нуклеотидных вариантов проводили в Platypus. Манипуляции с геномными координатами производили с помощью BEDtools. Сборку контигов проводили в IDBA.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1. Геномная организация tandemной повторяющейся ДНК стерляди

1.1. Идентификация и аннотация tandemноорганизованной ДНК стерляди

Геномы двух особей стерляди разного пола секвенировали на платформе Illumina HiSeq 2000. Прочтения кластеризовали и аннотировали с помощью программного инструмента Lyrebird для поиска tandemноорганизованных повторяющихся последовательностей (ТПП).

Были идентифицированы 13 самых представленных микросателлитных последовательностей в геноме стерляди. Содержание микросателлита ААТ оказалось больше в геноме самки, чем у самца.

Были обнаружены ТПП с большей длиной мономера (мини- и макросателлиты): Arut40A, Arut434A, Arut219A, Arut30A, Arut89A, Arut57A, Arut19A и Arut802A. Сателлиты ArutF26A и ArutF167A более представлены в геноме самки.

1.2. Сателлитная ДНК с множественными сайтами локализации

Мы приготовили зонды ТПП для их локализации на метафазных пластинках стерляди. сатДНК Arut19A, Arut30A, Arut40A и Arut57A локализуются в субтеломерных районах мелких пар хромосом (41-60) и на хромосомах 31 и 32, несущих район ядрышкового организатора (ЯОР).

1.3. Сателлитная ДНК с сайтами локализации на ЯОР-несущих хромосомах

ТПП Arut802A, ArutF26A и Arut219A в основном локализованы на хромосомах, несущих ЯОР. Картирование Arut802A выявило четыре сигнала, колокализирующихся с рДНК 18S/28S на двух из трёх пар ЯОР-несущих хромосом.

Сателлит ArutF26A также маркирует ЯОР-несущие хромосомы,

но располагается рядом с рибосомными генами в прицентромерных областях q-плеч, где были локализованы Arut19A, Arut30A, Arut40A и Arut57A.

Arut219A колокализован с 5S рДНК в прицентромерной области небольшой метацентрической хромосомы 41. Такое распределение согласуется с данными анализа последовательности Arut219A, который показал гомологию последовательности этого сателлита с частью гена 5S рРНК и межгенным спейсером.

В ходе физического картирования ТПП было выявлено, что две из трех несущих ЯОР пар хромосом насыщены практически всеми обнаруженными сателлитами с большой длиной мономера. Arut802A и Arut40A колокализируются с 18S/28S рДНК, тогда как остальные сателлиты располагаются вблизи кластеров рибосомных генов. Поиск гомологии в базе данных NCBI показал высокий уровень идентичности Arut802A транспозону озерного осетра, который расположен в межгенном спейсере рибосомных генов. Результаты картирования FISH показывают, что Arut802A также находится внутри кластера рибосомных генов.

1.4. Сателлиты с хромосомоспецифической локализацией

Сравнение ArutF167A с последовательностями из базы данных NCBI показало, что часть мономера сателлита ArutF167A гомологична ранее описанному семейству HindIII. ArutF167A локализовался в прицентромерной области хромосомы 14 и на мелкой хромосоме 60. Колокализация ArutF167A и сателлита HindIII показала их колокализацию только на хромосоме 14.

Arut434A локализовался в интерстициальных блоках крупных хромосом 1-7, 9 и 14. Из-за различий в интенсивности сигнала с помощью этого сателлита удалось четко различить паралоги — хромосомы 1 и 2, 3 и 4, 8 и 9, 7q и 14. Разница в интенсивности сигнала выявляет расхождение в количестве копий повторяющихся ДНК среди паралогов.

Локализация консервативного кластера генов малой ядерной РНК U2 была ограничена прицентромерным районом р-плеча паралогичных хромосом ARUT 10 и 12.

Характер распределения хромосомоспецифичных сателлитов схематично отображён на GTG-окрашенном кариотипе стерляди (Рисунок 1).

1.5. Хромосомное картирование микросателлитной ДНК

При картировании семи микросателлитных последовательностей — (AAT)_n, (AC)_n, (AAG)_n, (ACAG)_n, (ACAT)_n, (ATC)_n и (AAC)_n была выявлена схожая картина их распределения с преимущественной локализацией на нескольких парах хромосом, включая ЯОР-несущие

