

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

На правах рукописи

Нестеренко Максим Алексеевич

СРАВНИТЕЛЬНАЯ И ЭВОЛЮЦИОННАЯ ТРАНСКРИПТОМИКА РАЗНЫХ ФАЗ
СЛОЖНЫХ ЖИЗНЕННЫХ ЦИКЛОВ ДИГЕНЕТИЧЕСКИХ СОСАЛЬЩИКОВ

1.5.17. Паразитология (биологические науки)

Диссертация
на соискание ученой степени
кандидата биологических наук

Научный руководитель:
кандидат биологических наук
Виктор Вячеславович Старунов

Санкт-Петербург – 2022

СОДЕРЖАНИЕ

Введение.....	4
Глава 1 Обзор литературы.....	11
1.1 Общее описание фаз сложного жизненного цикла дигеней.....	11
1.2 Развитие представлений о структуре жизненных циклов дигеней	17
1.3 Описание схем реализации сложных жизненных циклов исследуемых видов дигеней	18
1.4 Применение методов биоинформатического анализа для исследования молекулярных основ реализации сложных жизненных циклов дигеней	22
1.5 Теоретические основы филостратиграфии и применение возрастных индексов транскриптомов в изучении жизненных циклов	32
Глава 2 Материалы и методы.....	39
Глава 3 Результаты.....	60
3.1 Подготовка библиотек коротких прочтений для <i>de novo</i> сборки транскриптомов	60
3.2 Сборка протяженных последовательностей из коротких прочтений <i>de novo</i>	62
3.3 Квантификация уровней экспрессии и определение кодируемых аминокислотных последовательностей.....	62
3.4 Подготовка «справочных» наборов последовательностей и их аннотация.....	62
3.5 Поиск и анализ гаплотипов <i>cox1</i> в «справочном» наборе генов <i>S. pseudoglobulus</i>	65
3.6 Определение и анализ групп ортологичных последовательностей.....	66
3.7 Определение молекулярных подписей фаз сложных жизненных циклов дигеней	71
3.8 Анализ дифференциальной экспрессии белок-кодирующих генов	77
3.9 Определение кластеров ко-экспрессии	78
3.10 Многомерное шкалирование молекулярных подписей фаз сложных жизненных циклов дигеней.....	79
3.11 Анализ обогащения терминов геномной онтологии	83
3.12 Определение и анализ наборов потенциальных секреторных белков	88

3.13 Реконструкция и анализ моделей геномов последних общих предков дигеней и свободноживущих плоских червей.....	89
3.14 Филостратиграфия и анализ филостратиграфического состава различных наборов белок-кодирующих генов	91
3.15 Определение и анализ возрастных индексов молекулярных подписей фаз сложных жизненных циклов дигеней	96
Глава 4 Обсуждение.....	102
Заключение	133
Выводы.....	135
Список литературы	136
Список публикаций по теме диссертации	152
ПРИЛОЖЕНИЕ 1. Список библиотек коротких парных прочтений, использованных в исследовании (страницы 42, 50)	154

Введение

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.

Стремительное развитие технологий высокопроизводительного секвенирования сделало возможным детальное исследование молекулярных основ биологии широкого круга живых организмов, в том числе и паразитических. 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 felinus* (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. Исследование транскриптома редии, одной из ключевых фаз сложного жизненного цикла дигеней, проведено только для *F. gigantea*. 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. gigantea* (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 дуплицированные 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 филостратиграфии 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 редий, 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 Multicollloquium of Parasitology, ЕМОРХII) (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 (VIATA2021, 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.

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. Халтурину 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.

Глава 1 Обзор литературы

В сложном жизненном цикле дигеней происходит последовательное и закономерное чередование разных поколений, каждое из которых обладает собственным онтогенезом и может быть разделено на несколько фаз (Гранович и др., 2010). В поколении материнской спороцисты можно выделить 3 фазы: личинка в яичевой оболочке, мирацидий и материнская спороциста. Дочернее партеногенетическое поколение в сложном жизненном цикле дигеней представлено только одной фазой – редией или дочерней спороцистой. Церкария, метацеркария и марита представляют собой 3 фазы амфимиктического поколения. В то же время каждая отдельная фаза состоит из стадий развития (Гранович и др., 2010). В качестве примера авторы указывают, что фаза метацеркарии включает в себя стадию, характеризующуюся интенсивными морфогенетическими перестройками и стадию относительного морфофункционального покоя. В зарубежной литературе практически нет разделения между «фазой» и «стадией» и для описания конкретного этапа сложного жизненного цикла дигеней используется термин «stage». В настоящей работе будут преимущественно рассмотрены фазы сложного жизненного цикла дигеней.

1.1 Общее описание фаз сложного жизненного цикла дигеней

Исследования сложных жизненных циклов дигеней охватывают сотни лет и одним из первых изученных видов стал печеночный сосальщик *Fasciola hepatica*. Самые ранние отсылки к печеночному сосальщику и к болезни печени у овец можно обнаружить в завершенном в 1379 году трактате «Le Bon Berger» за авторством де Бри (de Brie) (по Reinhard, 1957). Первое же узнаваемое описание дигеней вида *F. hepatica* приведено в книге сэра Фитцхерберта (Fitzherbert), опубликованной в 1523 (по Reinhard, 1957). Иллюстрация печеночного сосальщика была впервые представлена только в 1668 году Реди (Redi) (по Reinhard, 1957). Латинское название *Fasciola hepatica* паразиту было присвоено позже, в 1758 году Карлом Линнеем. Название же целой группе «перфорированных» плоских червей – «Trematoda», означающее «тело, пронзенное отверстиями» - дал Рудольфи (Rudolphi) в 1808 году (по Reinhard, 1957).

Яйцо и мирацидий

В материалах книги Бидлоо (Bidloo), напечатанной в 1698, описано первое наблюдение яиц в живых особях дигеней (по Reinhard, 1957). Исследователь также первым высказал предположение, что заражение происходит через поглощение самого паразита или его яиц.

Яйца дигеней эктолецитальные, снаружи покрыты скорлупой, обычно имеют овальную форму, от светло- до темно-коричневого цвета (Fried, Graczyk, 1997). Формирование сложных яиц у всех дигеней происходит схожим образом (Galaktionov, Dobrovolskij, 2003). Тем не менее количество желточных клеток и накопленных питательных веществ значительно различается и существует корреляцией с биологией мирацидиев (Galaktionov, Dobrovolskij, 2003).

Первым гельминтологом, увидевшим «ресничного эмбриона», покидающего яйцо дигеней, стал Мелис (Mehlis), опубликовавший свои наблюдения в 1831 году (по Reinhard, 1957). Впервые термин «мирацидий» для описания рассматриваемой фазы использовал Браун (Braun) только в 1892 (Fried, Graczyk, 1997). Принимая во внимание, что при поедании яиц хозяин не заражается, ван Нордманн (van Nordmann) первым в 1882 предположил, что мирацидии выполняют роль «охотника на хозяина» (по Reinhard, 1957). Сам же мирацидий представляет собой небольшую, прозрачную и продолговатую личинку, тело которой покрыто уплощенными ресничными эпидермальными клетками (Fried, Graczyk, 1997). Между последними находятся выросты тегумента материнской спороцисты (Galaktionov, Dobrovolskij, 2003). Многие мирацидии имеют конусообразный апикальный бугорок (или хоботок) на переднем конце тела, который не покрыт ресничными пластинками, способен сокращаться и содержит сенсорные рецепторы и отверстия желез (Fried, Graczyk, 1997). У мирацидиев, покидающих яйцо, также имеется два простых рабдомерных глазных пятна, расположенных непосредственно над мозговым ганглием со спинной стороны тела (Galaktionov, Dobrovolskij, 2003). Личинка обладает сложным поведением, направленным как на дисперсию во внешней среде, так и на поиск и заражение моллюска-хозяина (Атаев, 2017). Заднюю часть тела мирацидия занимает генеративный зачаток (Galaktionov, Dobrovolskij, 2003).

Материнская и дочерняя спороциста, реди

В роли первого промежуточного хозяина у большинства видов дигеней выступают брюхоногие моллюски (Galaktionov, Dobrovolskij, 2003). Проникновение мирацидия в хозяина-моллюска может происходить как активным, так и пассивным путем. Вне зависимости от характера проникновения мирацидия, его преобразование в молодую материнскую спороцисту протекает по типу регрессивного метаморфоза (Galaktionov, Dobrovolskij, 2003). Материнская

спороциста представляет собой взрослую особь первого партеногенетического поколения. Форма тела спороцисты варьирует и может быть сферической, яйцевидной, удлинённой и разветвлённой (Fried, Graczyk, 1997). Структура тегумента спороцисты варьирует в зависимости от возраста, развития и видовой принадлежности. Генеративный зачаток, содержащийся в мирацидии, также претерпевает изменения и их характер определяет особенности окончательной организации материнской спороцисты и её размножения (Galaktionov, Dobrovolskij, 2003).

Особь следующего партеногенетического поколения, рождаемые материнской спороцистой, могут быть представлены двумя различными морфологическими типами: редией и дочерней спороцистой, различающимися между собой как по морфологическим, так и по биологическим признакам (Galaktionov, Dobrovolskij, 2003). Первым, кто описал наличие «маточно-желтых червей» во вскрытых моллюсках, стал Боянус (Bojanus) в 1818 году (по Reinhard, 1957). Позже, в 1837, по предложению де Филиппи (de Filippi), обнаруженная фаза жизненного цикла получила название «редия» в честь Франческо Реди (по Reinhard, 1957). В настоящий момент известно, что редии характерны для более примитивных дигеней (в том числе, для представителей Fasciolidae, Echinostomatidae, Paramphistomidae, Psilostomidae, Philophthalmidae, Lepocreadiidae), но могут быть также обнаружены у нескольких специализированных групп, таких как Halipeginae, Apogocotylidae, Heterophyidae, Opisthorchiidae (Galaktionov, Dobrovolskij, 2003). Тело редии имеет удлинённую цилиндрическую форму с 3 коническими локомоторными выростами. Форма тела и степень развития различных структур обычно меняется в зависимости от возраста (Galaktionov, Dobrovolskij, 2003). Большая часть тела редии занята эмбриональной полостью, где происходит развитие новых особей. Сформированное потомство покидает тело материнской особи через родильную пору. В одном моллюске-хозяине обычно происходит смена нескольких поколений редий, количество которых варьирует между разными видами. После нескольких циклов самовоспроизведения редии переходят к формированию церкарий, личинок амфимиктического поколения (Galaktionov, Dobrovolskij, 2003).

Дочерние спороцисты характерны для более специализированных групп дигеней, в основном для представителей Strigeidida и Plagiorchiida (Galaktionov, Dobrovolskij, 2003). Они отличаются от редий более простой организацией: отсутствием локомоторных выростов и пищеварительной системы, родильная пора, если она есть, расположена терминально (Galaktionov, Dobrovolskij, 2003). Дочерние спороцисты Gymnophallidae, Schistosomatidae, Cyathocotylidae, Strigeidae, Diplostomatidae способны производить как новые поколения спороцист, так и церкарий (Galaktionov, Dobrovolskij, 2003). У всех остальных видов дочерние спороцисты представлены только одним поколением (Galaktionov, Dobrovolskij, 2003). Первым,

кто использовал термин «спороциста» для описания рассматриваемой фазы жизненного цикла, стал де Филиппи в 1854 году (Fried, Graczyk, 1997).

Вопрос о гомологии редий и дочерних спороцист впервые был поднят Лоосом (Looss) в 1892, а также в дальнейшем многократно обсуждался разными исследователями (Гинецинская, 1968). Сьюэлл (Sewell) в публикации 1922 года построил морфологический ряд, в котором были все переходные формы от типичной редии (*Echinostomatidae*, *Fasciolidae*) через мешковидную редию с развитым кишечником (*Notocotylidae*, *Lissorchiidae*) к мешковидному организму, лишённому кишечника (*Azygiidae*) и с трудом отличаемому от дочерней спороцисты (по Гинецинская, 1968). Позже было показано, что несмотря на морфобиологические отличия между редиями и дочерними спороцистами, начальные этапы развития у дочерних партенит совершенно одинаковы (Galaktionov, Dobrovolskij, 2003). В настоящее время наличие гомологии между спороцистами и редиями не вызывает сомнений (Гинецинская, 1968; Galaktionov, Dobrovolskij, 2003; Атаев, 2017).

Специализированным органом размножения дочерних партенит является герминальная масса (Galaktionov, Dobrovolskij, 2003). У спороцист и редий герминальная масса функционально заменяет гонады, объединяя в себе функции гонад (возобновление и «созревание» репродуктивных клеток) и выводковой камеры, в которой происходят ранние этапы эмбрионального развития потомства (Galaktionov, Dobrovolskij, 2003). Период и характер пролиферационной активности герминальной массы у разных групп дигеней могут существенно различаться и определяются общей тенденцией эволюции рассматриваемой партеногенетической фазы (Galaktionov, Dobrovolskij, 2003). По своей природе герминальные клетки редий и спороцист представляют собой ооциты первого порядка, которые приобрели способность к дроблению (Galaktionov, Dobrovolskij, 2003). Как следствие, размножение посредством этих клеток следует рассматривать, как настоящий апомиктический партеногенез (Galaktionov, Dobrovolskij, 2003).

Церкария

Первым, кто обнаружил церкарий в моллюсках, стал Сваммердам (Swammerdam) в 1737 (Fried, Graczyk, 1997). Термин «церкария» обязан своим появлением Мюллеру (Muller), который использовал его впервые в 1773 году как общее название для микроскопических животных, обладающих хвостом (по Reinhard, 1957). При вскрытии моллюсков Боянус в 1818 обнаружил не

только редий, но и обилие церкарий в них, которые будучи извлеченными из червя, активно двигали хвостами и кружили в воде (по Reinhard, 1957).

Церкарии представляют собой хвостатую личинку амфимиктического поколения, имеющую овальную или удлинённую форму тела, уплощённую в дорсовентральной плоскости (Galaktionov, Dobrovolskij, 2003). Церкарии объединяют в себе как признаки, характерные для представителей гермафродитного поколения, так и специфичные для личинок провизорные признаки, имеющие адаптивные функции (Galaktionov, Dobrovolskij, 2003). Разнообразие адаптивных приспособлений обуславливает разнообразие морфологических форм, наблюдаемых у церкарий (Galaktionov, Dobrovolskij, 2003). Покидая тело моллюска, церкарии попадают в воду, где после непродолжительного плавания заражают следующего хозяина либо путем прямого проникновения, либо через инцистирование на подходящем субстрате (Fried, Graczyk, 1997). Для передвижения в воде большинство церкарий используют свой хвост с хорошо развитой мускулатурой. В то же время, церкарии нескольких видов дигеней вовсе не покидают моллюска-хозяина (Fried, Graczyk, 1997). Тем не менее эмбриогенез и начальные этапы морфогенеза церкарий у представителей разных групп дигеней протекают схожим образом (Galaktionov, Dobrovolskij, 2003). Более того, дробление герминальной клетки и начальные этапы эмбрионального развития не отличаются от этих же процессов у мирацидиев и дочерних партенит (Galaktionov, Dobrovolskij, 2003).

Адолескария, метацеркария, шистосомула

Термины «метацеркария» и «адолескария» используют для обозначения покоящийся личиночной фазы гермафродитного поколения дигеней. Традиционно «метацеркарией» называют личинку, осевшую во втором промежуточном хозяине, тогда как личинку, сформировавшую цисту на различном субстрате в окружающей среде, принято называть «адолескарией» (Galaktionov, Dobrovolskij, 2003). Первое описание инцистирования церкарий сделал Нич (Nitzsch) в 1807 году, тогда как образование цист внутри моллюска первым обнаружил фон Зибольд (von Siebold) в 1837 году (Fried, Graczyk, 1997).

Несмотря на то, что метацеркария и адолескария соответствуют одной и той же фазе сложного жизненного цикла дигеней, функциональные и физиологические различия между ними огромны (Galaktionov, Dobrovolskij, 2003). Адолескария, лишённая внешних источников энергии, действительно представляет собой покоящуюся фазу, на которой не наблюдается никаких морфогенетических перестроек (Galaktionov, Dobrovolskij, 2003). Напротив, метацеркария в

большинстве случаев является настоящей паразитической фазой, установившей метаболические взаимодействия со своим хозяином (Galaktionov, Dobrovolskij, 2003). Как следствие, у дигеней с метацеркарией в сложном жизненном цикле на рассматриваемой фазе могут происходить существенные морфогенетические изменения (Galaktionov, Dobrovolskij, 2003). Наблюдаемое разнообразие метацеркарий у различных групп дигеней, с одной стороны, отражает уровень взаимоотношений личинок с окружающей средой или с организмом хозяина, с другой стороны, отражает тренды в эволюции морфологических признаков гермафродитного поколения (Galaktionov, Dobrovolskij, 2003). Так, к примеру, характер преобразования полового зачатка зависит от степени его развития у церкарий (Galaktionov, Dobrovolskij, 2003). Первые экспериментальные доказательства того, что не инцистированная церкария не может заразить животное и претерпеть в нем метаморфоз, в отличие от церкарии, сформировавшей цисту, привел Лавалетт Сен-Жорж (La Valette St. George) в 1855 году (Reinhard, 1957).

У представителей Schistosomatida церкария не формирует цисту, а проникает через кожу в кровеносный сосуд хозяина, и преобразуется в новую мигрирующую фазу, называемую «шистосомулой» (Galaktionov, Dobrovolskij, 2003). Указанный процесс, получивший название «трансформация», занимает около 3 часов и связан со значительными изменениями покровов личинки (Galaktionov, Dobrovolskij, 2003). По мере развития шистосомул происходит дифференциация полового зачатка, который у церкарий имеет вид клеточного скопления (Galaktionov, Dobrovolskij, 2003). На печеночной стадии миграции происходит попарное объединение шистосомул, причем в каждой паре у одной особи происходит развитие органов только мужской половой системы, а у другой – женской. Галактионов и Добровольский (Galaktionov, Dobrovolskij, 2003) считают, что формирование крупной группы кровяных паразитов – Schistosomatida – связано с абберацией в ходе онтогенеза гермафродитных особей. На основании анализа характеристик этой группы дигеней авторы предполагают, что появление такого жизненного цикла в ходе эволюции связано с выпадением окончательного хозяина, чью роль теперь исполняет второй промежуточный хозяин. Исходя из этого, шистосомула представляют собой метацеркарию, а марит шистосом следует рассматривать, как неотенических метацеркарий (Galaktionov, Dobrovolskij, 2003).

Марита

Марита представляет собой фазу сложного жизненного цикла дигеней, следующую за проникновением метацеркарии в окончательного хозяина (Galaktionov, Dobrovolskij, 2003).

Первым, кто продемонстрировал на примере *F. gigantea* развитие марит из метацеркарий, стал Лутц (Lutz) в 1892 году (Reinhard, 1957). Мариты имеют удлинённую форму тела, уплощённую в брюшной плоскости и выпуклую в спинной (Fried, Graczyk, 1997). У большинства дигеней онтогенез мариты разделен на два этапа: подготовительный (маритогония), во время которого происходит завершение развития definitivoных систем органов и/или физиологическое созревание, и функциональный – период репродукционной активности половозрелых особей (Galaktionov, Dobrovolskij, 2003). Продолжительность обоих этапов сильно различается у представителей разных групп дигеней (Galaktionov, Dobrovolskij, 2003). За исключением шистосом, мариты дигеней являются гермафродитными (Fried, Graczyk, 1997).

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). Сам исследователь определял этот феномен, как появление у животных потомков, которые не напоминают собственных родителей, но которые, в свою очередь, порождают потомство, схожее с исходными родительскими животными (Steenstrup, 1845). Тем не менее Стинструп полагал, что церкарии возникают в спороцистах и редиях в результате почкования, т. е. жизненный цикл дигеней может быть рассмотрен как пример метагенеза, чередования полового (гермафродитного) и бесполого (спороцисты и редии) поколений (Galaktionov, Dobrovolskij, 2003). Взгляды Стинструпа позднее поддержал Томас (Thomas) (Атаев, 2017), который параллельно с Лейкартом (Leukart), но независимо от него, впервые описал жизненный цикл дигеней на примере *F. hepatica* (Fried, Graczyk, 1997).

На рубеже XIX и XX столетий была сформулирована теория гетерогонии, согласно которой жизненный цикл дигеней представляет собой чередование полового и партеногенетического поколений (Атаев, 2017). Одними из первых сторонников указанной теории были Гроббен (Grobben), Лейкарт, Клаус (Claus) и Сеницын (Атаев, 2017). В дальнейшем теорию гетерогонии поддержали многие исследователи, в числе которых, Гинецинская (1965, 1968), Пирсон (Pearson, 1972), Добровольский (1983, 1998, 2003), Галактионов (1998, 2003, 2015). Основываясь на отсутствии экспериментальных доказательств мейотического деления в процессе созревания герминальных клеток (Dobrovolskij, Ataev, 2003), Галактионов и Добровольский (2003) делают вывод о том, что герминальные клетки дигеней являются ооцитами

первого порядка и, как следствие, размножение редий и спороцист происходит путем диплоидного (апомиктического) партеногенеза.

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. Стронниками указанной точки зрения являются, к примеру, Кларк (Clark, 1974), Шоп (Shoop, 1988) и Ванг с соавторами (Wang B. et al., 2013, 2018). На основании сходства молекулярных подписей герминальных клеток и необластов планарий, Ванг с соавторами предполагают, что древняя программа развития, основанная на стволовых клетках, поспособствовала эволюции сложного жизненного цикла паразитических плоских червей (Wang B. et al., 2013).

1.3 Описание схем реализации сложных жизненных циклов исследуемых видов дигеней

Эволюционная пластичность дигеней нашла отражение в разнообразии схем реализации их сложных жизненных циклов (Galaktionov, Dobrovolskij, 2003). В первую очередь это проявляется в изменении числа животных-хозяев, но может затрагивать и количество фаз жизненного цикла (Galaktionov, Dobrovolskij, 2003). Для типологизации жизненных циклов дигеней Догель (1947) предложил использовать наиболее общие признаки: количество хозяев, особенности процессов размножения паразитов и их накопления в организме хозяина. Два наиболее редко встречающихся варианта, характерных, главным образом, для представителей очень специализированных групп – одно- и четыреххозяинные циклы (Galaktionov, Dobrovolskij, 2003). Жизненные циклы с двумя (дисксенные) и тремя (триксенные) хозяевами распространены значительно шире (Galaktionov, Dobrovolskij, 2003). В отличие от триксенных жизненных циклов, которые довольно однородны по структуре, дисксенные циклы могут быть разделены на несколько подгрупп, различающихся, в основном, по тому, как развитие паразита распределено между промежуточным и окончательным хозяином (Galaktionov, Dobrovolskij, 2003). Галактионов и Добровольский (Galaktionov, Dobrovolskij, 2003) отмечают, что такие различия между дисксенными жизненными циклами убедительно свидетельствуют об их независимом возникновении в разных группах дигеней.

