

**Болбат Александр Васильевич**

**СТРУКТУРА И ЭВОЛЮЦИЯ МИТОХОНДРИАЛЬНОГО ГЕНОМА  
РЕЛИКТОВЫХ ПИЯВОК**

1.5.7. – Генетика

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата биологических наук

Новосибирск – 2022

Работа выполнена в лаборатории аналитической биоорганической химии  
ФГБУН «Лимнологический институт Сибирского отделения РАН», г. Иркутск

Научный **Кайгородова Ирина Александровна**  
руководитель: кандидат биологических наук, старший научный сотрудник  
лаборатории аналитической биоорганической химии,  
ФГБУН «Лимнологический институт СО РАН», г. Иркутск

Официальные **Политов Дмитрий Владиславович**  
оппоненты: доктор биологических наук, заведующий лабораторией  
популяционной генетики ФГБУН «Институт общей генетики  
им. Н.И. Вавилова», г. Москва

**Блинов Александр Геннадьевич**  
кандидат биологических наук, ведущий научный сотрудник  
лаборатории молекулярно-генетических механизмов  
регенерации, ФГБНУ «Федеральный исследовательский центр  
Институт цитологии и генетики СО РАН», г. Новосибирск

Ведущая **ФГБУН «Институт молекулярной и клеточной биологии СО**  
организация: **РАН», г. Новосибирск**

Защита диссертации состоится «\_\_» \_\_\_\_\_ 2022 г. на утреннем  
заседании Диссертационного совета 24.1.239.01 на базе ФГБНУ  
«Федеральный исследовательский центр Институт цитологии и генетики  
Сибирского отделения РАН» в конференц-зале Института по адресу: 630090,  
г. Новосибирск, проспект ак. Лаврентьева, 10, т. (383)363-49-06, факс (383)  
333-12-78, e-mail: [dissov@bionet.nsc.ru](mailto:dissov@bionet.nsc.ru).

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте  
Института <http://www.bionet.nsc.ru/>

Автореферат разослан «\_\_» \_\_\_\_\_ 2022 г.

Ученый секретарь  
диссертационного совета,  
доктор биологических наук

Т. М. Хлебодарова

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования и степень ее разработанности.**

Митохондриальный геном животных содержит информацию о структуре ключевых белковых компонентов дыхательной цепи, а также рибосомальных и транспортных РНК, необходимых для реализации этой информации. Несмотря на способность митохондрий импортировать отдельные белки, синтезируемые в цитоплазме, выраженная гидрофобность некоторых компонентов электрон-транспортной цепи затрудняет их перенос на внутреннюю мембрану митохондрий (Gearing and Nagley 1986). Это привело к сохранению у митохондрий части генома  $\alpha$ -протеобактерий с консервативным набором генов (Martijn et al. 2018). В связи с консервативностью, гаплоидным характером наследования и ограниченной рекомбинацией (Rokas et al. 2003), митохондриальный геном животных является важным объектом для изучения, поскольку реконструкция его эволюции может быть с высокой уверенностью интерпретирована как реальная эволюционная история организма. По сравнению с ядерным митохондриальный геном требует вчетверо меньшего эффективного размера популяции для сортировки линий (“lineage sorting”) (Moore 1997).

На сегодняшний день уровень развития молекулярно-генетических методов исследования и биоинформационного анализа предлагает относительно дешёвые и усовершенствованные технологии ДНК-секвенирования, позволяющие рутинно проводить полногеномные исследования, которые ранее требовали больших временных и экономических затрат. Реконструкция геномов организмов, исследование экспрессии их генов, получение сотен генных локусов для эволюционных исследований, изучение некультивируемых организмов в составе сообществ и даже обнаружение эпигенетических сигналов теперь ограничиваются лишь методиками обработки, а не получения данных (Meaburn et al. 2012, Mutz et al., 2013, Segata et al. 2013, Phillips et al. 2019, Nurk et al. 2021). Гораздо больше, чем ранее, стало известно о структуре наследственного материала, принципах его действия и передачи последующим поколениям. В связи с этим наиболее остро встаёт вопрос изучения генетической структуры таких ключевых групп организмов, как реликтовые пиявкоподобные паразиты (Acanthobdellida), являющихся одной из малоизученных фаунистических групп, испытывающих на себе сильное антропогенное влияние и сокращающих свой ареал в Европе (Dahm 1962, Kaygorodova et al. 2006).

Реликтовые пиявки вызывают научный интерес на протяжении столетия ввиду их предполагаемого промежуточного эволюционного положения между малощетинковыми червями (Oligochaeta) и пиявками (Hirudinea), что само по себе делает актуальным настоящее исследование. Вопрос точного эволюционного положения и классификации организмов данной группы невозможно разрешить с использованием одних морфологических признаков, в том числе из-за неоднозначности их интерпретации (Purschke et al. 1993, Brinkhurst et al. 1999). Сложности получения биологического материала ввиду труднодоступности мест обитания акантобделл, специфики их биологического цикла и экологических особенностей на данный момент позволили получить лишь

единичные сведения о первичной структуре отдельных фрагментов их генома, на основе которых оценка филогенетического положения акантобделлид оказалась неоднозначной (Siddal et al. 1998, 2001, Apakupakul et al. 1999, Trontelj 1999, Martin et al. 2000, 2001, Kaygorodova et al. 2006, Erseus et al. 2004, Rousset et al. 2008, Tessler et al. 2018, Phillips et al. 2019). На сегодняшний день полногеномных исследований акантобделлид не проводилось. Единственная работа, посвящённая изучению филогении акантобделлид с применением технологий секвенирования нового поколения (NGS), оперирует множеством нестандартизированных гомологичных локусов, в основном из ядерного генома (Phillips et al. 2019), что затрудняет использование этих данных в последующих исследованиях. В настоящее время остро ощущается недостаток информации о структуре генома *Acanthobdella peledina*, и полностью отсутствуют какие-либо генетические данные по виду *Paracanthobdella livanowi*, считающихся «живыми реликтами» и являющихся важными звеньями для понимания эволюции мировой фауны кольчатых червей.

Использование современных методов полногеномного секвенирования (NGS) даёт возможность в параллельном режиме расшифровать значительную долю всего генетического материала организма при отсутствии референсных данных о его структуре, что позволяет получить недостижимый ранее объём данных и значительно повысить достоверность эволюционных исследований. Эти условия идеально подходят для изучения реликтовых пиявок.

**Цели и задачи.** Целью данной работы является получение сведений о структуре полного митохондриального генома, степени генетической и таксономической дивергенции и эволюционной истории представителей отряда Acanthobdellida.

Для достижения цели были поставлены следующие задачи:

1. Провести полногеномное секвенирование образцов реликтовых пиявок из различных, географически разобщённых мест обитания и реконструировать первичную структуру их митохондриальных геномов.
2. Аннотировать полученные нуклеотидные последовательности.
3. Реконструировать эволюционную историю митохондриальных геномов реликтовых пиявок, оценить степень их генетической дивергенции и применимость полных митохондриальных геномов и их фрагментов для интерпретации филогении и молекулярной делимитации таксонов исследуемой группы.

