

**Нестеренко
Максим Алексеевич**

**СРАВНИТЕЛЬНАЯ И ЭВОЛЮЦИОННАЯ ТРАНСКРИПТОМИКА РАЗНЫХ ФАЗ
СЛОЖНЫХ ЖИЗНЕННЫХ ЦИКЛОВ ДИГЕНЕТИЧЕСКИХ СОСАЛЬЩИКОВ**

1.5.17. Паразитология (биологические науки)

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Санкт-Петербургский государственный университет»

Научный руководитель:

кандидат биологических наук,
Старунов Виктор Вячеславович

Официальные оппоненты:

Спиридонов Сергей Эдуардович
доктор биологических наук, Федеральное государственное бюджетное учреждение науки Институт проблем экологии и эволюции им. А. Н. Северцова Российской академии наук, заведующий лабораторией, главный научный сотрудник

Горбушин Александр Михайлович
кандидат биологических наук, Федеральное государственное бюджетное учреждение науки Институт эволюционной физиологии и биохимии им. И. М. Сеченова Российской академии наук, ведущий научный сотрудник

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский государственный педагогический университет им. А. И. Герцена»

Защита состоится «_____» _____ 2023 г. в _____ часов на заседании диссертационного совета 24.1.026.01 на базе Зоологического института РАН по адресу: 199034, г. Санкт-Петербург, Университетская наб., д. 1

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного учреждения науки Зоологический институт Российской академии наук <https://www.zin.ru/boards/24.1.026.01/theses.html>

Автореферат разослан «_____»

2022 г.

Ученый секретарь
диссертационного совета,
кандидат биологических наук

Ахметова Лилия Агдасовна

Общая характеристика работы

Актуальность работы. Digenea – дигенетические сосальщики – представляют собой одну из крупнейших групп паразитических организмов, входящих в Trematoda и обладающих большой медицинской и ветеринарной значимостью. Являясь ярким примером многоклеточных организмов со сложным жизненным циклом, дигеней и их жизненный цикл, построенный по типу гетерогонии, представляют большой интерес для фундаментальной науки. Тем не менее несмотря на долгую историю изучения молекулярных основ реализации жизненного цикла дигеней, вопрос разделения единого генома между разными поколениями и отдельными фазами одного жизненного цикла, отличающимися по морфологическим признакам, физиологическим особенностям и образу жизни, остается открытым.

Одним из перспективных направлений исследований является анализ транскриптомов. Представляя собой совокупность всех молекул РНК, присутствующих в рассматриваемом образце в конкретный момент времени и в определенных условиях, транскриптом не только позволяет определить вероятный набор генов исследуемого вида, но и отражает состояние исследуемой системы. Для описания набора последовательностей (генов, транскриптов или белков), который может быть использован в качестве маркера рассматриваемого фенотипа, используют термин «молекулярная подпись». В зависимости от выбранных исследователем критериев молекулярная подпись образца может включать как весь набор активных генов, так и только список генов со статистически значимым изменением экспрессии, т. е. дифференциально экспрессирующиеся гены.

В подавляющем большинстве публикаций, посвященных транскриптомному анализу дигеней, в качестве молекулярной подписи фазы сложного жизненного цикла рассматривается именно набор дифференциально экспрессирующихся генов. Для определения дифференциальной экспрессии генов проводят попарное сравнение фаз одного жизненного цикла. Как следствие, характеристики выявленной молекулярной подписи рассматриваемой фазы зависят от того, с какой фазой проводилось сравнение. Несмотря на всю важность получаемых результатов, необходимость использовать одну из контрастных фаз жизненного цикла в качестве «контроля» является одной из слабых сторон такого подхода. В настоящее время разработаны альтернативные методы анализа, способные одновременно рассматривать множество образцов и выявлять статистически значимое отклонение в экспрессии генов. Такие подходы позволяют снизить зависимость определения молекулярной подписи от выбранной пары для сравнения и предоставляют возможность исследовать изменение экспрессии генов на всем протяжении сложного жизненного цикла дигеней.

Изменение к подходу поиска дифференциально экспрессирующихся генов является не единственным примером стремительного развития целой междисциплинарной области, объединяющей общую и молекулярную биологию, математику и компьютерные науки – биоинформатики. Помимо новых версий и улучшений компьютерных программ, активно используемых научным сообществом, в настоящее время разрабатываются и применяются принципиально новые подходы к исследованию биоинформатических данных. В основе части из них лежит сравнительный анализ доступных геномных и транскриптомных данных филогенетически близких и далеких видов. Во-первых, наличие высококачественных сборок транскриптомов и геномов с разных ветвей филогенетического дерева открыло возможность реконструкции моделей геномов предков различных таксонов. Во-вторых, развитие получила и активно набирает популярность филостратиграфия, относительно новая область биоинформатического анализа, направленная на выявление и исследование филострат – групп генов с разным филогенетическим происхождением. В-третьих, объединение методов филостратиграфии и транскриптомики поспособствовало формированию эволюционной транскриптомики. Представляя собой новое направление транскриптомных исследований, эволюционная транскриптомика направлена на определение и количественную оценку эволюционной консервативности транскриптомов рассматриваемых образцов. Сравнительный анализ моделей геномов предков уже применялся в исследовании дигеней, но не был направлен

на определение молекулярных основ усложнения жизненного цикла последнего общего предка дигеней. Выявление филострат и определение их вкладов в молекулярные подписи фаз сложного жизненного цикла дигеней не проводилось, несмотря на большой научный интерес.

Степень разработанности темы исследования. В настоящее время методы сравнительной транскриптомики успешно применены в исследовании многих видов дигеней: *Clonorchis sinensis* (Digenea, Opisthorchiidae), *Echinostoma caproni* (Digenea, Echinostomatidae), *Eurytrema pancreaticum* (Digenea, Dicrocoeliidae), *Fasciola gigantica* (Digenea, Fasciolidae), *Fasciola hepatica* (Digenea, Fasciolidae), *Fascioloides magna* (Digenea, Fasciolidae), *Metorchis orientalis* (Digenea, Opisthorchiidae), *Microphallus livelyi* (Digenea, Microphallidae), *Opisthorchis felineus* (Digenea, Opisthorchiidae), *Opisthorchis viverrini* (Digenea, Opisthorchiidae), *Paragonimus skrjabini* (Digenea, Troglotrematidae), *Paragonimus westermani* (Digenea, Troglotrematidae), *Paramphistomum cervi* (Digenea, Paramphistomidae), *Schistosoma mansoni* (Digenea, Schistosomatidae), *Trichobilharzia regenti* (Digenea, Schistosomatidae), *Trichobilharzia szidati* (Digenea, Schistosomatidae). Большинство исследований направлено на изучение молекулярных подписей фаз амфимиктического поколения, в особенности марит, обитающих в телах окончательных хозяев. Результаты сравнительного транскриптомного анализа фаз одного жизненного цикла и/или близких видов приведены в 11 публикациях, посвященных 11 видам: *C. sinensis* (Young et al., 2010), *F. hepatica* (McNulty et al., 2017), *F. gigantica* (Zhang X.-X. et al., 2017, 2019), *M. livelyi* (Bankers, Neiman, 2017), *O. viverrini* (Young et al., 2010), *O. felineus* (Pomaznoy et al., 2016), *P. westermani* (Li et al., 2016), *P. skrjabini* (Li et al., 2016), *S. mansoni* (Protasio et al., 2012; Picard et al., 2016), *T. regenti* (Leontovyč et al., 2016, 2019), *T. szidati* (Leontovyč et al., 2019). Транскриптомы представителей Psilostomatidae, в том числе *Psilotrema simillimum* и *Sphaeridiotrema pseudoglobulus*, не были исследованы ранее. Более того, не проводилось комплексного сравнительного анализа молекулярных подписей фаз сложных жизненных циклов филогенетически близких и отдаленных видов дигеней.

Успешная попытка реконструировать модель генома предков дигеней, Plagiorchiida и Xiphidiata/Opisthorchiata предпринята Заяк с соавторами в 2021 году (Zajac et al., 2021). В исследовании авторы сфокусировали внимание на изучении того, как переход к паразитизму повлиял на геном предка дигеней. Сравнительный анализ моделей геномов предков свободноживущих плоских червей и дигеней, с целью описать молекулярные основы усложнения жизненного цикла, не проводился. К настоящему времени методы филостратиграфии и эволюционной транскриптомики в исследованиях дигеней не применялись.

Цель и задачи работы. Цель настоящего исследования заключается в определении молекулярных основ реализации сложных жизненных циклов дигеней и возможных путей их становления в ходе эволюции. Для реализации заявленной цели мы поставили следующие задачи:

1. Получить высококачественные справочные транскриптомы дигеней *P. simillimum* (Digenea: Psilostomatidae) и *S. pseudoglobulus* (Digenea: Psilostomatidae)
2. Определить и охарактеризовать молекулярные подписи разных фаз сложных жизненных циклов филогенетически близких и отдаленных видов дигеней
3. Установить филостратиграфическую принадлежность белок-кодирующих генов исследуемых видов плоских червей и вычислить возрастные индексы молекулярных подписей фаз сложных жизненных циклов дигеней
4. Реконструировать модели геномов последних общих предков плоских червей и определить биологические процессы, в которых принимают участие гены, появившиеся или дублированные у предка дигеней

Научная новизна исследования. Впервые получены «справочные» транскриптомы двух видов дигеней, *P. simillimum* и *S. pseudoglobulus*. Собранные транскриптомы являются первыми не только для двух исследуемых видов, но и для целого таксона Psilostomatidae. Сравнительный транскриптомный анализ для редий, церкарий и марит представителей Psilostomatidae проведен впервые. Предпринята первая попытка обобщить, что представляет собой молекулярная подпись фазы сложного жизненного цикла дигеней. Методы эволюционной транскриптомики ранее в

исследованиях дигеней не применялись. Впервые для *F. gigantea*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* достоверно показаны статистически значимые различия во вкладах групп генов с разным филогенетическим происхождением в молекулярные подписи фаз. Вслед за публикацией Заяк с соавторами (Zajac et al., 2021) предпринята попытка реконструировать модели геномов последних общих предков плоских червей. Результаты сравнения реконструированных моделей геномов предков впервые применены в исследовании молекулярных основ эволюционного усложнения жизненного цикла последнего общего предка дигеней.

Теоретическая и практическая значимость работы. Полученные для *P. simillimum* и *S. pseudoglobulus* транскриптомные данные и результаты их анализа имеют ключевое значение для глубокого понимания молекулярных основ реализации сложных жизненных циклов Psilostomatidae. Более того, принимая во внимание как эволюционно продвинутые, так и консервативные признаки исследованных видов дигеней, разработанные модели и опубликованные материалы представляют значимость для специалистов разных областей, заинтересованных в исследовании эволюции как отдельных систем паразит-хозяин, так и сложных жизненных циклов.

