



МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Николаева Ольга Владимировна

Дивергентные митогеномы беспозвоночных животных

03.01.09 – Математическая биология, биоинформатика

АВТОРЕФЕРАТ

на соискание ученой степени
кандидата биологических наук

Москва – 2022

Работа выполнена в отделе эволюционной биохимии Научно-исследовательского института физико-химической биологии имени А. Н. Белозерского Федерального государственного бюджетного учреждения высшего образования «Московский государственный университет имени М. В. Ломоносова».

Научный руководитель – *Алешин Владимир Вениаминович,*
доктор биологических наук

Официальные оппоненты – *Животовский Лев Анатольевич,*
доктор биологических наук, профессор, заведующий лабораторией генетических проблем идентификации Федерального государственного бюджетного учреждения науки «Институт общей генетики имени Н.И. Вавилова Российской академии наук».

Холодова Марина Владимировна,
доктор биологических наук, руководитель кабинета методов молекулярной диагностики Федерального государственного бюджетного учреждения науки «Институт проблем экологии и эволюции имени А.Н. Северцова Российской академии наук».

Римская-Корсакова Надежда Николаевна,
кандидат биологических наук, ведущий научный сотрудник кафедры зоологии беспозвоночных Биологического факультета Федерального государственного бюджетного учреждения высшего образования «Московский государственный университет имени М. В. Ломоносова».

Защита диссертации состоится 22 июня 2022 года в 15:00 на заседании диссертационного совета МГУ.03.04 Московского государственного университета имени М.В. Ломоносова по адресу: 119234, Москва, Ленинские горы, д. 1, стр. 73, Факультет биоинженерии и биоинформатики, ауд. 221.

E-mail: dissovet@belozersky.msu.ru

С диссертацией можно ознакомиться в отделе диссертаций Научной библиотеки МГУ имени М.В. Ломоносова (Москва, Ломоносовский просп., д. 27) и на сайте ИАС «ИСТИНА»: <https://istina.msu.ru/dissertations/450678663/>

Автореферат разослан «___» мая 2022 г.

Ученый секретарь диссертационного совета,
кандидат химических наук



Шаповалова И.В.

Общая характеристика работы

Актуальность и степень разработанности темы исследования.

Филогенетический анализ позволяет определять порядок ветвления и иногда даже время расхождения видов, благодаря чему возможно реконструировать сценарии происхождения различных таксономических групп от их предков. Молекулярная филогенетика исследует взаимоотношения между организмами или генами путем сравнения гомологичных нуклеотидных или аминокислотных последовательностей. Благодаря широкому внедрению методов высокопроизводительного секвенирования произошло резкое увеличение количества доступных для анализа данных, что открывает новые возможности для решения филогенетических задач.

Задача определения родства между различными организмами сложна, поскольку фенотипически похожие организмы могут иметь совершенно различные биохимические и молекулярные признаки. Поэтому при недостатке фенотипических признаков есть опасность, что биологически разные виды будут сведены в один таксономический вид либо наоборот, разные жизненные стадии одного организма будут описаны как разные виды. Использование молекулярных признаков в качестве филогенетических маркеров позволяет преодолевать возможные ошибки в систематике и таксономии. Иногда данных о первичной структуре оказывается недостаточно и приходится применять дополнительные методы, которые могли бы привнести новую информацию для анализа филогении. К ним можно отнести исследования вторичной структуры тРНК, интронов и рибосомных спейсеров, взаимного расположения интронов, порядка генов митохондриального генома, поиск синапоморфий и компенсаторных замен и т.д. К решению каждой конкретной задачи нужно применять творческий подход, выражающийся в выборе типа анализа, подборе видов для анализа и применяемой модели построения филогенетического дерева.

Митохондриальные геномы (митогеномы) широко используются в филогенетике ввиду ряда особенностей их структуры и эволюции. В меньшей степени митогеномы используют для реконструкции крупномасштабной филогении. Это обусловлено рядом причин, одна из которых – быстрая эволюция (дивергенция) митохондриальных геномов. То же обстоятельство, которое является преимуществом для анализа микроэволюционных событий, препятствует адекватному сравнению далеких видов. По сравнению с ядерными геномами, митохондриальные геномы доступны для большего числа видов, что позволяет

не только изучать структуру геномов отдельных видов, но и выдвигать гипотезы о функционировании мтДНК на основе сравнения больших выборок митогеномов, а также исследовать филогению крупных таксонов, таких как Metazoa. Хотя митогеномы животных содержат, как правило, один и тот же набор генов, порядок их расположения различен у крупных таксономических групп. Митохондриальные гены отличаются высокой консервативностью набора в сочетании с высокой частотой мутаций и наличием как консервативных, так и переменных сайтов, что упрощает подбор праймеров и обеспечивает значимый филогенетический сигнал. Особенно большие затруднения для филогенетического анализа вызывают т. н. дивергентные митогеномы – те, эволюция которых проходила быстрее, чем митогеномов «в среднем», что может отражаться в особенностях их структуры.

Целью данной работы является изучение структурных и эволюционных особенностей дивергентных митогеномов беспозвоночных животных и применимости различных типов анализа молекулярных данных для поиска родственных связей на основе митохондриальной ДНК.

В рамках работы были поставлены следующие задачи:

- Аннотировать митохондриальные геномы.
- Сравнить порядок генов дивергентных и прочих митогеномов.
- Провести филогенетический анализ по белок-кодирующим генам митохондриальных геномов.
- Провести тестирование различных дополнительных филогенетических методов, в том числе ориентированных на учет вторичной структуры нуклеиновых кислот.
- Определить положение группы Orthonectida на филогенетическом древе Metazoa и исследовать ее родственные связи с Dicyemida.

Объект и предмет исследования. Объектом исследования данной диссертационной работы являются восемь митохондриальных геномов беспозвоночных животных: киноринх (*Echinoderes svetlanae* и *Pycnophyes kielensis*), ортонектид (*Intoshia linei*), дициемид (*Dicyema* sp.) волосатиков (*Gordionus alpestris*, *Gordionus wolterstorffii*, *Gordius* sp. и *Chordodes* sp.). Предметом исследования в диссертации является аннотация митогеномов и проведение филогенетических исследований по данным их первичной и вторичной структур.

Научная новизна. В ходе работы были впервые собраны и аннотированы восемь полных или почти полных митохондриальных геномов беспозвоночных животных.

Полученные данные были использованы для предсказания аминокислотных последовательностей кодируемых белков, тРНК и рРНК, а также проведения филогенетического анализа.

Теоретическая и практическая ценность исследования. Полученные результаты расширяют выборку доступных для анализа последовательностей митохондриальных геномов беспозвоночных животных, в том числе предсказанных аминокислотных последовательностей митохондриальных белков и вторичных структур тРНК. Найден метод оценки частот гомоплазии одиночных аминокислот и порядка митохондриальных генов с применением базы данных RefSeq.

Методы исследования. В работе использованы молекулярно-генетические (секвенирование ДНК и РНК, обработка ДНК рестриктазами и нуклеазами), биоинформатические (аннотация митогеномов, определение первичных и вторичных нуклеотидных последовательностей, анализ порядка генов, филогенетический анализ, анализ синапоморфий) и статистические методы научного исследования.

Основные положения, выносимые на защиту:

1. Вероятность независимого появления одинаковых аминокислот в митохондриальных белках, оцененная по базе данных RefSeq, может быть применена в филогенетическом анализе.
2. Применение модели сайт-специфичных замен (CAT) в филогенетическом анализе позволяет снизить или преодолеть эффект притяжения длинных ветвей (LBA).
3. Синапоморфии (сходство по производному состоянию признака) одиночных аминокислот и порядка расположения митохондриальных генов могут применяться как дополнительный метод при анализе родства даже в случае сильно дивергированных последовательностей.
4. Ортонектиды и дициемиды не группируются в монофилетичный таксон Mesozoa.

Степень достоверности данных. Статистическая обработка данных проводилась с использованием программы RStudio (США). Результаты исследования доложены на международных конференциях и опубликованы в рецензируемых журналах.

Публикации. По материалам диссертации опубликовано 6 печатных работ, в том числе, 3 статьи в журналах Scopus и Web of Science, в т.ч. 2 статьи в журналах Q1.

Апробация диссертации. Результаты работы представлены на Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2016» (Москва, Россия, 2016), VI Всероссийской конференции с международным участием

«Школа теоретической и морской паразитологии» (Севастополь, Россия, 2016) и V Moscow International Conference «Molecular phylogenetics and biodiversity biobanking» (Москва, Россия, 2018).