**Научная новизна.** Впервые методом секвенирования нового поколения получены полногеномные данные для восьми образцов акантобделлид из географически разрозненных водных бассейнов Евразии от Швеции до Камчатки (*Acanthobdella peledina* и *Paracanthobdella livanowi*), а также четырех образцов эндемичных гирудинид оз. Байкал (*Baicalocleipsis echinulata*, *B. grubei* и два образца *Codonobdella* sp.). Кроме того, были собраны и аннотированы митохондриальные геномы восьми образцов из 7 видов аннелид (*Lumbriculus variegatus*, *Glossiphonia complanata*, *G. concolor*, *Theromyzon tessulatum*, *Piscicola geometra*, *Erpobdella octoculata* и два образца *Haemopsis sanguisuga*), необработанные прочтения которых доступны в международной базе коротких

прочтений SRA. Впервые полные митохондриальные геномы акантобделлид были использованы для реконструкции филогенетической истории кольчатых червей. Получены первые сведения о генетических дистанциях между представителями вида *A. peledina* из разных мест обитания, а также их дистанций до представителей вида *P. livanowi*. Результаты данной работы позволили разрешить спорные вопросы систематики кольчатых червей: уточнить положение акантобделлид в системе Clitellata и таксономический статус *A. peledina* и *P. livanowi*.

**Теоретическая и научно-практическая значимость работы.** Результаты данной работы дают первое представление о структуре митохондриального генома реликтовых пиявок и других кольчатых червей внесли существенный вклад в международные базы генетических данных, пополнив их достоверной информацией, и опубликованы в открытом доступе, что создаёт предпосылки для их использования всеми заинтересованными лицами, в частности для конструирования научно обоснованных филогений и систем молекулярной идентификации биологических образцов, в том числе частично сохранных. Данные о внутри- и межвидовой генетической вариабельности могут быть использованы в дальнейших исследованиях таксономического и филогенетического разнообразия кольчатых червей. Общий объём данных, полученных в ходе полногеномного секвенирования, значительно превышает объём данных, необходимых для реконструкции митохондриального генома, что позволяет использовать их в дальнейших исследованиях.

Материалы диссертации могут быть использованы при подготовке лекционных курсов и учебных пособий для учащихся ВУЗов и подготовки научных кадров высшей квалификации.

**Методология и методы исследования.** Данная работа выполнена с использованием технологии высокопроизводительного секвенирования Illumina. Анализ полученных данных проводился с использованием обширного комплекса современных компьютерных программ: FastQC, Trimmomatic v. 0.40, Mira v. 5, Tablet, Aragorn v 1.2.40, ExPasy Translate, Clustal Omega v. 1.2.4, Bowtie2 v. 2.3.5.1, barnap, Mafft v7.453, IQ-TREE v. 2.1.3, BEAST v. 2.6.4, PartitionFinder2, jModelTest 2, Tracer v. 1.7.1, TOPD/FMTS v. 4.6 и GMYC. Дополнительно был разработан собственный скрипт «fastq\_grep» на языке Python 3, позволяющий извлекать из сырых данных секвенирования отдельные прочтения и конвертировать их в формат Fasta.

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

1. Митохондриальный геном реликтовых пиявок *Acanthobdella peledina*, вида широко распространенного в Северной Евразии, содержит уникальные структурные элементы в виде тандемных повторов, размер которых варьирует у образцов из географически разрозненных мест обитания.
2. Реконструкция эволюционной истории акантобделлид при использовании полных митохондриальных геномов демонстрирует стабильную топологию филогенетических деревьев и доказывает происхождение реликтовых пиявок от олигохетного предка в качестве сестринской линии истинных пиявок.

**Степень достоверности результатов и апробация работы.** Достоверность полученных результатов подтверждается согласованностью структур реконструированных в данной работе митохондриальных геномов со структурами ранее опубликованных для других представителей Clitellata. Отдельная часть работы посвящена изучению влияния использованных данных и начальных параметров анализа на стабильность схемы эволюционных отношений.

Результаты работы были опубликованы в виде 12 печатных работ, в том числе пять статей в научных рецензируемых изданиях, включенных в список ВАК и приравненных к ним (в зарубежных и российских журналах системы Web of Science – три), а также представлены на 7 научных конференциях международного и национального уровня.

**Личный вклад.** Автором работы выполнены основные этапы исследования: экспедиционные работы по сбору образцов акантобделлид, морфологический анализ, экстракция тотальной ДНК, сборка контигов и аннотация митохондриального генома, филогенетический анализ, а также подготовка и публикация полученных результатов; разработана компьютерная программа для поиска прочтений с определённой последовательностью в необработанных данных секвенирования нового поколения. Выводы сделаны на основе собственных данных.

**Объём и структура диссертации.** Диссертация состоит из введения, четырёх глав, выводов, списка литературы и четырнадцати приложений. Работа изложена на 137 страницах, включая приложения, содержит 29 рисунков и 12 таблиц. Список цитируемой литературы включает 200 источников, из которых 189 зарубежных.

## СОДЕРЖАНИЕ РАБОТЫ

### ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

В разделе 1.1 приводятся обобщённые данные о современном представлении о пиявкоподобных паразитах отряда Acanthobdellida. Приводится характеристика морфологических особенностей видов *Acanthobdella peledina* и *Paracanthobdella livanowi*, их экологической ниши, ареала их распространения и лимитирующих факторов. Обобщены исторические работы, посвящённые изучению эволюционной истории акантобделлид на основе морфологических характеристик и молекулярно-генетических данных.

Раздел 1.2 посвящен обзору современных представлений о митохондриальном геноме животных. Приведены сведения о его структуре и проблемах аннотации функциональных генов на митохондриальных последовательностях. Рассмотрено применение теории тРНК-пунктуации для наиболее точной аннотации функциональных генов. Помимо этого, рассмотрен вопрос использования митохондриального генома в эволюционной биологии и таксономических исследованиях.

В разделе 1.3 проведён анализ современных методов эволюционных исследований. В нём были рассмотрены актуальные методы геномного секвенирования нового поколения, их классификация и области применения. Проведён анализ современного подхода к реконструкции молекулярной филогении, включая статистические методы

построения деревьев, модели молекулярной эволюции и дифференциальное применение этих моделей к различным сегментам ("partitions") набора данных. Рассмотрен вопрос использования молекулярных данных к разрешению спорных вопросов таксономии.

## ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ

**Сбор, идентификация и фиксация образцов.** Образцы акантобделлид были собраны в пресных водоемах Палеарктики (от Швеции до Камчатки) в период 2002-2018.

Образцы *Codonobdella* sp., *Baicalocleipsis grubei* и *B. echinulata* были собраны в оз. Байкал на глубине до 50 м. Идентификация образцов проводилась в соответствии с существующими таксономическими ключами (Лукин 1976, Эпштейн 1987). Образцы были фиксированы 70% этанолом и помещены для хранения в морозильную камеру при температуре минус 20°C.

**Выделение и секвенирование ДНК.** Тотальную ДНК выделяли с использованием набора DiaGene для выделения ДНК из культур клеток. Для отдельных образцов использовали фенол-хлороформный метод экстракции с использованием СТАВ. Очищенная ДНК была обработана ультразвуком на Covaris M200 для получения фрагментов длиной 300 п.н. Библиотеки были подготовлены и секвенированы при помощи наборов NEBNext Ultra II DNA Library Prep Kit и Illumina NextSeq 550 Mid Output Kit v2.5 для получения одиночных (SE) прочтений.

**Сборка митохондриальных геномов.** Контроль качества необработанных данных NGS проводили в FastQC (Andrews, 2010). Удаление адаптерных последовательностей и фильтрация качества прочтений осуществляли в Trimmomatic v. 0.40 (Bolger et al. 2014). Сборку контигов проводили с использованием Mira v. 5 (Chevreux et al. 1999). Визуализация полученных контигов проводилась в геномном браузере Tablet (Milne et al. 2013). Поиск митохондриальных контигов осуществлялся среди контигов с наибольшим покрытием и высоким содержанием А и Т путём поиска гомологии в GenBank с помощью BLASTn. Для разрешения сложных и повторяющихся участков был разработан собственный скрипт «fastq\_grep» на языке Python 3, позволяющий извлекать из сырых данных секвенирования отдельные прочтения и конвертировать их в формат Fasta. После чего прочтения выравнивались относительно друг друга с использованием Clustal Omega v. 1.2.4 (Sievers et al. 2011). Для оценки правильности сборки повторяющихся участков проводилось рекурсивное картирование прочтений на собранный митохондриальный геном с использованием Bowtie2 v2.3.5.1 (Langmead et al. 2012). Геном считался собранным правильно при достижении равномерного покрытия по всей длине.