С практической точки зрения, дигенеи *P. simillimum* и *S. pseudoglobulus*, представляют собой удобные и перспективные модели для исследования взаимоотношений в системе паразит-хозяин. В рамках представленной научной работы также проведен комплексный анализ ранее опубликованных геномных и транскриптомных данных разных видов дигеней. Полученные результаты являются ценным дополнением не только к теоретической, но и практической, ветеринарной трематодологии. В частности, расширен круг видов, для которых выявлены наборы генов, потенциально кодирующих секреторные белки. Принимая во внимание смену хозяев в ходе реализации жизненного цикла дигеней, подготовленные наборы белков могут быть рассмотрены в качестве объектов более детального анализа молекулярных основ паразитизма и/или разработки мишеней для лекарственных препаратов.

Методология и методы исследования. Теоретико-методологической базой диссертационной работы являются труды российских и зарубежных ученых, посвященные проблеме определения молекулярных основ реализации сложных жизненных циклов паразитических плоских червей. При решении поставленных задач использовались общенаучные (анализ, обобщение, идеализация, формализация, сравнение, моделирование, абстрагирование) и специальные методы исследования (методы биоинформатического анализа транскриптомов, методы анализа данных с использованием языков программирования R и Python, методы филогенетики и эволюционной транскриптомики, методы реконструкции моделей геномов предков исследуемых видов).

Положения, выносимые на защиту.

1. Молекулярная подпись фазы сложного жизненного цикла дигеней представляет собой уникальный набор белок-кодирующих генов, обладающих заметным уровнем экспрессии на рассматриваемом этапе реализации жизненного цикла. Характеристики молекулярной подписи определяются клеточным составом исследуемой фазы цикла и набором условий, воздействующих на организм паразита.

2. В большинстве случаев дифференциально экспрессирующиеся гены, т. е. белок-кодирующие гены, демонстрирующие статистически значимое повышение экспрессии, составляют относительно небольшую часть молекулярной подписи фазы жизненного цикла. Повышенная экспрессия генов соответствует активности биологических процессов, характерных для рассматриваемых фаз жизненных циклов.

3. Фазы сложного жизненного цикла дигеней различаются между собой по вкладам в их молекулярные подписи групп генов с разным филогенетическим происхождением. Высокая активность филогенетически более «молодых» генов, по всей видимости, связана с наличием у рассмотренных фаз сложных жизненных циклов специфичных и/или сильно преобразованных признаков.

4. Белок-кодирующие гены, появившиеся у последнего общего предка дигеней, принимают участие во многих биологических процессах, в том числе предположительно связанных с эмбриональным развитием и/или метаморфозом.

Степень достоверности и апробация результатов. В настоящем исследовании использованы стандартизированные методы пробоподготовки и секвенирования. Все образцы были выполнены в двух независимых биологических повторностях. Проведенный детальный биоинформатический анализ транскриптомных данных содержит подробное описание всех использованных методов и программ для анализа с указанием основных параметров запуска. В диссертации также приведены ссылки на репозиторий GitHub с программами, написанными соискателем и использованными в исследовании. Все библиотеки коротких парных прочтений образцов транскриптомов рений, церкарий и марит представителей Psilostomatidae, полученные в ходе исследования, опубликованы в открытом доступе в архиве прочтений последовательностей (Sequence Read Archive, SRA) Национального центра биотехнологической информации США NCBI и могут быть повторно проанализированы всеми желающими. Помимо собственных данных, автором диссертации тщательно проанализированы транскриптомные данные из литературных источников, ссылки на которые приведены в тексте диссертации. Достоверность полученных результатов, корректность их описания и полнота обсуждения проверены и подтверждены в ходе анонимного независимого рецензирования при публикации материалов исследования в научных изданиях.

Основные положения и научные итоги диссертации изложены в докладах на 12 Европейском мультиколлоквиуме по паразитологии (The 12th European Multicolloquium of Parasitology, EMOPXII) (20-24 июля 2016), VI Всероссийской конференции с международным участием «Школа по теоретической и морской паразитологии» (5-10 сентября 2016), 4 Международном конгрессе по морфологии беспозвоночных (The 4th International Congress on Invertebrate Morphology, ICIM4) (18-23 августа 2017), Международной конференции «Современная паразитология – основные тренды и вызовы» (VI Съезд Паразитологического общества, 15-19 октября 2018), Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2021» (12-21 апреля 2021), 9 конференции Скандинавско-Балтийского общества паразитологов (The 9th Conference of the Scandinavian-Baltic Society for Parasitology, CSBSP9) (21-23 апреля 2021), 5 международной конференции Bioinformatics: From Algorithms to Applications (BIATA2021, 12-15 июля 2021), 10 Московской конференции по вычислительной молекулярной биологии (The 10th Moscow Conference on Computational Molecular Biology, MCCMB'21) (30 июля – 2 августа 2021).

По материалам диссертации опубликовано 10 научных работ. Из них 2 научные статьи в журналах, индексируемых системами Web of Science и/или Scopus и рекомендованных ВАК РФ.

Структура и объем диссертации. Диссертация изложена на 159 страницах и включает введение, 4 главы, заключение, выводы, список литературы и 1 приложение. Объем основного текста диссертации составляет 135 страниц, включая 4 таблицы и 19 рисунков. Приложение содержит список библиотек коротких парных прочтений, использованных в исследовании. Список литературы включает 160 источников, из них - 9 на русском языке.

Благодарности. Выражаю огромную благодарность своему научному руководителю к.б.н. Виктору Вячеславовичу Старунову за важные советы, всестороннюю помощь в проведении исследований и колоссальную поддержку на всем протяжении обучения и выполнения представленной научной работы. Отдельную огромную благодарность выражаю своему учителю к.б.н. Андрею Александровичу Добровольскому за неоценимый вклад в опубликованные исследования и за активное участие в обсуждении как полученных результатов, так и общих вопросов биологии в целом и паразитологии в частности. Я крайне признателен к.б.н. Халтуруну Константину Викторовичу, Щенкову Сергею Владимировичу и Денисовой Софье Александровне за помощь и активное участие в проведении исследований. Выражаю глубокую благодарность сотрудникам кафедры Зоологии беспозвоночных СПбГУ за всестороннее обсуждение результатов исследований. Благодарю Маслову Анну Родионовну и Полева Дмитрия

Евгеньевича за помощь в синтезе библиотек и проведении высокопроизводительного секвенирования.

Работа поддержана грантом РФФИ 19-34-90111 Аспиранты «Геном трематоды семейства Psilostomatidae, как ключ к пониманию молекулярных основ реализации сложных жизненных циклов редиоидных трематод». Биоинформатический анализ данных частично выполнен на оборудовании Центра Коллективного Пользования «Биоинформатика» ИЦиГ СО РАН.

Содержание работы

Глава 1 Обзор литературы

1.1 Общее описание фаз сложного жизненного цикла дигеней

В подглаве приведены общие описания фазы яйца, мирацидия, материнской и дочерней спороцисты, редии, церкарии, адолескарии, метацеркарии, шистосомулы и мариты с указанием краткой истории исследования всех перечисленных фаз сложного жизненного цикла дигеней.

1.2 Развитие представлений о структуре жизненных циклов дигеней

Первым, кто использовал термин «чередование поколений» к описанию жизненного цикла дигеней, стал Стинструп (Steenstrup), опубликовавший свою научную работу в 1842 году (Reinhard, 1957). На рубеже XIX и XX столетий была сформулирована теория гетерогонии, согласно которой жизненный цикл дигеней представляет собой чередование полового и партеногенетического поколений (Атаев, 2017). Альтернативную точку зрения на структуру жизненного цикла дигеней представил Брукс (Brooks, 1930). Согласно Бруксу, жизненный цикл дигеней представляет собой удлинённый онтогенез особей одного поколения со сложным метаморфозом, а формирование дочерних партенит происходит вследствие полиэмбрионии.

1.3 Описания схем реализации сложных жизненных циклов исследуемых видов дигеней

В настоящем разделе приведены описания схем реализации сложных жизненных циклов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. pseudoglobulus*, *S. mansoni*, *T. regenti* и *T. szidati*.

1.4 Применение методов биоинформатического анализа для исследования молекулярных основ реализации сложных жизненных циклов дигеней

В подглаве описана краткая история применения технологий высокопроизводительного секвенирования в исследовании дигенетических сосальщиков и рассмотрены результаты транскриптомных анализов фаз жизненных циклов различных видов дигеней.

1.5 Теоретические основы филостратиграфии и применение возрастных индексов транскриптомов в изучении жизненных циклов

Представление о том, что геном каждого современного вида сохраняет части картины эволюционных эпох является одним из принципов филостратиграфии, статистического метода геномных исследований (Domazet-Lošo et al., 2007). В основе филостратиграфии лежит поиск гомологов генов исследуемого вида на всем филогенетическом дереве, ведущем от прокариотических организмов до рассматриваемого вида (Arendsee et al., 2019). В соответствии с вычисленными возрастными, все гены исследуемого вида распределяются по группам, названным «филостратами» (Domazet-Lošo et al., 2007).

Разработанный Домазет-Лошо и Таутцом (Domazet-Lošo, Tautz, 2010) «возрастной индекс транскриптома» (Transcriptome Age Index, TAI) объединяет возраст гена с уровнем его экспрессии в рассматриваемом образце и суммирует это значение по всем активным генам. Значение TAI тем выше, чем моложе исследуемый транскриптом. При расчете возрастного индекса транскриптома больший вес присваивается более «молодым» филостратам, тем самым компенсируя тот факт, что более «старые» филостраты обычно содержат большее количество

генов (Domazet-Lošo, Tautz, 2010). Таким образом, ТАИ представляет собой среднее арифметическое взвешенное значение вкладов филострат (Domazet-Lošo, Tautz, 2010).

В разделе продемонстрирована успешность применения анализа возрастных индексов транскриптомов в исследовании широкого круга живых организмов. Помимо этого, рассмотрены литературные данные, указывающие на возможность формирования генов *de novo*, а также описаны результаты реконструкции и анализа моделей геномов последних общих предков дигенеи *Atriophallophorus winterbourni* (син. *Microphallus* sp. или *Microphallus livelyi*) (Zajac et al., 2021).

Глава 2 Материалы и методы

Сложные жизненные циклы *S. pseudoglobulus* и *P. simillimum* частично поставлены в лабораторных условиях. Выделение образцов тотальной РНК из фиксаций редий, церкарий и марит *S. pseudoglobulus* и *P. simillimum* проводилось с помощью набора Quick-RNA™ Microprep Kit (R1050; Zymo Research, Калифорния, США) в соответствии с протоколом производителя. Создание библиотек для секвенирования на платформе Illumina было выполнено с помощью набора NEBNext Ultra Directional RNA Library Prep Kit (E7760; New England BioLabs, Массачусетс, США). Секвенирование подготовленных библиотек парных прочтений проводилось на приборе Illumina HiSeq2500. Подготовка библиотек для секвенирования и запуск секвенаторов были выполнены на базе Ресурсного Центра СПбГУ «Центр Биобанк».

Оценку качества полученных библиотек коротких парных прочтений мы выполнили с помощью программы FastQC (v0.11.5). Для исправления возможных ошибок секвенирования использовали программу Karet (v1.0) (Allam et al., 2015). Выявление и удаление возможной контаминации осуществлено с помощью набора программ BBTools (v37.02). Деконтаминированные библиотеки были загружены в базу данных NCBI SRA под номером БиоПроекта PRJNA516017. Далее библиотеки прочтений, очищенные от возможной контаминации, мы проанализировали с помощью программы Trimmomatic (v0.36) (Bolger et al., 2014).