В настоящем исследовании мы преимущественно рассматривали фазы жизненных циклов *F. gigantea*, *F. hepatica*, *P. simillimum*, *S. pseudoglobulus*, *S. mansoni*, *T. regenti* и *T. szidati*. С целью

сформировать более полное представление о перечисленных видах, далее будут приведены описания схем реализации их жизненных циклов.

Жизненный цикл *F. gigantea* является диксенным и включает моллюска и млекопитающее в качестве промежуточного и окончательного хозяев, соответственно (Zhang X.-X. et al., 2019). Яйца паразита попадают в окружающую среду вместе с фекалиями млекопитающего-хозяина, после чего начинается эмбриональное развитие мирацидия. Мирацидий покидает стенки яйца в ответ на воздействия температуры и света и активно плавает в поисках промежуточного хозяина. В природе *F. gigantea* заражает широкий круг брюхоногих моллюсков, в том числе евразийские виды *Radix auricularia* и *Austropeplea viridis*, а также африканский вид *Radix natalensis* (Zhang X.-X. et al., 2019). Внутри промежуточного хозяина обитает материнская спороциста и несколько поколений редий. В ходе исследований Динник и Динник (Dinnik, Dinnik, 1956, 1964) показали, что в моллюске происходит последовательная смена пяти поколений редий *F. gigantea*. Церкарии, покидающие тело моллюска, либо оседают на растительности, либо продолжают активно плавать в воде (Zhang X.-X. et al., 2019). После этого церкарии инцистируются и превращаются в адолескарию. Заражение окончательного хозяина, обычно крупного рогатого скота и буйволов, происходит при проглатывании млекопитающим адолескарий *F. gigantea* вместе с зараженной растительностью или водой. После разрушения стенок цисты паразит переходит в фазу ювенильной особи. Ювенили быстро мигрируют через стенку кишечника в паренхиму печени, где активно питаются кровью и тканью печени, что способствует их быстрому росту и развитию. Далее паразиты перемещаются в желчные протоки, где достигают половой зрелости и переходят к размножению (Zhang X.-X. et al., 2019).

В желчных протоках млекопитающих-хозяев, таких как овца, крупный рогатый скот и люди, обитают также мариты *F. hepatica* (Moazeni, Ahmadi, 2016). После размножения половозрелые особи переходят к формированию яиц. Яйца паразита попадают в двенадцатиперстную кишку с желчью и впоследствии покидают организм хозяина вместе с фекалиями. При благоприятных условиях температуры, влажности и содержания кислорода развитие мирацидия в яйце занимает 2–3 недели (Moazeni, Ahmadi, 2016). Полностью сформированный мирацидий покидает стенки яйца и активно плавает в поисках брюхоногого моллюска рода *Lymnaea* (обычно *Lymnaea truncatula*), который выступает в качестве промежуточного хозяина (Moazeni, Ahmadi, 2016). При проникновении в организм хозяина мирацидий теряет свои ресничные пластинки и преобразуется в материнскую спороцисту. Внутри материнской спороцисты содержится плотно упакованная масса зародышевых клеток, каждая из которых дает начало редиям. Редии растут до тех пор, пока не прорывают стенку материнской спороцисты и, таким образом, не попадают в печень моллюска-хозяина.

Размножаясь партеногенетически, редии формируют церкарий, личинок амфимиктического поколения, обладающих длинным хвостом для плавания. Полностью сформированные церкарии покидают организм промежуточного хозяина через 4–7 недель после заражения (Moazeni, Ahmadi, 2016). Личинки плавают в воде непродолжительное время (от нескольких минут до двух часов) и оседают на различные предметы, в основном на листья водных растений (Moazeni, Ahmadi, 2016). Далее церкарии отбрасывают свои хвосты и формируют цисты, превращаясь в адолескарий. Заражение окончательного хозяина происходит при поглощении им цисты. После попадания в организм хозяина ювенильные особи паразита выходят из стенок цисты. Ювенили *F. hepatica* проникают через стенку кишечника и оказываются в брюшной полости примерно через два часа (Moazeni, Ahmadi, 2016). Далее они достигают печени в течение 4–6 дней (Moazeni, Ahmadi, 2016). Молодые особи паразита мигрируют в паренхиме печени в течение 5–6 недель, вызывая обширные кровоизлияния и фиброз (Moazeni, Ahmadi, 2016). Примерно через 7 недель после заражения паразит попадает в желчные протоки, где достигает половой зрелости и начинает откладку яиц (Moazeni, Ahmadi, 2016).

Сложный жизненный цикл *P. simillimum* также является примером цикла фасциолидного типа с одним промежуточным хозяином (Белякова, 1978). В качестве окончательного хозяина у рассматриваемого вида выступают различные водоплавающие птицы, к примеру лебедь-шипун *Cygnus olor*, серая утка *Anas strepera*, домашние гусь и утка (Белякова, 1978). Мариты располагаются в тонком и верхнем отделе толстого кишечника (Белякова, 1978). В организме окончательного хозяина паразит достигает половой зрелости через 9 дней после заражения (Белякова, 1978). Из яиц, попавших в воду, через 19 дней вылупляется мирацидий (Белякова, 1978). Мирацидий активно плавает в поисках промежуточного хозяина – брюхоногих моллюсков *Bythinia* sp. Редии имеют вытянутую мешковидную форму тела и располагаются в пищеварительной железе моллюска. В каждой редии *P. simillimum* одновременно развивается до 20 церкарий и 6–8 зародышевых шаров (Белякова, 1978). После выхода из организма хозяина церкарии активно плавают непродолжительное время, а затем инцистируются на водной растительности близко к поверхности воды, иногда на поверхностной пленке, стараясь прикрепиться к уже инцистированным особям (Белякова, 1978). Адолескарии становятся инвазионными между 3-ми и 7-ми сутками после инцистирования, а не сразу, как это происходит у фасциолид (Белякова, 1978). Заражение водоплавающих птиц происходит через проглатывание адолескарий с кормом или водой (Белякова, 1978).

Близкий к *P. simillimum* вид, *S. pseudoglobulus*, описан Маклаfliном, Скоттом и Хаффманом в 1993 году, как крайне схожий с космополитичным видом дигеней, *S. globulus*, поражающим пищеварительный тракт водоплавающих птиц (McLaughlin et al., 1993). В качестве окончательных хозяев *S. pseudoglobulus* также, как и *S. globulus*, использует различных

водоплавающих птиц, а мариты локализуются в кишечнике. В отличие от вышеуказанных видов фасциол и *P. simillimum*, жизненный цикл *S. pseudoglobulus* является триксенным: как первым, так и вторым промежуточным хозяином рассматриваемого вида дигеней является моллюск *Bithynia tentaculata* (McLaughlin, et al., 1993; Sandland, Peirce, 2021). В первом промежуточном хозяине происходит формирование нескольких поколений редий, после чего партениты переходят к производству церкарий. Полностью сформированные церкарии покидают тело моллюска на 9–10 неделе после заражения (Sandland, Peirce, 2021). Метацеркарии *S. pseudoglobulus* обнаружены не только в живых *B. tentaculata*, но и в пустых раковинах моллюсков этого вида (Lepitzki, Bunn, 1994). Водоплавающие птицы заражаются, поедая жизнеспособных метацеркарий (Lepitzki, Bunn, 1994).

Диксенный сложный жизненный цикл кровяного сосальщика *S. mansoni* включает пресноводного моллюска *Biomphalaria* в качестве промежуточного хозяина, а приматов и различных видов грызунов – в роли окончательных хозяев (Picard et al., 2016). В отличие от подавляющего большинства видов дигеней, *S. mansoni* является раздельнополым видом (Galaktionov, Dobrovolskij, 2003; Picard et al., 2016). Различающиеся по многим признакам самцы и самки образуют пары и вместе мигрируют в мезентериальные вены, где интенсивно размножаются (Picard et al., 2016). Формирование пар имеет решающее значение для завершения созревания гонад обоих полов. В то же время, наблюдаемые различия по морфологическим признакам между особями разных полов способствуют обитанию паразита в кровеносных сосудах: мускулистый самец зажимает самку в своем гинекофорном канале, чтобы сопротивляться высокому кровяному давлению, а нитевидная самка способна проникать в мельчайшие вены кишечника для откладки яиц по одному за раз (Picard et al., 2016). Яйца паразита попадают в пресную воду с фекалиями окончательного хозяина. В пресной воде происходит вылупление свободноплавающего мирацидия из яйца, после чего личинка направляется на поиск промежуточного хозяина. В моллюске мирацидий превращается в материнскую спороцисту, которая дает начало дочерним спороцистам. После нескольких актов размножения спороцисты переходят к формированию церкарий. Покидая организм моллюска, церкарии активно плавают в поисках окончательного хозяина, в которого проникают через эпидермис. В ходе миграции через кожу церкария теряет свой хвост и претерпевает значительную морфологическую и физиологическую трансформацию. В течение двух часов свободноживущая личинка становится облигатной эндопаразитической шистосомулой, которая покидает кожу, попадает в кровоток и мигрирует в печень через легкие (Picard et al., 2016). После 2–5 недель пребывания в окончательном хозяине и достижения портальной системы печени, шистосомулы развиваются в мужских и женских особей (Picard et al., 2016).

Представители *Trichobilharzia*, *T. regenti* и *T. szidati*, характеризуются диксенным жизненным циклом и используют водоплавающих птиц в качестве окончательных хозяев (Leontovych et al., 2019). Подобно представителям *Schistosoma*, оба указанных вида *Trichobilharzia* являются раздельнополыми. Попадание мирацидиев в пресную воду происходит либо в яйце с фекалиями (*T. szidati*), либо напрямую со слизистой оболочкой носа (*T. regenti*) водоплавающей птицы (Leontovych et al., 2019). В пресном водоеме мирацидии активно плавают и заражают брюхоногих моллюсков. Оба рассматриваемых вида паразита различаются в том, какие виды моллюсков выступают в роли промежуточных хозяев: *T. regenti* заражает представителей *Radix* (Leontovych et al., 2016), а *T. szidati* – моллюсков вида *Lymnaea stagnalis* (Leontovych et al., 2019). Дочерние партеногенетические поколения у обоих видов представлены дочерними спороцистами. После партеногенетического размножения спороцисты переходят к формированию церкарий. Покидая тело промежуточного хозяина, церкарии активно плавают и ищут водоплавающих птиц в течение 1–1.5 дней (Leontovych et al., 2019). Как только личинки находят подходящего окончательного хозяина, церкарии проникают в его кожу и претерпевают метаморфоз. Во время проникновения церкарии теряют свой хвост, переключаются с аэробного на факультативно-анаэробный метаболизм и трансформируются в шистосомул (Leontovych et al., 2019). Оказавшись внутри птицы, шистосомулы *T. szidati* и *T. regenti* используют разные пути миграции (Leontovych et al., 2019). Шистосомулы *T. szidati* входят в кровеносную систему, достигают легких в течение двух недель после заражения и развиваются в раздельнополых взрослых особей в сосудах кишечной стенки, где и происходит размножение (Leontovych et al., 2019). Напротив, шистосомулы *T. regenti* мигрируют через периферические нервы, спинной и головной мозг к слизистой оболочке носа, где мариты формируют пары и размножаются (Leontovych et al., 2019). В отличие от *T. szidati*, шистосомулы *T. regenti* питаются нервными тканями, а не кровью (Leontovych et al., 2019).

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

Стремительное развитие технологий высокопроизводительного секвенирования сделало возможным детальное исследование молекулярных основ биологии широкого круга живых организмов, как свободноживущих (Fuchs et al., 2014; Wollenberg Valero et al., 2017), так и паразитических (Sahoo et al., 2013; Choi et al., 2017). Первые версии сборок ядерных геномов дигеней были представлены в 2009 году по результатам исследования *S. mansoni* (Berriman et al.,

2009) и *S. japonicum* (Zhou et al., 2009). Собранный для *S. mansoni* первая версия генома содержала 11809 генов и предоставила больше информации о дефиците липидного метаболизма червя, а также о наборах киназ и протеаз, используемых паразитом (Berriman et al., 2009). Улучшенная версия сборки генома *S. mansoni* представлена в публикации Протасио с соавторами (Protasio et al., 2012). В ней уточнено количество генов и приведены данные об изменении экспрессии большинства генов в ходе реализации сложного жизненного цикла *S. mansoni*. Результаты анализа 13469 генов *S. japonicum* позволили показать, что рассматриваемый вид дигеней имеет большой пул разнообразных протеаз и также неспособен *de novo* синтезировать жирные кислоты, стеролы и пурины (Zhou et al., 2009). В 2012 году была опубликована первая версия сборки генома *S. haematobium*, содержащая 13073 белок-кодирующих генов (Young et al., 2012). Сравнительный полногеномный анализ между шистосомами обнаружил значительно более высокую синтению (сохранение совместной локализации генов в хромосомах разных видов) между *S. haematobium* и *S. mansoni* (89,4%), чем между *S. haematobium* и *S. japonicum* (51,7%) или *S. mansoni* и *S. japonicum* (67%) (Young et al., 2012). Первые варианты сборок геномов представлены также для *Clonorchis sinensis* (Wang X. et al., 2011), *O. viverrini* (Young et al., 2014), *O. felinus* (Ershov et al., 2019). Результаты анализа синтении и реконструкции филогенетических взаимоотношений между *C. sinensis*, *O. viverrini* и *O. felinus* доказали, что *O. felinus* и *C. sinensis* являются близкородственными видами и, как следствие, разделение *C. sinensis* и *Opisthorchis* некорректно (Ershov et al., 2019). В 2015 (Cwiklinski et al., 2015) и 2017 (McNulty et al., 2017) годах были представлены первые сборки геномов двух разных изолятов *F. hepatica*. В отличие от шистосом, в геноме *F. hepatica*, как и у других видов печеночных сосальщиков, присутствуют гены ферментативного пути удлинения жирных кислот за счет обращения бета-окисления и катаболизма жирных кислот (McNulty et al., 2017). Позже были опубликованы результаты секвенирования и сравнения 81 генома гельминтов, включая и 11 видов дигеней, направленные на определение ключевых генов, имеющих отношение к паразитизму, модуляции иммунного ответа, миграции и питания паразитов (International Helminth Genomes Consortium, 2019). В публикации впервые были представлены сборки геномов *S. curassoni*, *S. margrebowiei*, *S. mattheei*, *S. rodhaini*, *T. regenti* и *E. caproni* (International Helminth Genomes Consortium, 2019). Накопление большого числа геномных данных дигеней привело к развитию таких ресурсов и баз данных, как SchistoDB (Zerlotini et al., 2012) и WormBase Parasite (Howe et al., 2017).

Первые результаты исследования активности генома в ходе реализации сложного жизненного цикла дигеней были опубликованы в 1995 году Франко с соавторами (Franco et al., 1995). В указанной работе представлены данные только для марит *S. mansoni*, однако уже в 2003 году Вержовски-Алмейда с соавторами опубликовали данные по шести разным фазам жизненного цикла шистосомы (Verjovski-Almeida et al., 2003). В публикации описаны наборы

генов, связанных с организацией тканей, развитием, проведением сигналов, половым диморфизмом, взаимодействием с хозяином и бегством от иммунного ответа (Verjovski-Almeida et al., 2003). Впервые результаты сравнительного анализа экспрессии генов между яйцом и маридами *S. japonicum* были опубликованы в 2003 году. Для марида показана активность генов, имеющих отношение к питанию кровью, проведению сигналов, развитию и избавлению от избытков воды и окислителей, а также описаны половые различия (Hu et al., 2003). В то же время для яиц *S. japonicum* обнаружены высокие уровни экспрессии различных антигенов, в том числе и от мирацидия, что указало на транскрипционную активность мирацидия в яйце (Hu et al., 2003). В 2003 году были также впервые опубликованы результаты анализа транскриптома марида *C. sinensis*, который включал транскрипты генов, ассоциированных с энергетическим обменом, сортировкой/метаболизмом белков и мембранным транспортом (Lee et al., 2003). Активность генов, вовлеченных в существенные для паразитизма физиологические процессы, такие как анаэробное дыхание, размножение, детоксикация и питание, обнаружена при исследовании транскриптома марида *O. viverrini* (Laha et al., 2007).

Все перечисленные исследования транскриптомов выполнены с помощью метода анализа экспрессируемых маркировочных последовательностей (Expressed sequence tags, ESTs). Указанный метод предоставляет возможность изучить набор мРНК, присутствующих в исследуемом образце, путем создания и секвенирования библиотеки кДНК, клонированных в соответствующие вектора (Parkinson, Blaxter, 2009). Одной из характерных черт такого метода является секвенирование не всей библиотеки, а только её части, отобранной, как правило, случайным образом (Parkinson, Blaxter, 2009). В настоящее время основным методом определения того, какие гены и на каком уровне экспрессируются в образце, является метод секвенирования РНК (RNA-seq). Преимуществами секвенирования РНК является высокая чувствительность и точность в количественной оценке уровней экспрессии генов (Wang Z. et al., 2009), что играет важную роль при исследовании транскриптомов и сравнительном анализе между ними.

Яйцо и мирацидий

Исследование транскриптомов яиц дигеней представляет большой интерес ввиду наличия в них сигнала от мирацидиев на разных этапах развития. В настоящее время доступны результаты анализов транскриптомов яиц дигеней *F. hepatica* (McNulty et al., 2017) и *F. gigantica* (Zhang X.-X. et al., 2019). Для яйца *F. hepatica* выявлена активность биологических процессов, связанных с

основными клеточными функциями, такими как трансляция, созревание РНК и везикулярный транспорт (McNulty et al., 2017). В сравнении с метацеркарий и маритой, для яйца *F. hepatica* удалось обнаружить относительно небольшое число генов с повышенной экспрессией (McNulty et al., 2017). Полученный результат авторы связали тем, что при определении дифференциальной экспрессии сравнение проводилось с маритой, которая также содержит яйца (McNulty et al., 2017). Наиболее значимое повышение экспрессии в яйце продемонстрировал ген, кодирующий глюкозо-6-фосфатдегидрогеназу (McNulty et al., 2017).

По результатам анализа транскриптома яйца *F. gigantica* отмечена обильная транскрипция генов, принимающих участие в кальций-зависимом связывании белков, транспорте и связывании гликолипидов, а также активности трансмембранных переносчиков глюконата (Zhang X.-X. et al., 2019). Среди генов с наибольшей экспрессией в яйце авторы отметили те, что кодируют белки теплового шока, ингибиторы широкого спектра протеаз, а также белки желтка. По сравнению с другими фазами жизненного цикла *F. gigantica* в яйце повышена активность нескольких биологических процессов, в том числе связанных с клеточным ростом и регуляцией оогенеза. Сравнительный анализ экспрессии ортологов *F. hepatica* и *F. gigantica* продемонстрировал, что в яйцах обоих фасциол повышена активность генов желточного ферритина, аспарагиновой протеазы, белков теплового шока, а также антиоксидантов, обычно обнаруживаемых в покоящихся/инцистированных метацеркариях (Zhang X.-X. et al., 2019).

Основными биологическими процессами, ассоциированными с мирацидием *F. gigantica*, оказались те, что могут быть связаны с поиском моллюска и проникновением в него (Zhang X.-X. et al., 2019). Так, к примеру, повышенную активность металлопептидаз и формирование внеклеточных везикул, авторы связывают с разрушением покровов моллюска (Zhang X.-X. et al., 2019). В сравнении с другими фазами жизненного цикла мирацидий *F. gigantica* характеризовался более низкими относительными уровнями транскрипции генов, ассоциированных с метаболизмом (Zhang X.-X. et al., 2019).

Редия и церкария

Результаты сравнительного анализа редии с другими фазами сложного жизненного цикла *F. gigantica* продемонстрировали, что у партениты повышен уровень транскрипции, трансляции и репликации / репарации (Zhang X.-X. et al., 2019). В качестве примера авторы указывают, что транскрипция генов и трансляция у редии увечены в 2–3,5 раз по сравнению с церкариями и

метацеркариями (Zhang X.-X. et al., 2019). Полученные результаты авторы связывают с активным формированием дочерних поколений редий внутри моллюска-хозяина.

В сравнении с редией церкария *F. gigantica* характеризовалась двукратным увеличением экспрессии генов, принимающих участие в транспорте и катаболизме (Zhang X.-X. et al., 2019). Повышенную экспрессию как на фазе редии, так и на фазе церкарии сложного жизненного цикла *F. gigantica* имели гены, белки которых способны связывать другие белки, АТФ, ДНК, ионы металлов, в том числе и ионы цинка. По мнению авторов, такие результаты указывают на высокую метаболическую активность рассмотренных фаз жизненного цикла.

На основании высоких уровней экспрессии генов, участвующих в гликолизе и в цикле трикарбоновых кислот, на фазе церкарии *S. mansoni*, было подтверждено, что свободноживущие личинки используют внутренние запасы гликогена во время активного движения при поиске хозяина (Protasio et al., 2012). Большой вклад митохондрий в транскриптом церкарии *S. mansoni* свидетельствовал о высоких энергетических потребностях активно плавающей личинки (Protasio et al., 2012).

Несмотря на отсутствие фенотипических различий между полами *S. mansoni* на фазе церкарии, между мужскими и женскими особями церкарий обнаружены молекулярные различия, в том числе и в экспрессии генов, имеющих отношение к гормональной и нервной системе (Picard et al., 2016). Более того показано, что гены с преобладающей экспрессией в женских особях вовлечены в реагирование личинки на световые и механические стимулы, в то время как гены с преобладающей экспрессией в мужских особях связаны с реакцией на химические раздражители (Picard et al., 2016).

Детальный анализ транскриптома церкарии *T. regenti* указал на вероятное участие углеводного обмена, трансляции и метаболизма аминокислот, а также кальциевой сигнальной системы в проникновении личинки в окончательного хозяина (Leontovyc̆ et al., 2016). Для церкарий *T. regenti* и *T. szidati* показано, что при заражении хозяев у личинок повышена активность процессов, связанных с метаболизмом углеводов (включая гликолиз и цикл Кребса), окислительным фосфорилированием и трансляцией белков, включая необходимые для формирования рибосом, производства экзосом и/или биогенеза липидов (Leontovyc̆ et al., 2019).

Метацеркария

Между метацеркариями, яйцом и маритой *F. hepatica* не выявлено значимых отличий в экспрессии генов, принимающих участие в процессах трансляции, созревания РНК и

везикулярного транспорта (McNulty et al., 2017). Наиболее заметное усиление экспрессии на фазе метацеркарии *F. hepatica* продемонстрировали гены, кодирующие цистеиновые протеазы.