**Аннотация митохондриальных геномов.** Процесс *de novo* аннотации митохондриального генома был разделён на несколько этапов. Первым этапом была аннотация транспортных РНК при помощи Aragorn v. 1.2.40 (Laslett et al. 2004) с опциями для поиска митохондриальных тРНК Metazoa. Согласно теории тРНК-пунктуации, перекрывающиеся гены аннотировали с сохранением целостности 5'-конца нижестоящего гена и сокращением 3'-конца вышестоящего гена. Вторым этапом являлась аннотация белок-кодирующих последовательностей, осуществлённая путём выявления

открытых рамок считывания с использованием веб-интерфейса программы ExPASy Translate и митохондриального генетического кода Metazoa (таблица трансляции № 5, GenBank). Полученные аминокислотные последовательности были аннотированы путём поиска гомологии в международной базе данных с использованием BLASTp. Белок-кодирующие последовательности были аннотированы с допущением перекрытия друг на друга, но без допущения перекрытия с генами тРНК. Последним этапом была проведена аннотация генов рибосомальной РНК с использованием программы barpar.

**Филогенетический анализ и молекулярная делимитация таксонов.** Для реконструкции эволюционной истории акантобделлид таксономически сбалансированная группа сравнения была выбрана на основе поиска в GenBank, состоящая из 69 последовательностей полных митохондриальных геномов длиной не менее 14 тыс. п.н. и с полным набором митохондриальных генов, из которых 46 составили геномы представителей Oligochaeta и 23 – представителей Hirudinea. Для повышения разнообразия группы сравнения нами реконструированы полные митохондриальные геномы еще восьми представителей Annelida, необработанные данные NGS которых найдены в свободном доступе. Последовательности митогеномов восьми представителей Polychaeta были заданы в качестве внешней группы во всех филогенетических реконструкциях.

При выравнивании набора данных в случае выявления генных перестановок, последовательности были отредактированы таким образом, чтобы сохранить единый порядок генов. После этого последовательности выравнивали с использованием Mafft v. 7.453 (Kato et al. 2013), с окончательным выравниванием сложных участков вручную.

Реконструкция филогении на основе фрагментов *cox1*, *12S* и полных митогеномов проводилась двумя методами: методом максимального правдоподобия (ML) в IQ-TREE v. 2.1.3 (Minh et al. 2020) и байесовским методом (BI) в BEAST v. 2.6.4 (Bouckaert et al. 2019). Определение оптимальной схемы сегментации набора данных производили с использованием PartitionFinder2 (Lanfear et al. 2017). Реконструкцию на основе несегментированных данных проводили с использованием моделей, рекомендованных jModelTest 2 (Darriba et al. 2012). Сходимость байесовской статистики сверяли в Tracer v. 1.7.1 (Rambaut et al. 2018). Консенсусные байесовские деревья были реконструированы в TreeAnnotator из пакета BEAST2. Сходство топологий полученных деревьев оценивали узловым («nodal») алгоритмом в TOPD/FMTS v. 4.6 (Puigbò et al. 2007). Для видовой делимитации использован алгоритм GMYC (Pons et al. 2006).

### **ГЛАВА 3. СТРУКТУРА МИТОХОНДРИАЛЬНОГО ГЕНОМА АННЕЛИД**

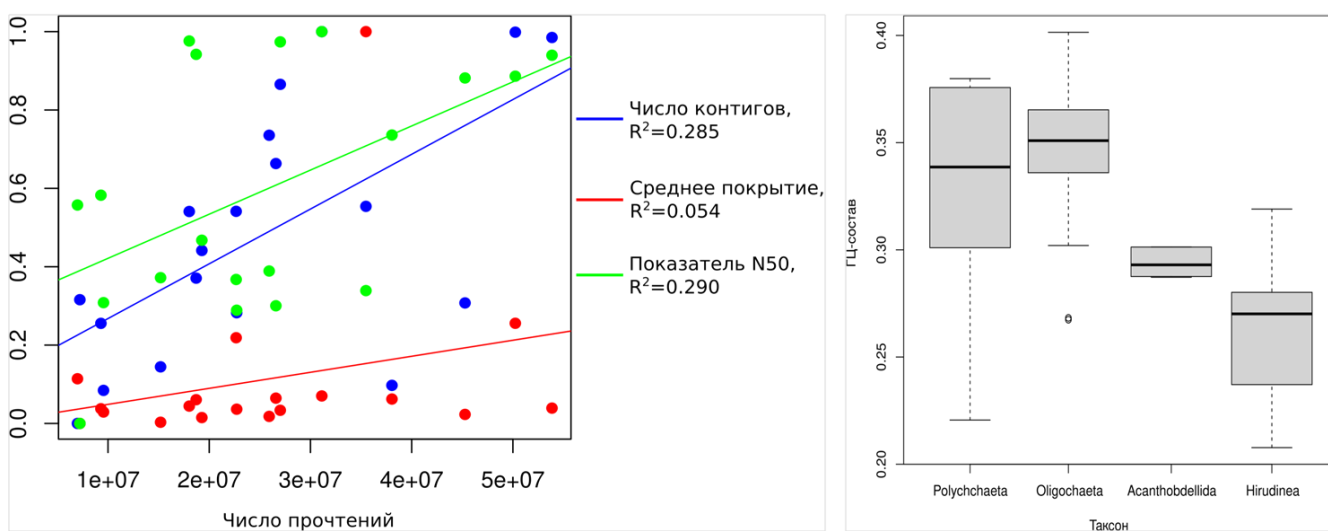
**Характеристика данных секвенирования нового поколения.** Ввиду отсутствия информации о размерах геномов исследуемых организмов секвенирование 12 образцов было проведено в два этапа. На первом этапе была секвенирована ДНК четырех образцов акантобделлид: Б35, Б52, Б53 и Б56. На втором этапе секвенирование проведено для остальных образцов акантобделлид и гирудинид: Б28, Б33, Б45, Б47, Б49, Б55, Б57 и Б58. Среднее число прочтений на образец на втором этапе секвенирования было в 2 раза



ниже, чем на первом, однако прочтений было достаточно, чтобы реконструировать полные митохондриальные геномы 11 из 12 образцов.

Для формирования максимально наполненной группы сравнения были загружены необработанные данные NGS 8 образцов из 7 видов поясковых кольцецов (*Lumbriculus variegates*, *Erpobdella octoculata*, *Glossiphonia complanata*, *G. concolor*, *Theromyzon tessulatum*, *Piscicola geometra* и два образца *Haemopsis sanguisuga*), митохондриальные геномы которых были опубликованы в GenBank в частичном виде, либо полностью отсутствовали. Митохондриальные геномы данных образцов были реконструированы в настоящей работе.

**Характеристика результатов геномной сборки.** Выходные данные сборки сырых прочтений в контиги были проанализированы на основные статистические показатели: число получившихся контигов, их среднее покрытие и показатель N50. Применение регрессионного анализа выявило, что только среднее покрытие контигов сильно зависит от числа прочтений в необработанных данных (Рисунок 1).