Личный вклад автора. Автору принадлежит ключевая роль в постановке целей и задач диссертационного исследования. Личный вклад автора в статьи, опубликованные по материалам диссертации:

1. **Popova¹ (Nikolaeva) O.V.**, Mikhailov K.V., Nikitin M.A., Logacheva M.D., Penin A.A., Muntyan M.S., Kedrova O.S., Petrov N.B., Panchin Y.V., Aleoshin V.V. «Mitochondrial genomes of Kinorhyncha: *trnM* duplication and new gene orders within animals» в журнале *PLoS ONE*, издательство *Public Library of Science (United States)*, 2016, том 11, № 10, с. e0165072. IF = 2.806, квартиль Q1 (здесь и далее квартиль указана на момент публикации). О.В. Попова аннотировала митохондриальные геномы двух видов киноринх, определила их нуклеотидные составы и аминокислотные составы предсказанных белок-кодирующих генов, провела анализ порядка генов, составила модель тРНК с учетом вторичной структуры, провела филогенетический анализ на основе данных по белок-кодирующим генам и генам тРНК, провела визуализацию генетических карт митохондриальных геномов и филогенетических деревьев, написала исходный черновик статьи, подготовила статью к подаче в журнал и участвовала в переписке с редактором и рецензентами статьи. К.В. Михайлов, М.А. Никитин, М.Д. Логачева, А.А. Пенин, М.С. Мунтян, О.С. Кедрова, Н.Б. Петров, Ю.В. Панчин и В.В. Алешин собрали материал, выделили ДНК, подготовили библиотеки для секвенирования, провели NGS, собрали митохондриальные геномы и проверили используемый генетический код.
2. Zverkov O.A., Mikhailov K.V., Isaev S.V., Rusin L.Y., **Popova (Nikolaeva) O.V.**, Logacheva M.D., Penin A.A., Moroz L.L., Panchin Y.V., Lyubetsky V.A., Aleoshin V.V. «Dicyemida and Orthonectida: Two stories of body plan simplification» в журнале *Frontiers in Genetics*, издательство *Frontiers Research Foundation. (Lausanne, Switzerland)*, 2019, том 10, № 443. IF = 3.260, квартиль Q2. О.В. Попова определила первичную структуру участка митохондриальной ДНК *Intoshia linei* с длинным инвертированным повтором (0.4 т.п.н.) с помощью капиллярного секвенирования, аннотировала митохондриальный геном *I. linei*, картировала границы генов и псевдогенов в кольцевых мини-репликонах диспергированного митохондриального генома *Dicyeta* spp., определила нуклеотидные составы геномов и аминокислотные составы предсказанных белок-кодирующих генов,

¹ Попова (Ророва) - девичья фамилия диссертанта до 25.08.2021.

предсказала вторичные структуры тРНК, провела анализ порядка генов *I. linei*, анализировала экспрессию митохондриальных генов *I. linei*, провела визуализацию генетических карт митохондриальных геномов, участвовала в подготовке статьи к подаче в журнал и переписке с редактором и рецензентами статьи. О.А. Зверков, К.В. Михайлов, С.В. Исаев, Л.Ю. Русин, М.Д. Логачева, А.А. Пенин, Л.Л. Мороз, Ю.В. Панчин, В.А. Любецкий и В.В. Алешин собрали материал, выделили ДНК и РНК, подготовили библиотеки для секвенирования, провели NGS, собрали митохондриальные геномы и транскриптом, провели филогенетический анализ.

3. Mikhailov K.V., Efeykin B.D., Panchin A.Y., Knorre D.A., Logacheva M.D., Penin A.A., Muntyan M.S., Nikitin M.A., **Popova (Nikolaeva) O.V.**, Zanagina O.N., Vyssokikh M.Y., Spiridonov S.E., Aleoshin V.V., Panchin Y.V. «Coding palindromes in mitochondrial genes of Nematomorpha» в журнале *Nucleic Acids Research*, издательство *Oxford University Press (United Kingdom)*, 2019, том 47, № 13, с. 6858-6870. IF = 11.502, квартиль Q1. О.В. Попова обрабатывала материал перед выделением ДНК, провела эксперимент по реассоциации ДНК *Gordionus alpestris* и обработке нуклеазой Mung bean с целью обогащения нативной ДНК инвертированными повторами, аннотировала митохондриальные геномы четырех видов волосатиков, определила их нуклеотидные составы и аминокислотные составы предсказанных белок-кодирующих генов, анализировала экспрессию митохондриальных генов *G. alpestris*, провела визуализацию генетических карт митохондриальных геномов, участвовала в подготовке статьи к подаче в журнал и переписке с редактором и рецензентами статьи. К.В. Михайлов, Б.Д. Ефейкин, А.Ю. Панчин, Д.А. Кнорре, М.Д. Логачева, А.А. Пенин, М.С. Мунтян, М.А. Никитин, О.Н. Занегина, М.Ю. Высоких, С.Э. Спиридонов, В.В. Алешин и Ю.В. Панчин собрали материал, выделили ДНК и РНК, подготовили библиотеки для секвенирования, провели NGS, собрали митохондриальные геномы и транскриптом, измерили дыхательную активность тканей, провели филогенетический анализ.

Объем и структура диссертации. Диссертация состоит из введения, списка сокращений, обзора литературы, описания материалов и методов, результатов собственных исследований и их обсуждения, заключения, выводов, списка литературы. Работа изложена на 151 странице, содержит 32 рисунка и 23 таблицы. Список литературы включает 472 источника, из них 9 отечественных и 463 зарубежных.

Результаты и обсуждение

1.1. Структурная организация митохондриальных геномов. Киноринхи *Echinoderes svetlanae* и *Pycnophyes kielensis*. Митогеномы *Echinoderes svetlanae* и *Pycnophyes kielensis* представляют собой кольцевые молекулы ДНК, содержащие типичный для Metazoa набор генов и по одному дополнительному гену метиониновой тРНК с идентичными антикодонами (Рис. 1).

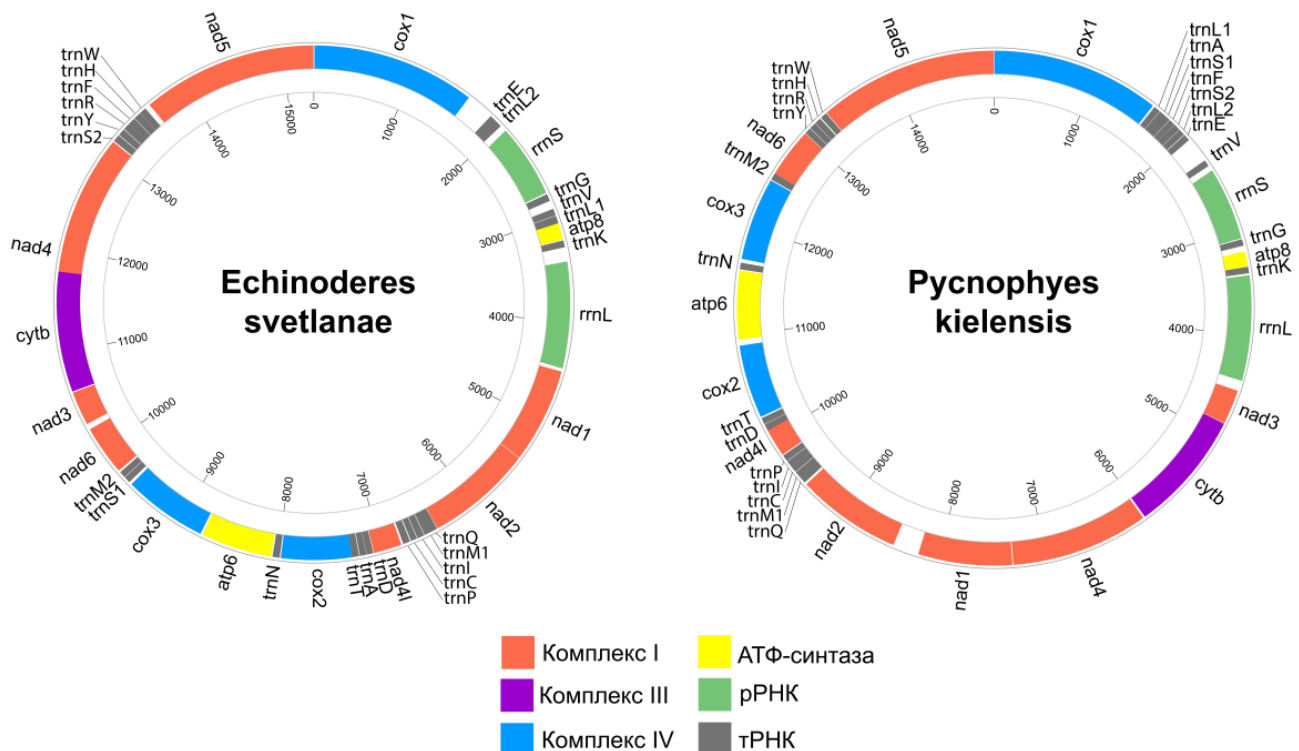


Рис. 1. Митохондриальные геномы *Echinoderes svetlanae* и *Pycnophyes kielensis*.