Результаты анализа набора генов, общих для церкарии и метацеркарии *F. gigantica*, показали, что на фазе метацеркарии происходит понижение экспрессии генов, связанных с синтезом нуклеотидов и метаболизмом (Zhang X.-X. et al., 2019). По мнению авторов, такие изменения указывают на то, что метацеркария находится в состоянии не полного покоя, а низкой метаболической активности. Активность биологических процессов, связанных с регуляцией транскрипции и трансляции генов, фосфорилирования белков и передачи сигналов, способствует продолжительному нахождению метацеркарии в окружающей среде (Zhang X.-X. et al., 2019). Основываясь на результатах анализа, авторы предполагают, что инцистированный паразит поддерживает свою низкую метаболическую активность путем регуляции кислотности и снижением эндопептидазной активности, чтобы избежать аутолитических процессов. Сравнительный анализ ортологов *F. hepatica* и *F. gigantica* показал, что на фазе метацеркарии обоих видов повышена экспрессия генов, кодирующих катепсины L и B, аспарагиновую протеазу, окислительно-восстановительные антиоксиданты, супероксиддисмутазы, а также белки, связывающие жирные кислоты (Zhang X.-X. et al., 2019).

Исследование транскриптома метацеркарии *O. felineus* позволило выяснить, что на рассматриваемой фазе жизненного цикла преимущественно идет транскрипция генов домашнего хозяйства, кодирующих рибосомальные белки, белки теплового шока и убиквитин (Pomaznou et al., 2016). Авторы считают, что такая транскрипционная активность отражает «дремлющее» состояние паразита, ожидающего встречи с окончательным хозяином. Для метацеркарии *O. felineus* также отмечен высокий уровень экспрессии цистеиновых протеаз, а также наличие большого числа сигнальных молекул, которые могут быть связаны с подготовкой к выходу из защитных оболочек цисты (Pomaznou et al., 2016).

Шистосомула

Шистосомула представляет собой облигатную эндопаразитическую фазу сложных жизненных циклов кровяных сосальщиков, мигрирующую в организме хозяина. В настоящее время доступны результаты исследования транскриптомов шистосомул *S. mansoni* (Protasio et al., 2012; Picard et al., 2016), *T. regenti* (Leontovyč et al., 2016) и *T. szidati* (Leontovyč et al., 2019).

В работе Протасио с соавторами (Protasio et al., 2012) представлены результаты сравнения транскриптомов шистосомул через 3 и 24 часа после заражения окончательного хозяина.

Исследователи выяснили, что в транскриптоме шистосомулы через 3 часа после заражения преобладают транскрипты генов, вовлеченных в регуляцию транскрипции, проведение сигналов через рецепторы, сопряженные с G-белком, сигнальный путь Wnt, клеточную адгезию и транспорт калия / натрия. Большинство процессов, активных через 3 часа после заражения, в том числе сигнальный путь через G-белок, сохраняли активность до 24 часа. Переход паразита на анаэробный метаболизм после проникновения в хозяина проявился в высокой экспрессии L-лактатдегидрогеназы и в значительном снижении транскрипции генов, участвующих в цикле трикарбоновых кислот. Повышение экспрессии выявлено и для генов белков теплового шока и антиоксидантов, что, по мнению авторов, может являться ответом на увеличение температуры окружающей среды при переходе из воды в тело хозяина и защитой от активных форм кислорода (Protasio et al., 2012). Сравнительный анализ транскриптомов мужских и женских особей шистосомул показал, что в отличие от самок развитие самцов характеризуется повышением экспрессии генов, вовлеченных в сборку нуклеосом и структурное поддержание хроматина (Picard et al., 2016).

В отличие от большинства шистосом, изученных на сегодняшний день, шистосомулы *T. regenti* не попадают в кровоток хозяина, а мигрируют в нервах, питаясь нервной тканью (Leontovuč et al., 2016). Исследование транскриптома шистосомулы *T. regenti* на седьмой день после заражения птицы показало, что при миграции паразита идет повышение экспрессии генов, кодирующих молекулы клеточной адгезии (Leontovuč et al., 2016). По мнению авторов, высокая экспрессия таких генов может указывать на быстрый рост и развитие различных структур органов внутри окончательного хозяина. Результаты анализа указывали также на то, что шистосомула *T. regenti* переходит на микроаэробный метаболизм через семь дней после заражения птицы, что совпадает с наблюдениями, сделанными для шистосом человека (Leontovuč et al., 2016). Миграция паразита сопровождается усилением экспрессии различных протеолитических ферментов, в том числе и цистеиновых протеаз (Leontovuč et al., 2016).

Шистосомулы *T. szidati*, в отличие от *T. regenti*, мигрируют к месту своего окончательного расположения через кровеносную систему своего хозяина (Leontovuč et al., 2019). На фазе шистосомулы как *T. szidati*, так и *T. regenti* основные активные биологические процессы связаны с проведением сигналов, клеточным ростом и смертью, клеточной подвижностью, транспортом и катаболизмом, мембранным транспортом, репликацией и репарацией ДНК. Различия между видами наблюдалось в специфичной для *T. regenti* активности молекулярных процессов, осуществляющих передачу сигналов и связанных с Ras, MAPK, Rap1, Wnt и ErbB. Авторы предположили, что такие различия могут быть связаны со стремительным ростом и развитием различных органов у шистосомулы *T. regenti* внутри окончательного хозяина и/или взаимодействием паразита и хозяина. Межвидовые различия также обнаружены и в наборах

экскретируемых белков, что отражает адаптацию висцеральных и нейротропных сосальщиков к различным способам получения питательных веществ (Leontovych et al., 2019).

Ювениль и марита

В ходе развития ювенили во взрослую особь *F. gigantica* потребляет кровь хозяина и ткани печени для получения энергии, о чем свидетельствует наличие транскриптов различных протеаз в транскриптоме ювенили (Zhang X.-X. et al., 2017). Активность биологических процессов, связанных с эндоцитозом, рибосомами, сплайсосомами, аденилатциклазным и окситоциновым путями проведения сигнала, обнаруженная в исследовании, также, по мнению авторов, указывает на участие метаболических процессов в развитии ювенили *F. gigantica* (Zhang X.-X. et al., 2017).

Сравнительный транскриптомный анализ 42- и 70-дневных ювенилей продемонстрировал схожий профиль экспрессии многих генов, что указало на небольшую разницу в транскрипции генов у паразитов, медленно перемещающихся в печени хозяина (Zhang X.-X. et al., 2019). Высокие уровни экспрессии генов, связанных с лизосомами и фагосомами, транспортными и катаболическими процессами, а также регуляцией актинового цитоскелета, выявленные у ювенилей через 42 и 70 дней после заражения хозяина, по мнению авторов, свидетельствовали об интенсивном питании паразита в ходе миграции через паренхиму печени. С миграцией авторы связали и обнаруженную как у ювенилей, так и у марит *F. gigantica* повышенную экспрессию генов акваглицеропоринов, способствующих транспорту воды и других растворов (Zhang X.-X. et al., 2019).

По количеству публикаций с результатами исследования транскриптомов наиболее изученной фазой сложного жизненного цикла дигеней является марита. В настоящее время проведен анализ транскриптомов марит *C. sinensis* (Young et al., 2010), *E. caproni* (Garg et al., 2013), *E. pancreaticum* (Liu G.-H. et al., 2016), *F. gigantica* (Young et al., 2011, Zhang X.-X. et al., 2017, 2019), *F. hepatica* (McNulty et al., 2017), *F. magna* (Cantacessi et al., 2012), *Metorchis orientalis* (Gao et al., 2018), *O. felineus* (Pomaznoy et al., 2016), *O. viverrini* (Young et al., 2010), *P. cervi* (Choudhary et al., 2015), *P. skrjabini* (Li et al., 2016), *P. westermani* (Li et al., 2016) и *S. mansoni* (Protasio et al., 2012, Picard et al., 2016).

Детальный анализ транскриптомов марит *C. sinensis* и *O. viverrini* показал наличие в них большого числа транскриптов генов, связанных с различными клеточными процессами (трансляция, фосфорилирование белков и т. д.), метаболическими процессами (протеолиз, метаболизм углеводов, окислительно-восстановительные реакции и т. д.), а также

биологическими регуляторными процессами, в числе которых регуляция транскрипции и проведения сигналов (Young et al., 2010). Наиболее активными на фазе мариты *O. felineus* являлись процессы, связанные с локализацией и транспортом липидов, негативной регуляцией протеолиза и эндопептидазной активностью (Pomaznoy et al., 2016). Исследование транскриптома мариты *P. cervi* также продемонстрировало обилие в нем транскриптов белков, связывающих жирные кислоты, что, вероятно, отражало способность паразита эффективно поглощать жирные кислоты из рубца своего хозяина (Choudhary et al., 2015). На основании обнаружения повышенной экспрессии генов, принимающих участие в удлинении жирных кислот, авторы предположили, что у марит *P. cervi*, в отличие от *Schistosoma*, указанный процесс активен (Choudhary et al., 2015).

В сравнении с ювенильными фазами марита *F. gigantica* характеризовалась более низким относительным уровнем метаболизма (Zhang X.-X. et al., 2019). Тем не менее для мариты *F. gigantica* отмечено усиление работы биологических процессов, ассоциированных с транспортом, катаболизмом, а также иммунной системой. По мнению авторов, повышенная активность последнего связана со взаимодействием паразита с иммунной системой его хозяина и манипуляцией ею. С иммунной системой связан и большой процент транскриптов мариты *M. orientalis*, что также может указывать на способность паразита модулировать иммунный ответ своих хозяев (Gao et al., 2018).

Сравнение транскриптомов марит *F. gigantica* и *F. hepatica*, доказало сходство экспрессии генов, участвующих во взаимодействии паразита с хозяином, в числе которых гены, кодирующие антиоксиданты, белки теплового шока и цистеиновые протеазы, между половозрелыми особями амфимиктического поколения филогенетически близких видов (Young et al., 2011). Помимо этого, для марит двух видов фасциол показана активность процессов, связанных с ингибиторами пептидаз и проведением сигналов через рецепторы, сопряженные с G-белком (Zhang X.-X. et al., 2019). Схожие уровни экспрессии многих ортологов обнаружены и между маритами *P. westermani* и *P. skrjabini* (Li et al., 2016). Тем не менее часть ортологов имели дифференциальную экспрессию: ортологи, обладающие повышенной экспрессией у *P. westermani*, связаны с активностью эндопептидаз, тогда как у *P. skrjabini* усиливали экспрессию ортологи, имеющие отношение к движению по микротрубочкам. Исследователи высказали предположение, что обнаруженные различия как в наборах генов двух видов, так и экспрессии ортологов могут объяснить различия в биологии филогенетически близких видов (Li et al., 2016).

Большой вклад митохондрий в транскриптом мариты *S. mansoni* указал на высокое потребление энергии половозрелой особью (Protasio et al., 2012). Результаты анализа транскриптома мариты *S. mansoni* показали повышение экспрессии генов антиоксидантов,

активность которых позволяет паразиту защищаться от потенциально опасных активных форм кислорода при нахождении в организме хозяина. Помимо этого, многие гены демонстрировали межполовые различия в экспрессии, как, например, гены, связанные с восприятием внешней среды, размножением и структурой хроматина (Picard et al., 2016).

В завершении обзора литературных данных, посвященных применению методов биоинформатического анализа в исследовании транскриптомов дигеней, следует отметить несколько важных моментов. Во-первых, большинство исследований направлено на поиск потенциальных мишеней для лекарственных препаратов ввиду медицинской и ветеринарной значимости дигеней. По результатам транскриптомного анализа марит *C. sinensis* (Young et al., 2010), *F. gigantica* (Zhang X.-X. et al., 2017, 2019), *F. hepatica* (McNulty et al., 2017), *F. magna* (Cantacessi et al., 2012.), *M. orientalis* (Gao et al., 2018), *O. felineus* (Pomaznoy et al., 2016), *O. viverrini* (Young et al., 2010), *P. cervi* (Choudhary et al., 2015), *P. westermani* и *P. skrjabini* (Li et al., 2016), в качестве наиболее перспективных кандидатов на специфичную мишень рассматривают различные протеазы. Во-вторых, сравнительный транскриптомный анализ различных фаз в пределах одного сложного жизненного цикла и/или между близкими видами проведен для широкого круга дигеней: *C. sinensis* (Young et al., 2010), *F. hepatica* (McNulty et al., 2017, Zhang X.-X. et al., 2019), *F. gigantica* (Zhang X.-X. et al., 2017, 2019), *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), *T. szidati* (Leontovyč et al., 2019). Полученные данные могут служить основой не только для определения сходств и различий между близкими видами дигеней, но и для выявления консервативных компонентов молекулярных основ реализации сложных жизненных циклов филогенетически отдаленных видов. В-третьих, критерии для определения генов с фаза-специфичным повышением уровней экспрессии различаются между публикациями. Так, в исследовании *F. hepatica*, проведенном Макналти с соавторами, при классификации учитывались результаты попарного анализа дифференциальной экспрессии (McNulty et al., 2017). Если ген значительно повышал свою экспрессию на одной фазе жизненного цикла относительно уровней его экспрессии на других фазах, но не демонстрировал статистически значимых различий в экспрессии в остальных парах сравниваемых фаз, он характеризовался фаза-специфичным повышением экспрессии. Напротив, в публикации Помазного с соавторами для классификации применялся следующий критерий: ген считался специфичным для одной фазы сложного жизненного цикла *O. felineus*, если его экспрессия была больше нуля на рассматриваемой фазе и равной нулю на другой фазе (Pomaznoy et al., 2016). Мы можем ожидать, что вне зависимости от того, какой критерий выбран исследователями, изменение набора сравниваемых фаз отразится и на результатах классификации. Как следствие, ген может быть рассмотрен как фаза-

специфичный или обладающий фаза-специфичным повышением экспрессии только в рамках конкретного эксперимента и анализа.

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

Большое число геномных и транскриптомных данных, доступных для широкого круга живых организмов, открыло возможность исследовать эволюционную историю трансформации наборов генов современных видов. 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). Как правило, для определения гомологов используют популярный метод поиска основного локального выравнивания (Basic Local Alignment Search Tool, BLAST), который, согласно результатам дополнительных исследований (Domazet-Lošo et al., 2017), является подходящим и достаточно чувствительным инструментом для филостратиграфического анализа. После того, как гомолог гена найден, становится возможным определить возраст гена. В филостратиграфическом контексте возраст гена – это не время, прошедшее с момента последней дупликации последовательности, а скорее время появления предка семейства генов, которому принадлежит рассматриваемая последовательность (Arendsee et al., 2019). Это означает, что определение возраста происходит относительно ветви филогенетического дерева, на которой, предположительно, появился основатель семейства генов (Arendsee et al., 2019). В соответствии с вычисленными возрастными, все гены исследуемого вида распределяются по группам, названным «филостратами» (Domazet-Lošo et al., 2007). Таким образом, каждая отдельная филострата представляет собой набор генов исследуемого вида, обладающих общим филогенетическим происхождением (Domazet-Lošo et al., 2007). В оригинальной публикации Домазет-Лошо, Брайковича и Таутца, в которой была представлена филостратиграфия, авторы показали на примере *Drosophila melanogaster*, что анализ филострат позволяет раскрыть следы важных адаптивных событий в эволюции вида.

Далее Домазет-Лошо и Таутц предложили использовать комбинацию данных филостратиграфии и результатов квантификации экспрессии генов для создания кумулятивного индекса, отражающего эволюционный возраст транскриптома исследуемого образца (Domazet-

Lošo, Tautz, 2010). Разработанный «возрастной индекс транскриптома» (Transcriptome Age Index, TAI) объединяет возраст гена с уровнем его экспрессии в рассматриваемом образце и суммирует это значения 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, TAI представляет собой среднее арифметическое взвешенное значение вкладов филострат (Domazet-Lošo, Tautz, 2010).

Филостратиграфия и основанный на её результатах анализ возрастных индексов транскриптомов активно применяются в исследовании широкого круга живых организмов. С помощью филостратиграфии удалось установить, что гены спорообразования *Bacillus subtilis* формируют несколько групп, возникших на разных этапах эволюции (Shi et al., 2020). Полученные результаты позволили авторам предположить, что молекулярные основы процесса образования спор у *B. subtilis* претерпели несколько стадий расширения (Shi et al., 2020). В исследовании наборов генов *Arabidopsis thaliana*, связанных с несколькими типами абиотических стрессов, показано не только то, что значительная часть этих генов имеет древнее происхождение, но и то, что гены одного возраста, как правило, связаны в сети (Mustafin et al., 2019). Более того авторам удалось установить, что «старые» гены, ассоциированные со стрессом, обычно участвуют как в ответе на различные виды стресса, так и во многих биологических процессах, не имеющих отношений к стрессу (Mustafin et al., 2019). Филостратиграфия применялась и в исследовании транскриптомов «неограниченных первичных стволовых клеток» («unlimited primordial stem cells», uPriSCs), т. е. стволовых клеток, которые сохраняют способность продуцировать как зародышевую линию, так и, по крайней мере, несколько типов соматических клеток взрослых организмов (Alié et al., 2015). Наличие таких клеток показано у губок, кишечнополостных, плоских червей и колониальных асцидий, что дает основание предположить, что uPriSC представляют собой предковый тип стволовых клеток животных (Alié et al., 2015). Результаты филостратиграфического анализа указали на то, что подавляющее большинство генов, активных в uPriSC, включая те, что кодируют белки, связывающие РНК, имеют древнее происхождение. На основании полученных результатов авторы предположили, что появление стволовых клеток у последнего общего многоклеточного предка в значительной степени было связано с использованием древних эукариотических РНК-связывающих белков (Alié et al., 2015). У предкового организма такие белки могли служить в качестве эффективного инструмента для поддержания баланса между самообновлением и дифференцировкой клеток. По мнению авторов, древние эукариотические РНК-связывающие белки могли принимать участие в организации

точной и синхронизированной регуляции активности генов во многих сосуществующих клеточных линиях (Alié et al., 2015).

Сравнительный анализ возрастных индексов транскриптомов в пределах одного жизненного цикла позволяет описать различия во вкладах филострат в молекулярные подписи различных фаз. Например, исследования, проведенные на *Danio rerio*, были направлены на изучение филотипической стадии, т. е. стадии в развитии, в которой виды, принадлежащие одному таксону, особенно похожи друг на друга (Domazet-Lošo, Tautz, 2010). Используя в качестве модели онтогенез данио и развитие взрослых особей, авторы обнаружили, что на филотипической стадии действительно экспрессируется самый «старый» набор генов, а более «молодые» гены экспрессируются на ранних и поздних этапах эмбрионального развития (Domazet-Lošo, Tautz, 2010). Наиболее «молодым» транскриптомом обладали ювенили и чем старше становится особь, тем больше экспрессируются «старые» гены (Domazet-Lošo, Tautz, 2010). Показано, что гены, возникшие до эволюции Metazoa, равномерно экспрессировались на протяжении всего онтогенеза, тогда как филостраты, появившиеся позже, демонстрировали более разнообразный вклад в возрастные индексы транскриптомов анализируемых фаз жизненного цикла. Схожие результаты получены для плодовой мушки *D. melanogaster*, нематоды *Caenorhabditis elegans* и комара *Anopheles gambiae* (Domazet-Lošo, Tautz, 2010). На основании выявленных общих закономерностей авторы предположили, что существует связь между эволюционными инновациями и появлением «новых» генов. По мнению авторов, «филотипической» можно назвать ту фазу онтогенеза, во время которой экспрессируется самый «старый» набор генов, что может быть связано либо с тем, что на такой фазе слабее всего проявляются специфичные адаптации, либо с тем, что фаза внутренне ограничена настолько, что возникшие гены не могут быть интегрированы в сформированные сети (Domazet-Lošo, Tautz, 2010).

Анализ возрастных индексов транскриптомов был применен в исследовании эволюции двухфазного жизненного цикла. Существуют две альтернативные гипотезы, касающиеся эволюции жизненного цикла многоклеточных животных. Первая гипотеза предполагает, что первые многоклеточные животные были похожи на современных личинок и добавление взрослых форм произошло позже в эволюции (Wang J. et al, 2020). Напротив, вторая гипотеза предполагает, что первые многоклеточные были схожи со взрослыми формами, а двухфазный жизненный цикл возник позже в результате интеркаляции личинок (Wang J. et al., 2020). В качестве модели в исследовании Ванг с соавторами выбрали моллюсков, так как Mollusca представляет собой один из крупнейших таксонов беспозвоночных животных и характеризуется как наличием трохофорных личинок в жизненном цикле, так и большим разнообразием взрослых форм (Wang J. et al., 2020). Основной акцент в исследовании сделан на анализе транскриптомов

морского гребешка *Patinopecten yessoensis*. Авторам удалось обнаружить молодость транскриптома личинки с необычно высоким включением в него «новых» генов (Wang J. et al, 2020). Результаты анализа также указали на то, что «новые» гены, транскрипты которых присутствуют в транскриптоме трохофоры, потенциально вносили вклад в эволюцию раковин взрослых форм, и что гены, связанные с проведением сигналов в клетке, вероятно играли важную роль в эволюции личинки. Авторы расширили анализ, включив транскриптомы с разных этапов онтогенеза губки *Amphimedon queenslandica*, кишечнополостной *Clytia hemisphaerica*, гребневика *Mnemiopsis leidyi*, двухстворчатого моллюска *Crassostrea gigas*, брюхоногого моллюска *Haliotis discus hannai*, эхиуриды *Urechis unicinctus*, голотурии *Apostichopus japonicus* и оболочника *Styela clava*. У большинства видов личинки имели самый высокий или почти самый высокий возрастной индекс транскриптома (Wang J. et al, 2020). По мнению Ванга с соавторами, полученные результаты являются аргументом в пользу гипотезы, согласно которой появление двухфазного жизненного цикла связано с интеркаляцией личинки в жизненный цикл предка, схожего с современными взрослыми формами. На основании того, что филострата, соответствующая появлению Metazoa, являлась самой старой из тех, что имели заметный вклад в возрастные индексы транскриптомов личинок, авторы предположили, что первое появление протоличинок, вероятно, произошло после отделения гребневиков с прямым развитием от многоклеточного предка (Wang J. et al, 2020).