Сборка протяженных последовательностей (контигов) из коротких прочтений производилось *de novo* с одновременным использованием всех подготовленных библиотек с трех фаз сложного жизненного цикла. Данные для каждого вида анализировались отдельно. В работе мы использовали программу Trinity (v2.3.2) (Grabherr et al., 2011). Кластеризацию изоформ последовательностей в обоих транскриптомах выполнили с помощью программы CD-HIT-est (v4.7) (Fu et al., 2012). Мы использовали программу TransRate (v1.0.1) (Smith-Unna et al., 2016) для проведения оценки качества и полноты итоговых сборок транскриптомов. В дальнейшем анализе мы использовали только те контиги, которые характеризовались высокими показателями качества и полноты сборки, согласно критериям TransRate.

Для квантификации уровней экспрессии отобранных последовательностей мы выбрали программу Salmon (v1.0.1) (Patro et al., 2017). Мы использовали таблицу соответствия контигов предполагаемым генам, созданную Trinity по результатам сборки транскриптома, для преобразования вычисленных значений уровней экспрессий контигов в уровни экспрессии их генов. Определение белок-кодирующих генов выполнили с помощью программы TransDecoder (v5.5.0).

Основное внимание в дальнейшем анализе мы сфокусировали на генах, удовлетворявших двум требованиям: 1) уровень экспрессии больше или равен 1 транскрипт-на-миллион как минимум на одной из рассматриваемых фаз сложного жизненного цикла дигеней; 2) кодирование белка длиной не меньше 100 аминокислот. В качестве репрезентативных последовательностей для каждого гена были отобраны наиболее длинный белок и контиг, кодирующий его.

Контроль полноты подготовленных «справочных» наборов белков был проведен с помощью BUSCO (v3.0.1) (Simão et al., 2015, Waterhouse et al., 2018) и 9 версии базы данных ортологов, присутствующих в геномах Metazoa в количестве одной копии. С целью сопоставить

полученные для *P. simillimum* и *S. pseudoglobulus* результаты с другими видами, анализ был проведен и для наборов белков дигеней, имеющих в открытом доступе.

Функциональная аннотация последовательностей из «справочных» наборов проводилась путем сравнения собранных последовательностей и белков с нуклеотидной (NCBI nt) и аминокислотными (NCBI nr, SwissProt, PfamA) базами данных. Белки двух видов Psilostomatidae были также исследованы с помощью Интернет-ресурса eggNOG-mapper с параметрами запуска по умолчанию.

По результатам аннотации из «справочного» набора *S. pseudoglobulus* мы отобрали последовательности, имеющие совпадения с первой субъединицей митохондриальной цитохром с-оксидазы (cox1). С целью определить различия между ними, баркодные регионы обнаруженных последовательностей cox1 были выравнены вместе с доступными для других видов дигеней в базе данных GenBank с помощью алгоритма MUSCLE (Edgar, 2004). Для Байесовского анализа мы использовали программу MrBayes (Ronquist, Huelsenbeck, 2003). Обнаруженные гаплотипы cox1 были загружены в базу данных GenBank под идентификаторами MT934401 и MT934400.

Для построения групп ортологичных последовательностей мы использовали OrthoFinder (v2.2.6) (Emms, Kelly, 2015, 2019) и OMA standalone (v2.5.0) (Altenhoff et al., 2019). В анализе были рассмотрены как паразитические, так и свободноживущие виды плоских червей. Для каждого исследуемого вида мы рассматривали только белки длиной не менее 100 аминокислот.

Для определения молекулярных подписей фаз сложных жизненных циклов дигеней мы использовали два пороговых значения экспрессии, равные 1 и 2 транскриптам-на-миллион. Ген был включен в молекулярную подпись фазы жизненного цикла только в том случае, если уровень его экспрессии на рассматриваемой фазе был равен или превышал используемый порог. Анализ был проведен для *F. gigantea*, *F. hepatica*, *P. simillimum*, *S. pseudoglobulus*, *S. mansoni*, *T. regenti* и *T. szidati*.

Определение статистически значимых различий в уровнях экспрессий генов выполнили с помощью библиотеки «RNentropy» (v1.2.2) (Zambelli et al., 2018) для языка программирования R. Определение кластеров ко-экспрессирующихся генов для *F. gigantea*, *F. hepatica*, *P. simillimum* и *S. mansoni* мы выполнили с помощью программы Clust (v1.10.8) (Abu-Jamous, Kelly, 2018).

Для проведения многомерного шкалирования мы подготовили матрицы присутствия (1) / отсутствия (0) по результатам анализа экспрессии генов. В качестве порогового значения экспрессии для классификации активных генов мы использовали 2 транскрипта-на-миллион. Во внутривидовом анализе все биологические повторности рассматривались независимо друг от друга. Межвидовое сравнение проводилось в два этапа: на первом этапе «редиоидные» (*F. gigantea*, *F. hepatica* и *P. simillimum*) и «спороцистоидные» (*S. mansoni*, *T. regenti* и *T. szidati*) виды рассматривались отдельно; на втором этапе анализа все виды были сравнены между собой одновременно. Многомерное шкалирование было выполнено с помощью функции metaMDS из библиотек «vegan» (v2.5-7) для языка программирования R.

Функциональная аннотация наборов белков *F. gigantea*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* была выполнена с помощью Интернет-ресурса eggNOG-mapper (v2) (Huerta-Cepas et al., 2017; Cantalapiedra et al., 2021) с параметрами запуска по умолчанию. Мы использовали библиотеку «topGO» (v2.40.0) для проведения анализа обогащения терминов геной онтологии (Gene Ontology, GO). Для каждой фазы жизненного цикла шести указанных видов дигеней в анализе рассматривались только наборы дифференциально экспрессирующихся генов. Для расчета статистической значимости мы использовали точный тест Фишера и среди полученных результатов (GO-терминов с P-значениями <0.01) были отобраны только GO-термины, которые включали не менее 10 значимых генов. Уменьшение избыточности мы выполнили с помощью библиотеки «gtvgo» (v1.0.2) для языка программирования R.

Определение потенциальных секреторных белков 11 видов дигеней мы выполнили *in silico* в соответствии с протоколом, описанным в публикации Гарга и Ранганатана (Garg, Ranganathan, 2011). Далее мы провели анализ обогащения терминов геной онтологии потенциальными «классическими» и «неклассическими» секреторными белками.

Реконструкция моделей геномов последних общих предков Platyhelminthes и Digenea, как и сравнительный анализ между ними были проведены с помощью библиотеки ruHAM (v1.1.10) (Train et al., 2019) для языка программирования Python. Мы исключили из моделей геномов те иерархические ортогруппы, которые включали менее 75% рассматриваемых видов плоских червей.

Анализ обогащения терминов генной онтологии выполнили для списков генов, появившихся и дублированных у последнего общего предка дигеней. Для проведения анализа мы использовали библиотеку «topGO» для языка программирования R.

Филостратиграфический анализ наборов белков 14 исследуемых видов плоских червей мы провели с помощью библиотеки «phylostratr» (v0.2.1) (Arendsee et al., 2019) для языка программирования R. Далее мы использовали результаты филостратиграфического анализа для определения композиции различных наборов белок-кодирующих генов.

Определение возрастных индексов транскриптомов (Transcriptome Age Index, TAI) провели для *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. Анализ был выполнен с помощью библиотеки «myTAI» (v0.9.3) (Drost et al., 2018) для языка программирования R. Далее для наборов генов с наибольшими вкладами в возрастные индексы молекулярных подписей фаз был проведен анализ обогащения терминов генной онтологии с использованием библиотеки «topGO» для языка программирования R.

Глава 3 Результаты

3.1 Подготовка библиотек коротких прочтений для *de novo* асемблирования транскриптомов

По результатам секвенирования подготовленных образцов транскриптомов для редий, церкарий и марит *P. simillimum* и *S. pseudoglobulus* получено по две библиотеки, каждая из которых включала не менее 23,3 миллионов коротких парных прочтений.

3.2 Сборка протяженных последовательностей из коротких прочтений *de novo*

С одновременным использованием всех подготовленных библиотек прочтений *de novo* собрано 247252 и 288060 протяженных последовательностей (контигов) для *P. simillimum* и *S. pseudoglobulus*, соответственно. У обоих исследованных видов Psilostomatidae более 95% контигов характеризовались высокими показателями качества и полноты сборки.

3.3 Квантификация уровней экспрессии и определение кодируемых аминокислотных последовательностей

При квантификации экспрессии контигов с помощью программы Salmon количество успешно картированных пар коротких прочтений в каждой из библиотек варьировало от 78 до 90 процентов. После исключения генов с уровнем экспрессии меньше 1 транскрипта-на-миллион во всех образцах таблицы содержали информацию о 144535 (*P. simillimum*) и 150816 (*S. pseudoglobulus*) последовательностях. Согласно результатам анализа последовательностей с помощью TransDecoder, 32986 (*P. simillimum*) и 69006 (*S. pseudoglobulus*) контигов кодировали белки длиной равной или большей 100 аминокислот.

3.4 Подготовка «справочных» наборов последовательностей и их аннотация

Подготовленные «справочные» наборы включали 21433 (*P. simillimum*) и 46424 (*S. pseudoglobulus*) белок-кодирующих генов. Большинство репрезентативных последовательностей имело совпадение как минимум с одной из баз данных NCBIInt/nr, SwissProt, PfamA или eggNOG.

На рисунке 1 представлены результаты сравнения наборов репрезентативных белков двух исследованных Psilostomatidae с базой данных ортологов, присутствующих в числе одной копии в геномах Metazoa.