При сравнении генов *trnM* между двумя видами киноринх в *trnM2* были найдены две компенсаторные замены, в *trnM1* — одна (Рис. 2a). Наличие компенсаторных замен свидетельствует в пользу того, что обе метиониновые тРНК каждого вида являются функциональными и кодируются не псевдогенами. Построенное по генам тРНК с учетом вторичной структуры байесово дерево показало сходство между большинством изоакцепторных тРНК обоих видов, включая две пары метиониновых тРНК (Рис. 2b). *trnM1* и *trnM2* группируются между собой у разных видов, что указывает на то, что появление дополнительного гена было обусловлено дубликацией гена метиониновой тРНК перед дивергенцией *P. kielensis* и *E. svetlanae* — представителей двух разных отрядов киноринх. Функциональная значимость дубликации не ясна, но она может быть связана с

разницей функций метиониновых тРНК в инициации трансляции и при элонгации синтезируемого белка.

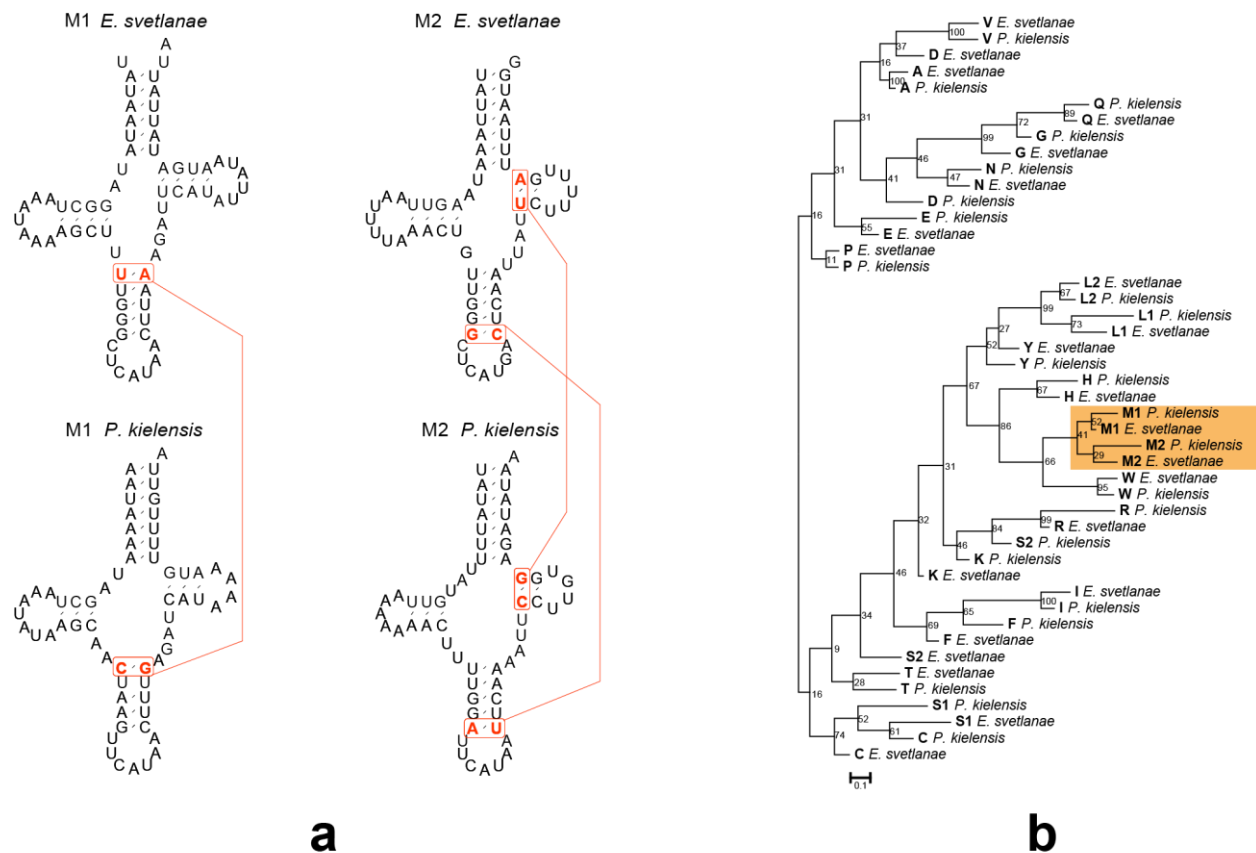


Рис. 2. Структура и филогения дублированных метиониновых тРНК киноринх.

- (a) Предсказанные вторичные структуры метиониновых тРНК *Echinoderes svetlanee* and *Pycnophyes kiensis*. Компенсаторные замены обозначены красным.
- (b) Байесово дерево, построенное по выравниванию генов тРНК киноринх. Числа на ветвях обозначают апостериорные байесовы вероятности. Гены метиониновых тРНК выделены оранжевым. тРНК обозначены в соответствии с однобуквенной кодировкой.

формирования шпильчатых структур *in vivo*. В некоторых случаях инвертированные повторы перекрываются, что может свидетельствовать о существовании сложных комплексных структур с конкурирующими шпильками и прямыми повторами, напоминающими структуру бактериального аттенуатора. Участки повторов имеют сниженное покрытие ридами (=чтениями при использовании методов геномного секвенирования) относительно других участков митогенома, что приводит к проблемам в процессе сборки. Поэтому само существование повторов нуждалось в дополнительном подтверждении. Для этого тотальная ДНК *Gordius* sp. была подвергнута быстрой денатурации-ренатурации с последующей обработкой нуклеазой Mung bean для удаления одноцепочечных фрагментов ДНК. Поскольку палиндромные участки восстанавливают двухцепочечную структуру гораздо быстрее, чем непалиндромные, они более устойчивы к обработке нуклеазой (Рис. 4). Результаты эксперимента подтвердили наличие палиндромных последовательностей в митохондриальной ДНК волосатиков.

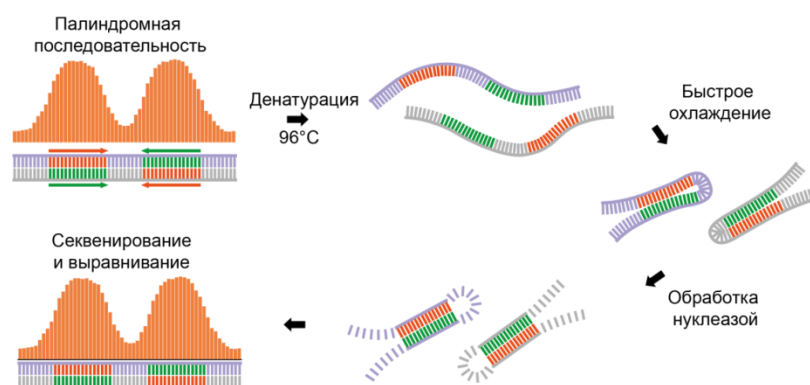


Рис. 4. Эксперимент с обработкой ДНК волосатиков нуклеазой Mung bean.

1.3. **Ортонектида *Intoshia linei*.** Митогеном *Intoshia linei* представляет собой кольцевую молекулу ДНК, содержащую типичный для Metazoa набор генов (Рис. 5). Предварительно митогеном был определен как один из контигов геномной сборки *I. linei* (Mikhailov et al., 2016); в настоящей работе методом Сэнгера уточнена первичная структура мтДНК и проведен ее анализ. Содержание GC% в митогеноме *I. linei* составляет около 17%, что лишь ненамного превышает минимальный показатель среди животных – 13% у наездника *Diadegma semiclausum*. Содержание GC% на третьей позиции кодона крайне низко и составляет всего 7%. Было проведено секвенирование кДНК, среди полученных чтений были отобраны картирующиеся на митохондриальный геном. Границы собранных контигов совпадают с транскриптами, ожидаемыми на основании

расположения генов. Было обнаружено семь полиаденилированных на 3'-концах транскриптов белок-кодирующих генов (*cox1*, *nad4*, *atp8*, *cox3*, *cytb*, *nad1* и *nad4l*).