Высокие уровни экспрессии «новых» генов на фазе трохофоры показаны и в исследовании Сю с соавторами (Xu et al., 2016). Авторы провели сравнительный анализ транскриптомов разных фаз жизненных циклов двухстворчатого моллюска *Crassostrea gigas*, брюхоногого моллюска *Haliotis discus hannai* и полихеты *Perinereis aibuhitensis* с целью определить их эволюционную историю. У всех трех рассмотренных видов самое низкое значение возрастного индекса транскриптома получено для трохофоры, что указывает на то, что эта личиночная фаза имеет самый «старый» транскриптом. Тем не менее в исследовании не выявлено большого сходства между транскриптомами трохофор, что указывало на значимые различия между трохофорами моллюсков и полихет (Xu et al., 2016). Обнаруженные различия между трохофорами двух типов могут быть результатом интеграции «новых» генов в регуляторные сети в ходе длительной эволюции (Xu et al., 2016). Авторы отмечают, что полученные результаты могут быть рассмотрены в пользу как гипотезы об интеркаляции личинки в жизненный цикл, так и гипотезы о более позднем появлении взрослых форм. Например, волна экспрессии генов, специфичных для Metazoa, на фазе трохофоры указывает на то, что личинка использует древние биологические процессы, что согласуется с гипотезой о том, что личинка была первой в жизненном цикле (Xu et al., 2016). Напротив, наблюдаемая у трохофоры повышенная экспрессия генов, специфичных для типа, класса и вида, может быть рассмотрена в пользу гипотезы об

интеркаляции личинки в предковый жизненный цикл. Сью с соавторами (Xu et al., 2016) подчеркивают, что поскольку морские личинки выполняют как функцию развития, так и функцию расселения, их эволюционная история может быть гораздо более сложной, чем простая дихотомия гипотез о том, появилась ли личинка позже или раньше взрослых форм (Xu et al., 2016).

Как видно из представленных данных, важное значение в эволюции таксонов играют гены, появившиеся относительно недавно. Вследствие чего возникает вопрос о том, как могут появляться «новые» гены в геноме. В течение длительного периода времени преобладала точка зрения, согласно которой все «новые» гены возникают в результате удвоения последовательностей, существующих у предка, с последующей дивергенцией копий и обретением ими новой функции (Tautz, 2014). В поддержку такой точки зрения собрано множество доказательств, в то время как альтернативная точка зрения, а именно появление «новых» генов *de novo* из межгенных областей, в значительной степени игнорировалась (Tautz, 2014). В настоящее время ситуация полностью изменилась и как повторный анализ существующих данных, так и огромное количество новых исследований указывают на то, что эволюция генов *de novo* может быть преобладающим механизмом появления генов (Tautz, 2014). Для обозначения генов, не имеющих сходства с последовательностями в геномах других организмов, используют термин «гены-сироты» (Orphan genes, ORFans) (Dujon, 1996) или «гены, ограниченные таксоном» (Taxonomically-restricted genes, TRGs) (Wilson et al., 2005). Второй термин является более точным определением «сиротских» или «новых» генов, так как указывает и на узкое филогенетическое распространение таких последовательностей (Khalturin et al., 2009). В контексте филостратиграфии, генами-сиротами могут считаться все гены, специфичные для определенного таксона, т. е. которые не могут быть сопоставлены с набором генов одноклеточного предка (Tautz et al., 2013). Потенциальными механизмами возникновения генов-сирот могут быть: дупликация предкового гена и быстрое расхождение его копий, в результате чего становится невозможным обнаружить гомологию между ними (Tautz, Domaset-Lošo, 2011), смена функции мобильных элементов (Toll-Riera et al., 2009), горизонтальный перенос генов (Husnik, McCutcheon, 2018), сдвиг рамки считывания, кодирующей белок, приводящий к появлению совершенно нового белка (Neme, Tautz, 2013) и фактическое появление генов *de novo* в ранее некодирующих областях генома (Schlötterer, 2015). Так, к примеру, согласно одной из моделей, ген может появиться *de novo* из протогена (Tautz et al., 2013). Эта модель предполагает, что «новый» ген может брать начало с ошибочно транскрибируемой области генома, которая приобретает некоторые дополнительные мутации для превращения полученной РНК в стабильный наследуемый транскрипт (Tautz et al., 2013). Первоначально такие протогены могут действовать как гены некодирующих РНК, ассоциированные с другими молекулярными

комплексами. Далее такие транскрипты могут быть транслированы, и как только получаемый пептид станет полезным для вида, он будет дополнительно оптимизирован и в итоге станет частью генного репертуара вида. Такие «новые» гены затем могут сформировать целые генные семейства посредством дупликации или, напротив, снова исчезнуть в межгенной области. Для всех указанных шагов от случайного транскрипта к стабильному «новому» гену получены конкретные примеры у различных видов (Tautz et al., 2013). Исследованные протогены обладают промежуточными характеристиками (уровень экспрессии, кодирующий потенциал открытой рамки считывания, длина, характер модификации хроматина, возможность белок-белковых и межгенных взаимодействий и др.) между генами и некодирующими участками генома (Cherezov et al., 2021).

В настоящее время стали возможными реконструкция и сравнительный анализ моделей геномов предков таксона с целью определить, какие гены у таксона дублировались, а какие появились впервые. Один из современных подходов для реконструкции модели предкового генома основан на анализе иерархических групп ортологов. Такие ортогруппы представляют собой наборы генов, которые, предположительно, произошли от общего предкового гена в пределах одной видовой клады (Train et al., 2019). Это означает, что набор всех иерархических ортогрупп одной клады соответствует набору предковых генов у последнего общего предка рассматриваемой клады. Кроме того, если данные об иерархических ортогруппах доступны для вложенных клад (например, млекопитающих по сравнению с позвоночными), то сравнение их наборов ортогрупп позволит определить различия между кладами: расщепление иерархической ортогруппы будет означать дубликацию генов, а исчезновение ортогруппы – потерю (Train et al., 2019).

Описанный подход использован в исследовании Заяк с соавторами для реконструкции моделей геномов предков современного вида дигеней *Atriophallophorus winterbourni* (син. *Microphallus* sp. или *Microphallus livelyi*) (Zajac et al., 2021). Выбранный вид дигеней характеризуется отсутствием в жизненном цикле таких фаз, как материнская спороциста, редия, церкария и, возможно, мирацидий (Zajac et al., 2021). Авторам удалось не только собрать геном *A. winterbourni*, но и построить модели геномов последних общих предков дигеней, Plagiorchiida, а также Opisthorchiata/Xiphidiata (Zajac et al., 2021). Реконструкция моделей предковых геномов была выполнена на основании анализа данных 14 видов дигеней, 3 видов цестод, 1 вида моногеней и 2 свободноживущих видов нематод. С целью увеличения надежности получаемых результатов сравнения между моделями геномов авторы ввели два критерия. Согласно первому критерию, ген классифицировался как «приобретенный», только если он присутствовал у большинства существующих видов рассмотренной клады. Согласно второму критерию, ген классифицировался как «дублированный» только если его последовательность

присутствовала по крайней мере в пяти копиях. С учетом введенных критериев сравнительный анализ между построенными моделями продемонстрировал высокую долю геномных инноваций при отхождении предка дигеней от общего предка дигеней и цестод: 37,2% генов оказались «новыми» и были приобретены последним общим предком дигеней (Zajac et al., 2021). В то же время только 660 генов отсутствовали в построенной модели предка дигеней в сравнении с моделью генома общего предка дигеней и цестод. Результаты анализа функциональной аннотации генов, сохраняющихся на всем протяжении эволюции от последнего общего предка паразитических плоских червей до современных видов дигеней, подтвердили, что они являются генами домашнего хозяйства и необходимыми для жизни, роста и размножения. Отдельное внимание авторы уделили исследованию «новых» (полученных) генов, так как они могут указать на те генетические инновации, которые имели решающее значение для адаптации к хозяевам, расширению/упорядочиванию жизненных циклов и адаптации к меняющимся условиям окружающей среды (Zajac et al., 2021). Удалось установить, что в числе биологических процессов, обогащенных приобретенными генами, были, к примеру, связанные с каталитической активностью, моторной активностью микротрубочек, а также организацией клеточных компонентов, включая организацию актиновых пучков филаментов и реакцию на стимулы. Авторы связывают эти процессы с устройством тегумента. В геноме же *A. winterbourni* в сравнении с предком Opisthorchiata/Xiphidiata удалось обнаружить массовое расширение 13 иерархических ортогрупп (Zajac et al., 2021). Исходя из результатов функциональной аннотации этих ортогрупп, Заяк с соавторами предположили, что гены, входящие в них, вероятно, участвуют в проникновении в ткани хозяина, избегании иммунного ответа и манипулировании поведением хозяина.

Представленный обзор литературных данных свидетельствует о том, что применение биоинформатических методов анализа в исследовании сложных жизненных циклов дигеней является одним из перспективных направлений и новых этапов в трехсотлетней истории изучения биологии дигенетических сосальщиков. Многочисленные опубликованные геномные и транскриптомные данные позволили детально изучить молекулярную биологию различных видов дигеней. Тем не менее до настоящего момента комплексного сравнительного анализа молекулярных подписей фаз жизненных циклов филогенетически близких и отдаленных видов проведено не было. Восполнить пробелы в исследовании молекулярных основ реализации сложного жизненного цикла дигеней, как общего признака для всего таксона, возможно не только вовлечением новых видов и транскриптомов, но и изменением подхода к анализу уже опубликованных данных.

Глава 2 Материалы и методы

Сбор материала

Приблизительно 7000 особей моллюсков вида *Bithynia tentaculata* были собраны с поверхностей водных растений и подводных камней реки Кристателька (Петродворцовый район Санкт-Петербурга, Россия) на территории парка Сергиевка (Рисунок 1) и содержались на кафедре Зоологии беспозвоночных Санкт-Петербургского государственного университета.

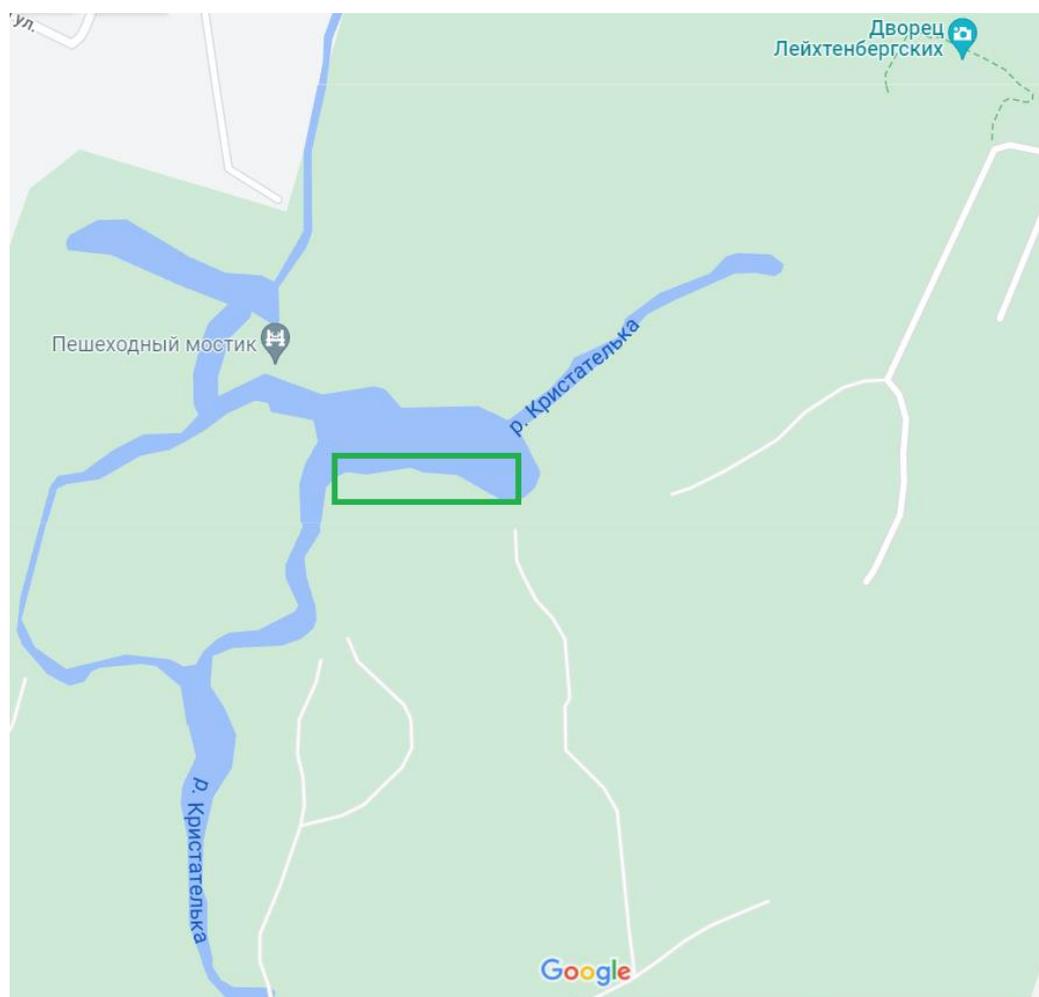


Рисунок 1 – Схематическая карта локации сбора моллюсков *Bithynia tentaculata*. Зеленым прямоугольником обозначен берег, вдоль которого проведен сбор моллюсков для дальнейшего лабораторного исследования на заражение дигенейями.

Общая схема реализации сложных жизненных циклов *S. pseudoglobulus* и *P. simillimum* в природных и лабораторных условиях представлена на рисунке 2.

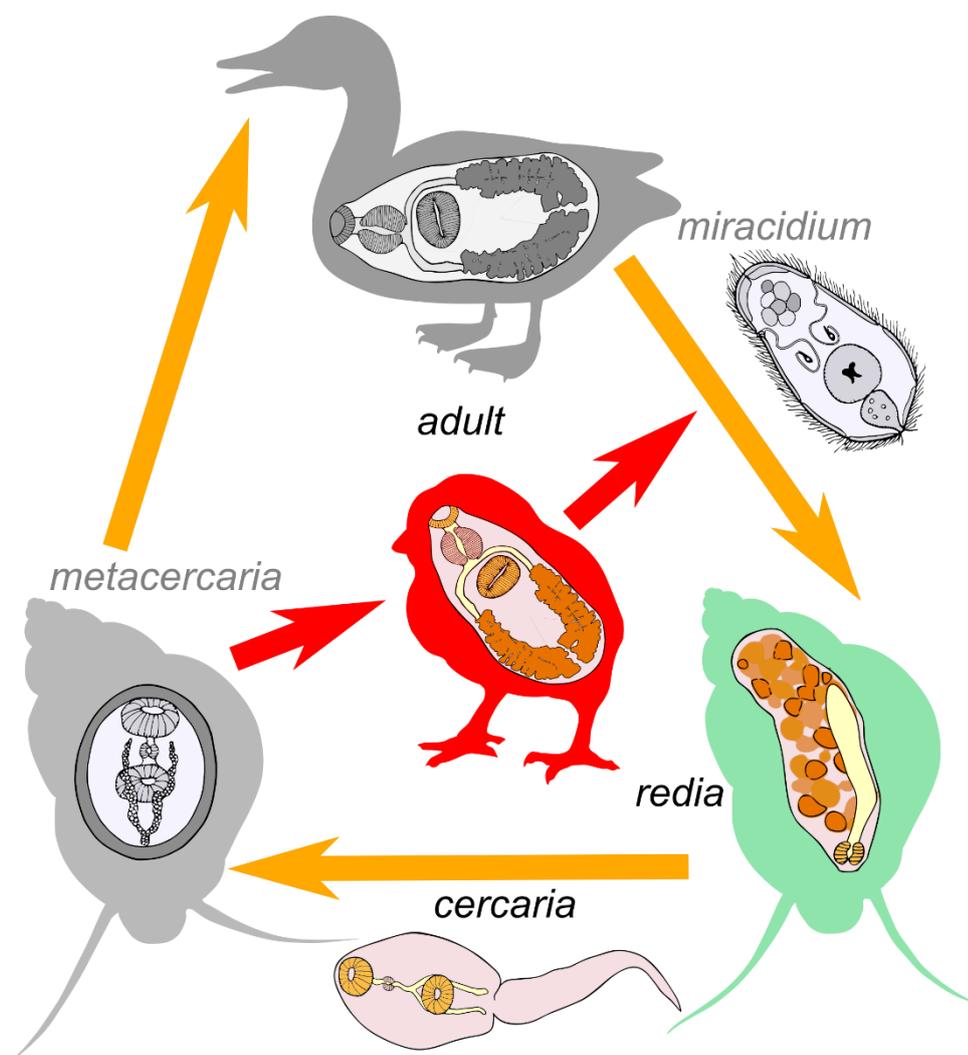


Рисунок 2 – Обобщенная схема реализации сложных жизненных циклов *Sphaeridiotrema pseudoglobulus* и *Psilotrema simillimum* в природных (желтые стрелки) и лабораторных (красные стрелки) условиях. В природе окончательным хозяином обоих исследованных видов Psilostomatidae являются водоплавающие птицы. В качестве окончательного хозяина дигеней в лабораторных условиях использовались цыплята. Промежуточными хозяевами *P. simillimum* и *S. pseudoglobulus* являются моллюски *Bithynia tentaculata*. Обозначения: Miracidium – мирацидий, Redia – редия, Cercaria – церкария, Metacercaria – метацеркария, Adult – марита. Редия, метацеркария и марита показаны внутри их хозяев. Окрашены только фазы сложного жизненного цикла, для которых проведен транскриптомный анализ.

Для выявления заражения паразитами каждый выловленный моллюск был помещен в отдельную бакпечатку, наполненную чистой пресной водой. Определение видовой принадлежности церкарий, покидающих зараженных моллюсков, проводили по морфологическим признакам (Bychovskaja-Pavlovskaja, Kulakova, 1971) с использованием светового микроскопа. Церкарии видов *P. simillimum* и *S. pseudoglobulus* были собраны в пробирки типа Эппендорф и зафиксированы с помощью реагента IntactRNA™ (Евроген, Москва,

Россия) согласно протоколу производителя для образцов транскриптомов. Для каждого вида мы подготовили по шесть фиксаций, каждая из которых содержала около сотни личинок.

Моллюски, зараженные исследуемыми видами дигеней, были отобраны и помещены на ночь в отдельные емкости с отфильтрованной водой. Извлечение редий производили путем вскрытия зараженных хозяев в охлажденном физрастворе. Отделение развивающихся эмбрионов от редий не проводилось ввиду практически полной невозможности сделать это без повреждения последних и потери материала. Все собранные редии были помещены в пробирки типа Эппендорф с буфером IntactRNA™ и заморожены согласно протоколу производителя для образцов транскриптомов. Для каждого вида мы подготовили четыре фиксации по 50 партенит.

Для получения адолескарий *P. simillimum* пустые раковины моллюсков и фрагменты водных растений помещали в емкости с зараженными моллюсками *B. tentaculata* на 24 часа. Метациркарии *S. pseudoglobulus* были собраны напрямую из моллюсков, зараженных этим видом дигеней. Собранный материал использовался для дальнейшего перорального заражения цыплят, специально выращенных в лабораторных условиях. На каждое экспериментальное заражение было использовано приблизительно до 200 адолескарий / метациркарий. Заражение цыплят проводили в период между вылуплением и первым кормлением. После заражения цыплята содержались в лаборатории в течение 5 дней. Далее производилось вскрытие и исследование пищеварительной системы цыпленка на предмет наличия марит. В среднем с одного цыпленка нам удалось получить по одной половозрелой особи паразита. Подтверждение видовой принадлежности дигеней производилось на основании морфологии их марит. После этого мариты *P. simillimum* и *S. pseudoglobulus* помещались в пробирки типа Эппендорф с реагентом intactRNA™, а затем фиксации переносились в холодильник в соответствии с протоколом производителя для образцов транскриптомов. Для каждого вида мы подготовили по три фиксации по пять марит.

Извлечение образцов транскриптомов и их секвенирование

Для каждой из исследуемых фаз сложных жизненных циклов *P. simillimum* и *S. pseudoglobulus* мы подготовили пулированные образцы фиксированного материала. Образцы марит содержали приблизительно по семь половозрелых особей паразита, образцы редий с эмбрионами церкарий состояли приблизительно из 100 партенит, а образцы церкарий содержали около 300 личинок из 15 моллюсков. Мы подготовили по две биологические повторности для каждой из исследуемых фаз сложного жизненного цикла *P. simillimum* и *S. pseudoglobulus*.

Выделение образцов тотальной РНК 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».

Подготовка библиотек коротких парных прочтений для *de novo* сборки транскриптомов

Оценку качества полученных библиотек коротких парных прочтений мы выполнили с помощью программы FastQC (v0.11.5). Для исправления возможных ошибок секвенирования использовали программу Korrect (v1.0) (Allam et al., 2015) со следующими параметрами запуска: --celltype=diploid --matchtype=hamming. С целью выявления и удаления возможной контаминации мы сравнили все библиотеки коротких прочтений со «справочным» транскриптомом моллюска *Bithynia siamensis goniomphalos* (Cantacessi et al., 2013), «справочной» сборкой генома *Gallus gallus* (Encode: Gallus gallus v5.0), а также «справочной» сборкой генома *Homo sapiens* (Encode: GRCh38, primary assembly), используя набор программ BBTools (v37.02). Деконтаминированные библиотеки были загружены в базу данных NCBI SRA под номером БиоПроекта PRJNA516017 (Приложение 1). Далее библиотеки прочтений, очищенные от возможной контаминации, мы проанализировали с помощью программы Trimmomatic (v0.36) (Bolger et al., 2014). Программа Trimmomatic удалила низкокачественные и адаптерные последовательности в соответствии с заданными параметрами: ILLUMINACLIP:\$ADAPTERS:2:30:10:2:TRUE SLIDINGWINDOW:4:20 MAXINFO:50:0.8 MINLEN:25.

Сборка протяженных последовательностей из коротких прочтений *de novo*

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), запуск которой проводился со

значением 25 для параметра размера k-мера и с минимальной длиной контига равной 200 нуклеотидам. К идентификаторам каждой собранной последовательности был добавлен уникальный трехбуквенный код: «Psi» для *P. simillimum* и «Sps» для *S. pseudoglobulus*.

Кластеризацию изоформ последовательностей в обоих транскриптомах выполнили с помощью программы CD-HIT-est (v4.7) (Fu et al., 2012) со следующими значениями для параметров: минимальное сходство последовательностей для объединения их в кластер должно было быть не менее 95% (-c 0.95), тщательный режим кластеризации (-g 1) и при выравнивании последовательностей допускалось ориентирование их как +/+, так и +/- (-r 1). Из каждого кластера было извлечено по одной репрезентативной последовательности.

Мы использовали программу TransRate (v1.0.1) (Smith-Unna et al., 2016) для проведения оценки качества и полноты итоговых сборок транскриптомов. Основываясь на результатах выравнивания коротких прочтений на собранные последовательности, TransRate определяла 4 разных типа ошибок:

1. наличие химерных участков, т. е. фрагментов разных последовательностей, объединенных в одну,
2. наличие структурных ошибок,
3. неполнота сборки протяженной последовательности,
4. наличие ошибок в расположении нуклеотидов.

Далее для каждого контига программа выдала заключение о качестве и полноте его сборки. Результаты всех последовательностей учитывались при оценке качества и полноты сборки всего рассматриваемого транскриптома целиком. В дальнейшем анализе мы использовали только те контиги, которые характеризовались высокими показателями качества и полноты сборки, согласно критериям TransRate.