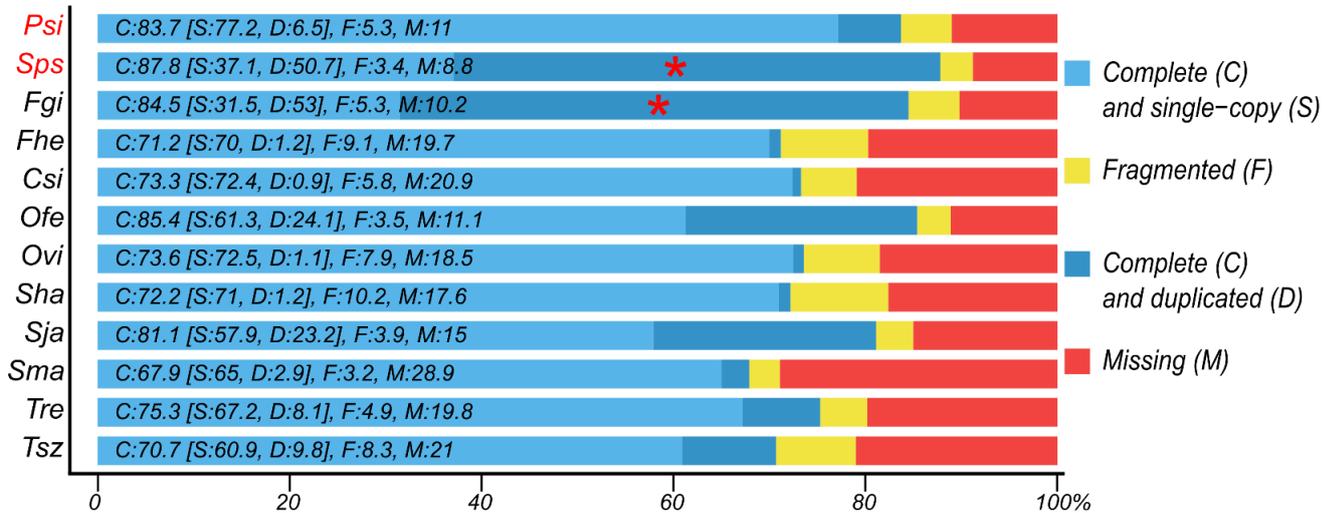


Рисунок 1 – Результаты анализа наличия последовательностей однокопийных ортологов Metazoa в наборах генов паразитических плоских червей. Обозначения: Complete (C) and single-copy (S) – количество последовательностей однокопийных ортологов, собранных полностью и присутствующих в числе одной копии в анализируемом наборе данных; Complete (C) and duplicated (D) – количество последовательностей однокопийных ортологов, собранных полностью и присутствующих в числе нескольких копий в анализируемом наборе данных; Fragmented (F) – количество последовательностей однокопийных ортологов, собранных только частично; Missing (M) – количество последовательностей однокопийных ортологов, отсутствующих в анализируемом наборе данных; красной звездочкой отмечено высокое число дублированных последовательностей у *F. gigantea* и *S. pseudoglobulus*; Psi – *Psilotrema simillimum*, Sps – *Sphaeriodotrema pseudoglobulus*, Fgi – *Fasciola gigantea*, Fhe – *Fasciola hepatica*, Csi – *Clonorchis sinensis*, Ofe – *Opisthorchis felineus*, Ovi – *Opisthorchis viverrini*, Sha – *Schistosoma haematobium*, Sja – *Schistosoma japonicum*, Sma – *Schistosoma mansoni*, Tre – *Trichobilharzia regenti*, Tsz – *Trichobilharzia szidati*.

3.5 Поиск и анализ гаплотипов *cox1* в «справочном» наборе генов *S. pseudoglobulus*

Результаты поиска последовательностей основной субъединицы цитохром с-оксидазы (*cox1*) указали на наличие двух копий гена в «справочном» наборе *S. pseudoglobulus*. Сравнение копий с аналогичными последовательностями из базы данных GenBank показало практически полную идентичность одной из них (GenBank: MT934401) с *cox1*, ранее опубликованной для *S. pseudoglobulus*. В то же время, другая обнаруженная последовательность (GenBank: MT934400) отличалась от всех рассмотренных *cox1* из базы данных.

3.6 Определение и анализ групп ортологичных последовательностей

Полученные результаты поиска и анализа ортогрупп указали на то, что по составу наборов белок-кодирующих генов *P. simillimum* и *S. pseudoglobulus* больше схожи с *F. gigantea* и *F. hepatica*, чем с другими исследованными видами дигеней.

3.7 Определение молекулярных подписей фаз сложных жизненных циклов дигеней

На рисунке 2 представлены диаграммы Венна для молекулярных подписей редий, церкарий и мариц *P. simillimum* (А) и *S. pseudoglobulus* (В).

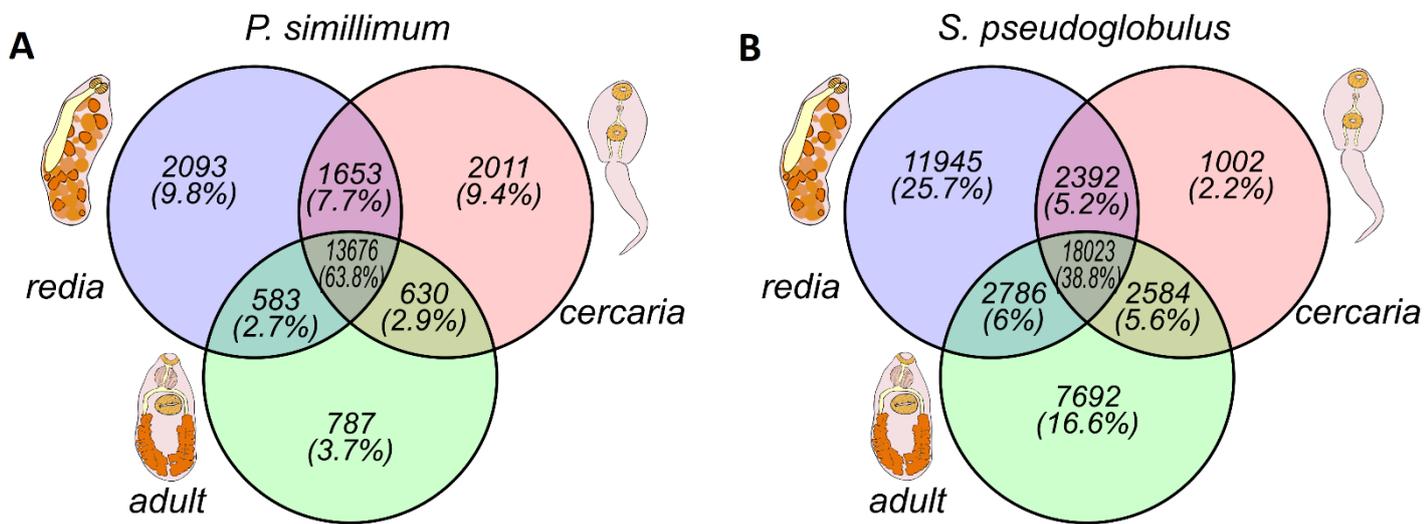


Рисунок 2 – Диаграммы Венна, построенные для молекулярных подписей фаз сложных жизненных циклов *Psilotrema simillimum* (A) и *Sphaeridiotrema pseudoglobulus* (B). Для включения гена в молекулярную подпись фазы ген должен был иметь уровень экспрессии ≥ 1 транскрипта-на-миллион на рассматриваемой фазе жизненного цикла. Числовые значения отражают количество общих (пересечения) или специфичных (участки без пересечений) генов в составе молекулярных подписей. Обозначения: redia – редия, cercaria – церкария, adult – марита.

Большая часть идентифицированных генов *P. simillimum* (63,8%, 13676 генов) (Рисунок 2A) активна на всех фазах цикла, тогда как у *S. pseudoglobulus* количество таких генов не превышало 39% (18023 генов) (Рисунок 2B).

С целью сфокусировать внимание только на генах, которые наиболее вероятно активно транскрибируются, мы повысили порог минимальной экспрессии для включения гена в молекулярную подпись до 2 транскриптов-на-миллион. Мы также исключили из анализа *S. pseudoglobulus* из-за предположительного влияния смеси видов на полученные результаты. Применение критерия минимальной экспрессии в 2 транскрипта-на-миллион позволило выявить, что каждая молекулярная подпись фазы жизненного цикла у *F. gigantea*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* включала минимум 60% от общего числа рассмотренных белок-кодирующих генов исследованного вида. Количество генов со стабильной экспрессией на всех рассмотренных фазах сложного жизненного цикла у 5 из 6 видов составило более 50% от общего числа анализируемых белок-кодирующих генов.

3.8 Анализ дифференциальной экспрессии белок-кодирующих генов

Более 1000 дифференциально экспрессирующихся генов обнаружено для 13 из 27 рассмотренных фаз сложных жизненных циклов. В большинстве случаев (18 из 27 образцов) доля генов со статистически значимым повышением экспрессии от размера молекулярной подписи рассматриваемой фазы не превышала 20%.

3.9 Определение кластеров ко-экспрессии

У *F. gigantea* и *F. hepatica* в кластеры ко-экспрессии вошло менее 50% белок-кодирующих генов, тогда как у *P. simillimum* и *S. mansoni* обнаруженные кластеры включали приблизительно 70% генов.

3.10 Многомерное шкалирование молекулярных подписей фаз сложных жизненных циклов дигеней

Определение сходства молекулярных подписей фаз как в пределах одного сложного жизненного цикла, так и между разными видами дигеней, было выполнено с помощью метода многомерного шкалирования. На рисунке 3 представлено распределение молекулярных подписей фаз одного сложного жизненного цикла на двумерной плоскости по результатам анализа подготовленных матриц. В большинстве случаев (11 / 18) кластеры содержали только биологические повторности одной фазы сложного жизненного цикла рассматриваемого вида

(Рисунок 3). Результаты сравнительного анализа разных видов дигеней представлены на рисунке 4.