**Рисунок 1** – Зависимость статистических показателей результатов геномной сборки от количества необработанных прочтений. Статистические показатели нормализованы от 0 до 1 (слева). Значения доли гуанина и цитозина в митохондриальных геномах (справа).

Реконструкция полных митохондриальных геномов *Codonobdella* sp. Б47 (собственные данные) и *Haemopsis sanguisuga* SRX9009400 оказалась невозможной ввиду разрушения митохондриального генетического материала и преобладания в данных прочтений ядерного генома, вследствие чего невозстановимые пробелы митогеномов составляли 5,58% и 0,33% соответственно.

Доля ГЦ-оснований в митогеномах сформированного набора данных оказалась неоднородной (Таблица 1, Рисунок 2). Значения ГЦ-состава митогеномов реликтовых пиявок заняли промежуточное положение между таковыми для олигохет и пиявок. При этом отмечается тренд к снижению процента ГЦ-оснований в предполагаемом эволюционном направлении от Oligochaeta к Hirudinea.

Отличительной особенностью генома *A. peledina* является наличие в нём многократно повторяющегося участка, содержащего псевдогенные последовательности

*atp6* и *trnR*. В остальном набор и порядок генов *A. peledina* сходен с характерным для большинства других аннелид. В геномной последовательности малощетинкового червя *L. variegatus* наблюдается смена порядка генов, а именно перестановка генов *cytB*, *nad6*, *trnR*, *trnH*, *trnQ* и *trnY*. В обоих реконструированных геномах *H. sanguisuga* имеется перестановка генов *trnG* и *trnY*, а также отсутствует последовательность *trnR*. Геном *T. tessulatum* характеризуется перестановкой гена *trnR* и наличием вставки после *nad3*, которая может быть некодирующей, либо нести в себе дополнительный ген *trnS*.

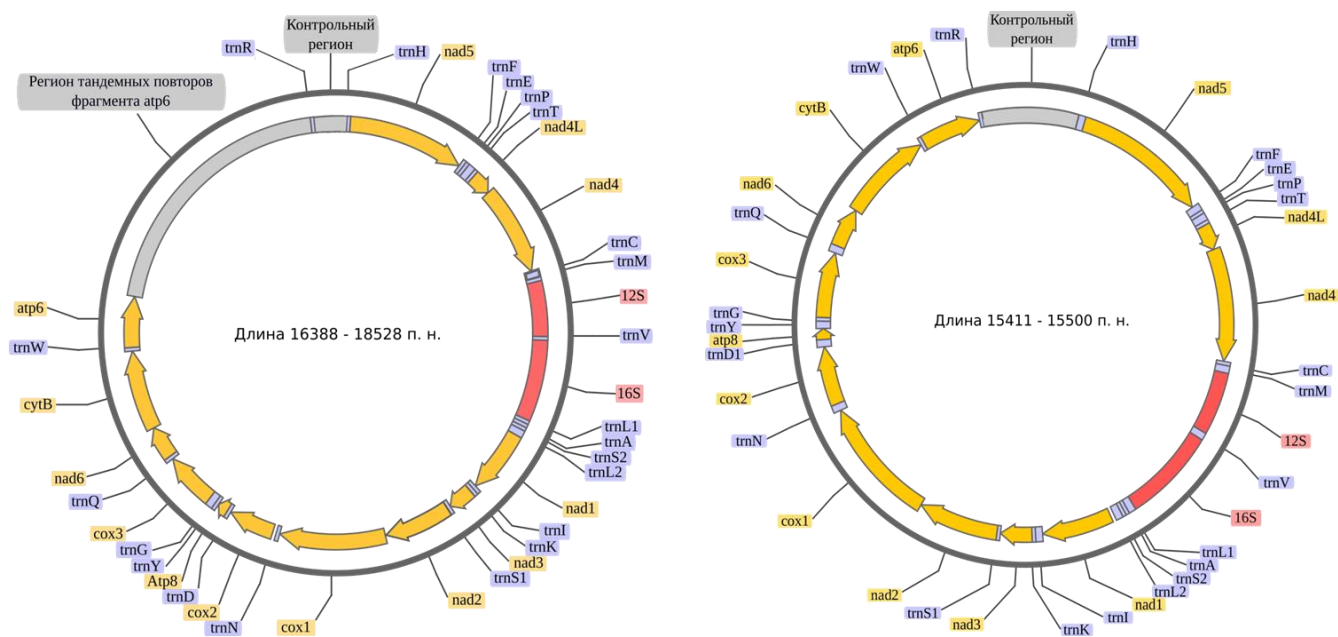
**Таблица 1** – Длина митохондриальных геномов исследуемых образцов

Вид	Номер образца	Длина генома	ГЦ-состав (%)
<i>Acanthobdella peledina</i> (Acanthobdellida)	Б28	18 528	30.13
	Б33	16 708	29.44
	Б35	17 499	29.55
	Б52	16 388*	28.76
	Б53	17 271*	28.72
	Б55	16 547*	29.06
<i>Paracanthobdella livanowi</i> (Acanthobdellida)	Б49	15 500	29.93
	Б56	15 411	29.94
<i>Codonobdella</i> sp.	Б45	14 486	24.48
	Б47	14 573**	23.63
<i>Baicalocleipsis grubei</i>	Б57	14 751	23.19
<i>Baicalocleipsis echinulata</i>	Б58	15 120	22.92
<i>Glossiphonia complanata</i>	SRX8928147	15 468	24.39
<i>Glossiphonia concolor</i>	SRX9009202	15 575	25.47
<i>Theromyzon tessulatum</i>	SRX8928146	15 913	23.63
<i>Piscicola geometra</i>	SRX9009199	14 788	21.73
<i>Erpobdella octoculata</i>	SRX9009198	15 580	27.50
<i>Haemopsis sanguisuga</i>	SRX9009141	14 530	22.15
<i>Haemopsis sanguisuga</i>	SRX9009400	14 462**	22.18
<i>Lumbriculus variegatus</i>	SRX9009164	15 883	31.94

\* - количество и последовательность повторяющихся регионов недостаточно разрешены; \*\* - последовательности имеют невосстановимые пробелы.

**Аннотация митогенома *Acanthobdella peledina*.** Митохондриальные геномы реликтовых пиявок *A. peledina* из географически разрозненных мест обитания имеют разную длину, варьирующую в диапазоне 16 388 – 18 528 п.н. Исследованные геномы содержат консервативный для большинства представителей Metazoa набор из 13 белок-кодирующих генов, 22 генов транспортных РНК, генов малой (*12S*) и большой (*16S*)

субъединиц рибосомальной РНК и контрольного региона. Все функциональные гены расположены на одной цепи (Рисунок 2).



**Рисунок 2** – Общая схема митохондриального генома представителей *A. peledina* (слева) и *P. livanowi* (справа). Жёлтым цветом обозначены белок-кодирующие гены, голубым – гены тРНК, красным – гены рРНК, серым – контрольный регион и регион псевдогенных тандемных повторов.

При сборке митохондриальных контигов вида *A. peledina* был обнаружен фрагмент с тандемными повторами фрагментов генов *atp6* и *trnR*. Подобные участки имеются во всех исследованных образцах *A. peledina*. Использование митохондриального генома образца Б28 в качестве референса показало, что соответствующий регион у разных образцов *A. peledina* имеет разную длину (зависит от количества копий) и крайнюю неконсервативность состава этих повторов. Исходя из этого, можно заключить, что повторы в составе этого региона, вероятнее всего, являются псевдогенами, появившимися в связи с неким рекомбинационным событием у общего предка вида *A. peledina*. Примечательно, что у близкородственного вида *P. livanowi* отсутствует такой участок, что, скорее всего, свидетельствует о появлении этой структуры уже после дивергенции этих двух видов. Другим потенциальным объяснением повышенного покрытия на данном участке может служить существование его копий в ядерном геноме (NUMTs). Однако среди прочтений, соответствующих данному региону, не было обнаружено таких, которые бы имели своим продолжением фрагмент, отличающийся от следующих повторяющихся прочтений или фрагментов функциональных генов *atp6* и *trnR*, что свидетельствует в пользу первого предположения.