Квантификация уровней экспрессии и определение кодируемых аминокислотных последовательностей

Для квантификации уровней экспрессии отобранных последовательностей мы использовали программу Salmon (v1.0.1) (Patro et al., 2017). Запуск программы был выполнен со следующими значениями параметров: -l ISF --discardOrphans --seqBias --gcBias --validateMappings. 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. Преобразование

выполнялось с помощью библиотеки «tximport» для языка программирования R. По результатам анализа мы подготовили таблицу с нормализованными и усредненными между биологическими повторностями уровнями экспрессии предполагаемых генов в транскриптах-на-миллион (Transcripts-Per-Million, TPM). Выбранная единица измерения отражает количество молекул РНК рассматриваемого гена на каждый миллион транскриптов в исследуемого образце. Из дальнейшего рассмотрения были исключены гены с уровнями экспрессии меньше 1 транскрипта-на-миллион на всех трех исследуемых фазах одного сложного жизненного цикла.

Определение белок-кодирующих генов выполнили с помощью программы TransDecoder (v5.5.0) в 3 этапа. На первом этапе среди отобранных ранее контигов были обнаружены те, что имели открытые рамки считывания (Open Reading Frames, ORF) длиной не менее 100 аминокислот. Далее продукты трансляции найденных ORF были сравнены с аминокислотной базой данных NCBI (NCBI non-redundant) с помощью программы DIAMOND BLASTp (v0.9.22.123) (Buchfink et al., 2015). При запуске программы в качестве пороговой величины для параметра «e-value», описывающего количество случайных совпадений, которое можно было бы ожидать при сравнении с базой данных заданного размера, установили значение $1e^{-3}$. Дополнительно мы использовали программу HMMscan (v3.1b2) для сравнения продуктов трансляции ORF с базой данных белковых доменов PfamA (El-Gebali et al., 2019; Mistry et al., 2021). На третьем этапе анализа результаты сравнения с базами данных были использованы при повторном запуске TransDecoder для уточнения аминокислотных последовательностей кодируемых белков.

Подготовка «справочных» наборов последовательностей и их аннотация

Основное внимание в дальнейшем анализе мы сфокусировали на генах, удовлетворявших два требования: 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. Полученные таким образом наборы нуклеотидных и аминокислотных последовательностей рассматривали в качестве «справочных» для *P. simillimum* и *S. pseudoglobulus*. Все этапы подготовки «справочных» наборов выполнялись с помощью написанных нами программ на языке программирования Python (v3.6).

Контроль полноты подготовленных «справочных» наборов белков был проведен с помощью BUSCO (v3.0.1) (Simão et al., 2015, Waterhouse et al., 2018) (e-value = $1e^{-3}$, mode = proteins) 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. Мы использовали наборы идентифицированных *in silico* белков для *Clonorchis sinensis* (UP000286415), *Fasciola hepatica* (UP000230066), *Opisthorchis viverrini* (UP000054324), *Opisthorchis felineus* (UP000308267), *Schistosoma mansoni* (UP000008854), *Schistosoma haematobium* (UP000054474) и *Schistosoma japonicum* (UP000311919) из базы данных UniProt (The UniProt Consortium, 2021), а также опубликованные данные для *Fasciola gigantica* (Zhang X.-X et al., 2019), *Trichobilharzia regenti* (Leontovyč et al., 2016) и *Trichobilharzia szidati* (Leontovyč et al., 2019). Транскриптомные данные по последним трем видам были предоставлены нам Романом Леонтовичем (Roman Leontovyč) и Кристиной Цвиклински (Krystyna Cwiklinski) по личному запросу. Для корректного сравнения мы исключили из рассматриваемых наборов белков дигеней все последовательности, имеющие длину меньше 100 аминокислот, как это было сделано ранее для *P. simillimum* и *S. pseudoglobulus*. Все результаты анализов сводились в единую таблицу с помощью написанной нами программы на языке программирования Python. Если однокопийный ортолог отсутствовал во всех рассмотренных видах дигеней, то мы рассматривали его в качестве потенциальной специфичной для дигеней потери.

Функциональная аннотация последовательностей из «справочных» наборов проводилась путем сравнения контигов и белков с нуклеотидной и аминокислотной базами данных NCBI, а также с базой данных белков SwissProt. Поиск совпадений с нуклеотидной базой данных NCBI был выполнен с помощью программы BLASTn (Camacho et al., 2009). Для сравнения с белками из NCBI и SwissProt мы использовали программу DIAMOND BLASTp, запущенную в «чувствительном» (--sensitive) режиме. В качестве порогового значения для параметра «e-value» во всех сравнениях мы выбрали величину $1e^{-3}$. Дополнительно все белки из «справочных» наборов *P. simillimum* и *S. pseudoglobulus* проанализировали программой HMMscan на наличие в их аминокислотных последовательностях профилей белковых доменов из базы данных PfamA. Белки двух видов Psilostomatidae были также исследованы с помощью Интернет-ресурса eggNOG-mapper с параметрами запуска по умолчанию.

Результаты аннотации, полученные с помощью eggNOG-mapper, использовались для идентификации в «справочных» наборах ферментов, принимающих участие в биосинтезе жирных кислот. Из таблиц с аннотацией мы извлекли все уникальные идентификаторы ферментов, связанных с биосинтезом жирных кислот (Kyoto Encyclopedia of Genes and Genomes

(KEGG) (Kanehisa et al., 2007): map00061). Исследование доменной архитектуры обнаруженных ферментов было выполнено на основе результатов, полученных с помощью HMMscan.

Поиск и анализ гаплотипов *cox1* в «справочном» наборе *Sphaeridiotrema pseudoglobulus*

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 (GenBank: GQ890329.1, GQ890328.1, FJ477222.1, KM538101.1, KM538091.1, KM538104.1, KM538090.1, KM538104.1, MN748721.1, MK982785.1, KY636236.1, KY636202.1, KY636234.1 и MG964028.1). Выравнивание последовательностей производилось автоматически с помощью алгоритма MUSCLE (Edgar, 2004), имплементированного в программе SeaView 4.0 (Gouy et al., 2010). Для Байесовского анализа мы использовали программу MrBayes (Ronquist, Huelsenbeck, 2003) на мощностях XSEDE 3.2.7a, предоставленных на портале CIPRES (Miller et al., 2010). В анализе мы использовали модель GTR с гамма-коррекцией межсайтовой вариации скорости (8 категорий) и модель ковариации. Построение деревьев запускали как две отдельные цепочки (параметры «нагрева» по умолчанию) для 15 миллионов поколений. Итоговое среднее стандартное отклонение частот разделения было менее 0.01. Качество полученных результатов оценивалось с помощью программ, встроенных в MrBayes. Обнаруженные гаплотипы *cox1* были загружены в базу данных GenBank под идентификаторами MT934401 и MT934400.

Определение и анализ групп ортологов

Для построения групп ортологичных последовательностей мы использовали две программы, различающиеся по алгоритмам, лежащих в их основе. Первая программа, OrthoFinder (v2.2.6) (Emms, Kelly, 2015, 2019), характеризуется высокой скоростью и простотой использования, что играло важную роль для первичного определения ортологов между разными видами плоских червей. В анализ были включены не только подготовленные «справочные» наборы белков *P. simillimum* и *S. pseudoglobulus*, но и данные, находящиеся в открытом доступе:

1. наборы белков *Clonorchis sinensis* (UP000286415), *Fasciola hepatica* (UP000230066), *Opisthorchis viverrini* (UP000054324), *Opisthorchis felinus* (UP000308267), *Schistosoma*

mansoni (UP000008854), *Schistosoma haematobium* (UP000054474), *Schistosoma japonicum* (UP000311919) из базы данных UniProt,

2. опубликованные и предоставленные нам наборы белков *Fasciola gigantica* (Zhang X.-X. et al., 2019), *Trichobilharzia regenti* (Leontovyč et al., 2016) и *Trichobilharzia szidati* (Leontovyč et al., 2019),
3. наборы белков свободноживущих плоских червей *Schmidtea mediterranea* (PRJNA379262) и *Macrostomum lignano* (PRJNA371498) из базы данных WormBase Parasites (Howe et al., 2017).

Для каждого исследуемого вида отобрали только белки длиной не менее 100 аминокислот. Визуализацию полученных результатов построения ортогрупп выполнили с помощью библиотек «ggplot2» (v3.3.5), «pheatmap» (v1.0.12) и «RColorBrewer» (v1.1-2) в окружении RStudio для языка программирования R.

Дополнительно мы разделили все рассмотренные виды плоских червей на четыре группы в соответствии с особенностями их жизненных циклов для того, чтобы определить количество общих и специфичных ортогрупп:

1. *P. simillimum*,
2. *S. pseudoglobulus*,
3. «редииоидные» виды дигеней (*C. sinensis*, *F. gigantica*, *F. hepatica*, *O. felineus* и *O. viverrini*),
4. «спороцистоидные» виды дигеней (*S. haematobium*, *S. japonicum*, *S. mansoni*, *T. regenti* и *T. szidati*),
5. свободноживущие виды плоских червей (*M. lignano* и *S. mediterranea*).

Если выделенная группа включала больше 1 вида, то мы рассматривали только те ортогруппы, которые содержат белки всех объединенных видов. Построение и визуализацию диаграмм Венна для подготовленных наборов ортогрупп выполнили с помощью Интернет-ресурса InteractiVenn (Heberle et al., 2015).

С целью уточнить результаты, полученные программой OrthoFinder, и построить иерархичные ортогрупп (hierarchical orthologous groups, HOG), мы провели повторный поиск ортологов с использованием программы OMA standalone (v2.5.0) (Altenhoff et al., 2019). В отличие от OrthoFinder, OMA standalone может использовать филогенетическое дерево исследуемых видов для коррекции результатов поиска ортологов. Анализ был выполнен в несколько этапов.

На первом этапе мы обновили наборы исследуемых данных. Для этого мы заменили анализируемые ранее наборы белков теми, что были получены для последних на момент обращения к базе данных WormBase Parasite (19 августа 2021) версий сборок геномов. В результате были обновлены наборы белков *C. sinensis* (PRJNA386618; ASM360417v1,

GCA_003604175.1), *F. gigantea* (PRJNA230515; F_gigantica_1.0.allpaths, GCA_006461475.1), *F. hepatica* (PRJEB25283; Fasciola_10x_pilon, GCA_900302435.1), *M. lignano* (PRJNA371498; Mlig_3_7, GCA_002269645.1), *O. felineus* (PRJNA413383; ICG_Ofel_1.0, GCA_004794785.1), *O. viverrini* (PRJNA222628; OpiViv1.0, GCF_000715545.1), *S. haematobium* (PRJNA78265; SchHae_2.0, GCA_000699445.2), *S. japonicum* (PRJNA520774; ASM636876v1, GCA_006368765.1), *S. mansoni* (PRJEA36577; Smansoni_v7, GCA_000237925.3), а также *S. mediterranea* (PRJNA379262; ASM260089v1, GCA_002600895.1). Мы рассматривали только виды плоских червей, для которых доступны высококачественные сборки геномов: значение N50 > 175 Кб и более 50% успешно собранных однокопийных ортологов, присутствующих у более чем 90% животных, согласно информации в базе данных WormBase Parasite. Несмотря на отсутствие собранных геномов с высокими показателями протяженности и полноты сборки для *T. regenti* и *T. szidati*, мы оставили оба указанных вида в анализе ввиду наличия для них высококачественных сборок транскриптомов и библиотек коротких прочтений для церкарий и шистосомул. В анализируемый набор данных был также добавлен идентифицированный *in silico* протеом свободноживущего плоского червя *Prostheceraeus vittatus* из базы данных PlanMine (bg_Pvit_v1) (Brandl et al., 2016) на основании высокой молекулярной консервативности его последовательностей (Martín-Durán et al., 2017). Принимая во внимание возможное влияние смеси криптических видов *Sphaeridiotrema* на результаты поиска ортогрупп, мы исключили из рассмотрения набор белков *S. pseudoglobulus*. Таким образом, итоговый набор исследуемых видов состоял из 11 дигеней (*C. sinensis*, *F. gigantea*, *F. hepatica*, *O. felineus*, *O. viverrini*, *P. simillimum*, *S. haematobium*, *S. japonicum*, *S. mansoni*, *T. regenti* и *T. szidati*) и трех свободноживущих плоских червей (*S. mediterranea*, *M. lignano*, *P. vittatus*).

На втором этапе анализа использовали программу MMseqs2 (Mirdita et al., 2019) для кластеризации схожих последовательностей в данных по *P. simillimum*, *P. vittatus*, *T. regenti* и *T. szidati*. В отличие от других рассматриваемых видов плоских червей, для указанных четырех видов нет геномных данных и доступны только транскриптомы, которые могут содержать изоформы. Запуск MMseq2 для кластеризации изоформ проводился для каждого вида в отдельности со следующими параметрами: mmseqs easy-cluster -cov-mode 0 -min-seq-id 0.9.

На третьем этапе анализа использовали все белки длиной ≥ 100 аминокислот от 14 видов плоских червей для построения ортогрупп с помощью OMA standalone. Первый запуск программы был выполнен с параметрами по умолчанию и «bottom-up» алгоритмом для поиска иерархических групп ортологов. Мы не предоставляли филогенетического дерева, но указали три свободноживущих вида (*M. lignano*, *P. vittatus*, *S. mediterranea*) в качестве аутгруппы.

Полученные результаты мы использовали на четвертом этапе анализа для реконструкции филогенетического дерева согласно протоколу Дилуса с соавторами (Dylus et al., 2020). С

помощью программы `filter_groups.py`, представленного в публикации, были отобраны группы ортологов (ОМА группы), которые включали белки минимум 13 из 14 исследуемых видов плоских червей. Далее в пределах каждой ОМА группы мы произвели множественное выравнивание белков с помощью программы MAFFT (v7.487) (Katoh, Standley, 2013), используя следующие параметры: `--maxiterate 1000 --localpair`. Все полученные выравнивания были объединены в единую суперматрицу программой `concat_alignments.py`. Выбор подходящих сайтов для анализа производился программой `trimAl` (Capella-Gutiérrez et al., 2009) в автоматическом (`-automated1`) режиме. Мы использовали программу `ProtTest` (v3.4.2) (Darriba et al., 2011) для определения наиболее подходящей модели эволюции исследуемых последовательностей. Реконструкцию филогенетического дерева выполнили с помощью IQ-TREE (v2.1.4-beta) (Nguyen et al., 2015; Minh et al., 2020), запущенной со следующими параметрами: `-m LG + I + G + F --seed 12345 -B 1000 --nmax 1000`. Укоренение консенсусного дерева было выполнено с помощью библиотеки «ape» (Paradis, Schliep, 2019) для языка программирования R, а в качестве аутгруппы мы указали 3 свободноживущих вида плоских червей.

Итоговое укорененное дерево было использовано на пятом этапе анализа при повторном запуске ОМА standalone с параметрами по умолчанию и «bottom-up» алгоритмом для поиска иерархических групп ортологов. Построение тепловой карты с количеством общих ОМА групп между исследуемыми видами плоских червей выполнили в RStudio с помощью библиотек «ggplot2», «pheatmap» и «RColorBrewer».

Определение молекулярных подписей фаз сложных жизненных циклов дигеней

Для определения молекулярных подписей фаз сложных жизненных циклов дигеней мы использовали два пороговых значения экспрессии. В качестве первого порога мы выбрали 1 транскрипт-на-миллион, то есть ген был включен в молекулярную подпись фазы жизненного цикла только в том случае, если уровень его экспрессии на рассматриваемой фазе был ≥ 1 транскрипту-на-миллион. Указанный порог использовался для внутривидового анализа, выполненного для *P. simillimum* и *S. pseudoglobulus*, с целью наиболее полно охватить весь набор активных генов для каждой из рассматриваемых фаз сложного жизненного цикла исследуемого вида. Анализ проводился на значениях экспрессии, полученных для «справочных» наборов последовательностей. Построение и визуализацию диаграмм Венна для молекулярных подписей фаз одного сложного жизненного цикла мы провели с помощью Интернет-ресурса Venny (v2.1.0).

Гомеобокс-содержащие гены были идентифицированы на основании результатов сравнения кодируемых белков с базой данных белковых доменов PfamA. Для построения тепловых карты значения экспрессии преобразовывались с помощью функции $\log_2(\text{TPM}+1)$, где TPM соответствовало уровню экспрессии гена в транскриптах-на-миллион. Визуализацию выполнили с помощью библиотек ggplot2, pheatmap и RColorBrewer в RStudio.

Далее мы подняли порог минимальной экспрессии до 2 транскриптов-на-миллион с целью уточнить полученные результаты определения молекулярных подписей фаз. Указанный порог экспрессии был выбран в соответствии с результатами исследований Вагнера (Wagner), Кина (Kin) и Линча (Lynch) (2013), согласно которым гены с уровнями экспрессии больше, чем 2 транскрипта-на-миллион наиболее вероятно являются активно транскрибируемыми. В описываемом анализе мы не рассматривали *S. pseudoglobulus* из-за возможного влияния смеси криптических видов на определение молекулярных подписей разных фаз сложного жизненного цикла дигеней.

Для того, чтобы иметь возможность сопоставить результаты анализов молекулярных подписей фаз жизненных циклов филогенетически близких и отдаленных видов дигеней, мы включили в исследование не только данные по *P. simillimum*, но и по *F. gigantica*, *F. hepatica*, *S. mansoni*, *T. regenti* и *T. szidati*. Для проведения исследования мы использовали «сырые» библиотеки коротких парных прочтений для транскриптомов следующих фаз сложных жизненных циклов:

1. яйцо, мирацидий, редия, церкария, метацеркария, 42- и 70-дневная ювениль, марита *F. gigantica* (Zhang X.-X. et al., 2019),
2. яйцо, метацеркария, ювениль через 1, 3 и 24 часа после выхода из оболочки цисты, 21-дневная ювениль и марита *F. hepatica* (Cwiklinski et al., 2015; McNulty et al., 2017),
3. церкария, шистосомула через 3 и 24 часа после заражения хозяина, 7-недельные мариты разных полов *S. mansoni* (Protasio et al., 2012),
4. церкария и шистосомула *T. regenti* (Leontovyč et al., 2016),
5. церкария и шистосомула *T. szidati* (Leontovyč et al., 2019).

Дополнительная информация об использованных библиотеках представлена в Приложении 1. С целью исключения возможного влияния небиологических факторов на результаты определения и сравнения молекулярных подписей фаз, все «сырые» библиотеки прочтений *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* были подготовлены к дальнейшему анализу с помощью единого протокола. Составленный протокол включал поиск и удаление адаптерных и низкокачественных последовательностей в парах коротких прочтений, определение и исключение из библиотек возможной контаминации, а также квантификацию уровней экспрессии последовательностей.

Удаление низкокачественных и адаптерных последовательностей было выполнено с помощью программы fastP (Chen et al., 2018). Запуск fastP проводился со следующими параметрами: `--cut_right --cut_window_size 4 --cut_mean_quality 20 --qualified_quality_phred 20 --length_required 25`. Далее все библиотеки сравнивались с собранной нами базой данных, основу которой составляли «справочные» наборы последовательностей архей, бактерий, грибов, плазмид, протистов, векторов, вирусов и *H. sapiens*, подготовленные для Kraken2 (Wood, Salzberg, 2014; Wood et al., 2019). Принимая во внимание специфику работы с паразитическими организмами, в базу были также добавлены геномы их хозяев или близких к ним видов. Мы добавили в базу данных геномы 8 видов Gastropoda: *Aplysia californica*, *Biomphalaria glabrata*, *Chrysomallon squamiferum*, *Elysia chlorotica*, *Gigantopelta aegis*, *Haliotis discus hannai*, *Lottia gigantea*, *Pomacea canaliculate*. Все перечисленные геномы моллюсков получены из базы данных MolluscDB (Liu F. et al., 2021). Помимо этого, в локальную базу данных для деконтаминации мы включили геномы окончательных хозяев исследуемых видов дигеней: коровы *Bos taurus* (ARS-UCD1.2; GCA_002263795.2), курицы *Gallus gallus* (GRCg6a; GCA_000002315.5), утки *Anas platyrhynchos platyrhynchos* (CAU_duck1.0; GCA_002743455.1), мыши *Mus musculus* (GRCm39; GCA_000001635.9), и овцы *Ovis aries* (Oar_rambouillet_v1.0; GCA_002742125.1) из базы данных Ensembl (Yates et al., 2020). Сравнение библиотек парных прочтений с подготовленной базой данных было выполнено с помощью программы Kraken2. В дальнейший анализ мы включили только неклассифицированные пары коротких прочтений.

Мы использовали программу Salmon (v1.2) (Patro et al., 2017) для квантификации экспрессии белок-кодирующих генов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. Построение индексов с длиной k-мера равной 25 нуклеотидам выполнялось либо на последовательностях мРНК из базы данных WormBase Parasite (*F. gigantica*, *F. hepatica*, *S. mansoni*), либо на собранных контигах (*P. simillimum*, *T. regenti*, *T. szidati*). Выравнивание парных прочтений на индексы проводилось с использованием следующих параметров запуска Salmon: `-l A --seqBias --gcBias --minScoreFraction 0.50 --softclip --validateMappings`. Для каждого вида было подготовлено по две таблицы: одна таблица содержала уровни экспрессии в транскриптах-на-миллион для каждой биологической повторности в отдельности, другая - усреднённые между повторностями значения экспрессии в транскриптах-на-миллион.

Принимая во внимание, что данные по *F. hepatica* были получены в разных публикациях, мы использовали программу ComBat-seq (Zhang Y. et al., 2020) из библиотеки «sva» (v3.36.0) для языка программирования R с целью минимизации возможного влияния небиологических факторов на результаты квантификации экспрессии. Анализ проводился на количественных значениях выравненных прочтений (NumReads), полученных Salmon. Подправленные ComBat-seq значения мы преобразовали в транскрипты-на-миллион. Для разделения образцов

транскриптомов метацеркарий, полученных в разных исследованиях, мы переименовали их следующим образом: образцы «met0h», представленные в публикации Цвиклински с соавторами (Cwiklinski et al., 2015), и образцы «meta», проанализированные в исследовании Макналти с соавторами (McNulty et al., 2017) были переименованы в «раннюю» и «позднюю» метацеркарию, соответственно.

На основании результатов квантификации экспрессии генов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* для каждой фазы сложного жизненного цикла шести видов дигеней мы определили молекулярную подпись. В молекулярную подпись фазы жизненного цикла были включены только гены, имеющие уровень экспрессии ≥ 2 транскрипта-на-миллион на рассматриваемой фазе. Если ген имел уровень экспрессии ≥ 2 транскрипта-на-миллион на всех исследуемых фазах одного сложного жизненного цикла, мы классифицировали его, как «стабильно экспрессирующийся».