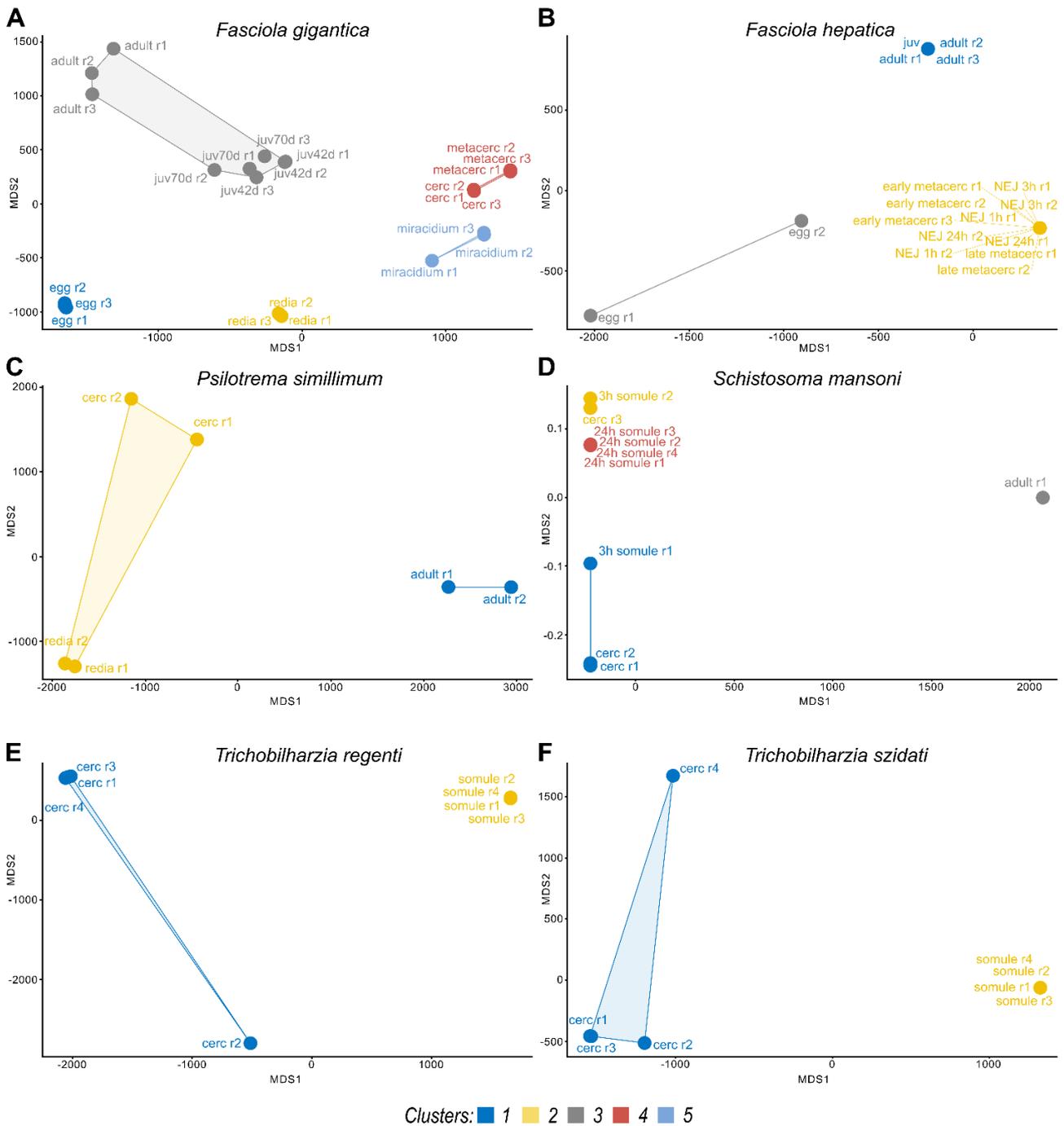


Рисунок 3 – Результаты внутривидового многомерного шкалирования молекулярных подписей фаз. Схожесть молекулярных подписей обратно пропорциональна расстоянию между ними: чем ближе располагаются молекулярные подписи, тем больше схожи они между собой. Разными цветами отмечены обнаруженные кластеры. Наименования и цветовые идентификаторы кластеров представлены в нижней части рисунка. Обозначения: cerc – церкария; metacerc – метацеркария; juv – ювенили; juv42/70d – 42- и 70-дневные ювенили, соответственно; NEJ 1/3/24h – ювенили через 1, 3 и 24 часа после покидания цисты, соответственно; somule – шистосомула; 3/24h somule – шистосомула через 3 и 24 часа после заражения, соответственно; r1/2/3/4 – идентификаторы биологических повторностей.

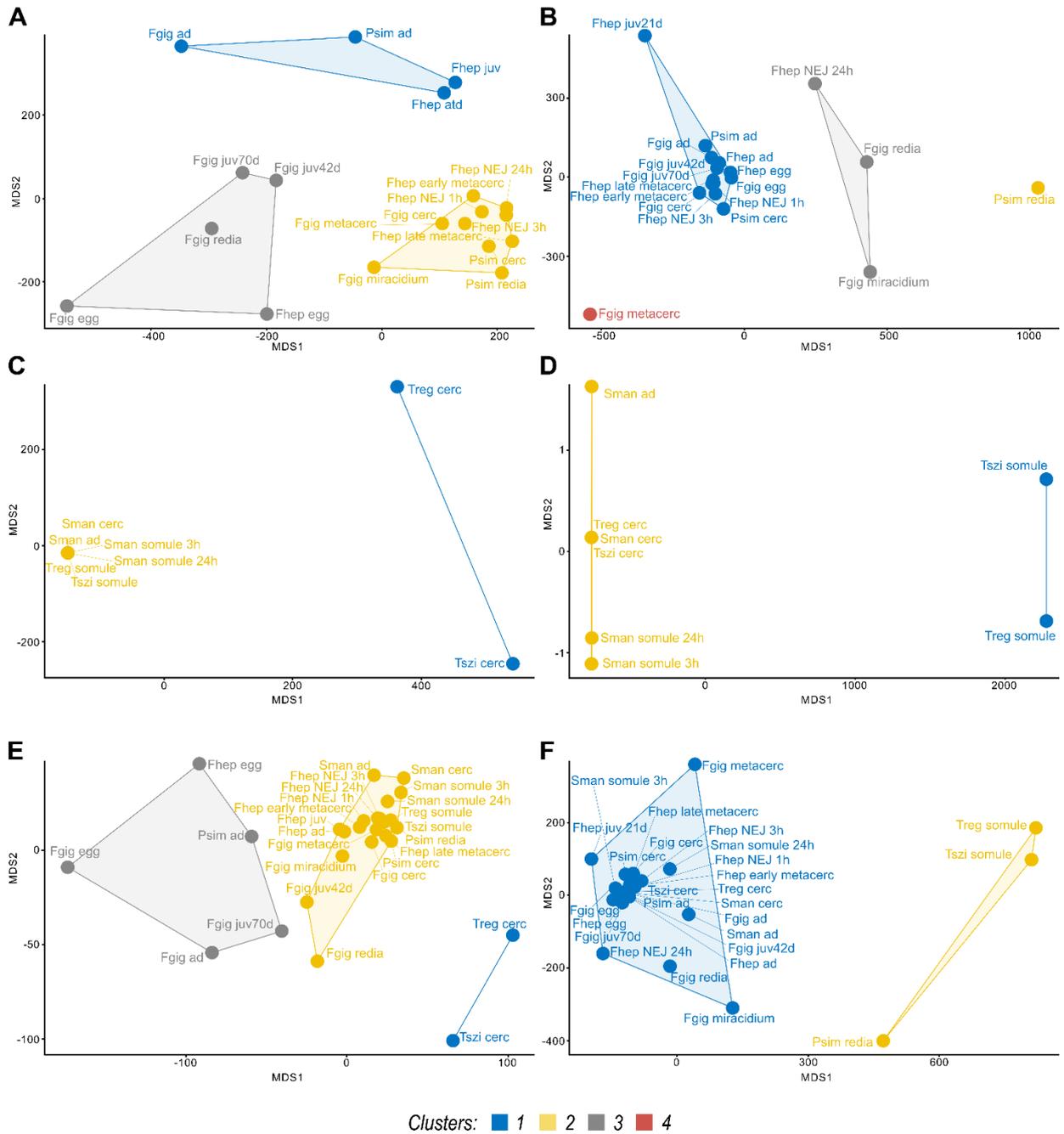


Рисунок 4 – Результаты межвидового многомерного шкалирования молекулярных подписей фаз. Были рассмотрены «редииоидные» (A, B), «спороцистоидные» (C, D), а также все виды дигеней одновременно (E, F). Многомерное шкалирование было проведено на матрицах присутствия/отсутствия ортолога в молекулярной подписи фазы жизненного цикла (A, C, E) и повышенной экспрессии ортолога на рассматриваемой фазе (B, D, F). Схожесть фаз жизненного цикла обратно пропорциональна расстоянию между ними: чем ближе располагаются фазы, тем больше схожи они между собой. Разными цветами отмечены обнаруженные кластеры. Наименования и цветные идентификаторы кластеров представлены в нижней части рисунка. Обозначения: cerc – церкария, metacerc – метацеркария, juv – ювениль, juv42/70d – 42- и 70-дневные ювенили, соответственно; NEJ 1/3/24h – ювинили через 1, 3, 24 часа после покидания цисты, соответственно; somule – шистосомула, somule 3/24 h – шистосомула через 3 и 24 часа после заражения, соответственно; ad – марита.

3.11 Анализ обогащения терминов геномной онтологии

Согласно полученным результатам, для каждой из рассмотренных фаз обнаружено не менее 80 биологических процессов, обогащенных генами со статистически значимым повышением экспрессии на исследуемой фазе сложного жизненного цикла.

Транскриптомные данные по яйцам дигеней доступны только для фасциолид. В числе общих процессов обнаружены «регуляция клеточной смерти» и «пигментация в ходе развития».

Данные по мирацидиям доступны только для *F. gigantica* и для личинки материнской спороцисты выявлена активность таких биологических процессов, как «клеточное дыхание», «транспорт токсинов», «организация митотического веретена», «регенерация печени», а также процессов, ассоциированных с метаболизмом и процессингом РНК.

Для редий *F. gigantica* и *P. simillimum* обнаружено 7 общих обогащенных биологических процессов: «биогенез рибонуклеопротеинового комплекса», «метаболический процесс клеточных соединений азота», «процессинг РНК», «сборка рибонуклеопротеинового комплекса», «ядерно-транскрибируемый катаболический процесс мРНК», «трансляция в митохондриях», «метаболический процесс ДНК». У редий *P. simillimum* удалось обнаружить активность биологических процессов «ДНК-зависимая репликация ДНК», «G1/S переход в митотическом клеточном цикле», «определение пола», «поддержание клеточного состава», «определение судьбы эндодермальных клеток», «дифференцировка дофаминергических нейронов среднего мозга».

Данные по церкариям доступны для *F. gigantica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. В большинстве случаев (10 из 15) общие биологические процессы ассоциированы с метаболизмом, в том числе с метаболическими процессами АТФ, НАДН, НАД. В то же время, у церкарий нескольких видов в числе обогащенных терминов обнаружены биологические процессы, связанные с мышечным движением.

Для ранней и поздней метацеркарий *F. gigantica* и метацеркарии *F. hepatica* удалось обнаружить 20 общих биологических процессов. Ни один из них не был характерен для всех трех образцов одновременно, а большинство (18 из 20) биологических процессов оказалось общими для метацеркарии *F. gigantica* и поздней метацеркарии *F. hepatica*. Среди последних были такие процессы, как «ответ на абиотический раздражитель», «проведение сигнала», «локомоция», «межклеточная передача сигнала», «развитие тканей», «миграция эпителия», «поведение», «ответ на внешний раздражитель», «защитная реакция на другой организм», «постэмбриональное развитие органов животного».

Наличие транскриптомов ювенилей *F. hepatica*, собранных через один, три и 24 часа после покидания цисты, позволило проследить как меняется активность биологических процессов у паразита в ходе миграции в организме хозяина. Нам удалось обнаружить активность 86 биологических процессов, среди которых 11 являлись общими как минимум для двух образцов с различных временных точек. В числе общих биологических процессов были «окислительно-восстановительные процессы», «регуляция клеточной смерти», «локомоция», «негативная регуляция биологических процессов», «гликозилирование» и «проведение сигнала через рецептор на клеточной поверхности».

Сравнительный анализ списков обогащенных терминов геномной онтологии 21-дневной ювенили *F. hepatica* и 42- и 70-дневных ювенилей *F. gigantica* обнаружил, что из 52 биологических процессов большинство обогащено только на одной из рассмотренных ювенилей паразита. Среди них обнаружены такие процессы как «организация ресничек» (21-дневная ювениль *F. hepatica*), «подвижность клеток, зависящая от ресничек или жгутиков» (21-дневная ювениль *F. hepatica*), «материнский процесс, связанный с женской беременностью» (42-дневная ювениль *F. gigantica*), «женская беременность» (70-дневная ювениль *F. gigantica*).