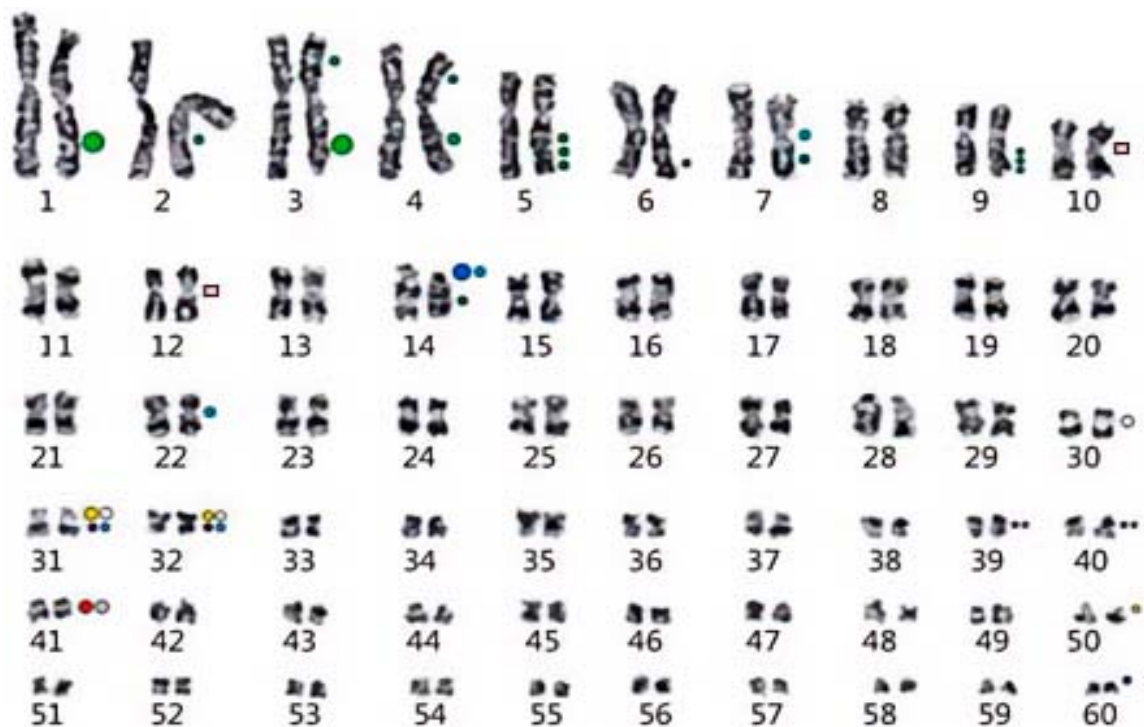


Рисунок 1: Схема расположения ДНК-зондов на GTG-окрашенном кариотипе стерляди. Цвета соответствуют разным зондам: *Arut434A* (зеленый), *ArutF167A* (темно-синий), *HindIII* (синий), *18S/28S* рДНК (белый), *Arut802A* (желтый), *Arut26A* (фиолетовый), $(AAT)_n$ (светло-фиолетовый), *Arut219A* (красный), *5S* рДНК (серый), *U2* (розовый).

хромосомы 31 и 32, а также множество слабых сигналов на субтеломерных областях малых хромосом.

2. Хромосомное картирование тандемноорганизованных повторяющихся ДНК сибирского осетра

2.1. Хромосомное картирование микросателлитных ДНК на хромосомах *A. baerii*

При картировании пяти микросателлитных последовательностей — $(AC)_n$, $(AAG)_n$, $(ACAG)_n$, $(ACAT)_n$ и $(AAC)_n$ выявляли около 80–100 сигналов в основном в субтеломерных регионах мелких хромосом *A. baerii*.

Другая картина локализации выявлена на хромосомах, несущих ЯОР, где все ТПП расположены в прицентромерных областях.

2.2. Локализация сателлитных ДНК-зондов *Arut19A*, *Arut30A*, *Arut40A*, *Arut57A* и *Arut167A* на хромосомах сибирского осетра

ТПП *Arut19A*, *Arut30A*, *Arut40A* и *Arut57A* распределены по множеству районов мелких хромосом сибирского осетра. *Arut40A* и *Arut57A* накапливались предпочтительно в субтеломерных областях q-плеч, *Arut19A* имел, в основном, перицентромерное расположение, а *Arut30A* располагался преимущественно на терминальных областях

p-плеч. Количество сигналов на метафазу варьировало от 12–14 для Arut19A до примерно 80 для Arut30A, Arut40A и Arut57A. Эти сателлиты (кроме Arut19A), также локализованы в прицентромерных областях хромосом, несущих ЯОР.

По сравнению со стерлядью число сигналов гибридизации ТПП на хромосомах сибирского осетра, как и ожидалось, увеличилось примерно вдвое после полногеномной дупликации Ac2R, кроме зонда Arut19A. Он выявлял одинаковое число областей у обоих видов. Этот результат может указывать на снижение количества копий после Ac2R.