Анализ дифференциальной экспрессии белок-кодирующих генов

Определение статистически значимых различий в уровнях экспрессий генов выполнили с помощью библиотеки «RNentropy» (v1.2.2) (Zambelli et al., 2018) для языка программирования R. В отличие от многих широко используемых методов для дифференциальной экспрессии, выбранный подход может одновременно анализировать любое количество образцов и условий. RNentropy оценивает отличается ли и насколько экспрессия гена в любом количестве образцов от заданного фонового (например, равномерного) распределения, вычисляя вероятность получения того же значения, если наблюдаемая картина экспрессии на самом деле является результатом фонового распределения (Zambelli et al., 2018). В результате RNentropy способен обнаружить гены или транскрипты со значительными вариациями в экспрессии среди всех рассматриваемых условий и определить образцы, в которых последовательности демонстрируют значимое повышение или понижение экспрессии. В качестве порогового значения как для теста специфичности глобальной выборки, так и для теста локальной выборки мы использовали $P < 0.01$. Результаты теста специфичности глобальной выборки были скорректированы по методу Бенджамини-Хохберга (Benjamini-Hochberg). Анализ проводился для *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* на таблицах с не усреднёнными между биологическими повторностями уровнями экспрессии в транскриптах-на-миллион.

Определение кластеров ко-экспрессии

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) на наборах генов с уровнями экспрессии ≥ 2 транскриптов-на-миллион как минимум на одной из рассматриваемых фаз сложного жизненного цикла. Для анализа использовались таблицы с усредненными уровнями экспрессии в транскриптах-на-миллион.

Многомерное шкалирование молекулярных подписей фаз сложных жизненных циклов дигеней

Для проведения многомерного шкалирования (Multidimensional Scaling, MDS) мы подготовили матрицы присутствия (1) / отсутствия (0) по результатам анализа экспрессии генов. Матрицы имели следующую структуру: в названиях строк были указаны уникальные идентификаторы генов, в названиях столбцов – наименования образцов, а каждая ячейка матрицы содержала значение 0 или 1, в зависимости от того, активен ли анализируемый ген в рассматриваемом образце (1) или нет (0). 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. Строчки (гены), содержащие только одинаковые символы (инвариантные) были исключены из матрицы.

Межвидовой анализ проводился на результатах определения ОМА групп, одним из свойств которых является то, что все члены одной группы являются ортологами по отношению друг к другу. Названия строк матрицы содержали уникальные идентификаторы ОМА групп, а названия столбцов – наименования образцов. В отличие от внутривидового анализа были рассмотрены значения экспрессии, усредненные между биологическими повторностями, но, как и в первом анализе, мы использовали 2 транскрипта-на-миллион в качестве порогового значения экспрессии для разделения активных (1) и неактивных (0) генов. Для межвидового сравнения мы также составили дополнительную матрицу, содержащую информацию о том, имеет (1) или нет (0) ортолог статистически значимое повышение экспрессии на рассматриваемой фазе сложного жизненного цикла. Если вид не представлен в ОМА группе, в соответствующую ячейку матрицы было вписано обозначение «NA». Строчки (ортологи), содержащие только одинаковые символы (инвариантные) или как минимум одну «NA», были исключены из дальнейшего рассмотрения. Межвидовое сравнение проводилось в два этапа: на первом этапе «редииодные» (*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.

Оптимальное число кластеров во внутри- и межвидовом анализе мы определили с помощью метода «силуэт» («silhouette») из библиотеки «factoextra» (v1.0.7) для языка программирования R. Многомерное шкалирование было выполнено с помощью функции metaMDS из библиотек «vegan» (v2.5-7) и с использованием следующих значений параметров запуска: distance = «manhattan», try = 100, trymax = 1000, autotransform = FALSE, binary = TRUE, k = выявленное оптимальное число кластеров. Как при определении оптимального числа кластеров, так и при проведении многомерного шкалирования, мы использовали значение seed равное 1234. Визуализацию результатов выполнили с помощью функции ggscatter из библиотеки «ggpubr» (v0.4.0) для языка программирования R.

Анализ обогащения терминов генной онтологии

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 генной онтологии (Gene Ontology, GO). Анализ обогащения GO терминов рассматривает целый список генов («образец»), учитывая также и гены, не вошедшие в этот список («фон»). По результатам сравнения двух этих множеств генов возможно определить какие GO термины появляются в «образце» чаще, чем можно было бы ожидать при случайном совпадении, основываясь на частоте встречи терминов в «фоне».

Для каждой фазы жизненного цикла шести указанных видов дигеней в анализе рассматривались только наборы дифференциально экспрессирующихся генов. В качестве фона выступали все гены, не имеющие повышенную экспрессию на исследуемой фазе сложного жизненного цикла. Мы изучали только термины генной онтологии, описывающие биологические процессы. 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.

С целью снизить избыточность списка биологических процессов, мы использовали иерархичность базы данных генной онтологии для определения «родительских» процессов. Термин «родительский» биологический процесс в данном случае означает более общий термин

генной онтологии, тогда как его «дети» описывают более специфичные процессы. Уменьшение избыточности мы выполнили с помощью библиотеки «grvgo» (v1.0.2) для языка программирования R. Анализ был проведен на $-\log_{10}$ -трансформированных P-значениях, а в качестве порогового значения для функции `reduceSimMatrix` использовали значение 0.7.

Определение и анализ наборов потенциальных секреторных белков

Определение потенциальных секреторных белков мы выполнили *in silico* в соответствии с протоколом, описанным в публикации Гарга и Ранганатана (Garg, Ranganathan, 2011). Анализ был проведен для 11 видов дигеней: *C. sinensis*, *F. gigantica*, *F. hepatica*, *O. felineus*, *O. viverrini*, *P. simillimum*, *S. haematobium*, *S. japonicum*, *S. mansoni*, *T. regenti* и *T. szidati*. На первом этапе анализа мы проанализировали все белки с помощью программы SignalP (v5.0b) (Almagro Armenteros et al., 2019). На основании полученных результатов, белки были разделены на 2 группы: потенциальные «классические» (значение SP ≥ 0.5) и «неклассические» (SP < 0.5) секреторные белки. Ключевым отличием между последовательностями «классических» и «неклассических» секреторных белков является наличие у первых специального N-концевого сигнального пептида.

Мы использовали программу SecretomeP (v1.0) (Bendtsen et al., 2004) на втором этапе анализа для исследования последовательностей «неклассических» секреторных белков. Среди потенциальных «неклассических» секреторных белков мы отобрали те, которые, согласно результатам SecretomeP, имели NN-счет ≥ 0.9 и не содержали сигнальных пептиды. На третьем этапе анализа все потенциальные секреторные белки были проверены с помощью TargetP (v2.0) (Emanuelsson et al., 2000) на наличие митохондриального транзитного пептида. Мы исключили из рассмотрения белки, содержащие такой сигнал в последовательности. Для поиска трансмембранных доменов в исследуемых белках использовали ТМНММ (v2.0c) (Krogh et al., 2001). В качестве потенциальных секреторных белков мы рассматривали только белки, не имеющие никаких трансмембранных доменов.

Далее для каждого из 11 исследуемых видов дигеней мы провели анализ обогащения терминов генной онтологии генами потенциальных «классических» и «неклассических» секреторных белков. Анализ выполнили с помощью библиотеки «topGO» для языка программирования R. В качестве фона использовали все гены, которые не были отнесены к кодирующим потенциальные секреторные белки. Мы рассматривали только GO-термины, описывающие биологические процессы. 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 выполнили с помощью библиотеки «rrvgo» (v1.0.2) для языка программирования R. Анализ был проведен на $-\log_{10}$ -трансформированных P-значениях, а в качестве порогового значения для функции reduceSimMatrix использовали значение 0.7.

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

Реконструкция моделей геномов последних общих предков Platyhelminthes и Digenea, как и сравнительный анализ между ними были проведены с помощью библиотеки ruNAM (v1.1.10) (Train et al., 2019) для языка программирования Python. В качестве входных данных для анализа мы использовали построенное ранее укорененное филогенетическое дерево всех 14 анализируемых видов плоских червей, а также таблицу HierarchicalGroups.orthoxml, содержащую информацию о иерархических ортогруппах, полученные по результатам работы OMA standalone.

Принимая во внимание возможное отсутствие некоторых последовательностей в анализируемых наборах генов, мы исключили из моделей геномов те иерархические ортогруппы, которые включали менее 75% рассматриваемых видов плоских червей. Итоговая модель «Platyhelminthes» соответствовала геному предполагаемого последнего общего предка трех анализируемых видов свободноживущих плоских червей и 11 видов дигеней, тогда как модель «Digenea» описывала геном последнего общего предка 11 исследуемых видов дигеней. В сравнительном анализе реконструированных моделей ген классифицировался как «дублированный» у последнего общего предка дигеней, если его последовательность присутствовала в модели последнего общего предка дигеней минимум в 2 копиях, а в модели последнего общего предка плоских червей в количестве одной копии. Мы использовали «справочную» сборку генома *S. mansoni* (PRJEA36577; Smansoni_v7, GCA_000237925.3) из базы данных WormBase Parasite для извлечения наборов генов в соответствии с реконструированными моделями геномов. Программный код, использованный в исследовании, был написан на языке программирования Python (v3.6) (Nesterenko, 2022).

Анализ обогащения терминов генной онтологии выполнили для списков генов, появившихся и дублированных у последнего общего предка дигеней, согласно результатам сравнения построенных моделей геномов последних общих предков Platyhelminthes и Digenea.

Для проведения анализа мы использовали библиотеку «topGO» для языка программирования R. В качестве фона были рассмотрены гены *S. mansoni*, включенные в модель последнего общего предка Digenea, но не классифицированные как «появившиеся» или «дублированные». Мы рассматривали только GO-термины, описывающие биологические процессы. 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 списков обогащенных терминов генной онтологии и визуализацию результатов выполнили с помощью библиотеки «rtvgo» для языка программирования R.

Филостратиграфия и анализ филостратиграфического состава различных наборов белок-кодирующих генов

Филостратиграфический анализ наборов белков исследуемых видов плоских червей мы провели с помощью библиотеки «phylostratr» (v0.2.1) (Arendsee et al., 2019) для языка программирования R. Анализ был выполнен для 11 видов дигеней (*C. sinensis*, *F. hepatica*, *F. gigantica*, *O. viverrini*, *O. felinus*, *P. simillimum*, *S. mansoni*, *S. haematobium*, *S. japonicum*, *T. regenti* и *T. szidati*) и 3 (*S. mediterranea*, *M. lignano*, *P. vittatus*) видов свободноживущих плоских червей. На первом этапе анализа для каждого вида плоских червей было автоматически построено филогенетическое дерево, ведущее от прокариот до рассматриваемого вида. На филогенетическое дерево были также добавлены рекомендуемые разработчиками программы виды прокариот (функция «use_recommended_prokaryotes»), дрожжи *Saccharomyces cerevisiae* (функция «add_taxa(4932)»), а также *H. sapiens* (функция «add_taxa(9606)»). Далее для каждого вида на филогенетическом дереве был загружен его набор белков из базы данных UniProt. Для тех видов плоских червей, для которых нет данных в базе UniProt или доступны более актуальные данные в базе WormBase Parasite, наборы белков были добавлены или обновлены нами вручную. На втором этапе анализа мы использовали алгоритм BLASTp (v2.6.0+) (Camacho et al., 2009) для поиска совпадений между всеми подготовленными данными. Полученные таблицы с результатами сравнения в формате «б» мы использовали в качестве входных данных для phylostratr на третьем этапе анализа. На основании выявленного сходства между аминокислотными последовательностями программа разделила белки исследуемых видов плоских червей по следующим неперекрывающимся группам генов с разным филогенетическим происхождением, «филостратам»: 1) «Cellular organisms», 2) «Eukaryota», 3) «Opisthokonta», 4)

«Metazoa», 5) «Eumetazoa», 6) «Bilateria», 7) «Protostomia», 8) «Spiralia», 9) «Lophotrochozoa», 10) «Platyhelminthes», 11) «Class» («Rhabditophora» / «Digenea»), 12) «Order» («Plagiorchiida»), 13) «Family» («Echinostomatoidea» / «Opisthorchiidae» / «Schistosomatidae»), 14) «Genus» («Fasciola» / «Opisthorchis» / «Schistosoma» / «Trichobilharzia»), 15) «Species». Визуализацию результатов анализа выполнили в RStudio с помощью библиотек «ggplot2», «viridis» и «reshape» для языка программирования R.

Далее мы использовали результаты филогенетического анализа для определения композиции следующих наборов генов:

1. генов *S. mansoni*, включенных в модели геномов последних общих предков Platyhelminthes и Digenea,
2. генов, обладающим уровнем экспрессии ≥ 2 транскриптам-на-миллион на всех рассмотренных фазах сложного жизненного цикла («стабильно экспрессирующиеся»),
3. генов, включенных в предполагаемые кластеры ко-экспрессии,
4. генов, кодирующих потенциальные «классические» и «неклассические» секреторные белки.

Визуализацию результатов анализа мы выполнили в RStudio с помощью библиотек «ggplot2», «viridis» и «reshape» для языка программирования R.

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

Определение возрастных индексов транскриптомов (Transcriptome Age Index, TAI) провели для *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* на основе результатов филогенетики и квантификации экспрессии генов. Анализ был выполнен с помощью библиотеки «myTAI» (v0.9.3) (Drost et al., 2018) для языка программирования R на таблицах с усредненными между биологическими повторностями уровнями экспрессии в транскриптах-на-миллион. Мы исключили из анализа гены с уровнями экспрессии < 2 транскрипта-на-миллион на всех исследуемых фазах сложного жизненного цикла. Вычисленные значения уровней экспрессии оставшихся генов были преобразованы с помощью функции $\log_2(x + 1)$, где x – уровень экспрессии гена в транскриптах-на-миллион. Для расчета статистической значимости отклонения общего профиля TAI от плоской линии мы использовали функцию FlatLineTest из myTAI. Для анализа относительных экспрессий все филогенетические ветви были поделены на две группы: «до» (1–11 филогенетических ветвей) и «после» (от 12 до видоспецифичной филогенетической ветви).

расхождения дигеней. Наибольший уровень экспрессии был взят за «1», наименьший – за «0», а все остальные значения экспрессии были преобразованы в этом диапазоне. Для *P. simillimum* весь анализ провели дважды: сначала с рассмотрением всех генов, а затем без генов, отнесенных к видоспецифичной филострате.

Для определения вклада каждого гена в возрастные индексы молекулярных подписей фаз сложных жизненных циклов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* мы использовали функцию `rMatrix` из `myTAI`. Далее для фаз с наибольшим и наименьшим TAI в пределах одного жизненного цикла мы отобрали 500 генов, имеющих наибольший вклад в возрастные индексы молекулярных подписей фаз. Рассматривались только вклады генов, обладающих функциональной аннотацией по базе данных генной онтологии. Для полученных наборов генов был проведен анализ обогащения терминов генной онтологии с использованием библиотеки «`topGO`» для языка программирования R. Мы рассматривали только GO-термины, описывающие биологические процессы. Расчет статистической значимости проводился с помощью точного теста Фишера, а затем среди полученных результатов (GO-терминов с P-значениями <0.01) были отобраны только те термины, которые включали не менее 10 значимых генов. В качестве фона при анализе каждой фазы жизненного цикла мы использовали все гены из её молекулярной подписи за исключением отобранных 500 последовательностей. Уменьшения избыточности списков обогащенных терминов генной онтологии были выполнены с помощью библиотеки «`gtvgo`» для языка программирования R.

Глава 3 Результаты

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

По результатам секвенирования подготовленных образцов транскриптомов для рений, церкарий и марит *P. simillimum* и *S. pseudoglobulus* получено по две библиотеки, каждая из которых включала не менее 23,3 миллионов коротких парных прочтений. Потенциальная контаминация материалом моллюска *B. siamensis goniomphalos*, цыпленка *G. gallus* или человека *H. sapiens* не превышала 5,4% от общего числа прочтений в каждой из библиотек. После удаления адаптерных и низкокачественных последовательностей библиотеки содержали не менее 93% от количества коротких парных прочтений в деконтаминированных библиотеках. Обобщение результатов подготовки библиотек пар коротких прочтений для *P. simillimum* и *S. pseudoglobulus* представлено в таблице 1.

Таблица 1: Обобщение результатов подготовки библиотек коротких парных прочтений. Обозначения: Кол-во пар «сырых» прочтений – исходное количество пар коротких прочтений в библиотеках, полученных после секвенирования образцов транскриптомов; *Homo sapiens* (%), *Bithynia siamensis* (%), *Gallus gallus* (%) – процент пар коротких прочтений, отнесенных к потенциальной контаминации материалом человека, моллюска и цыпленка, соответственно; Кол-во прочтений в деконтам. библиотеках – количество пар коротких прочтений в библиотеках после удаления потенциальной контаминации (деконтаминированные библиотеки); Кол-во прочтений требуемого качества – количество пар коротких прочтений в библиотеке, соответствующих заданным критериям качества (в скобках приведен процент от количества пар прочтений в деконтаминированных библиотеках); Цифрами указаны данные для первой и второй биологических повторностей; X – анализ не проводился.

<i>Psilotrema simillimum</i>						
Метрики\Образцы	Редия (1)	Редия (2)	Церкария (1)	Церкария (2)	Марита (1)	Марита (2)
Кол-во пар «сырых» прочтений	29876923	30629840	30613174	24004011	23397182	23760802
<i>Homo sapiens</i> (%)	0,91	0,88	0,93	1,47	0,65	0,71
<i>Bithynia siamensis</i> (%)	4,43	5,40	X	X	X	X
<i>Gallus gallus</i> (%)	X	X	X	X	0,99	0,29
Кол-во прочтений в деконтам. Библиотеках	28281391	28703356	30327891	23651019	23015763	23524547
Кол-во прочтений требуемого качества	26879426 (95,04%)	26722840 (93,10%)	28762520 (94,84%)	22467004 (94,99%)	21885128 (95,09%)	22296310 (94,78%)
<i>Sphaeridiotrema pseudoglobulus</i>						
Метрики\Образцы	Редия (1)	Редия (2)	Церкария (1)	Церкария (2)	Марита (1)	Марита (2)
Кол-во пар «сырых» прочтений	28192806	24510459	32082446	24737749	29331490	25717702
<i>Homo sapiens</i> (%)	0,49	0,51	0,85	0,85	0,61	0,41
<i>Bithynia siamensis</i> (%)	4,36	4,28	X	X	X	X
<i>Gallus gallus</i> (%)	X	X	X	X	0,70	0,22
Кол-во прочтений в деконтам. Библиотеках	26827138	23337565	31809701	24526828	28949753	25556505
Кол-во прочтений требуемого качества	25519328 (95,13%)	22008828 (94,31%)	30161988 (94,82%)	23173379 (94,48%)	27274987 (94,21%)	24195574 (94,67%)

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. После кластеризации схожих последовательностей транскриптомы содержали 188588 (*P. simillimum*) и 211698 (*S. pseudoglobulus*) контигов.

Длины наибольших из собранных протяженных последовательностей у *P. simillimum* и *S. pseudoglobulus* составляли 37339 и 45140 нуклеотидов, соответственно. У обоих исследованных видов Psilostomatidae более 95% контигов характеризовались высокими показателями качества и полноты сборки: 180786 (*P. simillimum*) и 201608 (*S. pseudoglobulus*) контигов.

3.3 Квантификация уровней экспрессии и определение кодируемых аминокислотных последовательностей

При квантификации экспрессии собранных последовательностей с помощью программы Salmon количество успешно картированных пар коротких прочтений в каждой из библиотек варьировало от 78 до 90 процентов. Полученные значения в транскриптах-на-миллион были преобразованы в вероятные уровни экспрессии генов с использованием данных, полученных в ходе работы программы-сборщика Trinity. 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*) последовательностей.

Для определения кодируемых аминокислотных последовательностей собранными контигами мы провели трехэтапный анализ с помощью TransDecoder. Согласно полученным результатам, 32986 (*P. simillimum*) и 69006 (*S. pseudoglobulus*) протяженных последовательностей кодировали белки длиной равной или большей 100 аминокислот.

3.4 Подготовка «справочных» наборов последовательностей и их аннотация

В дальнейшем анализе мы приняли решение сфокусировать внимание только на белок-кодирующих генах с заметными уровнями экспрессии, поскольку белки и взаимодействия между

ними являются основой всех биологических процессов, протекающих в живых организмах. Основываясь на результатах определения кодируемых белков и квантификации экспрессий генов, мы подготовили «справочные» наборы последовательностей для двух исследованных видов псилостоматид. В наборы вошли только гены, кодирующие белки длиной ≥ 100 аминокислот и имеющие уровни экспрессии ≥ 1 транскрипта-на-миллион минимум на 1 фазе жизненного цикла. После применения указанных фильтров наборы включали 21433 и 46424 гена для *P. simillimum* и *S. pseudoglobulus*, соответственно. В качестве репрезентативных последовательностей для каждого гена были выбраны самый длинный белок и кодирующий его контиг.

На рисунке 3 представлены результаты сравнения наборов репрезентативных белков двух исследованных Psilostomatidae с базой данных ортологов, присутствующих в числе одной копии в геномах Metazoa. Для сравнения на рисунке приведены также аналогичные данные для 10 видов дигеней, взятые из базы данных UniProt или предоставленные нам по запросу (*F. gigantea*, *T. regenti* и *T. szidati*). Как у *P. simillimum*, так и у *S. pseudoglobulus* более 83,5% однокопийных ортологов собрано полностью в подготовленных «справочных» наборах. На рисунке 3 хорошо видно, что полнота наборов генов двух видов Psilostomatidae сопоставима с результатами анализа доступных наборов белков других 10 видов дигеней. Сравнение полученных нами результатов с другими видами дигеней показало, что 57 генов из использованной базы данных однокопийных ортологов Metazoa отсутствуют у всех 12 рассмотренных видов паразитических плоских червей.