Транскриптомы шистосомул опубликованы для *S. mansoni* (шистосомула через 3 и 24 часа после заражения), *T. regenti* и *T. szidati*. Большинство биологических процессов (33 из 40) оказались общими для двух рассмотренных видов *Trichobilharzia*. В их числе были «клеточное общение», «миграция клеток», «определение клеточной судьбы», «пролиферация клеточной популяции», «эмбриональное развитие», «процессы иммунной системы», «размножение»,

«нейрогенез». Для шистосомул *S. mansoni*, собранных через 3 и 24 часа после заражения, не найдено ни одного общего обогащенного биологического процесса.

Среди 11 биологических процессов, общих для марит различных видов, обнаружены такие, как «организация реснички» (*F. gigantica*, *P. simillimum*), «движение реснички» (*F. gigantica*, *P. simillimum*), «подвижность сперматозоидов» (*F. gigantica*, *P. simillimum*) и «ремоделирование тканей» (*F. gigantica*, *F. hepatica*). Напротив, 77 биологических процессов классифицированы как «обогащённые» только в одном из рассмотренных видов.

3.12 Определение и анализ наборов потенциальных секреторных белков

Для каждого из 11 рассматриваемых видов дигеней удалось *in silico* обнаружить сотни секреторных белков. Все потенциальные секреторные белки были разделены на две группы на основании присутствия («классические») или отсутствия («неклассические») в последовательности классического N-концевого сигнального пептида.

Согласно результатам проведенных анализов обогащения терминов генной онтологии, потенциальные «классические» секреторные белки вовлечены в такие процессы, как «организация внеклеточного матрикса», «ремоделинг тканей», «регуляция хемотаксиса». Среди обогащенных биологических процессов четыре являлись общими для более чем половины исследованных видов дигеней: «протеолиз», «ответ на стресс эндоплазматического ретикулума», «ответ на раздражитель» и «дегрануляция лейкоцитов». Сравнительный межвидовой анализ списков биологических процессов, обогащенных потенциальными «неклассическими» секреторными белками, не обнаружил никаких пересечений.

3.13 Реконструкция и анализ моделей геномов последних общих предков дигеней и свободноживущих плоских червей

После исключения иерархичных ортогрупп, включавших менее 75% рассматриваемых видов плоских червей, обновленные модели геномов последних общих предков Platyhelminthes и Digenea состояли из 2579 и 4622 групп ортологов, соответственно. Результаты сравнительного анализа построенных моделей геномов последних общих предков указали на то, что в модели Digenea 2258 генов осталось от Platyhelminthes, 64 гена дублированы, а 1850 генов появились у последнего общего предка дигеней.

Согласно результатам анализа обогащения терминов генной онтологии, гены, появившиеся у последнего общего предка дигеней, принимают участие в процессах, связанных с проведением сигналов, функционированием клеток, развитием, размножением и регуляцией экспрессии генов. В то же время, гены, дублированные у последнего общего предка дигеней, вовлечены в развитие органов животных, развитие женской половой системы, клеточный ответ на различные раздражители, фагоцитоз, метаболизм эстрогена и жирных кислот, регуляцию различных процессов и миграцию эпителиальных клеток.

3.14 Филостратиграфия и анализ филостратиграфического состава различных наборов белок-кодирующих генов

Согласно результатам филостратиграфического анализа, обнаружены следующие филостраты: 1) «Cellular organisms», 2) «Eukaryota», 3) «Opisthokonta», 4) «Metazoa», 5) «Eumetazoa», 6) «Bilateria», 7) «Protostomia», 8) «Spiralia», 9) «Lophotrochozoa», 10) «Platyhelminthes», 11) «Class», 12) «Order», 13) «Family», 14) «Genus», 15) «Species». Наиболее крупной филостратой у 12 из 14 видов плоских червей оказалась «Cellular organisms». Наименьшей филостратой у всех видов оказалась «Spiralia».

Сравнительный анализ построенных моделей геномов последних общих предков показал, что размер каждой филостраты в модели Digenea превышал размер этой же филостраты в модели Platyhelminthes.

Более 55% генов, имеющих заметную экспрессию на всех рассмотренных фазах сложных жизненных циклов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*, принадлежало филостратам «Cellular organisms» и «Eukaryota».

Сложный филостратиграфический состав обнаружен и у ранее выявленных кластеров ко-экспрессии *F. gigantica*, *F. hepatica*, *P. simillimum* и *S. mansoni*.

Филостратиграфический состав наборов генов, кодирующих как «классические», так и «неклассические» секреторные белки, различался у всех 11 рассмотренных видов дигеней.

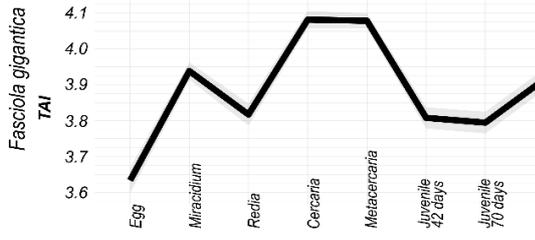
3.15 Определение и анализ возрастных индексов молекулярных подписей фаз сложных жизненных циклов дигеней

Наименьшие значения возрастных индексов получены для молекулярных подписей яиц *F. gigantica* (Рисунок 5A1) и *F. hepatica* (Рисунок 5B1), редии *P. simillimum* (Рисунок 5C1), мариты *S. mansoni* (Рисунок 5D1), а также церкарий *T. regenti* (Рисунок 5E1) и *T. szidati* (Рисунок 5F1). Напротив, наибольшими значениями возрастных индексов обладали церкария *F. gigantica* (Рисунок 5A1), 21-дневная ювениль *F. hepatica* (Рисунок 5B1), марита *P. simillimum* (Рисунок 5C1), шистосомула *S. mansoni* через 3 часа после заражения (Рисунок 5D1), шистосомулы *T. regenti* (Рисунок 5E1) и *T. szidati* (Рисунок 5F1).

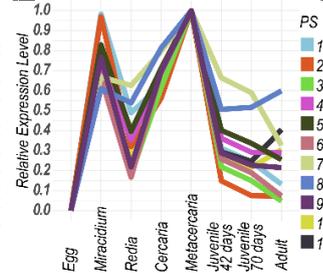
Гены с наибольшими вкладами в возрастные индексы молекулярных подписей перечисленных фаз принимают участие в различных биологических процессах, в том числе и связанных с развитием.

Большинство филострат демонстрировали наименьшие относительные уровни экспрессии на фазах яиц *F. gigantica* (Рисунок 5A2,3) и *F. hepatica* (Рисунок 5B2,3), марит *P. simillimum* (Рисунок 5C2,3) и *S. mansoni* (Рисунок 5D2,3), а также церкарий *T. regenti* (Рисунок 5E2,3) и *T. szidati* (Рисунок 5F2,3). В то же время, наибольший относительный уровень экспрессии у большинства филострат обнаружен на фазах метацеркарии *F. gigantica* (Рисунок 5A2,3), ювенили *F. hepatica* (Рисунок 5B2,3), редии *P. simillimum* (Рисунок 5C2,3), шистосомулы *S. mansoni* через 3 часа после заражения (Рисунок 5D2,3), шистосомул *T. regenti* (Рисунок 5E2,3) и *T. szidati* (Рисунок 5F2,3).