**Аннотация митогенома *Paracanthobdella livanowi*.** Митохондриальные геномы двух образцов вида *P. livanowi*, обитающего на Камчатке, составляют в длину 15 411 и 15 500 п.н. Геномы *P. livanowi*, как и у сестринского вида *A. peledina*, характеризуются консервативным для большинства представителей Metazoa набором, состоящим из 37

функциональных генов, 13 из которых являются белок-кодирующими генами, 22 – генами транспортных РНК, 2 – генами малой (12S) и крупной (16S) субъединиц рибосомальной РНК, и контрольного региона. Все функциональные гены расположены на одной цепи (Рисунок 2). В отличие от митохондриального генома *A. peledina* в геноме камчатского вида отсутствует регион тандемных повторов фрагмента генов *atp6* и *trnR*.

Как и в случае митохондриальных геномов *A. peledina*, в геномах *P. livanowi* были обнаружены множественные перекрытия генов: гены *trnE* и *trnP* перекрываются на три нуклеотида; гены *trnC* и *trnM*, а также гены *trnY* и *trnG* перекрываются на один нуклеотид; гены *trnA* и *trnS2* перекрываются не только между собой, но и с окружающими их *trnL1* и *trnL2*; гены *nad4L* и *nad4* перекрываются на семь нуклеотидов, *nad2* и *cox1* – на 23 нуклеотида, а *nad6* и *cytB* – на 8 нуклеотидов.

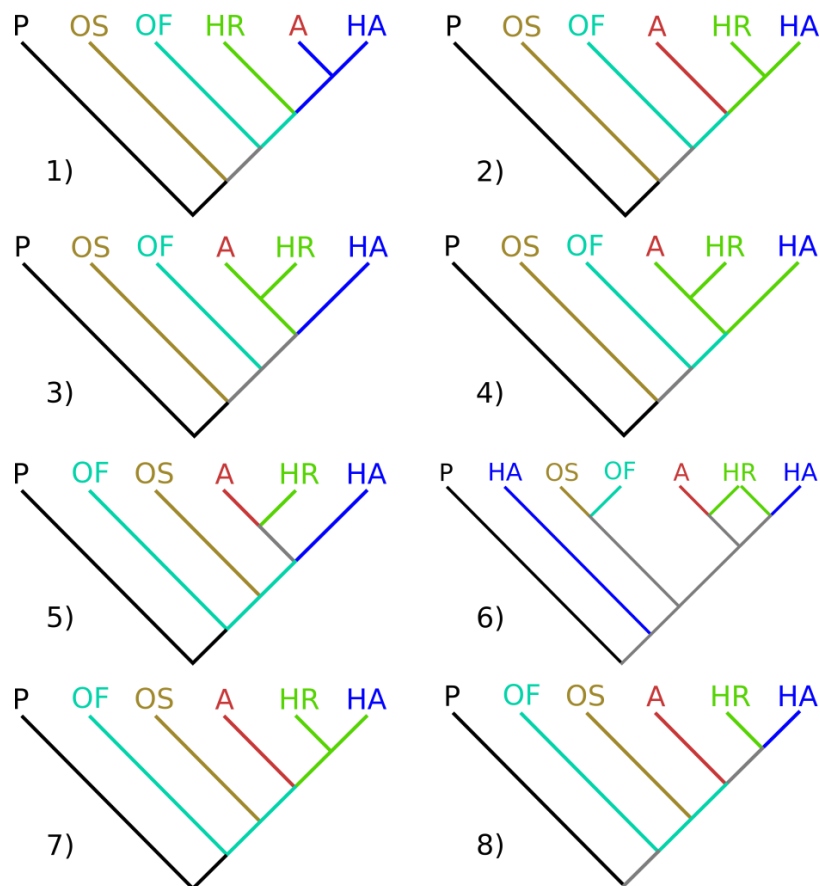
Порядок генов в митохондриальных геномах реликтовых пиявок *A. peledina* и *P. livanowi* отличается от порядка генов у *L. variegatus*, представителя сем. Lumbriculidae, потенциально наиболее близкой предковой группы относительно паразитических аннелид (пиявок и пиявкоподобных червей). Тем не менее, их порядок генов аналогичен характерному для большинства представителей класса Clitellata, включая почвенных олигохет. Таким образом, порядок генов может демонстрировать стабильность внутри родов и семейств, однако корреляция между ним и принадлежностью к более крупным таксонам (отрядам, классам) не очевидна или отсутствует.

#### ГЛАВА 4. ЭВОЛЮЦИЯ МИТОХОНДРИАЛЬНОГО ГЕНОМА АКАНТОБДЕЛЛИД

**Филогения на основе маркерных фрагментов.** В данном разделе был проверен результат реконструкции молекулярной филогении на основе коротких маркерных фрагментов митохондриальных генов *cox1* и *12S*. Получившиеся филогенетические схемы сильно различаются по топологии между собой и с современными представлениями о систематике и эволюции поясковых кольцецов (Рисунок 3.1-3.7). Большое количество узлов на всех филогенетических деревьях имеют низкие показатели бутстреп и апостериорных поддержек узлов на таксономическом уровне выше родов (Рисунок 5.1), что свидетельствует о нестабильности полученных топологий и неприменимости данных маркерных фрагментов генов для анализа древней филогении.

Применение алгоритма видовой делимитации GMYC к ультраметрическим байесовским деревьям на основе *cox1* показало наибольшее соответствие предсказанных таксономических групп идентифицированным видам. Применение этого алгоритма к деревьям на основе *12S* показало сдвиг порогового значения внутривидовой вариабельности к корню, и как следствие, чрезмерное объединение видов, в том числе, из разных родов, в оперативные таксономические единицы (ОТЕ).