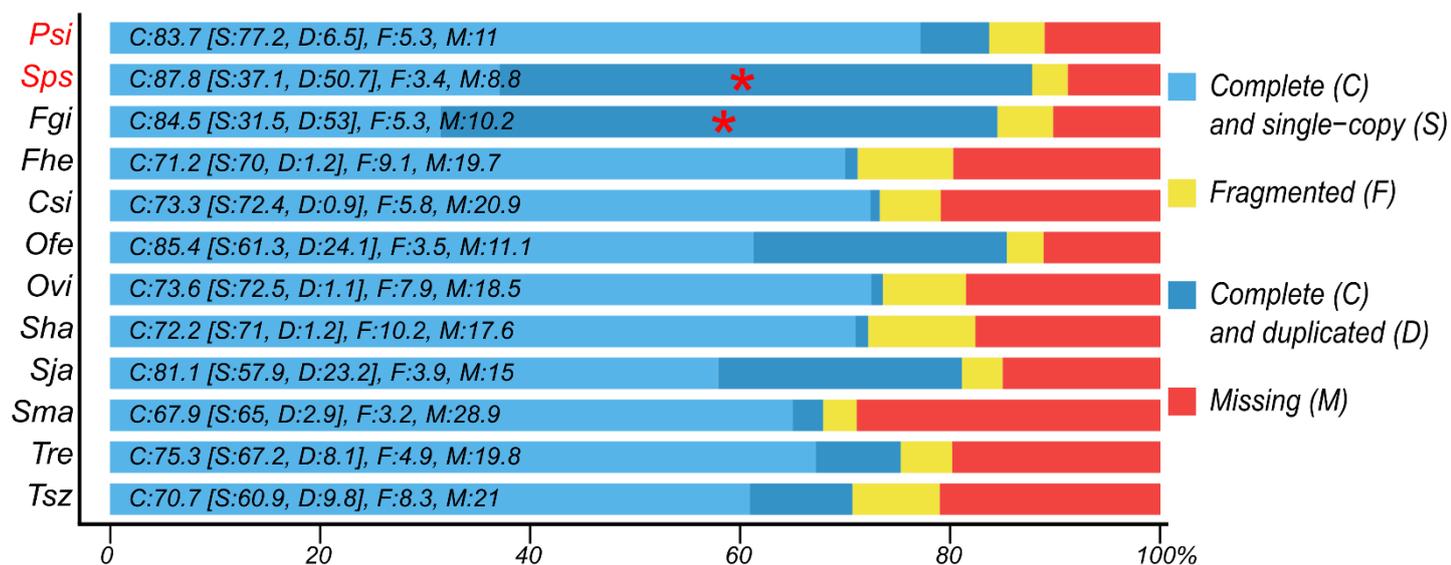


Рисунок 3 – Результаты анализа наличия последовательностей однокопийных ортологов Metazoa в наборах генов паразитических плоских червей. Обозначения: Complete (C) and single-copy (S) – количество последовательностей однокопийных ортологов, собранных полностью и присутствующих в числе одной копии в анализируемом 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 – *Sphaeridiotrema 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*.

Для функциональной аннотации репрезентативных последовательностей было проведено их сравнение как с нуклеотидной (NCBI nt) и аминокислотными (NCBI nr и SwissProt) базами данных, так и с базами данных ортологов (eggNOG) и профилей белковых доменов (PfamA). Согласно полученным результатам, большинство репрезентативных последовательностей имело совпадение как минимум с одной из баз данных: 74% (15865 генов) и 82,5% (38316 генов) у *P. simillimum* и *S. pseudoglobulus*, соответственно.

Принимая во внимание литературные данные о нарушенном биосинтезе жирных кислот у дигеней (Wang X. et al., 2011), мы провели поиск транскриптов ферментов, принимающих участие в указанном процессе, в собранных транскриптомах *P. simillimum* и *S. pseudoglobulus*. У обоих видов обнаружены совпадения с последовательностями 7 ферментов: ацетил-КоА-карбокситрансфераза (EC 2.1.3.15), β -кетоацил-АЦП-синтаза FabF (EC 2.3.1.179), малонил-трансфераза FabD (EC 2.3.1.39), синтаза жирных кислот FasN (EC 2.3.1.85), ацил-КоА-синтетазы (EC 6.2.1.3), биотин карбоксилаза (EC 6.3.4.14), ацетил-КоА-карбоксилаза (EC 6.4.1.2). Белки *P. simillimum* и *S. pseudoglobulus*, имеющие совпадения с синтазой жирных кислот FasN, отличались по доменной архитектуре: в белке *P. simillimum* найдена одна копия ацил-трансферазного домена (Acyl_transf_1), а белок *S. pseudoglobulus* содержал один домен дегидрогеназы с короткой цепью (adh_short). У *P. simillimum* обнаружено 11 белков, схожих с ацил-КоА-синтетазой, тогда как у *S. pseudoglobulus* выявлено 25 таких белков.

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

При подготовке «справочных» наборов последовательностей обнаружено почти двукратное различие в их размерах между *S. pseudoglobulus* и *P. simillimum*. В то же время сравнение с базой данных однокопийных ортологов Metazoa доказало, что примерно половина (50,7%) таких белков в «справочном» наборе *S. pseudoglobulus* имеют дубликаты.

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 GenBank. На рисунке 4 представлено филогенетическое дерево, построенное для последовательностей митохондриальной *cox1*. На основании полученных результатов сделано предположение, что образцы *S. pseudoglobulus* могут содержать материал не одного вида, а смеси криптических видов.

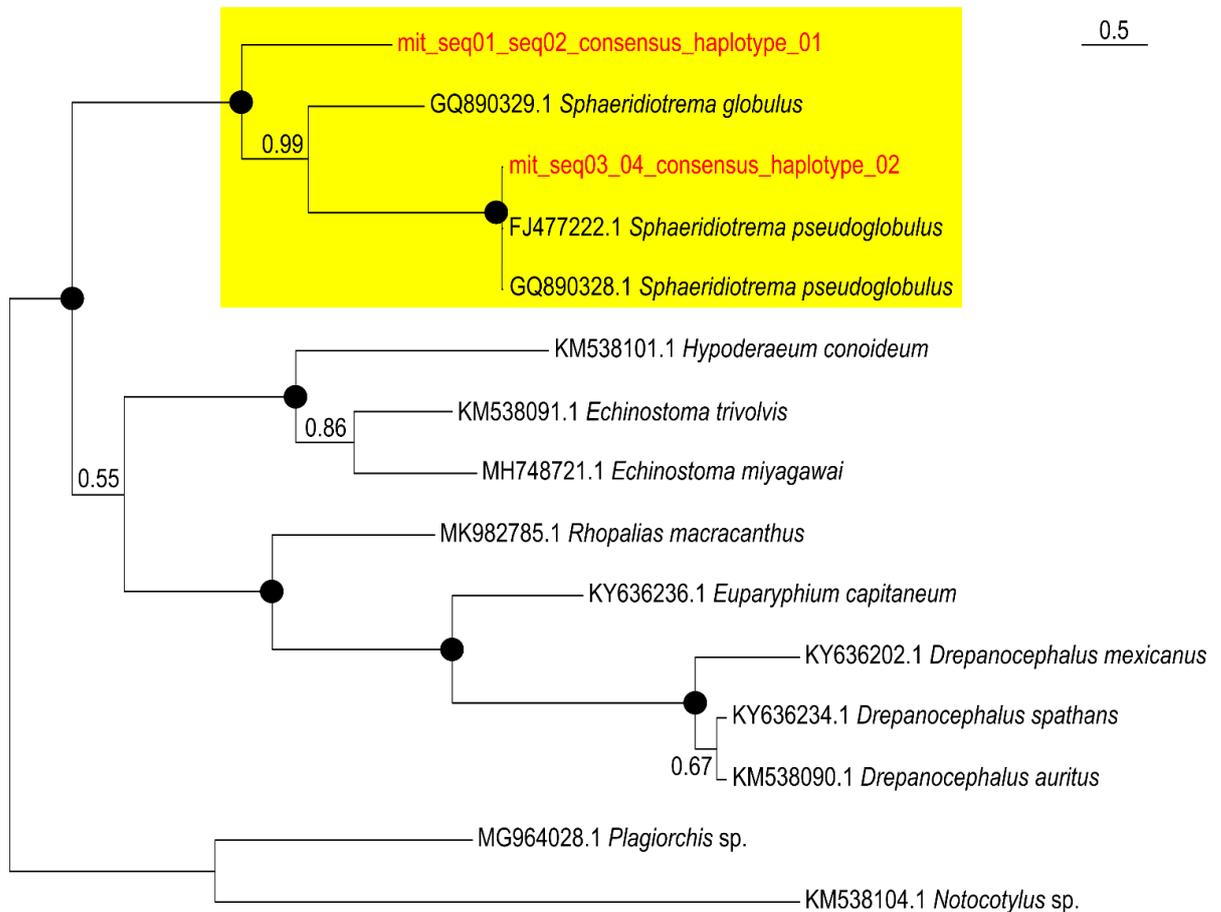


Рисунок 5 – Тепловая карта, построенная на основе результатов попарного сравнения списков ортогрупп, идентифицированных программой OrthoFinder. Справа представлена цветовая шкала, отображающая количество общих ортогрупп между парами видов: чем больше количество общих ортогрупп, тем теплее цвет в ячейке, находящейся на пересечении строки и столбца, соответствующих сравниваемым видам. В верхней части карты представлена кладограмма, построенная по результатам кластеризации видов по количеству общих ортогрупп. Обозначения: Csi – *Clonorchis sinensis*, Fgi – *Fasciola gigantica*, Fhe – *Fasciola hepatica*, Mli – *Macrostomum lignano*, Ofe – *Opisthorchis felineus*, Ovi – *Opisthorchis viverrini*, Psi – *Psilotrema simillimum*, Sha – *Schistosoma haematobium*, Sja – *Schistosoma japonicum*, Sma – *Schistosoma mansoni*, Sme – *Schmidtea mediterranea*, Sps – *Sphaeridiotrema pseudoglobulus*, Tre – *Trichobilharzia regenti*, Tsz – *Trichobilharzia szidati*.

На тепловой карте видна совместная кластеризация *P. simillimum* и *S. pseudoglobulus* с филогенетически близкими видами *F. gigantica* и *F. hepatica*. Для двух исследованных видов Psilostomatidae обнаружено более 10000 общих ортогрупп и приблизительно 7000–8000 общих ортогрупп выявлено для псилостоматид и фасциOLID. Помимо кластера Psilostomatidae и Fasciolidae были также определены две группы видов, соответствующих Opisthorchiidae (*C. sinensis*, *O. felineus* и *O. viverrini*) и Schistosomatidae (*S. mansoni*, *S. japonicum*, *S. haematobium*, *T. regenti* и *T. szidati*). Рассмотренные виды свободноживущих плоских червей сформировали кластер, отдельный от паразитических видов.

Принимая во внимание, что все рассмотренные виды плоских червей различаются между собой по множеству признаков, в том числе по типам и характеру морфологических и физиологических адаптаций, особенностям жизненных циклов, мы провели дополнительный анализ для определения количества общих и специфичных ортогрупп. С этой целью мы условно разделили все виды на группы в зависимости от того, является ли вид свободноживущим или паразитическим и какой фазой, редией или дочерней спороцистой, представлены дочерние партеногенетические поколения. Оба исследуемых вида Psilostomatidae были рассмотрены отдельно не только от «редиидных» видов, но и от друг друга для того, чтобы определить сходства и различия между ними. На рисунке 6 представлена диаграмма Венна, построенная для множеств ортогрупп, общих для всех видов, объединенных в одну группу.

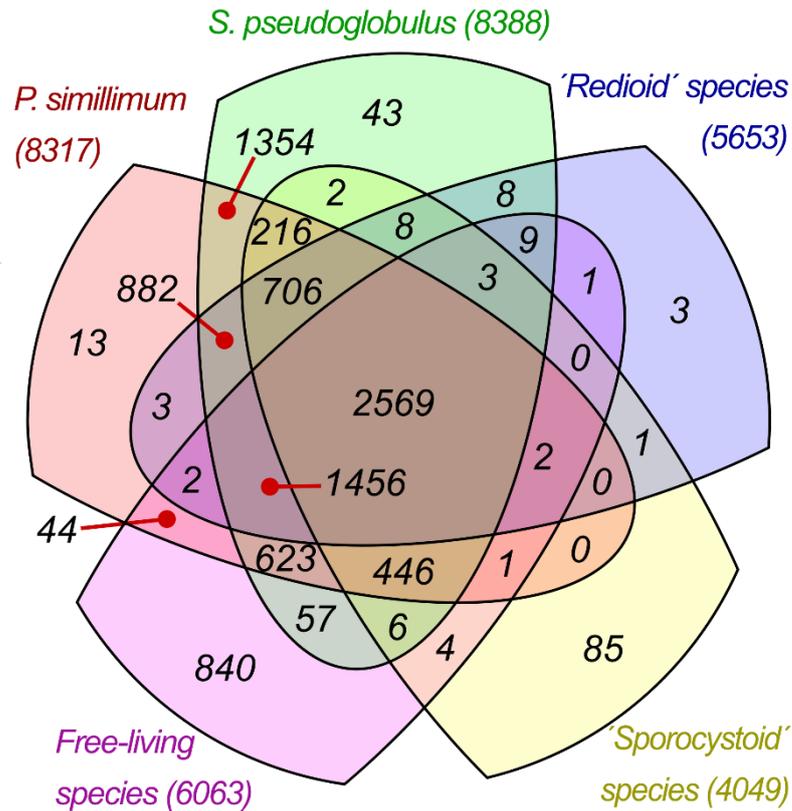


Рисунок 6 – Диаграмма Венна, построенная для множеств ортогрупп. Были рассмотрены только ортогруппы, в которые вошли белки всех видов, объединенных в одну группу. Числовые значения отражают количество общих (пересечения) или специфичных (участки без пересечений) ортогрупп. Обозначения: 'Redioid' species – множество ортогрупп, общих для видов дигеней с редией в сложном жизненном цикле (*C. sinensis*, *F. gigantica*, *F. hepatica*, *O. viverrini*, *O. felinus*); 'Sporocystoid' species – множество ортогрупп, общих для видов дигеней с дочерней спороцистой в сложном жизненном цикле (*S. mansoni*, *S. haematobium*, *S. japonicum*, *T. regenti* и *T. szidati*); Free-living species – множество ортогрупп, общих для *M. lignano* и *S. mediterranea*.

На диаграмме явно видно, что последовательности всех исследуемых видов плоских червей вошли в 2569 ортогрупп. Проведенный сравнительный анализ показал, что в 840 ортогруппах представлены белки только двух свободноживущих видов, тогда как 706 ортогрупп включали исключительно белки дигеней. Общими для всех рассмотренных «редииоидных» видов оказались 882 группы ортологов, тогда как 85 ортогрупп являлись специфичными для «спороцистоидных» видов. Уникальными для двух исследуемых видов Psilostomatidae оказались 1354 ортогруппы. Среди них только 13 и 43 являлись специфичными для *P. simillimum* и *S. pseudoglobulus*, соответственно. У первого вида эти ортогруппы включали 34 белка, такие, как белок, содержащий повтор F-box/WD (F-box/WD repeat-containing protein 10), и рецептор для нетринов (Netrin receptor UNC5B). В то время как специфичные для *S. pseudoglobulus* ортогруппы содержали 435 белков, в числе которых кубилин-подобный белок (cubilin-like protein) и АТФ-

связывающий кассетный транспортный белок (ABC transporter). Большинство белков из уникальных для *P. simillimum* (18 белков, 52,9%) и *S. pseudoglobulus* (307 белков, 70,5%) ортогрупп не имели совпадений ни с одной из использованных для функциональной аннотации баз данных. Схожим образом, 50,1% белков из 1354 ортогрупп, специфичных только для двух исследуемых видов Psilostomatidae, не имели совпадений с базами данных NCBI nr, SwissProt и PfamA. Результаты аннотации генов, входящих в состав ортогрупп, специфичных для *P. simillimum* и/или *S. pseudoglobulus*, приведены в Приложении 8 к публикации Nesterenko et al., 2020.

Для уточнения результатов, полученных программой OrthoFinder, и построения иерархических ортогрупп, мы провели повторный поиск ортогрупп с использованием программы OMA standalone. На первом этапе мы обновили набор анализируемых белков *C. sinensis*, *F. gigantea*, *F. hepatica*, *M. lignano*, *O. felineus*, *O. viverrini*, *S. haematobium*, *S. japonicum*, *S. mansoni*, *S. mediterranea*, а также добавили данные по свободноживущему виду *P. vittatus*. Нами также было принято решение не включать в повторный анализ набор белков *S. pseudoglobulus* ввиду возможного влияния смеси криптических видов на результаты анализа. С помощью проведенной на втором этапе кластеризации схожих последовательностей у видов, для которых доступны только транскриптомы (*P. simillimum*, *P. vittatus*, *T. regenti* и *T. szidati*), мы успешно снизили количество изоформ, которые могли оказывать влияние на результаты построения ортогрупп (в скобках приведено количество белков до и после кластеризации): *P. simillimum* (21433 и 21257 белков), *P. vittatus* (89863 и 60763 белков), *T. regenti* (12705 и 11456 белков), *T. szidati* (13007 и 11718 белков).

Все подготовленные наборы белков были использованы на третьем этапе для построения ортогрупп с помощью OMA standalone. По результатам анализа обнаружено 36486 OMA групп ортологов. Среди них белки всех 14 рассмотренных видов плоских червей одновременно были представлены только в 175 ортогруппах, тогда как белки всех видов 11 дигеней одновременно присутствовали в 438 группах ортологов.

На четвертом этапе анализа мы отобрали 614 OMA ортогрупп, включавших белки как минимум 13 из 14 исследуемых видов плоских червей. После проведения множественного выравнивания аминокислотных последовательностей в каждой ортогруппе, объединения результатов и их фильтрации, итоговая суперматрица содержала 345931 позиций. Согласно результатам анализа, наиболее подходящей моделью эволюции исследуемых последовательностей является «LG + I + G + F». Топология реконструированного филогенетического дерева имеет полную поддержку на всех основных ветвях (Рисунок 7) и соответствует современным представлениям о филогенетических отношениях исследуемых видов (De León, Hernández-Mena, 2019).

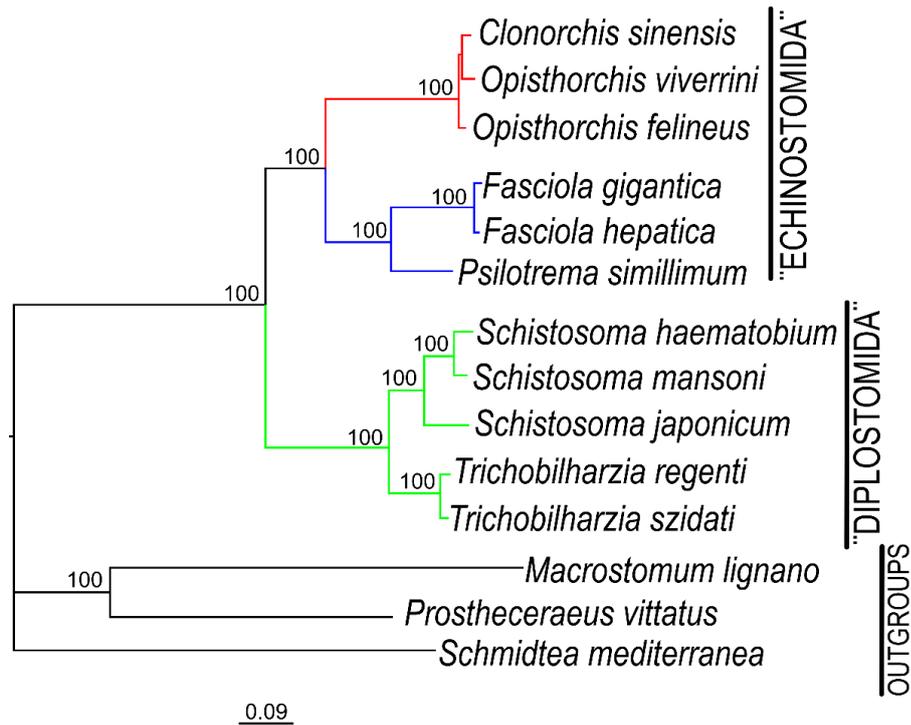


Рисунок 7 – Филогенетическое дерево плоских червей, построенное на основе результатов сравнения последовательностей из 614 ортогрупп. Разными цветами обозначены следующие таксоны: Opisthorchiidae (*C. sinensis*, *O. viverrini*, *O. felineus*), Echinostomatoidea (*F. gigantica*, *F. hepatica*, *P. simillimum*), Schistosomatidae (*S. haematobium*, *S. mansoni*, *S. japonicum*, *T. regenti*, *T. szidati*). Названия клад даны в соответствии с публикацией Одининга (Odening, 1974).

Построенное филогенетическое дерево было использовано на пятом этапе анализа для повторного запуска поиска ортологов с помощью OMA standalone. Согласно полученным результатам, обнаружено 36486 OMA групп ортологов и 23852 иерархические ортогруппы. На рисунке 8 представлена тепловая карта с количеством общих OMA групп для разных пар исследуемых видов плоских червей. На тепловой карте видна совместная кластеризация *P. simillimum* с филогенетически близкими фасциолидами. Помимо кластера, объединяющего *F. gigantica*, *F. hepatica* и *P. simillimum*, были также определены две группы видов, соответствующих Opisthorchiidae (*C. sinensis*, *O. felineus* и *O. viverrini*) и Schistosomatidae (*S. mansoni*, *S. japonicum*, *S. haematobium*, *T. regenti* и *T. szidati*) (Рисунок 8). Три рассмотренных вида свободноживущих плоских червей сформировали отдельный кластер.

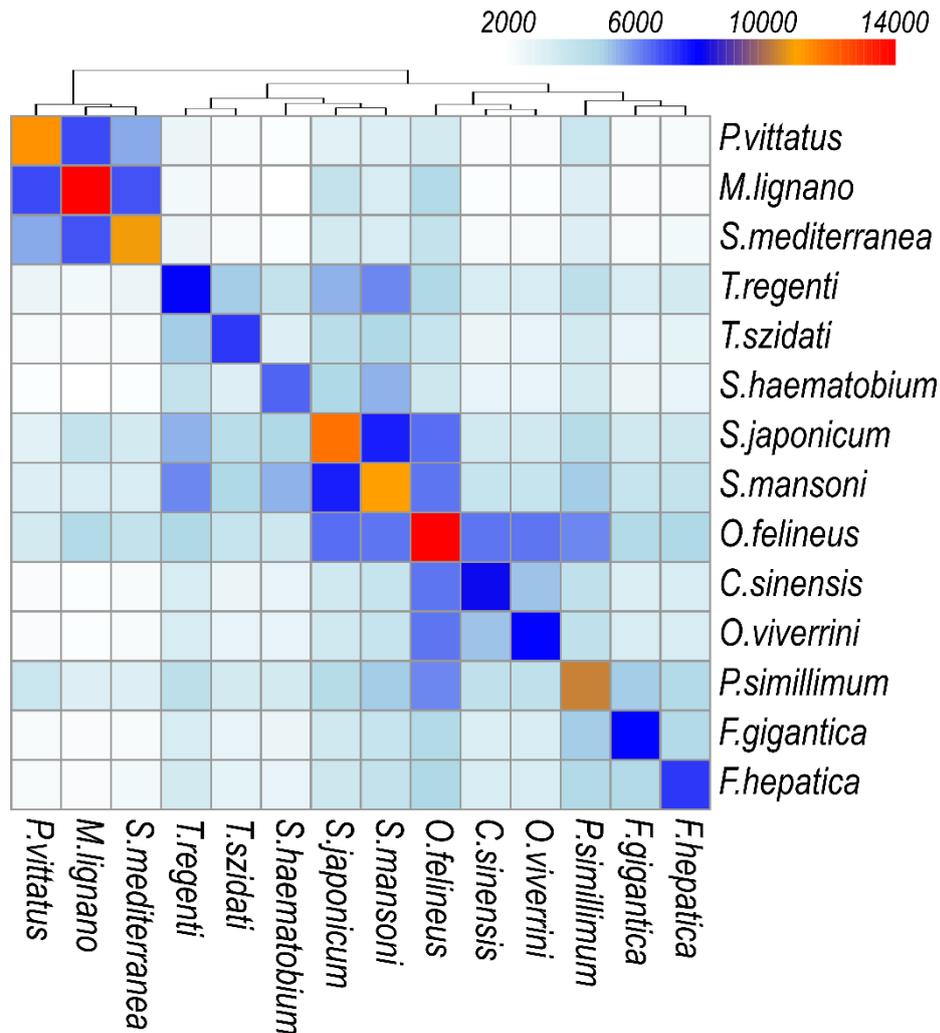


Рисунок 8 – Тепловая карта, построенная на основе результатов попарного сравнения списков ОМА групп. Цветовая шкала в верхнем правом углу карты отображает количество общих ортогрупп между парами видов: чем больше количество общих ортогрупп, тем теплее цвет в ячейке, находящейся на пересечении строки и столбца, соответствующих сравниваемым видам. В верхней части карты представлена кладограмма, построенная на результатах кластеризации видов по количеству общих ортогрупп.