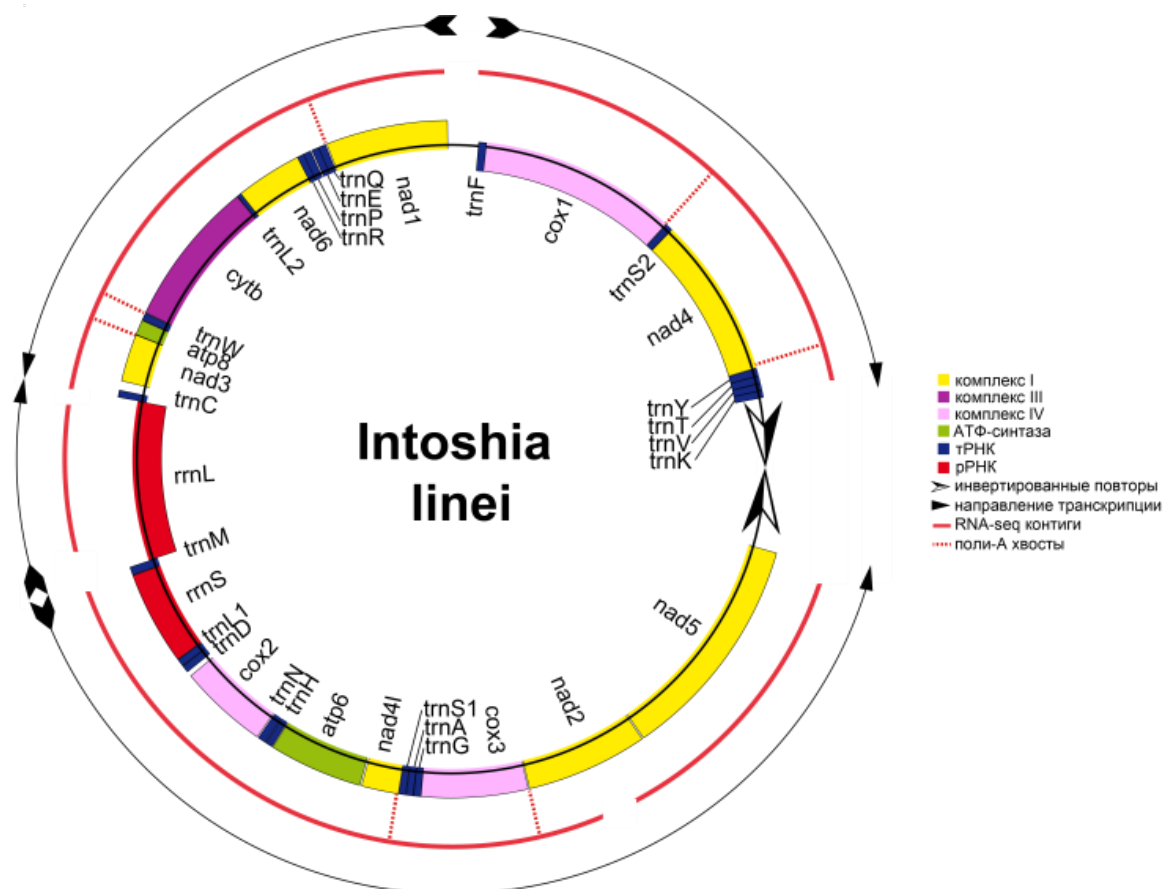


Рис. 5. Митохондриальный геном *Intoshia linei*.

1.4. **Дициемида *Dicyema* sp.** В результате сборки генома *Dicyema* sp. было обнаружено 24065 «кольцевых» контигов, фланкированных прямыми повторами. На 21 из этих контигов длиной от 344 п.н. до 1605 п.н. были обнаружены следующие митохондриальные гены: *cox1-3*, *cob*, *nad1-5*, *atp6*, *rrnL*, *rrnS*, *trnH*, *trnI*, *trnK*, *trnL1*, *trnN*, *trnP*, *trnQ*, *trnR*, *trnS2* и *trnY*, а также псевдогены *cox3*, *nad2* и *nad4* (Рис. 6). По сравнению со стандартным для Metazoa набором митохондриальных генов, не были обнаружены гены *atp8*, *nad4l*, *nad6* и некоторые гены тРНК. В отличие от предыдущих работ, в данной работе было обнаружено миникольцо с белок-кодирующим геном *cox2* и геном рРНК *rrnS*, а также три миникольца с двумя генами тРНК на каждом. Гены *nad2*, *atp6* и *trnQ* были обнаружены в двух различных вариантах–паралогах. Также были выявлены три кольцевых контига, содержащие митохондриальные псевдогены.

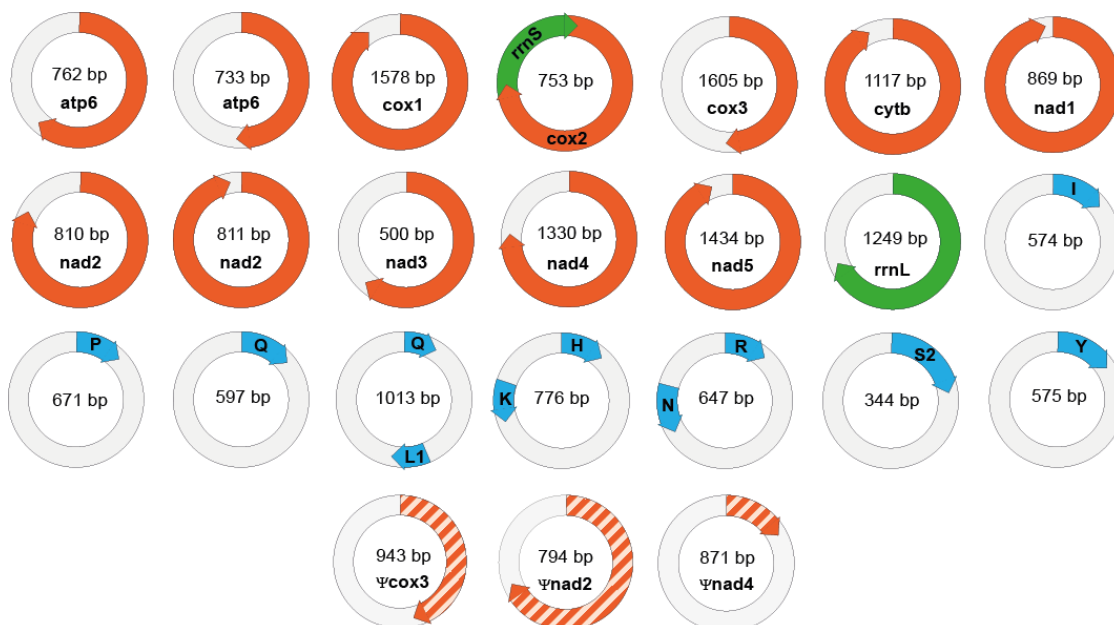


Рис. 6. Миникольца ДНК *Dicyeta* sp., кодирующие митохондриальные гены. Длина миникольца показана в его центре. Белок-кодирующие гены показаны красным, гены рРНК – зеленым, гены тРНК – синим, псевдогены – штрихом, а некодирующий регион показан светло-серым цветом.

2. **Эволюция порядка генов в митогеномах Ecdysozoa.** На основании реконструированных в данной работе порядков генов предков киноринх и волосатиков, а также ранее опубликованных порядков генов для *Deuterostomia*, *Lophotrochozoa*, *Panarthropoda* и *Priapulida* были реконструированы предположительные порядки генов для *Protostomia* и *Ecdysozoa*, а также возможные сценарии эволюции порядка генов в митогеномах *Bilateria* в соответствии с гипотезой о консервативных блоках генов (Рис. 7).