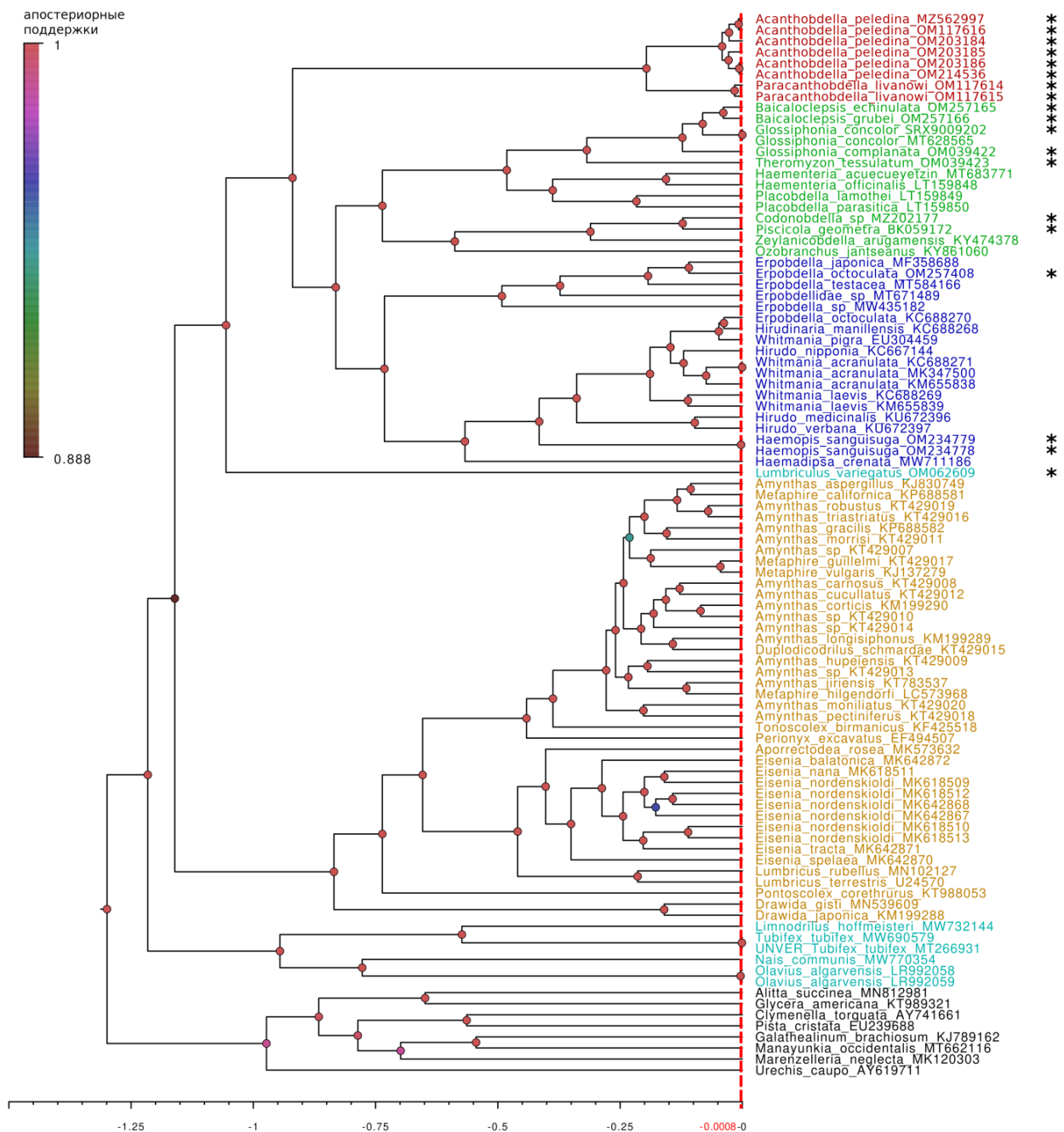
**Филогения митохондриальных геномов.** Реконструкция филогении на основе последовательностей полных митохондриальных геномов показала наибольшее сходство с современными представлениями об эволюции поясковых кольцецов (Рисунки 3.8, 4).



**Рисунок 3** – Упрощенные схемы кластеризации ключевых групп кольчатых червей (А – акантобделлиды, P – полихеты, OS – почвенные олигохеты, OF – водные олигохеты, HR – хоботные пиявки, HA – бесхоботные пиявки), выведенные из ML-деревьев, полученных на основе сравнения фрагментов различных геномных фрагментов: 1 – *cox1* ML; 2 – *cox1* ML+сегментация; 3 – *cox1* BI; 4 – *cox1* BI+сегментация; 5 – *12S* BI; 6 – *12S* BI+сегментация; 7 – *12S* ML и ML+сегментация; 8 – полный митохондриальный геном (все методы). Цвет ветвей отображает соответствие таксономическим группам и эволюционные гипотезы (моно-, пара- и полифилия).

Все полученные деревья имели максимально сходную топологию вне зависимости от использованного метода реконструкции и схемы расчета эволюционной модели (Рисунки 5.2). Их различия заключались лишь в порядке ветвления близкородственных организмов, тогда как кластеризация крупных таксонов оставалась неизменной. Значения бутстреп и апостериорных поддержек узлов при использовании полных митохондриальных геномов близки к 100%, что говорит о высокой стабильности топологии полученных деревьев.

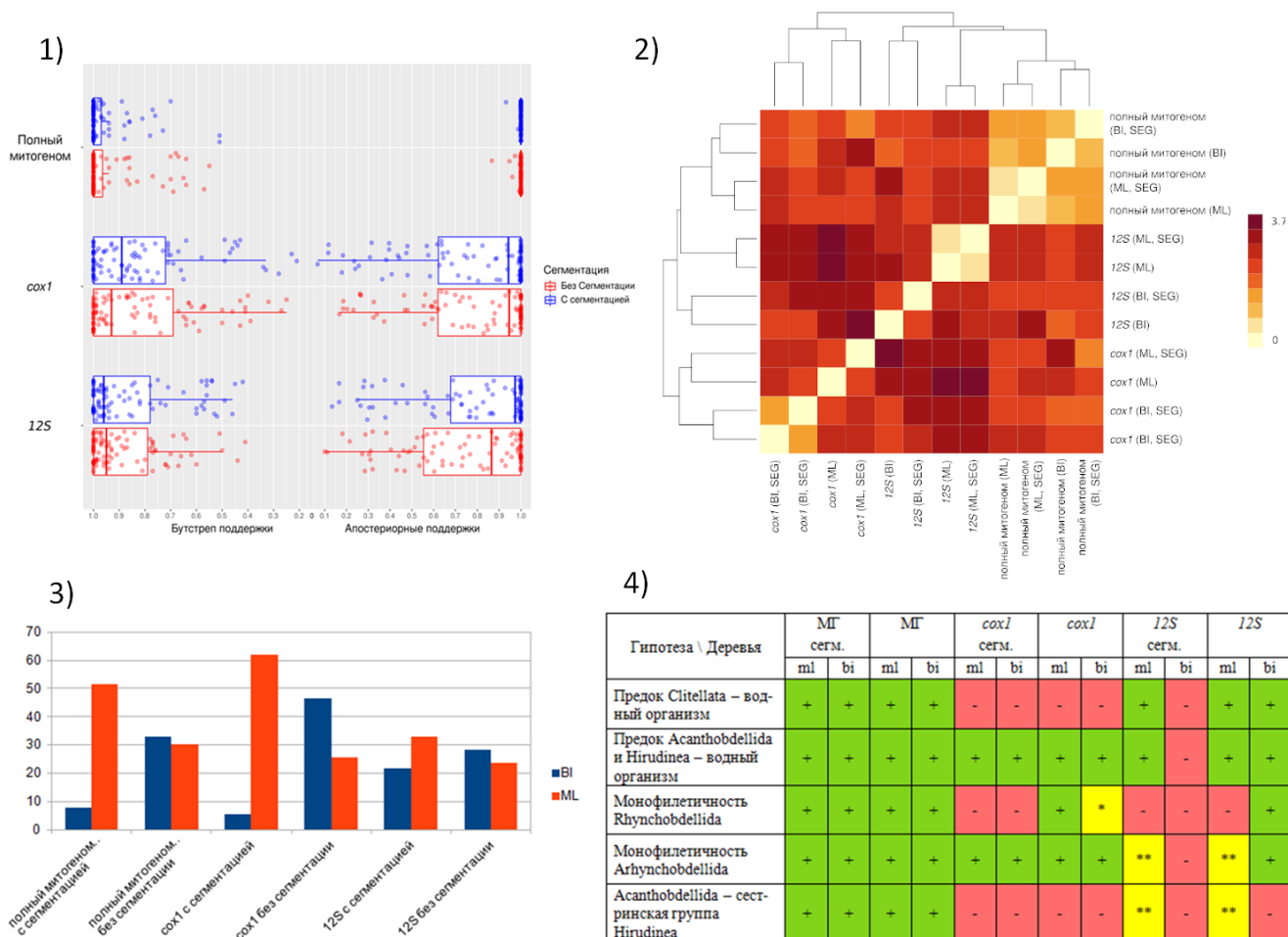
Применение GMYC к байесовским деревьям как на основе сегментированного (Рисунок 4), так и на основе несегментированного выравнивания показало чрезмерный сдвиг порогового значения внутривидовой варибельности к листьям, и как следствие, чрезмерное дробление организмов на ОТЕ. В группы были объединены только организмы, для которых длина разделяющих их ветвей близка к нулю. Таким образом, деревья, реконструированные на основе *cox1*, остаются наиболее оптимальными для делимитации видов, несмотря на то, что они не способны достоверно и стабильно отразить эволюцию таксонов высокого ранга.