ArutF167A локализуется в виде прицентромерных блоков на более чем 40 акроцентриках. У стерляди ArutF167A выявлялся в прицентромерной области двух хромосом: акроцентрической хромосомы 14 и небольшой хромосомы 60. Таким образом, сателлит ArutF167A амплифицирован в геноме сибирского осетра. Хромосомное распределение этого повтора коррелирует с изменениями в структуре кариотипа после последней полногеномной дупликации (ПГД) – нелинейным увеличением числа акроцентриков. Таким образом, изменения в структуре кариотипа после ПГД сопровождались как накоплением, так и редукцией некоторых ТПП.

2.3. Сателлитная ДНК с хромосомоспецифическим расположением

У *A. baerii* сателлит Arut434A был локализован на 20 парах хромосом: АВАЕ 1–11, 13, 15, 16, 23, 101–104 и 108. Интенсивность сигнала Arut434A, как и у стерляди, отличалась на паралогах сибирского осетра.

Двухцветная FISH Arut434A в сочетании с пэйнтинг-пробами стерляди ARUT 1–7, 9 и 14 выявила ортологичные хромосомы *A. baerii*. Как и ожидалось, каждый из этих зондов гибридизовался с двумя инпаралогичными хромосомами *A. baerii*.

Колокализация ТПП Arut434A с хромосомными зондами позволила реконструировать перестройки первых пяти хромосом предполагаемого предка осетровых (AcA) в геномах *A. ruthenus* и *A. baerii*. Ортологи трех из них (AcA 1, 3 и 5) остались неизменными в обоих кариотипах, а ортологи хромосомы AcA 2 претерпели разрыв центромеры в кариотипе *A. baerii*. Наиболее перестроенными оказались ортологи хромосомы AcA 4, претерпевшие ряд последовательных центромерных делений и слияний в кариотипах обоих видов.

Кластер генов U2 локализован в прицентромерных областях двух пар паралогов: АВАЕ 19 и 20; 25 и 35, соответственно.

Таким образом, локализация хромосомоспецифичных повторов и

зондов показала, что эволюция кариотипа сибирского осетра после ПГД включала множество сложных меж- и внутрихромосомных перестроек.

2.4. Сателлитная ДНК с сайтами локализации на ЯОР-несущих хромосомах

Сателлиты Arut802A, ArutF26A и микросателлит (AAT)_n локализованы на ЯОР-несущих хромосомах сибирского осетра. У стерляди сигналы гибридизации этих ТПП ограничены двумя из трех пар ЯОР-несущих хромосом, в то время как рибосомный зонд 18S/28S рДНК был локализован на всех трех парах хромосом. У сибирского осетра этот зонд пометил семь пар хромосом вместо ожидаемых шести пар, которые должны присутствовать после ПГД.

Колокализация зондов 18S/28S рДНК и 5S рДНК с хромосомоспецифическими пробами показала сложные перестройки малых хромосом. В отличие от крупных хромосом, все пять небольших предковых хромосом (AcA 15, 16, 20, 21 и 29) претерпели в кариотипе осетровых сложные перестройки, включающие не только разделения и слияния, но и другие трансформации, приведшие к появлению гетероморфных гомологов на хромосоме 50 и появлению кластеров рибосомных генов на хромосоме 63.

Все использованные хромосомоспецифические повторы были нанесены на карту хромосом сибирского осетра (Рисунок 2).

3. Идентификация и аннотация повторяющихся элементов в сборке генома стерляди

В результате анализа повторяющихся последовательностей в сборке генома самца стерляди выяснили, что на них приходится 40,3% генома, при этом 34,94% занимают рассеянные последовательности.

МГЭ класса 2 являются самыми представленными в геноме стерляди, занимая 13,78%. Наибольшую долю генома занимают представители non-LTR ретротранспозонов – LINE (5,82%) и SINE (2,85%). Наиболее распространенными суперсемействами порядка LINE являются CR1, L2 и Hero, а SINE – MIR. LTR-ретроэлементы занимают 5,11% генома, самыми представленными суперсемействами оказались Gypsy и DIRS.

ДНК-транспозоны занимают 9,83% генома стерляди. Представители подкласса 1 занимают 9,36% геномной сборки. Самыми распространёнными суперсемействами порядка TIR оказались Mariner и hAT. Транспозоны из порядка Crypton, принадлежащие одноимённому суперсемейству, составляют 0,36%. Представители подкласса 2 — Helitron и Maverick — представляют 0,45 и 0,02% генома.