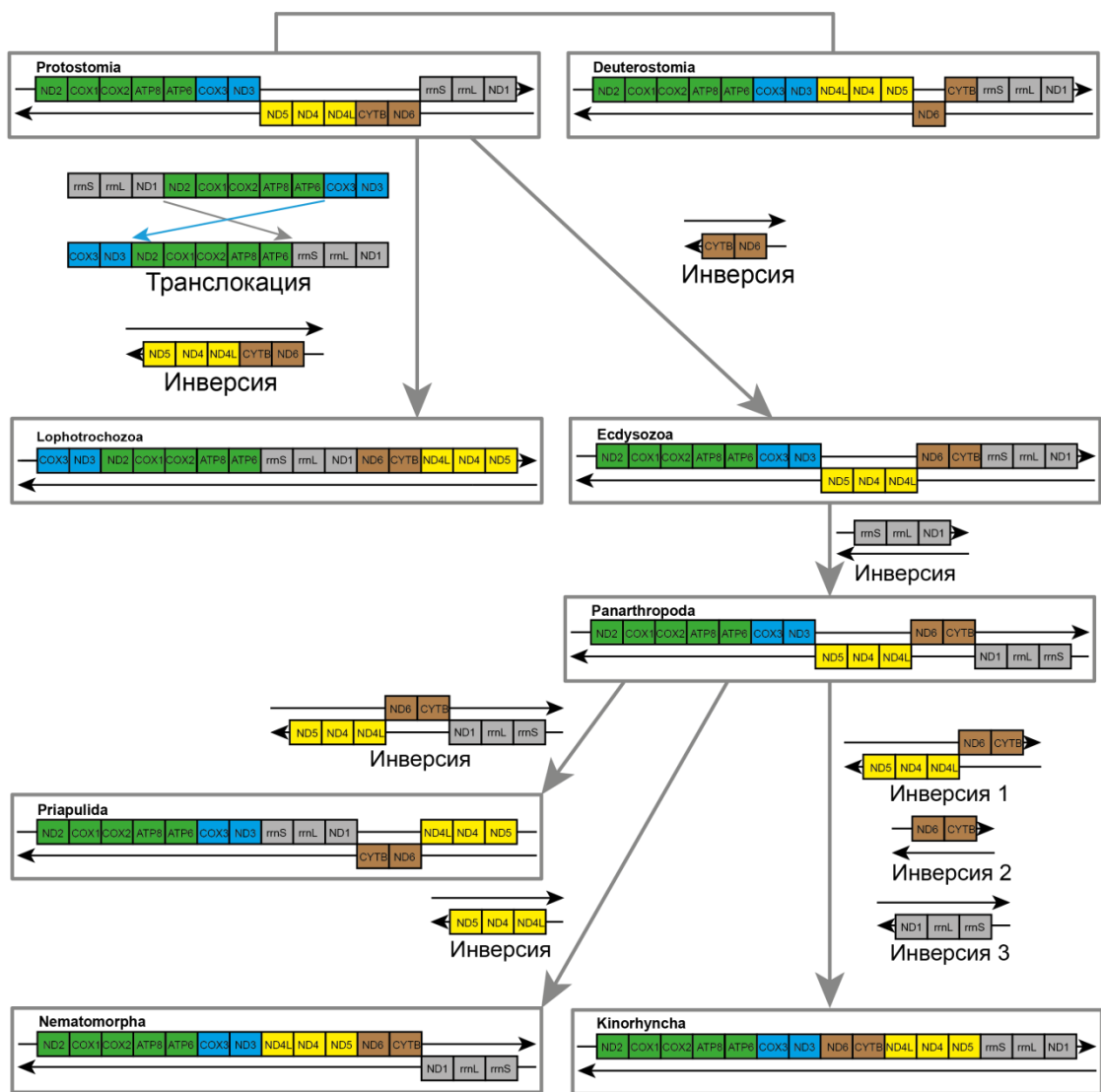


Рис. 7. Реконструкция порядка генов в митогеномах Bilateria.

Добавление к общей схеме порядка генов киноринх и волосатиков позволило определить, что порядок генов Panarthropoda является более плезиоморфным, чем порядок генов Priapulida, и реконструировать порядок генов в митохондриальном геноме предка Ecdysozoa в соответствии с принципом наибольшей экономии.

3. **Филогения Ecdysozoa на основании митохондриальных белок-кодирующих генов.** На основе исходного выравнивания митохондриальных белков после удаления 25% наиболее быстро эволюционирующих сайтов было построено байесово дерево по основным крупным таксонам Ecdysozoa по модели GTR+CAT+Г для 12 категорий (Рис. 8).

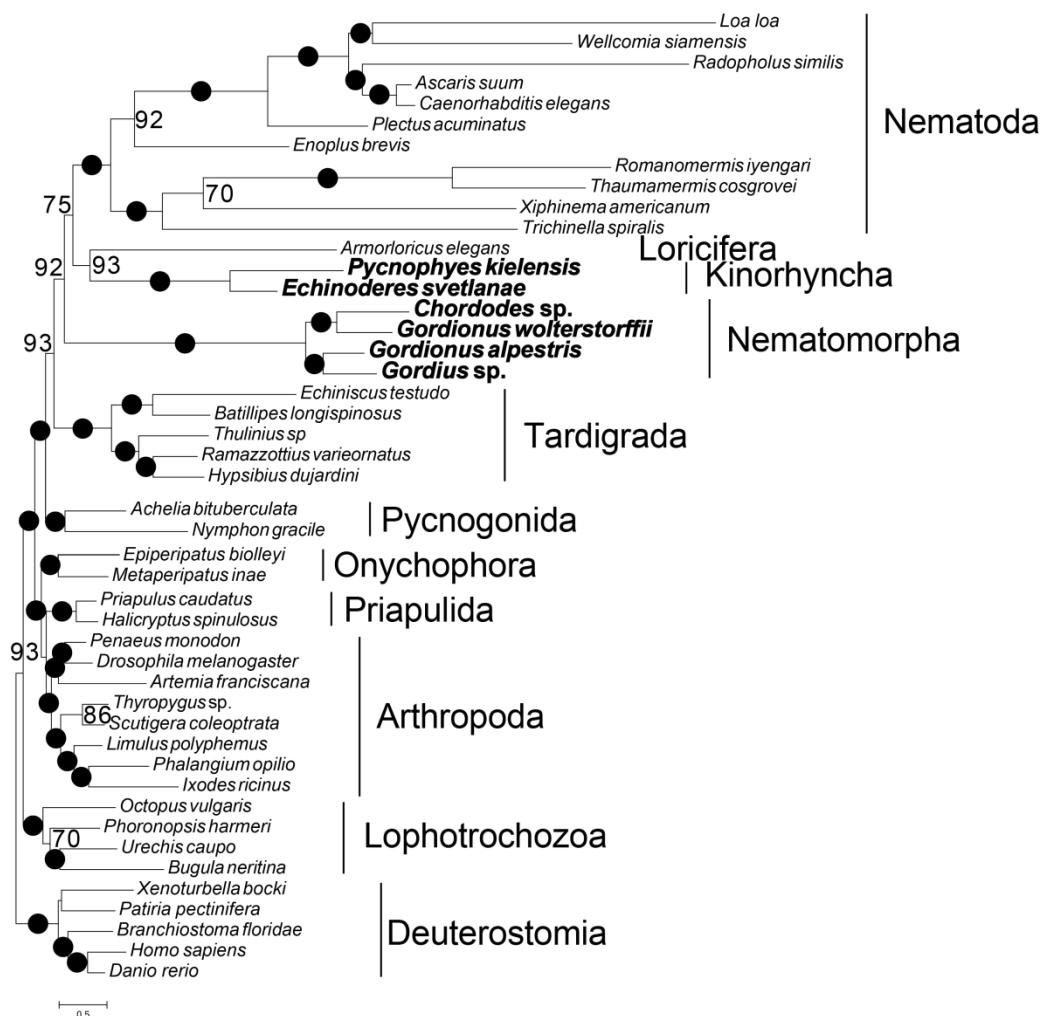
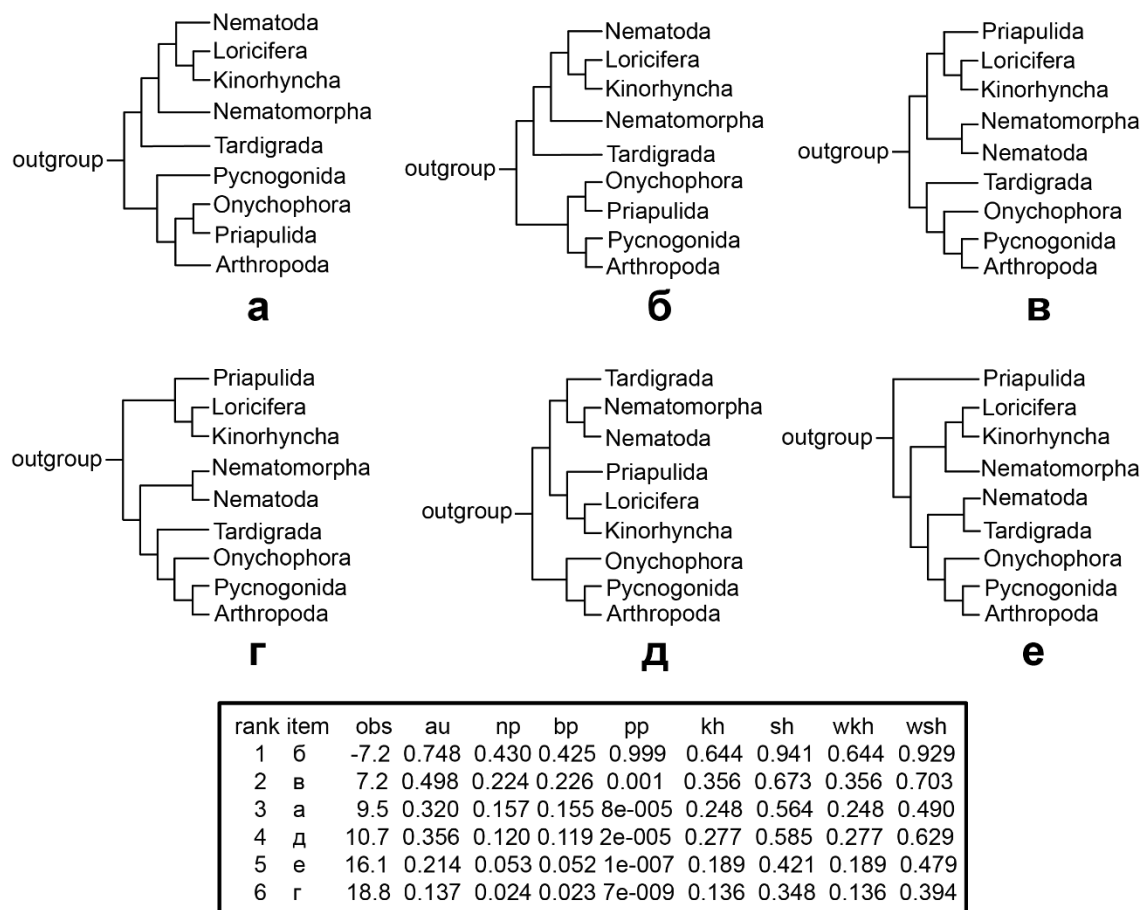