**Рисунок 4** – Байесовское древо на основе несегментированного набора полных митогеномов. Красной линией отмечен порог внутривидовой варируемости по GMYC. Ярлыки образцов выделены цветом: красным – реликтовые пиявки, зеленым – хоботные, синим – бесхоботные (плоточные + челюстные), голубым – водные олигохеты, коричневым – почвенные олигохеты, черным – полихеты (внешняя группа). Звездочки указывают на данные, полученные в этой работе.

Важным результатом, полученным в ходе реконструкции, является монофилия отрядов Rhynchobdellida и Arhynchobdellida (Рисунок 4.8), которая отвергалась авторами ранних работ (Siddal et al. 1998, Arakupakul et al. 1999, Trontelj et al. 1999, Tessler et al. 2018) и была единожды подтверждена с использованием мультилокусного анализа на основе данных секвенирования нового поколения (Phillips et al. 2019). Данный факт указывает на повышение сходства и стабильности топологий филогенетических деревьев с

увеличением объёма использованных для реконструкции данных. В свою очередь, отряд *Arhynchobdellida* разделился на две клады, соответствующие челюстным и глоточным пиявкам (Рисунок 4). Кластеризация некоторых образцов внутри отряда *Arhynchobdellida* не соответствует их таксономической идентификации, а топология и длины ветвей внутри данной клады указывают на необоснованность деления группы челюстных на отдельные семейства.



**Рисунок 5** – Сравнение различных показателей реконструированных деревьев. 1 – бутстреп (слева) и байесовские (справа) поддержки узлов на различных деревьях; 2 – сходство топологий на основе узловых дистанций; 3 – суммы длин ветвей при использовании разных методов; 4 – соответствие деревьев различным гипотезам об эволюции поясковых колец.

**Сравнение результатов филогенетической реконструкции.** Реконструкция молекулярной филогении на основе трёх разных генетических фрагментов показала различные топологии филогенетических деревьев и различные уровни их статистической поддержки. Наибольшее сходство с современными представлениями об эволюции и таксономии поясковых колец продемонстрировало древо на основе полных митохондриальных геномов. В этих же деревьях значения бутстреп и байесовских апостериорных поддержек узлов имели наивысшие значения в сравнении с остальными (Рисунок 5.1).

Реконструкция филогенетических деревьев на основе коротких маркерных генов *cox1* и *12S* показала расхождение с современной систематикой в топологии всех

деревьев. Данные маркерные фрагменты показывают высокую разрешающую способность на уровне видов и родов, как видно из высоких бутстреп и апостериорных поддержек узлов на этих уровнях, однако неприменимы для анализа древней филогении.

Сегментация данных и расчёт отдельных моделей замен для сегментов, сгруппированных программой PartitionFinder2, оказывает значительно больший эффект на длины ветвей деревьев, чем на их топологию. При реконструкции филогении по байесовскому методу сегментация значительно сокращает длины ветвей для всех использованных генных фрагментов. Противоположный эффект наблюдается при реконструкции по методу максимального правдоподобия (Рисунок 5.3).

Полученные филогенетические деревья в разной мере соответствуют различным гипотезам об эволюции кольчатых червей. Соответствие деревьев разным гипотезам суммированы на рисунке 5.4.

При применении алгоритма GMYC предсказанные оперативные таксономические единицы (ОТЕ) соответствовали морфологически идентифицированным видам только в случае деревьев на основе маркерного фрагмента *cox1*. Анализ деревьев на основе маркерного фрагмента *12S* показал чрезмерную кластеризацию образцов, а результат на основе деревьев по полным митохондриальным геномам – чрезмерное их разделение.

## ЗАКЛЮЧЕНИЕ

В рамках данной работы проведено полногеномное секвенирование, реконструкция и аннотация последовательностей митохондриальных геномов 12 образцов, принадлежащих видам *Acanthobdella peledina*, *Paracanthobdella livanowi*, *Codonobdella* sp., *Baicaloclepsis grubei* и *Baicaloclepsis echinulata*. Дополнительно заново реконструированы и аннотированы митохондриальные геномы 8 аннелид, необработанные данные которых доступны в базе SRA.

Большая часть доступных на сегодняшний день митогеномов поясковых кольцецов (Clitellata) имеет сходную длину (около 16 тыс. нуклеотидов) и консервативный набор из 37 генов, характерный для большинства Metazoa. У некоторых рассмотренных групп выявлены перестановки функциональных генов, по-видимому, имеющие таксон-специфический паттерн. Во всех аннотированных геномах обнаружены случаи перекрытия функциональных генов, что согласуется с литературными данными.

В митохондриальных геномах реликтовых пиявок *A. peledina* обнаружен уникальный регион с многократными тандемными повторами фрагментов генов *atp6* и *trnR*, имеющий неконсервативную длину и структуру у образцов из разных географических районов Северной Евразии, что, скорее всего, свидетельствует о его нефункциональности. Подобный регион отсутствует у сестринского вида *P. livanowi*, что указывает на его формирование после дивергенции этих видов.

На основе полученных данных был проведён анализ эволюционных взаимоотношений различных таксономических групп кольчатых червей, а также влияния заданных начальных параметров анализа на характеристики конечных филограмм. Использование коротких маркерных последовательностей генов *cox1* и *12S* показало



расхождение топологий полученных деревьев с рядом представлений об эволюции поясковых кольцецов. При этом ряд значений бутстреп и байесовских апостериорных поддержек имеют очень низкие значения, что указывает на недостоверность полученных топологий. Наибольшее соответствие современным представлениям о систематике и эволюции кольчатых червей показали деревья на основе полных митохондриальных геномов. Независимо от метода реконструкции и применения или неприменения сегментации набора данных, филогенетические деревья различались лишь положением близкородственных организмов. Примечательно, что несоответствие положения образца *Erpobdella octoculata* КС688270 своему таксономическому статусу на данных филограммах согласуется с результатами анализа порядка генов, указывающего на большее родство данного образца с пиявками семейства Hirudinidae, что говорит о высокой вероятности его неверной идентификации авторами последовательности, выложенной в GenBank. Кроме того, топология всех деревьев на основе полных митогеномов поддерживает монофилию отрядов Rhynchobdellida и Arhynchobdellida, оспаривая результаты ранних филогенетических исследований на основе малого количества коротких генетических маркеров и подтверждая выводы, полученные на основе анализа массивных данных (свыше 50 тыс. нуклеотидов). Анализ полных митохондриальных геномов подтвердил подразделённость отряда бесхоботных пиявок (Arhynchobdellida) на монофилетические клады, соответствующие подотрядам плоточных (Erpobdelliformes) и челюстных (Hirudiniformes) пиявок. На уровне семейств выявлены проблемы классификации последних. Так, согласно современной классификации, роды *Haemopsis* и *Whitmania* принадлежат семейству Haemoridae. Однако при реконструкциях на основе полных митохондриальных геномов они формируют полифилетическую группу, разделённую представителями рода Hirudinidae.