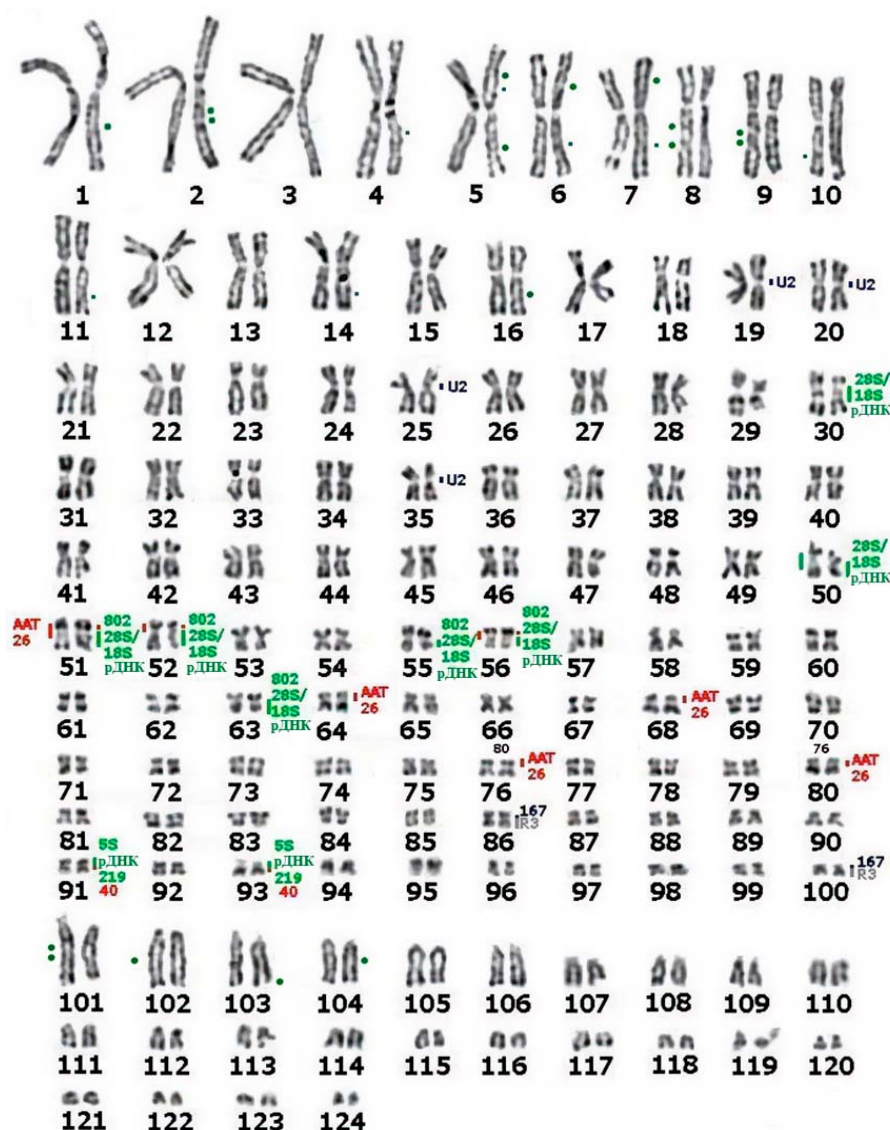


Рисунок 2: Схема расположения ДНК-зондов на GTG-окрашенном кариотипе сибирского осетра. Обозначения: Arut434A (зеленый круг), кластер генов U2 (U2), кластер генов большой субъединицы рибосом (28S/18S pДНК), Arut26A (26), Arut802A (802), (AAT)_n (AAT), 5S рибосомная ДНК (5S pДНК), Arut40A (40), Arut167A (167), Arut219A (219).

Содержание повторяющихся элементов и состав мобильных элементов в геноме стерляди сравнимы с таковыми у других рыб. Таким образом, мобилом стерляди похож на геномы многих других современных рыб, в том числе быстро эволюционирующих костистых.

Используя анализ дивергенции, основанный на вычисленном расстоянии Кимуры, удалось детектировать недавнюю волну экспансии МГЭ (Рисунок 3). Эта волна активности МГЭ включает в себя, в основном, ДНК-транспозоны Mariner и CMC и ретротранспозоны LTR/Gypsy, LINE/Hero, LINE/L2, LINE/CR1 и SINE/MIR. Таким образом, несмотря на то, что стерлядь является представителем древней, медленно эволюционирующей линии