Рис. 8. Байесово дерево, построенное по конкатенированным данным 12 митохондриальных белков представителей Ecdysozoa. Виды, полученные в данной работе, выделены полужирным шрифтом. Числа на ветвях указывают байесовы постериорные вероятности в процентах, значения ниже 70 не показаны, значения 100 показаны точками.

Реконструированная топология подтверждает монофилию крупных таксономических групп среди Ecdysozoa. В целом, таксономические группы на дереве можно разделить на две категории: представленных длинными либо короткими ветвями. К первым относятся нематоды, лорициферы, киноринхи, волосатики и тихоходки; ко вторым – морские пауки, онихофоры, приапулиды и членистоногие. Таксоны, представленные длинными ветвями, группируются между собой в порядке уменьшения длины ветвей, что в различных группах может как отражать истинные родственные связи, так и являться следствием LBA (артефакта притяжения длинных ветвей). Киноринхи группируются с лорициферой, но не группируются с приапулидами, что подтвердило бы

монофилию Scalidophora. Поскольку эффект LBA ранее неоднократно выявлялся при анализе митохондриальных белок-кодирующих генов других экдизозой, возможно, полученное положение киноринх и волосатиков на дереве по митохондриальным белкам связано с их высокими скоростями эволюции. Используя полученное выравнивание по митохондриальным белкам Ecdysozoa, был проведен анализ альтернативных топологий как по полученному байесову дереву (Рис. 8), так и по некоторым распространенным



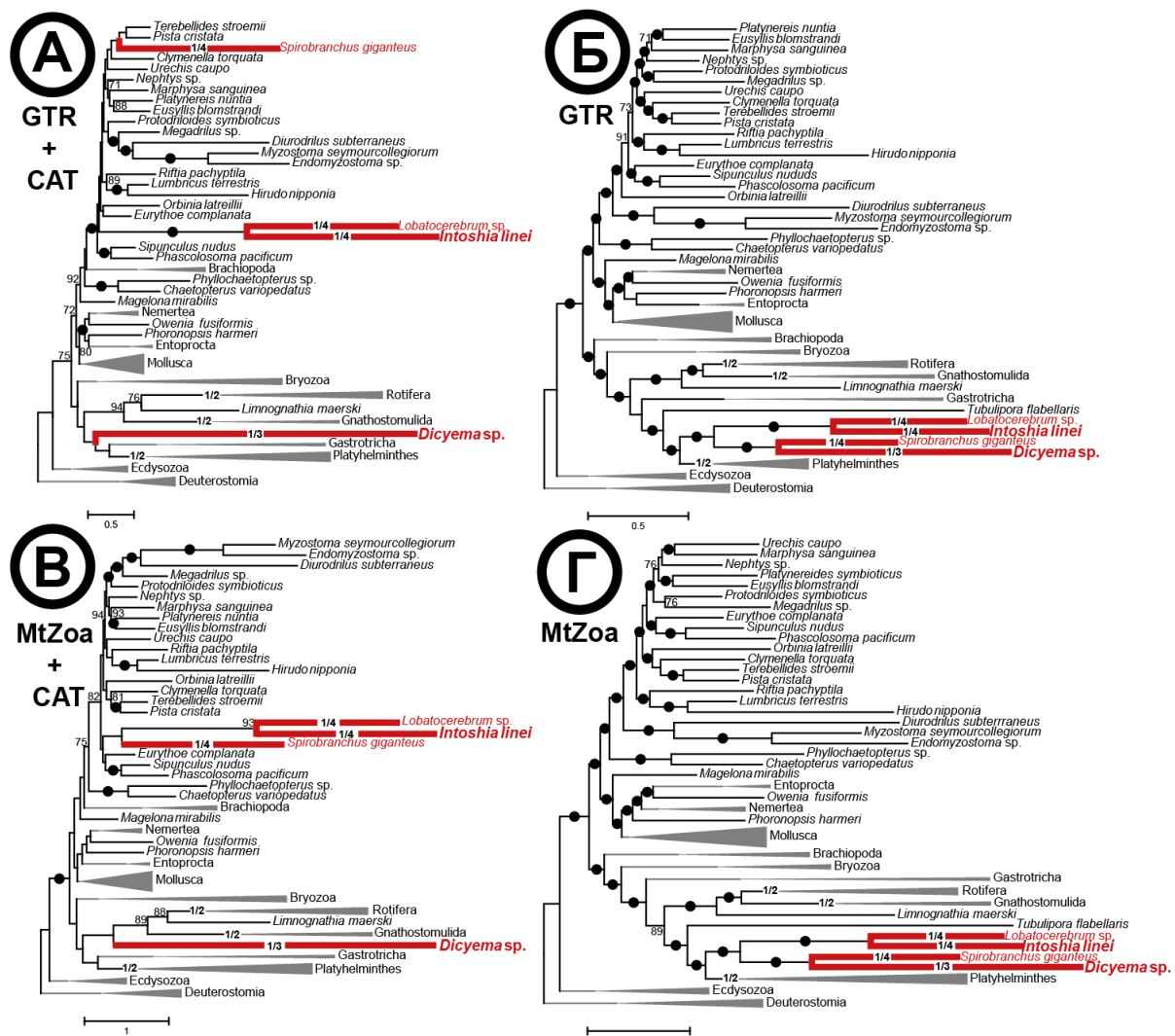
гипотезам о ветвлении крупных таксонов внутри Ecdysozoa (Рис. 9).

Рис. 9. Альтернативные топологии Ecdysozoa для конкатенированного набора 12 митохондриальных белков. а) Исходное байесово дерево (Рис. 19); б) Байесово дерево с группировкой Arthropoda + Pycnogonida; в-е) Топологии по литературным данным. Результаты тестирования топологий в программе CONSEL показаны в рамке.

Все шесть альтернативных топологий шесть преодолели 5%-ный порог значимости согласно AU-тесту (au), тесту Кишины-Хасегавы (kh) и тесту Шимодаиры-Хасегавы (sh). Таким образом, в рамках проведенного исследования принятые на сегодня статистические тексты не позволяют отвергнуть ни одну из тестируемых топологий. Лучшей (не значимо)

на основании статистических параметров оказалась топология б), представляющая собой исходное байесово дерево, модифицированное путем объединения Arthoropoda и Rysnagonida.