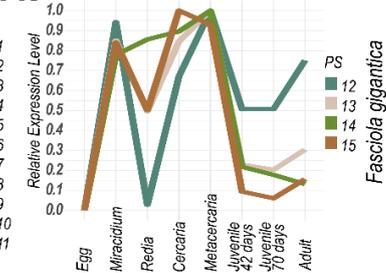
A1



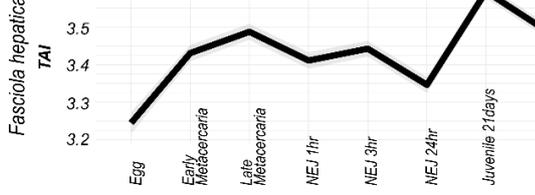
A2



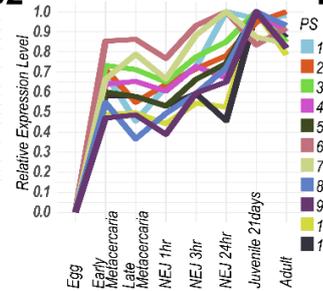
A3



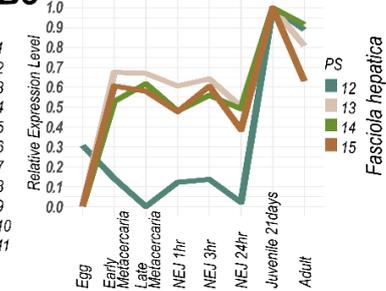
B1



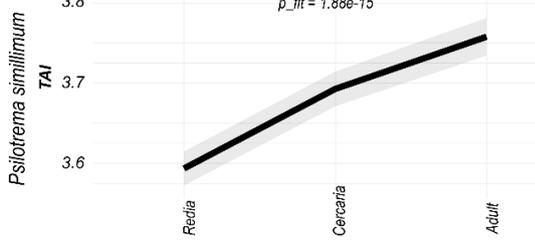
B2



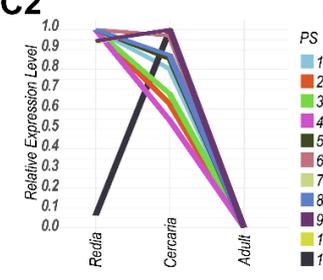
B3



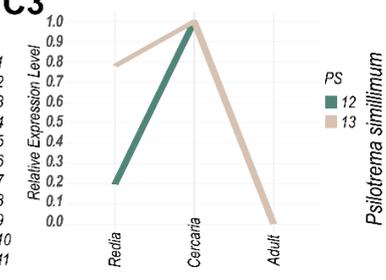
C1



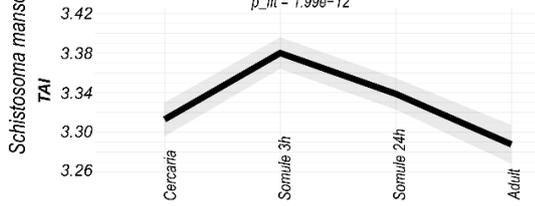
C2



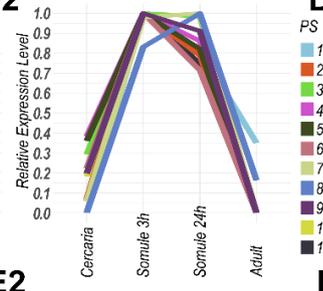
C3



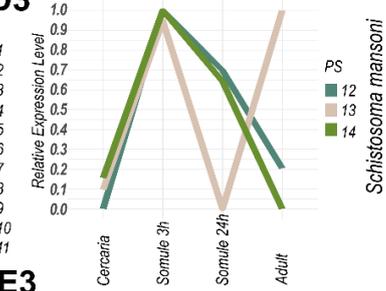
D1



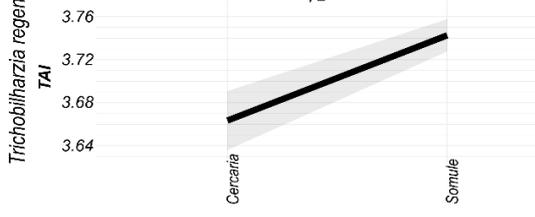
D2



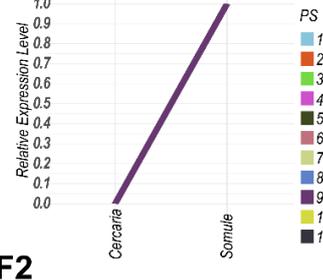
D3



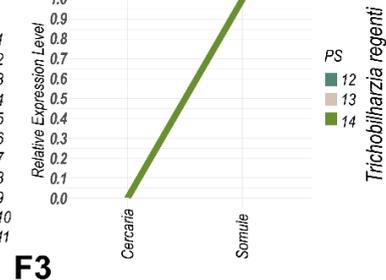
E1



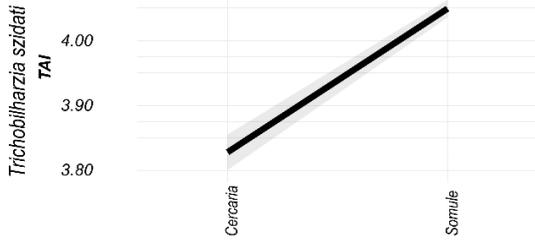
E2



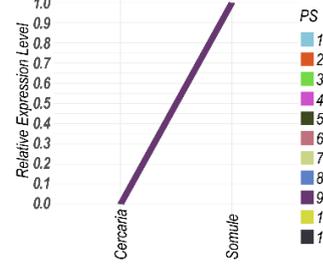
E3



F1



F2



F3

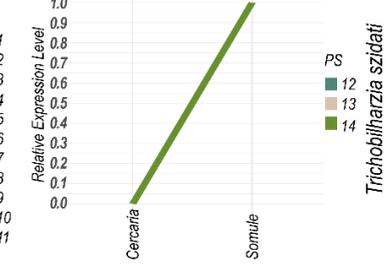


Рисунок 5 – Возрастные индексы транскриптомов фаз сложных жизненных циклов дигеней и относительные уровни экспрессии филострат. Представлены результаты анализов для *F. gigantica* (A1-3), *F. hepatica* (B1-3), *P. simillimum* (C1-3), *S. mansoni* (D1-3), *T. regenti* (E1-3) и *T. szidati* (F1-3). Значения возрастных индексов соответствуют взвешенным вкладам различных филострат: низкие значения возрастных индексов указывают на больший вклад более древних филострат в сравнении со вкладами более молодых филострат. В верхних частях профилей вариации возрастных индексов (A1, B1, C1, D1, E1, F1) приведены P-значения (p_ft), рассчитанные при анализе статистической значимости отклонения филотранскриптомного профиля от плоской линии. Серая зона вокруг профиля соответствует стандартному отклонению (A1, B1, C1, D1, E1, F1). При визуализации относительной экспрессии все филостраты были разделены на две группы: первая группа включала филостраты от «Cellular organisms» до «Class» (A2, B2, C2, D2, E2, F2), тогда как вторая группа содержала филостраты от «Order» до «Species» (A3, B3, C3, D3, E3, F3). Наименьший и наибольший уровни экспрессии среди всех филострат были приняты за «0» и «1», соответственно, а все остальные значения экспрессии были преобразованы в этом диапазоне. Для *P. simillimum* представлены результаты анализа без рассмотрения видоспецифичной филостраты. Обозначения: TAI – возрастные индексы транскриптомов; Egg – яйцо; Miracidium – мирацидий; Redia – редия; Cercaria – церкария; (Early/ Late) Metacercaria – (ранняя / поздняя) метациркария; NEJ1/3/24hr – ювенили через 1, 3 и 24 часа после покидания цисты, соответственно; Juvenile 21/42/70 days – ювенили через 21, 42 и 70 дней после заражения; Somule – шистосомула; Somule 3/24hr – шистосомула через 3 и 24 часа после заражения, соответственно; Adult – марита.

Глава 4 Обсуждение

Сложный жизненный цикл дигеней с последовательным и закономерным чередованием половых поколений является ярким примером формирования и функционирования множества фенотипов путем регуляции работы одного генома. Для описания набора последовательностей, который может быть использован в качестве маркера рассматриваемого фенотипа, используют термин «молекулярная подпись».

Высокая эволюционная пластичность дигеней нашла отражение в разнообразии их жизненных циклов. Как следствие, для успешного сравнительного анализа и определения эволюционно консервативных компонентов молекулярных основ сложных жизненных циклов, необходимо вовлечение геномных и транскриптомных данных с разных ветвей филогенетического дерева дигеней. Перспективными моделями для изучения являются *P. simillimum* и *S. pseudoglobulus*, ввиду как особенностей их биологии, так и возможности постановки их жизненных циклов в лабораторных условиях (Nesterenko et al., 2020).

В ходе наших исследований для *P. simillimum* и *S. pseudoglobulus* впервые выделены и проанализированы образцы транскриптомов фаз партеногенетического (редия) и амфимиктического (церкария и марита) поколений. Представленные транскриптомы являлись первыми не только для указанных видов, но и для всего таксона Psilostomatidae (Nesterenko et al., 2020).

Ввиду отсутствия «справочных» геномов нами было принято решение проводить сборку протяженных последовательностей *de novo* с одновременным использованием библиотек коротких парных прочтений с трех фаз жизненного цикла. Выбранный нами подход позволил не только получить наиболее полное представление о транскриптах, присутствующих на трех анализируемых фазах жизненных циклов двух псилостоматид, но и корректно восстановить их нуклеотидные последовательности. В дальнейшем анализе мы решили сфокусировать внимание только на белок-кодирующих генах с заметными уровнями экспрессии, поскольку белки и взаимодействия между ними являются основой подавляющего большинства биологических процессов, протекающих в живых организмах.

Принимая во внимание результаты поиска однокопийных ортологов Metazoa (Рисунок 1) и сравнения с базами данных, мы утверждаем, что подготовленные наборы белок-кодирующих генов *P. simillimum* и *S. pseudoglobulus* сопоставимы по качеству и полноте с данными по другим видам дигеней и могут быть использованы для внутри- и межвидового сравнительного анализа (Nesterenko et al., 2020). Тем не менее прямое сравнение *S. pseudoglobulus* с другими видами

плоских червей осложнено возможным наличием материала криптических видов *Sphaeridiotrema* в анализируемых образцах (Nesterenko et al., 2020).

По результатам поиска и анализа ортогрупп, все рассмотренные виды плоских червей объединялись в кластеры в соответствии с современными представлениями об их филогенетическом родстве (Nesterenko et al., 2020, 2022).

Для определения множеств активных генов для каждой из рассмотренных фаз сложных жизненных циклов, т. е. для выявления их молекулярных подписей, мы использовали два порога минимальной экспрессии, равные 1 (Рисунок 2) и 2 транскриптам-на-миллион. У большинства исследованных видов дигеней большая часть белок-кодирующих генов демонстрировала заметную экспрессию на всех рассмотренных фазах сложных жизненных циклов. Полученные результаты означают, что фенотип фазы жизненного цикла, включающий также морфологические признаки, физиологические особенности и характерные черты поведения, формируется не только активностью фаза-специфичных генов, но и фаза-специфичным изменением экспрессии множества генов, активных и на других этапах сложного жизненного цикла дигеней.

Результаты проведенного анализа доказали, что в ходе реализации сложных жизненных циклов дигеней многие белок-кодирующие гены статистически значимо усиливают свою экспрессию на разных фазах (Nesterenko et al., 2022). В большинстве случаев количество генов со статистически значимым повышением экспрессии не превышало 20% от размера молекулярной подписи фазы (Nesterenko et al., 2022). Результаты анализа обогащения терминов генной онтологии позволили определить, что многие дифференциально экспрессирующиеся гены принимают участие в ключевых для исследованных фаз жизненных циклов биологических процессах. В качестве примеров могут быть рассмотрены регуляция мышечного сокращения у активно плавающих церкарий или развитие тканей у метацеркарий и шистосомул, подготавливающих к обитанию и размножению в организме окончательного хозяина.

При обсуждении результатов определения дифференциальной экспрессии необходимо учитывать, что при выделении образцов мРНК фазы жизненного цикла мы частично или полностью охватываем транскриптомы всех клеточных типов, составляющих организм паразита. Следовательно, мы можем заключить, что молекулярная подпись фазы сложного жизненного цикла – это многокомпонентная система, составными частями которой являются не столько гены сами по себе, сколько уровни их экспрессии в определенных условиях, в том числе и в клеточных типах, формирующих организм (Nesterenko et al., 2022).

Мы можем предположить, что чем больше схож клеточный состав сравниваемых фаз сложного жизненного цикла дигеней или чем больше схожи условия их обитания, тем больше схожи будут их молекулярные подписи (Nesterenko et al., 2022). Подтверждением нашего предположения может выступать совместная кластеризация схожих фаз как в пределах одного сложного жизненного цикла (Рисунок 3), так и между видами (Рисунок 4), выявленная по результатам многомерного шкалирования молекулярных подписей. Основываясь на полученных результатах, мы можем сделать вывод, что в молекулярных подписях фаз жизненных циклов присутствуют эволюционно консервативные модули, общие как для филогенетически близких, так и отдаленных видов дигеней.

С целью ответить на вопрос, какие изменения в геноме предка дигеней сделали возможным эволюционное усложнение жизненного цикла, мы провели реконструкцию моделей генов последних общих предков Digenea. Результаты проведенного анализа обогащения терминов генной онтологии указали на большой вклад как появившихся, так и дублированных генов во многие биологические процессы последнего общего предка дигеней, в том числе и в его эмбриональное развитие и/или метаморфоз (Nesterenko et al., 2022). Мы предполагаем, что добавление «новых» генов во многие биологические процессы послужило преадаптацией для усложнения жизненного цикла последнего общего предка дигеней. В качестве одного из возможных механизмов формирования нескольких контрастных фенотипов может быть рассмотрено «закрепление» полифенизма, т. е. наличия множества альтернативных фенотипов, вызванных действием различных условий окружающей среды (Minelli, Fusco, 2010).

Большое число доступных геномных и транскриптомных данных открыло возможность не только реконструировать модели геномов предков, но и разделить набор генов исследуемого таксона на группы с разным филогенетическим происхождением, называемых филостратами. В рамках проведенного исследования практически все анализируемые белок-кодирующие гены 14 видов плоских червей были успешно распределены по 15 филостратам от «Cellular organisms» до «Species». Основываясь на полученных результатах анализа филостратиграфического состава различных наборов генов, мы предполагаем, что усложнение жизненного цикла последнего общего предка дигеней потребовало не только добавления «новых» генов в уже имеющиеся генетические сети, но и согласования экспрессии генов из разных филострат.

Используя информацию о филостратиграфической принадлежности генов, возможно как исследовать профили экспрессии отдельных филострат, так и измерить их вклад в молекулярные подписи фаз (Рисунок 5). Несмотря на то, что в пределах жизненных циклов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* обнаружены различия в возрастных индексах молекулярных подписей фаз, полученные результаты не могут быть напрямую использованы для реконструкции последовательности включения фаз в жизненный цикл (Nesterenko et al., 2022). Высокий возрастной индекс молекулярной подписи фазы, указывает не столько на позднее «появление» фазы в жизненном цикле, сколько на наличие специфичных и/или сильно преобразованных признаков у рассматриваемой фазы.