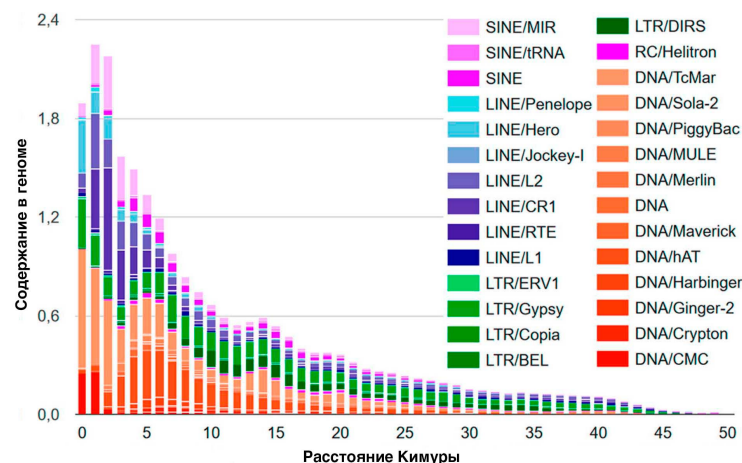


Рисунок 3: Анализ дивергенции копий классов МГЭ стерляди, основанный на посчете двухпараметрических расстояний Кимуры. Содержание МГЭ в геноме откладывается по оси ординат, а значения расстояния Кимуры по оси абсцисс. Более древние копии МГЭ расположены в правой части графиков, а недавние инсерции – в левой.

осетрообразных, предполагаемая активность транспозонов обнаруживает недавнюю экспансию всех основных классов МГЭ.

Сравнение графиков дивергенции МГЭ на паралогичных хромосомах стерляди показало, что каждая пара паралогов имеет почти идентичное содержание МГЭ со схожей эволюционной историей экспансии (Рисунок 4). Таким образом, геном стерляди не обнаруживает признаков наличия дивергировавших субгеномов, следовательно, полиплоидизация возникла в результате дублирования всего генома в одном организме (автополиплоидии).

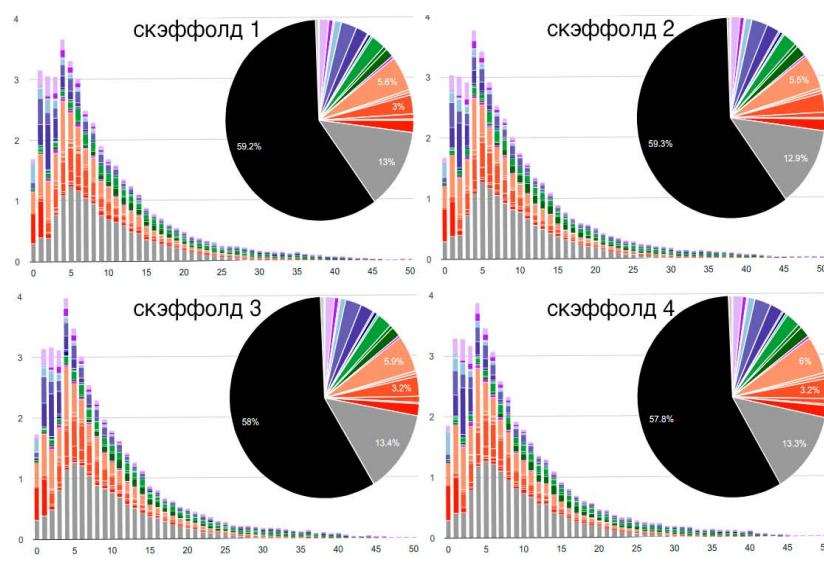


Рисунок 4: Графики Кимуры повторяющихся элементов для паралогичных скэффолдов 1-2, 3-4. В верхней правой части графиков Кимуры изображены диаграммы, отражающие содержание МГЭ в этом скэффолде.

Анализ транскриптомов восьми тканей показал значительную экспрессию для членов некоторых суперсемейств МГЭ (Рисунок 5). Наиболее выраженная экспрессия наблюдалась для суперсемейств LTR/ERV1, SINE/tRNA, SINE/MIR, LINE/Tx1 и DNA/Chaparev. Экспрессия суперсемейств LTR/Copia, LINE/I, DNA/Merlin и DNA/Maverick была понижена, что свидетельствует о механизмах сайленсинга этих МГЭ. Экспрессия МГЭ варьировала между тканями. Так, например, суперсемейство LTR/ERV1 показывало относительно высокий уровень экспрессии в семенниках, недифференцированных гонадах и яичниках, а представители суперсемейства SINE/tRNA относительно высоко экспрессировались печени и селезёнке, а в остальных тканях транскрипция была снижена. Присутствие МГЭ в транскриптомах стерляди указывает на то, что стерлядь сохраняет некоторые активные транспозоны.