4. Филогения Lophotrochozoa на основании митохондриальных белок-кодирующих генов. Для выявления монофилии/полифилии Mesozoa и определении местоположения ортонекид и дициемид среди различных групп Lophotrochozoa по митохондриальным данным, было построено байесово дерево по 12 митохондриальным белкам в соответствии с моделью GTR+CAT+ Γ после удаления 25% наиболее быстро эволюционирующих сайтов (Рис. 10А). На байесовом дереве все крупные таксоны, кроме Annelida, монофилетичны. Ортонекида *Intoshia linei* внутри основного кластера аннелид группируется с апостериорной вероятностью 1.0 с *Lobatocerebrum* sp., свободноживущим мейобентосным червем с необычным планом строения, принадлежность которого к аннелидам была показана ранее на основании транскриптомных данных. Дициемиды формируют одиночную длинную ветвь высокого ранга и неясного филогенетического родства. Объединение *I. linei* и *Lobatocerebrum* sp. сохраняется при построении дерева при любой из тестируемых моделей (Рис. 10), в то время как их группировка внутри аннелид присутствует только при применении модели сайт-специфичных замен (CAT). Так, эта группировка присутствует на байесовом дереве, построенном с моделью MtZoa+ Γ +CAT (Рис. 10В). В отсутствие модели CAT вне зависимости от использования модели GTR или MTZoa, все представители Lophotrochozoa группируются в две большие группы по длинам ветвей – в одну попадают виды с короткими ветвями, в другую – с длинными (Рис. 10Б, Г). В целом, кластер длинных ветвей не находит поддержки в свете современных представлений о филогении Bilateria и является артефактом LBA. Примечательно, что на обоих деревьях без CAT (Рис. 10Б, Г) поддержки узлов значительно выше, а поскольку формируемые при этом группы признаются ошибочными, использование CAT при построении филогенетических деревьев по митохондриальным данным оказывается особенно важным для устранения ошибочных группировок таксонов.



rank	item	obs	au	np	bp	pp	kh	sh	wkh	wsh
1	B	-25.0	0.821	0.630	0.625	1.000	0.719	0.913	0.719	0.922
2	A	25.0	0.424	0.214	0.219	1e-011	0.281	0.606	0.281	0.602
3	Г	46.3	0.222	0.079	0.075	8e-021	0.156	0.376	0.156	0.396
4	Б	63.1	0.103	0.036	0.035	4e-028	0.089	0.231	0.089	0.240

Рис. 10. Байесовы деревья, построенные по конкатенированным данным 12 белок-кодирующих генов представителей Lophotrochozoa. Числа на ветвях указывают байесовы постериорные вероятности в процентах. Виды, полученные в данной работе, выделены полужирным шрифтом. Числа на ветвях указывают байесовы постериорные вероятности в процентах, значения ниже 70 не показаны, значения 100 показаны точками. Результаты тестирования топологий в программе CONSEL показаны в рамке.

5. **Синапоморфии *I. linei* и аннелид.** Порядок митохондриальных генов ортонектиды *I. linei* уникален среди Bilateria. Однако в митогеноме *I. linei* имеются два признака в порядке генов, общие с аннелидами – соседствующие гены *trnN-cox2* и *cytb-trnW*. 79% аннели, имеют признак *trnN-cox2*, а 88% – *cytb-trnW*. За исключением аннелид, только 3%

Lophotrochozoa имеют пару *trnN-cox2*, а 1.5% – *cytb-trnW*. Таким образом, признаки *trnN-cox2* и *cytb-trnW* можно считать синапоморфиями аннелид, возникшими в ранней радиации базальных линий и демонстрирующих низкие уровни гомоплазии в 3% и 1.5%, соответственно. Эти числа могут быть переоцененными, так как базируются на предположении о независимом возникновении каждого случая гомоплазии и не учитывают совместного наследования признака от общего предка. Представители базальных линий аннелид не имеют признака *trnN-cox2*. Отсутствие этой синапоморфии у других аннелид может быть истолковано как результат вторичной модификации порядка генов митохондриальных тРНК. В связи с отсутствием признака *cytb-trnW* у большинства представителей базальных линий аннелид, можно предположить, что он возник у аннелид позднее, чем признак *trnN-cox2*.

Филогенетический сигнал может быть найден и среди точечных замен. Для поиска специфичного филогенетического сигнала в митохондриальных белок-кодирующих генах все аннелиды были подразделены на базальных (*Owenia fusiformis*, *Magelona mirabilis*, Chaetopteridae), крону (Pleistoannelida + Amphinomidae + Sipunculidae) и плейстоаннелид. Всего было найдено 11 аминокислотных признаков с низким уровнем гомоплазии (менее 3%) и высоким уровнем стабильности внутри группы (встречается более чем у 70% представителей группы) (Табл. 1).

Таблица 1. Встречаемость синапоморфий аминокислотного состава аннелид по базе данных RefSeq среди Metazoa.

Признак	Ген	Встречаемость среди Annelida, %	Встречаемость вне Annelida, %
1	<i>atp6</i>	85	0.6
2	<i>atp6</i>	91	0.7
3	<i>cox1</i>	74	0.8
4	<i>cox1</i>	71	0.1
5	<i>cox1</i>	91	0.9
6	<i>cox1</i>	85	0.3
7	<i>cox3</i>	78	0.6
8	<i>cox3</i>	97	1.5
9	<i>cox3</i>	95	2.5
10	<i>cytb</i>	94	0.2
11	<i>nad6</i>	95	2.8

I. linei обладает четырьмя синапоморфиями, присущим аннелидам кроны филогенетического дерева (Табл. 2). Встречаемость этих признаков у плеистоаннелид варьирует от 96% до 100%, в то время как у других Metazoa – от 0.2% до 2.8% в зависимости от сайта, что может интерпретироваться как низкие уровни гомоплазии соответствующих сайтов. Все плеистоаннелиды имеют синапоморфию в nad6. *Marphysa sanguinea* и *Endomyzostoma* sp. не имеют синапоморфии в atp6, *Goniada japonica* и *Whitmania laevis* – в cox1. Предположительно, эти три синапоморфии, имеющиеся у большинства плеистоаннелид, были унаследованы плеистоаннелидами от общего предка плеистоаннелид, амфиномид, сипункулид, лобатоцеребрид и ортонектид, а синапоморфия в nad6 – от общего предка всех аннелид, но потеряны перечисленными видами.

Таблица 2. Синапоморфии *Intoshia linei* и Annelida или Annelida кроны дерева.

Таксон	Белок, аминокислотный мотив и встречаемость, %			
	cox1	cytb	atp6	nad6
	LSNMA– L –TLFMW	DWFL– W –AYAIL	PLTLS– F –RICAI	SIYIF– F –ISSGG
Annelida	89	93	93	96
Базальные аннелиды (<i>Owenia</i> + <i>Magelona</i> + Chaetopteridae)	0	0	0	33
Аннелиды кроны (Pleistoannelida + Lobatocerebridae + Amphinomidae + Sipuncula)	96	99	96	99
Pleistoannelida	94	97	96	100
Metazoa без Annelida	0.9	0.2	0.7	2.8

Аминокислотный мотив указан для *I. linei*.

С использованием значений частот встречаемости 11 признаков и с учетом условных вероятностей для трех пар зависимых признаков (5+11, 8+9 и 8+11) была вычислена вероятность независимого возникновения 4 или более из этих признаков. Полученные значения составило 0.000019. В базе данных (11082 видов Metazoa на момент анализа) был найден только один вид с 4 гомоплазиями и ни одного с большим числом гомоплазий, что хорошо подтверждает проведенные вычисления. Принимая во внимание независимые перестройки порядка генов (пары *trnN-cox2* и *cytb-trnW*) со значениями гомоплазии 3% и 1.5%, итоговая вероятность эволюционно независимого возникновения синапоморфий у *I. linei* и аннелид составляет 8.6×10^{-9} .

После установления родства ортонектид и аннелид в целом возникает задача более точного определения положения ортонектид внутри дерева аннелид. В связи с этим, был

проведен поиск потенциальных синапоморфий ортонектид и различных таксономических групп кроны аннелид. Потенциальные синапоморфии, прошедшие отбор по уровню гомоплазии и стабильности внутри группы, были обнаружены с тремя группами: Clitellata, Hirudinea и Siboglinidae. Во всех случаях *I. linei* обладает лишь небольшой частью синапоморфий данной группы. Поэтому была рассмотрена гипотеза, согласно которой *I. linei* утратила большую часть синапоморфий, полученную от общего предка группы, в связи с высокой скоростью молекулярной эволюции. Таким образом, кроме вычисления вероятности независимого (т.е. благодаря гомоплазии) возникновения синапоморфий, также была вычислена вероятность утраты *I. linei* синапоморфий, в случае ее действительной принадлежности к данной группе (Табл. 3).

Таблица 3. Вероятность независимого возникновения i и более синапоморфий и вероятность утраты *I. linei* $n-i$ синапоморфий.