Важно отметить, что обогащение биологических процессов, связанных с регуляцией развития, выявлено у фаз жизненных циклов как с наименьшими (яйца *F. gigantica* и *F. hepatica*, редия *P. simillimum*, марита *S. mansoni*), так и с наибольшими (церкария *F. gigantica*, шистосомула *S. mansoni* через 3 часа после заражения, шистосомулы *T. regenti* и *T. szidati*) значениями возрастных индексов молекулярных подписей. Значение возрастного индекса будет тем выше, чем больше вклад в молекулярную подпись более молодых филострат. Исходя из этого, мы можем предположить, что и в регуляцию процессов развития, протекающих на фазах с высокими возрастными индексами, вовлечено больше «новых» генов, чем на фазах жизненного цикла с низкими значениями возрастных индексов.

Проведенный анализ относительных уровней экспрессии филострат доказал, что группы генов с разным филогенетическим происхождением изменяют активность на разных фазах сложного жизненного цикла. Полученные результаты подтверждают, что для формирования последовательного и закономерного чередования контрастных фаз и стадий в сложном жизненном цикле дигеней требуется наличие комплексной системы регуляции, обеспечивающей согласованную экспрессию различных групп генов.

Заключение

В рамках проведенного исследования мы применили современные методы биоинформатического анализа как к полученным нами, так и опубликованным ранее транскриптомным данным дигеней с целью определить молекулярные основы реализации сложных жизненных циклов дигенетических сосальщиков (Nesterenko et al., 2020, 2022). Ранее детального сравнительного анализа молекулярных подписей фаз жизненных циклов филогенетически близких и отдаленных видов дигеней не проводилось.

Подготовленные и представленные в настоящей работе транскриптомы *Psilotrema simillimum* и *Sphaeridiotrema pseudoglobulus* являются первыми секвенированными транскриптомами для исследованных видов и целого таксона Psilostomatidae (Nesterenko et al., 2020). По качеству и полноте сборки транскриптомы обоих видов соответствуют данным, опубликованным для многих видов дигеней. Результаты поиска и анализа ортогорупп доказали, что по составу наборов белок-кодирующих генов *P. simillimum* и *S. pseudoglobulus* больше схожи с *F. gigantica* и *F. hepatica*, чем с другими исследованными видами плоских червей.

В отличие от *P. simillimum*, прямой сравнительный анализ транскриптома *S. pseudoglobulus* с различными видами плоских червей осложнен возможным наличием материала криптического вида *Sphaeridiotrema* в собранных нами образцах и отсекуемых

библиотеках (Nesterenko et al., 2020). Предположение о наличии близкого к *S. globulus* и *S. pseudoglobulus* криптического вида, информация о котором отсутствует в базе данных GenBank, основано на результатах анализа копий последовательностей основной субъединицы цитохром с-оксидазы. Опубликованные нами данные станут важным дополнением в исследовании молекулярных сходств и различий между криптическими видами *Sphaeridiotrema*.

Результаты проведенного анализа экспрессии генов позволили установить, что в ходе реализации сложных жизненных циклов дигеней происходит изменение экспрессии многих белок-кодирующих генов, что приводит к формированию фаза-специфичных молекулярных подписей. Составными компонентами молекулярной подписи фазы жизненного цикла являются не столько гены сами по себе, сколько уровни их экспрессии в определенных условиях, в том числе и в клеточных типах, формирующих организм (Nesterenko et al., 2020, 2022). Совместная кластеризация схожих фаз жизненных циклов подтвердила наше предположение о наличии зависимости между схожестью молекулярных подписей и сходством клеточного состава и/или воздействующих условий между фазами жизненных циклов. Выявленные сходства между молекулярными подписями фаз филогенетически близких и отдаленных видов дигеней могут служить основой для детального исследования эволюционно консервативных модулей регуляции активности генома в ходе реализации сложного жизненного цикла.

Не менее важной частью нашего исследования стала реконструкция моделей геномов предков дигеней. Сравнительный анализ моделей геномов последних общих предков Platyhelminthes и Digenea показал, что количество генов, появившихся у последнего общего предка дигеней, превышает количество дублицированных генов. Тем не менее как появившиеся, так и дублицированные гены принимали участие во многих биологических процессах, в том числе и связанных с развитием и/или метаморфозом. На основании полученных данных мы можем рассматривать расширение молекулярных основ многих биологических процессов в качестве одной из основных преадаптаций для усложнения жизненного цикла последнего общего предка Digenea (Nesterenko et al., 2022).

В рамках нашего исследования впервые применены методы филостратиграфии и эволюционной транскриптомики для анализа данных по дигенетическим сосальщикам (Nesterenko et al., 2022). Нам удалось успешно разделить наборы белок-кодирующих генов у 11 паразитирующих и 3 свободноживущих видов плоских червей на группы с разным филогенетическим происхождением. Использование информации о филостратиграфической аффилиации генов позволило нам продемонстрировать сложный филостратиграфический состав разных групп генов, в том числе и кодирующих потенциальные секреторные белки. Более того, с помощью методов эволюционной транскриптомики нам удалось установить, что фазы сложного жизненного цикла дигеней различаются между собой по вкладам в их молекулярные подписи групп генов с разным филогенетическим происхождением (Nesterenko et al., 2022). Полученные нами результаты анализа возрастных индексов молекулярных подписей фаз сложных жизненных циклов дигеней ставят новые вопросы и создают основу для дальнейших исследований. Одним из перспективных направлений возможных исследований является секвенирование и анализ транскриптомов отдельных клеток, формирующих организм контрастных фаз сложного жизненного цикла дигеней, и применение методов эволюционной транскриптомики в исследовании полученных данных.

Выводы

- 1) В ходе реализации сложных жизненных циклов дигеней происходит согласованное изменение экспрессии многих белок-кодирующих генов, что приводит к формированию фаза-специфичных молекулярных подписей. Молекулярная подпись фазы жизненного цикла дигеней представляет собой уникальный набор активных белок-кодирующих генов и включает последовательности, экспрессирующиеся как на одном, так и нескольких этапах жизненного цикла.

- 2) Обнаруженная схожесть молекулярных подписей фаз как во внутривидовом, так и межвидовом анализе свидетельствует о схожести клеточного состава и/или условий, воздействующих на организм паразита.
- 3) Основным направлением эволюционной трансформации генома последнего общего предка дигеней было, вероятно, возникновение новых белок-кодирующих генов, а не дупликация уже присутствующих последовательностей.
- 4) Фазы одного сложного жизненного цикла статистически значимо различаются по вкладам в их молекулярные подписи групп генов с разным филогенетическим происхождением.

Список публикаций по теме диссертации

Публикации в изданиях, рекомендуемых ВАК:

Nesterenko M. A. Molecular signatures of the rediae, cercariae and adult stages in the complex life cycles of parasitic flatworms (Digenea: Psilostomatidae) / M. A. Nesterenko, V. V. Starunov, S. V. Shchenkov, A. R. Maslova, S. A. Denisova, A. I. Granovich, A. A. Dobrovolskij, K. V. Khalturin // *Parasites & Vectors*. – 2020. – Vol. 13. – №. 1. – P. 1–21.

Nesterenko M. The digenean complex life cycle: phylostratigraphy analysis of the molecular signatures / M. Nesterenko, S. Shchenkov, S. Denisova, V. Starunov // *Biological Communications* – 2022. – Vol. 67. – №. 2. – P. 65–87.

Публикации в прочих журналах, трудах, сборниках и материалах конференций:

Nesterenko M. A. Comparative transcriptome analysis of parthenogenetic and amphimictic generations in *Sphaeridiotrema pseudoglobulus* (Trematoda: Psilostomatidae) / M. A. Nesterenko, V. V. Starunov, S. V. Shchenkov, A. A. Dobrovolskij, A. I. Granovitch, A. G. Gonchar, K. V. Khalturin // *The 12th European Multicolloquium of Parasitology*. [Электронный ресурс] – 2016. – 1 электрон. опт. диск (DVD-ROM).

Нестеренко М. А. Сравнительный анализ транскриптомов партеногенетического и амфимиктического поколений трематоды *Sphaeridiotrema globulus* / М. А. Нестеренко, В. В. Старунов, А. А. Добровольский, С. В. Щенков, К. В. Халтурин // *Современные проблемы теоретической и морской паразитологии: сборник научных статей* / Ред. К. В. Галактионов, А. В. Гаевская – Севастополь: Изд-ль Бондаренко Н. Ю. – 2016. – С. 242.

Nesterenko M. Does the blind trematod`s larvae see the light? / M. Nesterenko, S. Shchenkov, V. Starunov, V. Khabibulina // 4-й Международный конгресс по морфологии беспозвоночных (4th International Congress on Invertebrate Morphology) / Отр. ред. Е. Н. Темерева – М.: Издательство «Перо» – 2017. – С. 398.

Нестеренко М. А. Сложный жизненный цикл: множество фенотипов на базе одного генома / М. А. Нестеренко, В. В. Старунов, С. В. Щенков, А. А. Добровольский, К. В. Халтурин // *Современная паразитология – основные тренды и вызовы. Материалы VI Съезда Паразитологического общества: Международная конференция (15–19 октября 2018 г., Зоологический институт РАН, Санкт-Петербург)* / Ред. К. В. Галактионов, С. Г. Медведев, А. Ю. Рысс, Ф. О. Фролов – Санкт-Петербург: Издательство «Лема» – 2018. – С. 298.

Нестеренко М. А. Сложный жизненный цикл трематод: филостратиграфический анализ молекулярных подписей стадий / М. А. Нестеренко // *Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2021»* / Отв. ред. И. А. Алешковский, А. В. Андриянов, Е. А. Антипов, Е. И. Зимакова. [Электронный ресурс] – М.: МАКС Пресс – 2021. – 1 электрон. опт. диск (DVD-ROM).

Nesterenko M. The digenean complex life cycle: the phylostatigraphy analysis of the molecular signatures / M. Nesterenko, V. Starunov, S. Shchenkov, S. Denisova // 9th Conference of the Scandinavian - Baltic Society for Parasitology: Abstract Book [Электронный ресурс] – 2021. – 1 электрон. опт. диск (DVD-ROM).

Nesterenko M. A. Evolution of complexity: transcriptome age indexes analysis in a complex life cycle / M. A. Nesterenko, S. V. Shchenkov, S. A. Denisova, V. V. Starunov // Bioinformatics: From Algorithms to Applications: Conference proceedings [Электронный ресурс] – 2021. – 1 электрон. опт. диск (DVD-ROM).

Nesterenko M. A. The digenean complex life cycle: phylostratigraphy analysis of molecular signatures / M. A. Nesterenko, V. V. Starunov, S. V. Shchenkov, S. A. Denisova // 10th Moscow Conference on Computational Molecular Biology MCCMB'21: Proceedings [Электронный ресурс] – М.: ИППИ РАН – 2021. – 1 электрон. опт. диск (DVD-ROM).