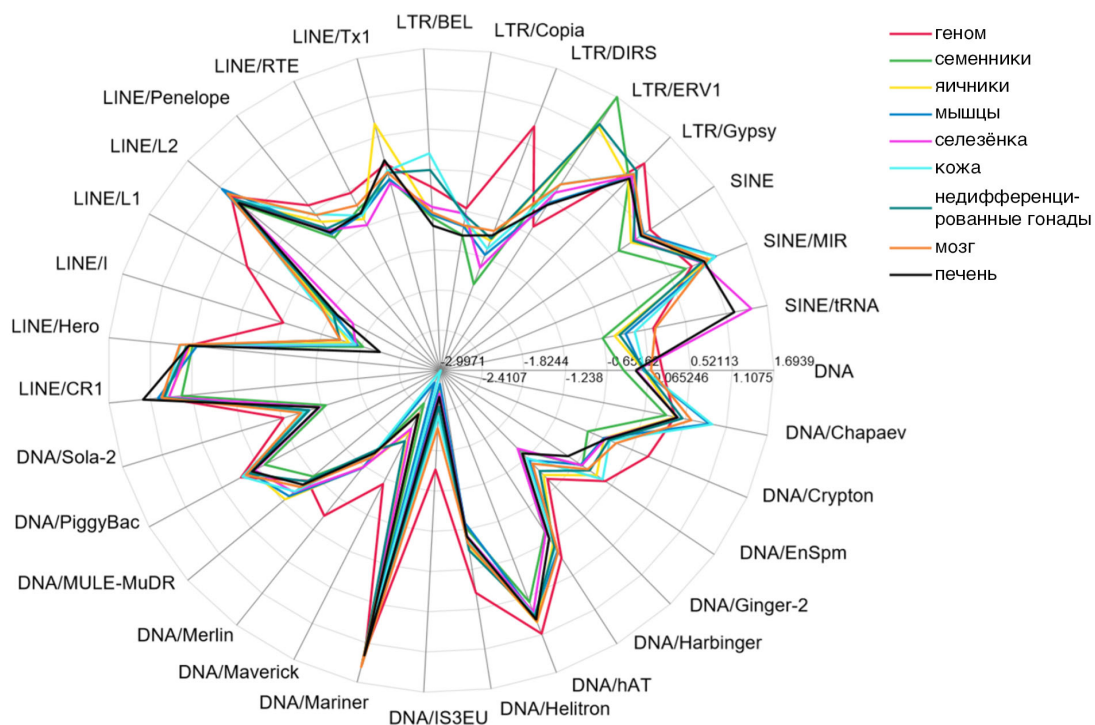


Рисунок 5: Представленность суперсемейств МГЭ в геноме и транскриптомах стерляди.

4. Идентификация полоспецифичного маркера у осетрообразных

4.1. Секвенирование ДНК самцов и самок стерляди

После секвенирования смеси ДНК (пулов) 31 самки и 30 самцов *A. ruthenus* существенной разницы между обоими пулами в процентном соотношении выравненных на геном самца прочтений обнаружено не было (96,9%).

Для поиска специфичных для самок последовательностей

использовали две биоинформатические стратегии. Сначала были проанализированы полоспецифичные варианты с использованием референсного генома самца для картирования последовательностей пулов. Число специфичных для самок или самцов вариантов было низким (1682 специфичных для самок варианта и 318 специфичных для самцов вариантов, соотношение самки/самцы = 5,3), однако повышенное содержание уникальных для самок замен указывало на систему определения пола ZZ/ZW. Наибольшее число вариантов у самок наблюдалось в районе скэффолда HiC_scaffold_4 (хромосома 4) по координатам 45149338–45165057.

Для дальнейшего анализа были изучены специфические для самок прочтения, которые не выравнивались на геном самца. С помощью этих прочтений была проведена сборка контигов. Поиск белок-кодирующих последовательностей с помощью BlastX не обнаружил последовательностей, связанных с известными генами определения пола.

4.2. Идентификация локуса, специфичного для самок у *A. ruthenus* с использованием референсного генома самки

Данные секвенирования ДНК самцов и самок стерляди картировали на геном самки стерляди для поиска районов с различающимся покрытием между самцами и самками. При строгих параметрах фильтрации картирования выравнивания прочтений самок попали в район SASTIG010000179.1: 61229779–61285849 (включая 20 т.п.н. перед и после этого района). Эта область соответствует области HiC_scaffold_4: 43 938 061–43 987 561 в сборке самца. Специфичный для самок локус расположен на расстоянии 1,162 т.п.н. от идентифицированного кластера из 27 специфичных для самок вариантов, которые были описаны ранее. Более того, один контиг (559 п.н.) из 1067 специфичных для самок контигов, собранных из данных пула самок, соответствовал идентифицированному локусу. Таким образом, три независимых анализа данных — поиск специфичных для самок вариантов нуклеотидов, сборка специфичных для самок прочтений и различий в покрытии подтверждают, что скэффолд хромосомы 4 содержит полоспецифичную последовательность и, таким образом, является половой хромосомой стерляди.