Таксон	Число синапоморфий (n)	Число синапоморфий у ортонектид (i)	Вероятность возникновения	Вероятность утраты
Annelida	11	4	1.9×10^{-5}	0.24
Clitellata	12	3	0.0012	0.015
Hirudinea	45	3	0.014	2.3×10^{-6}
Siboglinidae	102	4	0.017	3.3×10^{-11}

Несмотря на сопоставимое число (три или четыре) синапоморфий ортонектид и четырех исследованных групп аннелид, вероятность их независимого возникновения в случае аннелид на два или три порядка ниже по сравнению с тремя группами плейстоаннелид. Основной вклад в это различие вносит разница в уровнях гомоплазии для этих признаков – для аннелид она в среднем составляет 1.15% (Табл. 1), а для остальных групп она варьирует от 1.85% до 2.33%. Поскольку на момент анализа в базе данных RefSeq было 78 аннелид, за счет гомоплазии ожидается обнаружить 0.09, 1.1 и 1.3 аннелид, независимо приобретших такое же или большее число синапоморфий, как *I. linei*. Поэтому сходство *I. linei* с Hirudinea и Siboglinidae может быть обусловлено гомоплазиями. В то же время, вероятность утраты *I. linei* 7 из 11 аннелидных синапоморфий, в случае ее принадлежности к этой группе, составляет 24%, что не противоречит гипотезе о том, что оставшиеся 4 признака были унаследованы от общего предка. В случае анализа синапоморфий Clitellata, вероятность утраты *I. linei* 9 из 12 признаков, в случае ее принадлежности к этой группе, составляет 1.5%. Вероятность же утраты *I. linei* синапоморфий групп Hirudinea и Siboglinidae, в случае ее принадлежности к

этим группам, составляет гораздо более низкие $2.3 \times 10^{-40}\%$ и $3.3 \times 10^{-9}\%$, соответственно. На основании этих вычисленных вероятностей гипотеза о принадлежности ортонектид к пиявкам отвергается. Однако принадлежность ортонектид к Clitellata и родство ортонектид с пиявками внутри Clitellata не может быть отвергнуто согласно этим данным.

Заключение

Дивергентность в митохондриальных геномах в первую очередь проявляется в высоких уровнях отличия нуклеотидных последовательностей по сравнению с другими организмами. Этот фактор приводит к возникновению длинных ветвей на филогенетических деревьях и ошибочной группировке видов с длинными ветвями. Все представленные в данной работе митогеномы образуют длинные ветви на сконструированных деревьях, и разрешение их положения на общем филогенетическом дереве затруднено.

При реконструкции филогении Ecdysozoa по предсказанным митохондриальным белкам и проверке принятыми на сегодня статистическими тестами (AU и другими) не удалось отвергнуть ни один из альтернативных сценариев радиации Ecdysozoa. Несколько большего успеха удалось достичь, применяя эти методы к митогеномам дивергентных Lophotrochozoa, таких как Mesozoa. Mesozoa – исторически выделенный таксон исключительно просто организованных животных – ортонектид и дициемид, – рассматривавшийся как промежуточное звено между одноклеточными и многоклеточными. Реконструкция филогении Lophotrochozoa указывает на полифилию Mesozoa, однако, AU-тест и другие статистические тесты не выявляют значимых различий между гипотезами о монофилии и полифилии этой группы. Принадлежность ортонектид к Annelida, в отличие от дициемид, была установлена ранее по ядерным генам. В рамках данного исследования найдено 11 аминокислотных признаков с низким уровнем гомоплазии, которые были количественно оценены с использованием базы данных Refseq и использованы для частичной реконструкции радиации Annelida. Ортонектида *Intoshia linei* обладает четырьмя из этих признаков и ответвляется от общего древа аннелид после базальных таксонов. Дициемида *Dicyema* sp. формирует одиночную длинную ветвь высокого ранга внутри Lophotrochozoa.

Все исследованные дивергентные митогеномы отличаются порядком генов от других Bilateria. Митогеном *Dicyema* sp. организован в виде отдельных миниколец и поэтому как такового порядка генов не имеет. Однако различия в порядке генов представителей киноринх и волосатиков позволяют реконструировать порядок генов

предков соответствующих групп. Учет порядка генов в митогеномах предков киноринх и волосатиков, в свою очередь, позволяет реконструировать порядок генов предков Ecdysozoa и Protostomia.

Использование модели сайт-специфических замен (CAT) при построении филогенетических деревьев по митохондриальным данным оказывается важным для устранения ошибочных группировок таксонов. В данном анализе модель CAT оказалась необходима для разрешения положения ортонектид как части аннелидной радиации и для преодоления ошибочной группировки с Rousphozoa (Platyhelminthes и Gastrotricha) и Gnatifera. Однако применение модели CAT не позволяет преодолеть эффект притяжения длинных ветвей (LBA) для митогеномных данных представителей Ecdysozoa, и вопрос о родственных отношениях между разными типами внутри этой группы остается открытым.

Выводы

1. Вероятность независимого появления одинаковых аминокислот (гомоплазий) в митохондриальных белках может быть оценена по базе данных RefSeq и применена в филогенетическом анализе.
2. Применение модели сайт-специфических замен (CAT) в филогенетическом анализе аминокислотных последовательностей митохондриальных белок-кодирующих генов беспозвоночных животных позволяет снизить или преодолеть эффект притяжения длинных ветвей (LBA).
3. Синапоморфии (сходства по производному состоянию признака) одиночных аминокислот и порядка расположения митохондриальных белок-кодирующих генов могут применяться как дополнительный метод при анализе филогенетического родства даже в случае сильно дивергированных последовательностей.
4. По результатам анализа митохондриальных геномов ортонектиды являются аннелидами, в отличие от дициемид, и потому таксон Mesozoa является полифилетичным.

Научные статьи по теме диссертационной работы, опубликованные в журналах SCOPUS, WOS, RSCI:

1. **Popova (Nikolaeva) O.V.**, Mikhailov K.V., Nikitin M.A., Logacheva M.D., Penin A.A., Muntyan M.S., Kedrova O.S., Petrov N.B., Panchin Y.V., Aleoshin V.V. «Mitochondrial genomes of Kinorhyncha: *trnM* duplication and new gene orders within animals» в журнале

PLoS ONE, издательство *Public Library of Science (United States)*, 2016, том 11, № 10, с. e0165072. IF = 2.806, квартиль Q1³.

2. Zverkov O.A., Mikhailov K.V., Isaev S.V., Rusin L.Y., **Popova (Nikolaeva) O.V.**, Logacheva M.D., Penin A.A., Moroz L.L., Panchin Y.V., Lyubetsky V.A., Aleoshin V.V. «Dicyemida and Orthonectida: Two stories of body plan simplification» в журнале *Frontiers in Genetics*, издательство *Frontiers Research Foundation. (Lausanne, Switzerland)*, 2019, том 10, № 443. IF = 3.260, квартиль Q2.

3. Mikhailov K.V., Efeykin B.D., Panchin A.Y., Knorre D.A., Logacheva M.D., Penin A.A., Muntyan M.S., Nikitin M.A., **Popova (Nikolaeva) O.V.**, Zanegina O.N., Vyssokikh M.Y., Spiridonov S.E., Aleoshin V.V., Panchin Y.V. «Coding palindromes in mitochondrial genes of Nematomorpha» в журнале *Nucleic Acids Research*, издательство *Oxford University Press (United Kingdom)*, 2019, том 47, № 13, с. 6858-6870. IF = 11.502, квартиль Q1.

Доклады на конференциях по теме диссертационной работы, опубликованные в сборниках тезисов:

1. **Попова О.В.** «Митохондриальные геномы двух видов Kinorhyncha (Ecdysozoa)» в сборнике Ломоносов – 2016: Международная конференция студентов, аспирантов и молодых ученых: секция «Биология», место издания Москва: Товарищество научных изданий КМК Москва, МГУ имени М.В. Ломоносова, биологический факультет, тезисы, с. 170.

2. Ефейкин Б.Д., Михайлов К.В., **Попова О.В.**, Алёшин В.В., Спиридонов С.Э., Панчин Ю.В. «Новый взгляд на систематику волосатиков в свете использования молекулярных данных» в сборнике Современные проблемы теоретической и морской паразитологии. Сборник научных статей. VI Всероссийская конференция с международным участием «Школа теоретической и морской паразитологии», место издания Издательство Бондаренко Н.Ю. Севастополь, тезисы, с. 28-30.

3. Aleoshin V.V., Slyusarev G.S., Mikhailov K.V., Rusin L.Y., **Popova O.V.**, Zverkov O.A., Lyubetsky V.A. «Orthonectida and Dicyemida: Two stories of body plan simplification» в сборнике Contributions to the 5th Moscow International Conference “Molecular Phylogenetics and Biodiversity Biobanking”, серия Molecular Phylogenetics, место издания Torus Press, Moscow, Russia, тезисы, с. 3-4.

³ параметры IF и квартили указаны на момент публикации.