Топологии филограмм на основе фрагмента гена *cox1* показали несоответствие систематике высоких таксонов поясковых кольцецов. Применение алгоритма видовой делимитации GMYC к данным деревьям выявило наибольшее соответствие предсказанных таксономических единиц идентифицированным видам. Следовательно, фрагмент *cox1* может быть успешно использован для разграничения криптических видов, однако он должен с осторожностью применяться для филогенетического анализа.

Установлено, что главным фактором, влияющим на топологию филогенетических деревьев, является использованный для их реконструкции метод. Деревья на основе одинаковых наборов молекулярно-генетических данных, реконструированные разными методами, могут больше отличаться друг от друга, чем от деревьев, реконструированных на основе других маркеров. Вторым по важности фактором оказалась природа использованного генетического фрагмента. Применение сегментации оказывало минимальное влияние на топологию филограмм.

В данной работе были использованы все доступные митогеномные данные организмов, принадлежащих к потенциально предковым группам относительно пиявок и пиявкоподобных паразитов. Однако на сегодняшний день представленность этой таксономической группы в GenBank остаётся довольно скудной. В перспективе для

построения более подробной картины эволюции Clitellata требуется пополнение базы данных последовательностями митохондриальных геномов водных олигохет. Особый интерес вызывают митогеномы представителей сем. Lumbriculudae, морфологически близких реликтовым пиявкам. Получение отсутствующих на сегодняшний день данных о структуре митогеномов Branchiobdellida даст более детальное представление об эволюции поясковых кольцецов.

Использованный в данной работе метод секвенирования геномов не специфичен к происхождению генетического материала, поэтому полученный массив данных, кроме митохондриальных последовательностей, включает и массу прочтений ядерного генома. Эти материалы не были использованы в данной работе, но в дальнейшем могут быть пригодны для получения более детальных сведений о механизмах адаптации организмов к их среде обитания и экологической нише, а также для поиска метаболических путей, образующих биологически-активные соединения, потенциально полезные для медицины и биотехнологии.

### ВЫВОДЫ

1. Впервые у кольцецов обнаружен регион с тандемными повторами псевдогенных последовательностей *atp6* и *trnR*, который является уникальной структурной особенностью вида *Acanthobdella peledina* и причиной вариабельности размера митохондриального генома у образцов из географически разрозненных популяций.
2. Показано, что различия в длине нуклеотидных последовательностей митохондриальных геномов камчатских образцов реликтовых пиявок *Paracanthobdella livanowi* связаны с индивидуальной изменчивостью в контрольном регионе митогенома и не затрагивают функциональных генов.
3. Впервые отмечено, что, несмотря на крайне консервативный набор функциональных митохондриальных генов у поясковых кольцецов, порядок их расположения имеет таксон-специфический паттерн на уровне семейств.
4. Использование полных митохондриальных геномов для построения филогении демонстрирует увеличение стабильности топологий филограм, что отражается в повышении статистических поддержек узлов, по сравнению с короткими фрагментами маркерных генов вне зависимости от метода реконструкции и сегментации набора данных.
5. Реконструкция эволюционной истории на основе полных митохондриальных геномов подтвердила гипотезу о древних щетинконосных пиявках (*Acanthobdellida*) как промежуточной форме между *Oligochaeta* и *Hirudinea*.

### СПИСОК ОСНОВНЫХ РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Болбат А. В.**, Кайгородова И. А., Букин Ю. С., Федорова Л. И., Сороковикова Н. В. Применение биоинформационных методов для определения границ видов пиявок рода *Erpobdella*. Известия Иркутского Государственного Университета. Серия «Биология. Экология» 2017; 20: 3-13. IF=0,257 (РИНЦ)

2. **Болбат А. В.**, Сороковикова Н. В., Кайгородова И. А. Оценка эффективности использования митохондриального маркера 12S для реконструкции филогении акантобделлид. Генетика 2019; 55: 1461-1465. IF=1.188 (РИНЦ)  
**Bolbat A. V.**, Sorokovikova N. V., Kaygorodova I. A. Assessing efficiency of the mitochondrial 12S marker fragment for the use in reconstruction of the phylogeny of Acanthobdellids. Russian Journal of Genetics 2019; 55: 1554-1558 IF=0.581 (WoS, Scopus)
3. Kaygorodova I., Bolbat N., **Bolbat A.** Species delimitation through DNA barcoding of freshwater leeches of the *Glossiphonia* genus (Hirudinea: Glossiphoniidae) from Eastern Siberia, Russia. Journal of Zoological Systematics and Evolutionary Research 2019; 58: 1437-1446. IF=2.288 (WoS, Scopus)
4. **Bolbat A.**, Vasiliev G., Kaygorodova I. The first mitochondrial genome of the relic *Acanthobdella peledina* (Annelida, Acanthobdellida). Mitochondrial DNA Part B: Resources 2020; 5: 3300-3301. IF=0.658 (WoS, Scopus)
5. **Bolbat A.**, Matveenko E., Dzyuba E., Kaygorodova I. The first mitochondrial genome of *Codonobdella* sp. (Hirudinea, Piscicolidae), a new endemic leech species from Lake Baikal, Russia and reassembly of the *Piscicola geometra* data from SRA. Mitochondrial DNA Part B: Resources 2021; 6: 3112-3113. IF=0.658 (WoS, Scopus)

#### МАТЕРИАЛЫ И ТЕЗИСЫ КОНФЕРЕНЦИЙ

1. **Болбат А. В.**, Кайгородова И. А., Букин Ю. С. Применение биоинформационных методов в молекулярной экологии. Вестник Иркутского университета 2017; 20: 39-40.
2. Mandzyak N. B., Kaygorodova I. A., **Bolbat A. V.** DNA barcoding of Eastern Siberian flat leeches (Hirudinea: Glossiphoniidae). Материалы 5-ой межд. конф. «Molecular Phylogenetics and Biodiversity Biobanking»; Москва, 2018; С. 93.
3. **Bolbat A.**, Sorokovikova N., Kaygorodova I. Estimation of 12s marker fragment effectiveness for ancient phylogeny reconstruction. Материалы 11-ой межд. школы молодых учёных «Systems Biology and Bioinformatics (SBB-2019)»; Новосибирск, 2019; С. 10.
4. Bolbat N., **Bolbat A.**, Kaygorodova I. DNA barcode-based delimitation of the *Glossiphonia* species. Материалы 11-ой межд. школы молодых учёных «Systems Biology and Bioinformatics (SBB-2019)»; Новосибирск, 2019; С. 11.
5. **Bolbat A.**, Vasiliev G., Kaygorodova I., Bogdanova V. New data on Acanthobdellida phylogeny based on complete mitochondrial genomes. Материалы 12-ой межд. мультikonференции «Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2020)»; Новосибирск, 2020; С. 202-203.
6. **Болбат А. В.**, Болбат Н. Б., Васильев Г. В., Богданова В. С., Матвеев Е. Ю., Кайгородова И. А. Эффект использования полных митохондриальных геномов и их отдельных фрагментов для делимитации видов. Материалы 12-ой межд. школы молодых учёных «Systems Biology and Bioinformatics (SBB-2020)»; Ялта-Севастополь, 2020; С. 65.
7. **Болбат А. В.**, Кайгородова И. А. 170 лет изучения реликтовых паразитов: итоги и актуальные вопросы. Материалы 4-ой всерос. конф. с межд. участием «Разнообразие почв и биоты Северной и Центральной Азии»; Улан-Удэ, 2021; С. 76.