Полоспецифичный локус содержит 43% повторяющихся последовательностей, в том числе 8% сателлитных последовательностей, 1% микросателлитов, 4% LINE и 13% LTR-ретротранспозонов, 17% ДНК-транспозонов.

Таким образом, была обнаружена специфическая для самок область W-хромосомы, соответствующая потенциальной системе ZZ/ZW у осетровых. Этот вывод согласуется с предыдущими

предположениями и косвенными свидетельствами того, что осетровые обладают системой определения пола с гетерогаметными самками.

4.3. Проверка ПЦР-маркера на разных видах осетровых рыб

Нами был разработан ПЦР-маркер AllWSex2 длиной 108 п.н., праймеры которого фланкируют полоспецифичный район генома самки. ПЦР с использованием этого маркера выявила специфические для самок продукты амплификации для всех самок у шести видов осетровых, охватывающих все филогенетическое древо осетровых, как тетраплоидных (*A. sturio*, *A. oxyrinchus*, *A. ruthenus*, *H. huso*), так и октаплоидных (*A. gueldenstaedtii* и *A. baerii*). Электрофореграмма приведена на Рисунке 6.

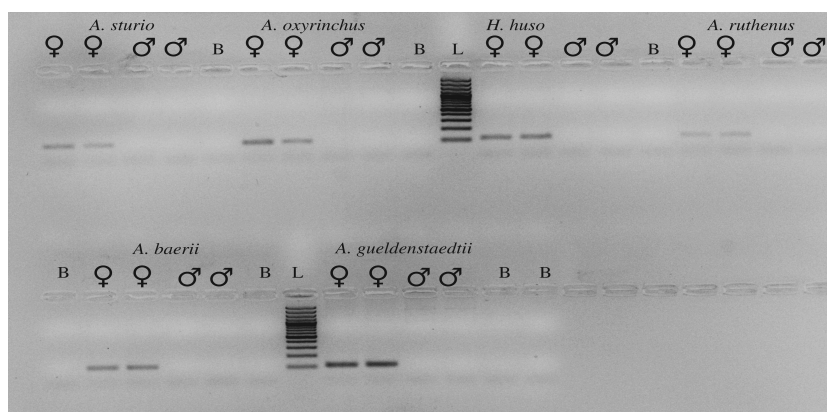


Рисунок 6: Электрофореграмма с визуализацией ПЦР-продуктов, полученных с использованием маркера AllWSex2 у шести видов осетровых в идентичных условиях ПЦР для двух самок и двух самцов каждого вида. Символы и сокращения: ♀: самка; ♂: самец; B: отрицательный контроль; L: маркер длины ДНК с шагом 100 п.н.

Полученные данные выявили самую старую из известных систем определения пола позвоночных с недифференцированными половыми хромосомами. Область, специфичная для самок, сохраняется на протяжении 180 млн лет эволюции осетровых, в том числе после по крайней мере одного раунда полиплоидизации.

ЗАКЛЮЧЕНИЕ

В настоящем исследовании охарактеризована основная фракция ТПП стерляди и сибирского осетра. Картирование на хромосомах стерляди показало, что ТПП неравномерно распределены в геноме и имеют тенденцию к кластеризации в определенных областях. Некоторые из сателлитных ДНК могут быть использованы в качестве маркеров для идентификации отдельных хромосом и их паралогов.

Выявленные в геноме стерляди ТПП были применены для анализа кариотипа сибирского осетра. Событие дупликации генома сибирского осетра сопровождалось как экспансией, так и редукцией ТПП. Некоторые из зондов позволяют различать паралогичные хромосомы и выявлять предковые синтенные районы. Была

представлена физическая карта хромосом генома сибирского осетра, подтвержденная молекулярными маркерами.

Анализ сборки генома стерляди показал, что геном стерляди содержит 40,3% повторяющихся элементов. Выявлена недавняя волна экспансии повторяющихся элементов. Сравнение содержания МГЭ и их эволюционной истории на паралогичных хромосомах выявило автополиплоидное происхождение генома стерляди. Некоторые МГЭ экспрессируются в геноме стерляди, а экспрессия других снижена.

Был разработан маркер пола осетровых рыб и тест ПЦР-генотипирования, дающий специфические для самок продукты у шести видов осетровых.

ВЫВОДЫ

1) Анализ данных секвенирования генома стерляди позволил идентифицировать в нем 24 типа тандемноорганизованных повторяющихся последовательностей (ТПП), для 17 из которых был выбран специфический зонд. Локализация зондов на хромосомах стерляди показала, что ТПП имеют неравномерное распределение с тенденцией к кластеризации в перицентромерных, субтеломерных и интерстициальных районах хромосом.

2) Локализация 16 зондов к ТПП стерляди на хромосомах сибирского осетра показала, что для большинства ТПП число сигналов увеличено примерно вдвое, что подтверждает предположение о дупликации генома сибирского осетра. Последнее событие дупликации генома у сибирского осетра сопровождалось экспансией сателлитной последовательности Arut167A и редукцией Arut19A.

3) Показано, что использование 10 из 17 разработанных зондов к ТПП позволяют идентифицировать 20 хромосом и их паралоги в кариотипе стерляди. В случае сибирского осетра применение зондов к ТПП позволяют идентифицировать 36 хромосом, а также выявлять предковые синтенные районы.

4) Анализ сборки генома стерляди показал, что повторяющиеся элементы по большей части рассеяны по геному, а их содержание составляет 40,3%. Для ДНК-транспозонов Mariner и CMC, а также ретротранспозонов LTR/Gypsy, LINE/Her0, LINE/L2, LINE/CR1 и SINE/MIR показана недавняя волна экспансии. Схожий состав и паттерн дивергенции рассеянных повторяющихся элементов на паралогичных хромосомах указывает на автополиплоидное происхождение генома стерляди. Большинство мобильных генетических элементов в геноме стерляди, за исключением представителей суперсемейств LTR/ERV1, SINE/tRNA, SINE/MIR, LINE/Tx1 и DNA/Chaparev, транскрипционно неактивны.

5) Анализ данных секвенирования самцов и самок стерляди позволил выявить полоспецифичный маркер W-хромосомы самок, который локализован на одном из гомологов четвертой хромосомы и имеет длину 16 т.п.н. Маркер содержит 43% повторяющихся последовательностей, в том числе 9% тандемноорганизованных последовательностей и по 17% ретротранспозонов и ДНК-транспозонов. На основе найденного маркера был разработан ПЦР-тест, позволяющий специфично выявлять ДНК самок шести видов осетровых, как тетраплоидных (*Acipenser sturio*, *Acipenser oxyrinchus*, *Acipenser ruthenus*, *Huso huso*), так и октоплоидных (*Acipenser gueldenstaedtii*, *Acipenser baerii*). Наблюдаемая консервативность полоспецифичного маркера у представителей всех основных клад семейства осетровых позволяет говорить о том, что система определения пола осетровых не менялась последние 180 млн. лет.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Biltueva L.S., **Prokopov D.Yu.**, Makunin A.I., Komissarov A.S., Kudryavtseva A.V., Lemskaya N.A., Vorobieva N.V., Serdyukova N.A., Romanenko S.A., Gladkikh O.L., Graphodatsky A.S., Trifonov V.A. Genomic organization and physical mapping of tandemly arranged repetitive DNAs in sterlet (*Acipenser ruthenus*) //Cytogenetic and Genome Research. – 2017. – Т. 152. – №. 3. – С. 148-157.

2. Du K., Stöck M., Kneitz S., Klopp C., Woltering J., Adolphi M., Feron R, **Prokopov D.**, Makunin A., Kichigin I., Schmidt C., Fischer P., Kuhl H., Wuertz S., Gessner J., Kloas W., Cabau C., Lampietro C., Parrinello H., Tomlinson C., Journot L., Postlethwait J., Braasch I., Trifonov V., Warren W.C., Meyer A., Guiguen Y., Scharl M. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization //Nature ecology & evolution. – 2020. – Т. 4. – №. 6. – С. 841-852

3. Biltueva L.S., **Prokopov D.Yu.**, Romanenko S.A., Interesova E.A., Scharl M., Trifonov V.A. Chromosome distribution of highly conserved tandemly arranged repetitive DNAs in the siberian sturgeon (*Acipenser baerii*) //Genes. – 2020. – Т. 11. – №. 11. – С. 1375.

4. Kuhl H., Guiguen Y., Höhne C., Keuz E., Klopp C., Du K., Lopez-Roques C., Yebra-Pimentel E.S., Ciorpac M., Gessner J., Holostenco D., Kleiner W., Kohlmann K., Lamatsch D., **Prokopov D.**, Bestin A., Bonpunt E., Debeuf B., Haffay P., Morvezen R., Patrice, P. Suci R., Dirks R., Wuertz S., Kloas W., Scharl M., Stöck M. A 180 Myr-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes //Philosophical Transactions of the Royal Society B. – 2021. – Т. 376. – №. 1832. – С. 20200089.