3.7 Определение молекулярных подписей фаз сложных жизненных циклов дигеней

При определении молекулярной подписи фазы сложного жизненного цикла дигеней мы ставили перед собой цель наиболее полно охватить набор генов, активных на рассматриваемой фазе. Для реализации поставленной цели мы выбрали два пороговых значения экспрессии – 1 и 2 транскрипта-на-миллион. Первый порог был использован для первичного внутривидового анализа при рассмотрении жизненных циклов псилостоматид. На рисунке 9 представлены

диаграммы Венна для молекулярных подписей редий, церкарий и марит *P. simillimum* (A) и *S. pseudoglobulus* (B).

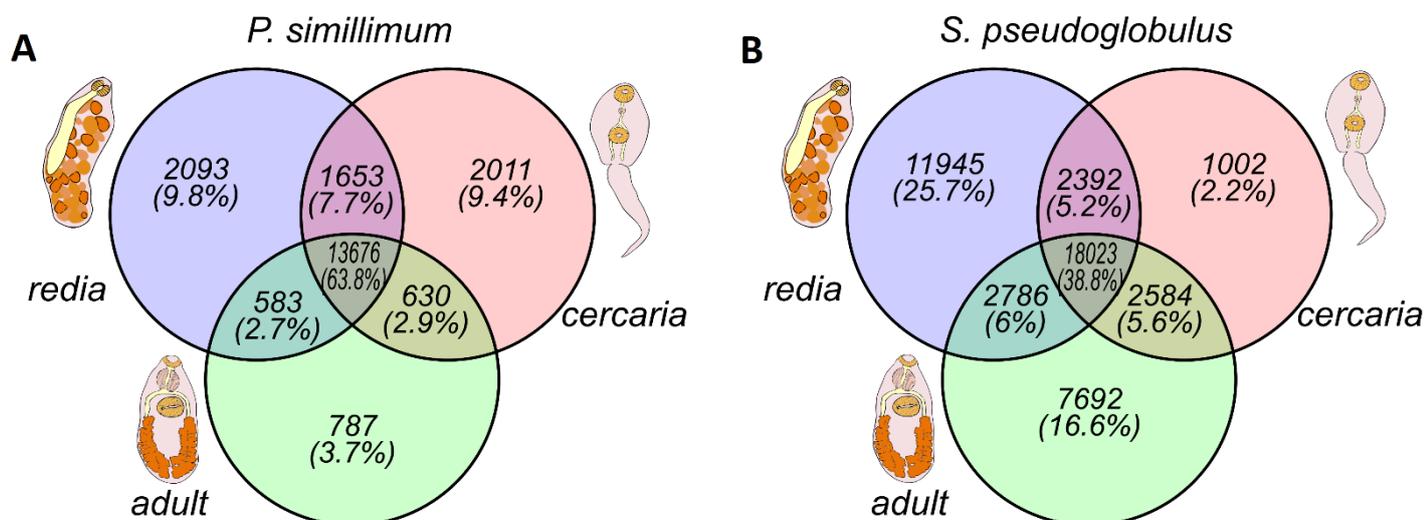


Рисунок 9 – Диаграммы Венна, построенные для молекулярных подписей фаз сложных жизненных циклов *Psilotrema simillimum* (A) и *Sphaeridiotrema pseudoglobulus* (B). Для включения гена в молекулярную подпись фазы ген должен был иметь уровень экспрессии ≥ 1 транскрипта-на-миллион на рассматриваемой фазе 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 9А) активна на всех фазах цикла, тогда как у *S. pseudoglobulus* количество таких генов не превышало 39% (18023 генов) (Рисунок 9В). На диаграммах Венна (Рисунок 9) неперекрывающимися участками окружностей представлены подмножества генов, активных только на одной фазе сложного жизненного цикла. Гены с таким профилем экспрессии мы условно классифицировали как «фаза-специфичные». У *P. simillimum* количество марита-специфичных генов (3,7%, 787 генов) меньше количества генов, специфичных для церкарии (9,4%, 2011 ген) и редии (9,8%, 2093 генов) (Рисунок 9А). Суммарное количество «фаза-специфичных» генов у *P. simillimum* составило почти 23 процента от общего числа идентифицированных генов. У *S. pseudoglobulus* к числу генов с таким профилем экспрессии отнесено 44,5% генов из «справочного» набора. Среди «фаза-специфичных» генов *S. pseudoglobulus* преобладали гены, характеризующиеся экспрессией только на рассмотренных паразитических фазах сложного жизненного цикла: 7692 и 11945 генов демонстрировали марита- и редия-специфичную экспрессию, соответственно.

Мы предположили, что наиболее заметные различия в экспрессии между рассматриваемыми фазами сложного жизненного цикла будут проявлять гены, принимающие участие в формировании специфичных признаков. Для проверки предположения мы провели визуализацию экспрессии гомеобокс-содержащих генов. Гомеодомен-содержащие белки являются факторами транскрипции, регулирующими экспрессию различных генов и играющими важную роль в эмбриональном развитии (Bürglin, Affolter, 2016).

На рисунках 10 представлены тепловые карты экспрессии гомеобоксных генов. Большинство гомеобокс-содержащих генов демонстрировали повышенную экспрессию на фазах редии и церкарии (Рисунок 10). Визуализация экспрессии позволила обнаружить высокую степень согласованности в профилях активности гомеобоксных генов между *P. simillimum* и *S. pseudoglobulus* (Рисунок 10).

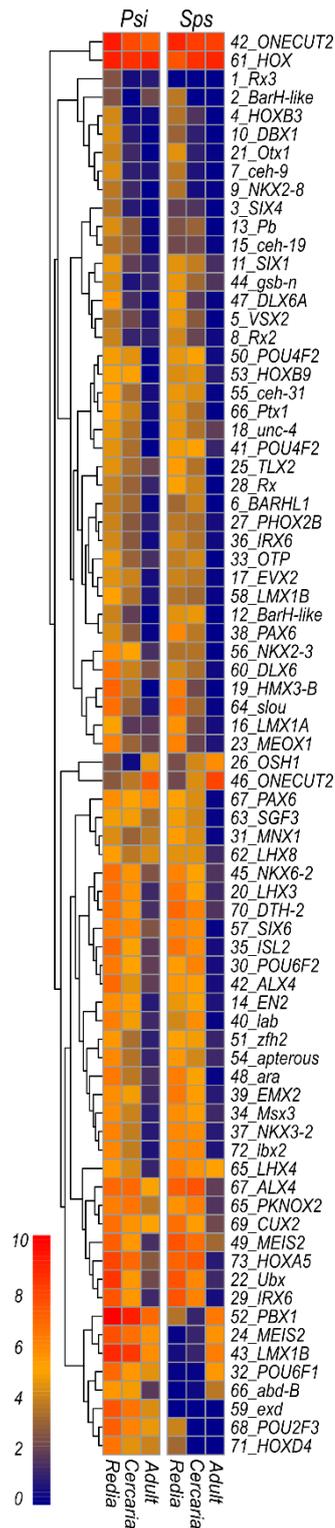


Рисунок 10 – Тепловая карта экспрессии гомеобокс-содержащих генов, обнаруженных у *P. simillimum* (Psi) и *S. pseudoglobulus* (Sps). Цветовая шкала соответствует преобразованным ($\log_2(\text{TPM}+1)$) уровням экспрессии: чем больше уровень экспрессии, тем теплее цвет. С левой стороны карты представлена кладограмма, построенная на результатах кластеризации гомеобоксных генов в соответствии с их профилями активности. Обозначения: Redia – редия, Cercaria – церкария, Adult – марита.

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, но включили данные по *F. gigantica*, *F. hepatica*, *S. mansoni*, *T. regenti* и *T. szidati*. Для проведения корректного сравнения между *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* все доступные данные были проанализированы с помощью единого протокола, включающего как поиск и удаление низкокачественных и адаптерных последовательностей в парах коротких прочтений, так и поиск возможной контаминации в библиотеках. В каждой из рассмотренных библиотек выставленным критериям качества соответствовало от 79,9% (третья биологическая повторность шистосомулы *T. szidati*) до 95,88% (вторая биологическая повторность редии *F. gigantica*) пар коротких прочтений. К возможной контаминации тканями хозяев отнесено от 1,59% (первая биологическая повторность мариты *F. hepatica*) до 30,17% (четвертая биологическая повторность церкарии *T. szidati*) пар коротких прочтений.

По результатам применения критерия в минимальную экспрессию в 2 транскрипта-на-миллион, каждая молекулярная подпись фазы жизненного цикла у всех шести проанализированных видов дигеней включала минимум 60% от общего числа рассмотренных белок-кодирующих генов исследованного вида. В каждом случае молекулярная подпись состояла из нескольких тысяч генов, а минимальные и максимальные размеры молекулярных подписей представлены в таблице 2. 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. Единственным исключением являлся вид *F. gigantica*, у которого доля генов с такой динамикой экспрессии составила 48,89%.

Таблица 2: Обобщение результатов анализа молекулярных подписей. Обозначения: Кол-во генов со стаб. экспр. – количество генов с уровнями экспрессии ≥ 2 транскрипта-на-миллион на всех рассмотренных фазах сложного жизненного цикла; Min / Max размер мол. подписи (фаза жиз. цикла) – минимальный и максимальный размеры молекулярной подписи в генах, соответственно, с указанием фазы жизненного цикла, для которой получено это значение; метацерк. – метацеркария, шистосомула, 3ч – шистосомула *S. mansoni* через 3 часа после заражения хозяина.

Метрики \ Виды	<i>F. gigantica</i>	<i>F. hepatica</i>	<i>P. simillimum</i>	<i>S. mansoni</i>	<i>T. regenti</i>	<i>T. szidati</i>
Кол-во генов со стаб. экспр.	5104	5698	8116	8212	6766	6837
Min размер мол. подписи (фаза жиз. цикла)	6349 (яйцо)	6421 (яйцо)	9938 (марита)	9480 (марита)	6948 (церкария)	6958 (церкария)
Max размер мол. подписи (фаза жиз. цикла)	9273 (метацерк.)	7999 (ювениль)	11481 (редия)	10353 (шистосомула, 3ч)	9197 (шистосомула)	9393 (шистосомула)

3.8 Анализ дифференциальной экспрессии белок-кодирующих генов

Определение статистически значимых различий в уровнях экспрессии белок-кодирующих генов было выполнено для *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. Общее количество генов с повышенной экспрессией, а также наименования фаз жизненных циклов, для которых выявлено минимальное и максимальное количество таких генов, представлены в таблице 3.

Таблица 3: Обобщение результатов анализа дифференциальной экспрессии генов. Обозначения: Общее кол-во генов с повышенной экспрессией – общее количество генов со статистически значимым повышением как минимум на одной из рассмотренных фаз сложного жизненного цикла; Min / Max кол-во генов с повышенной экспрессией (фаза жиз. цикла) – минимальное и максимальное количество генов со статистически значимым повышением экспрессии, соответственно, с указанием фазы жизненного цикла, для которой получено это значение; метацерк. – метацеркария; ювениль, 70д – 70-дневная ювениль *F. gigantica*.

Метрики \ Виды	<i>F. gigantica</i>	<i>F. hepatica</i>	<i>P. simillimum</i>	<i>S. mansoni</i>	<i>T. regenti</i>	<i>T. szidati</i>
Общее кол-во генов с повышенной экспрессией	7538	5757	7171	4539	6578	6104
Min кол-во генов с повышенной экспрессией (фаза жиз. цикла)	352 (ювениль, 70д)	446 (марита)	1137 (марита)	585 (церкария)	535 (церкария)	449 (церкария)
Max кол-во генов с повышенной экспрессией (фаза жиз. цикла)	3466 (метацерк.)	2648 (ювениль)	4415 (редия)	1944 (марита)	6043 (шистосомула)	5655 (шистосомула)

Более 1000 дифференциально экспрессирующихся генов обнаружено для 13 из 27 рассмотренных фаз сложного жизненного цикла: мирацидия (2551 ген), редии (1641), церкарии (1029), метацеркарии (3466) *F. gigantica*; ювенили *F. hepatica* через 24 часа после покидания цисты (1644 генов) и в конце миграции (2648); редии (4415 генов), церкарии (1760) и мариты (1137) *P. simillimum*; шистосомулы через 24 часа после заражения (1486 генов) и мариты (1944) *S. mansoni*; шистосомулы *T. regenti* (6043 гена) и *T. szidati* (5655). В большинстве случаев (18 из 27 образцов) доля генов со статистически значимым повышением экспрессии от размера

молекулярной подписи рассматриваемой фазы не превышала 20%. Наибольший вклад дифференциально экспрессирующихся генов в молекулярную подпись обнаружен у шистосомул *Trichobilharzia*: процент генов со статистически значимым повышением экспрессии составлял 60,2% у *T. szidati* и 65,71% у *T. regenti*. Более подробные результаты определения дифференциальной экспрессии генов приведены в Приложении 3 к публикации Nesterenko et al., 2022.

3.9 Определение кластеров ко-экспрессии

Для выявления групп генов, изменяющих свою экспрессию согласованно в ходе реализации сложного жизненного цикла дигеней, мы провели поиск ко-экспрессирующихся генов. Мы не рассматривали данные по *Trichobilharzia*, поскольку для *T. regenti* и *T. szidati* доступны только транскриптомы двух фаз жизненного цикла, церкарии и шистосомулы, и, как следствие, профили экспрессии обнаруженных кластеров, вероятнее всего, совпадали бы с результатами поиска дифференциальной экспрессии. Обобщение результатов поиска кластеров ко-экспрессирующихся генов для *F. gigantica*, *F. hepatica*, *P. simillimum* и *S. mansoni* представлены в таблице 4.

Таблица 4: Обобщение результатов поиска кластеров ко-экспрессирующихся генов. Обозначения: Кол-во кластеров – количество кластеров ко-экспрессирующихся генов; Общее кол-во генов в кластерах – общее количество генов, включенных в кластеры ко-экспрессии; Кол-во генов, не включенных в кластеры – количество генов, не включенных ни в один из кластеров ко-экспрессии; Min / Max / Средний размер кластера – минимальный, максимальный или средний размер кластеров ко-экспрессии (в генах), соответственно.

Метрики\Виды	<i>F. gigantica</i>	<i>F. hepatica</i>	<i>P. simillimum</i>	<i>S. mansoni</i>
Кол-во кластеров	16	9	12	16
Общее кол-во генов в кластерах	5125 (49,09%)	3201 (37,03%)	9789 (70,02%)	8039 (69,46%)
Общее кол-во генов, не включенных в кластеры	5314	5443	4192	3534
Min размер кластера	13	106	426	36
Max размер кластера	998	897	1898	1471
Средний размер кластера	320.3125	355.6667	815.75	502.4375

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

генов (Таблица 4). Средний размер кластера ко-экспрессирующихся генов у *P. simillimum* превышал значения, полученные для *F. gigantica*, *F. hepatica* и *S. mansoni* (Таблица 4).

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. Для внутривидового анализа мы подготовили матрицы присутствия/отсутствия, содержащие информацию о том, входит (1) или нет (0) рассматриваемый ген в молекулярную подпись конкретной фазы. После исключения строк (генов), содержащих только одинаковые символы (инвариантные), итоговые матрицы включали следующее количество строк: 5630 – *F. gigantica*, 3694 – *F. hepatica*, 7175 – *P. simillimum*, 4154 – *S. mansoni*, 3963 – *T. regenti*, 3638 – *T. szidati*.

На рисунке 11 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 11). У *F. gigantica* третий кластер включал повторности 42- и 70-дневных ювенилей, а также мариты, тогда как четвертый кластер содержал повторности церкарий и метацеркарий (Рисунок 11А). Первый кластер *F. hepatica* объединял повторности ювенилей и марит, тогда как второй кластер включал как ранних и поздних метацеркарий, так и ювенилей, недавно покинувших цисты (Рисунок 11В). Редии и церкарии *P. simillimum* также были объединены в единый кластер (Рисунок 11С). Первый и второй кластер *S. mansoni* имели схожий состав и включали одну повторность шистосомулы через 3 часа после заражения и одну (второй кластер) или две (первый кластер) повторности церкарий (Рисунок 11D).

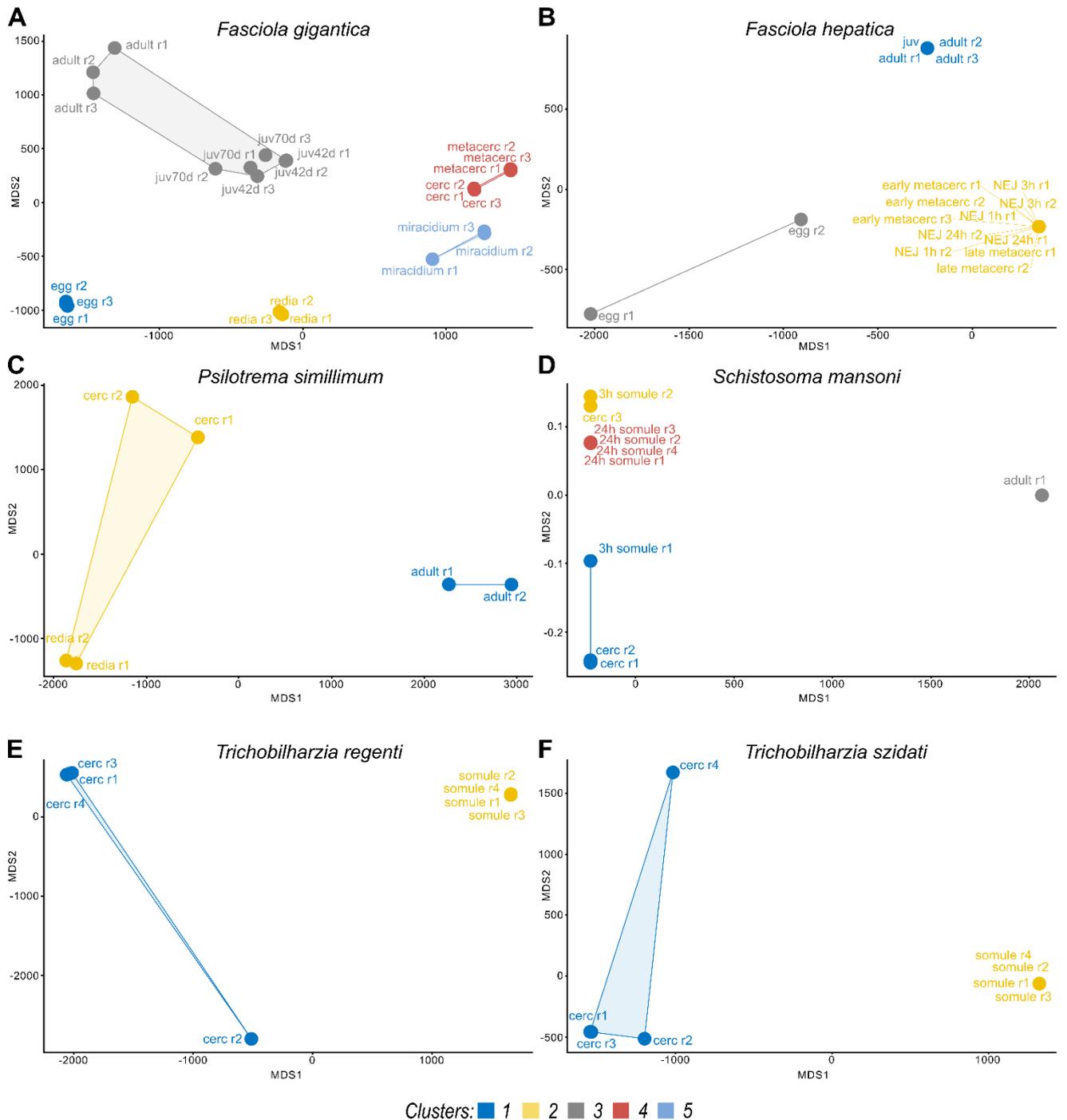


Рисунок 11 – Результаты внутривидового многомерного шкалирования молекулярных подписей фаз. Схожесть молекулярных подписей обратно пропорциональна расстоянию между ними: чем ближе располагаются молекулярные подписи, тем больше схожи они между собой. Разными цветами отмечены обнаруженные кластеры. Наименования 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.

Для сравнительного анализа между видами дигеней были подготовлены два типа матриц присутствия / отсутствия. Первый тип матриц содержал информацию о том, входит (1) или нет (0) ортолог в молекулярную подпись фазы жизненного цикла. Второй тип матриц содержал информацию о том, имеет (1) или нет (0) ортолог статистически значимое повышение экспрессии на рассматриваемой фазе жизненного цикла. Межвидовое сравнение провели на обоих типах матриц в два этапа: на первом этапе «редииодные» (*F. gigantica*, *F. hepatica* и *P. simillimum*) и «спороцистоидные» (*S. mansoni*, *T. regenti* и *T. szidati*) виды рассматривались отдельно, а на втором этапе анализа все виды были сравнены между собой одновременно. Результаты сравнительного анализа разных видов дигеней представлены на рисунке 12.

По результатам сравнения молекулярных подписей фаз сложных жизненных циклов «редииодных» видов дигеней обнаружено три кластера (Рисунок 12А). Первый кластер включал схожие фазы различных видов: ювениль *F. hepatica* и мариты *F. gigantica*, *F. hepatica* и *P. simillimum*. Состав двух других кластеров был более разнообразным, но также включал схожие фазы жизненных циклов близких видов. Четыре кластера выявлено при сравнении результатов анализа дифференциальной экспрессии ортологов (Рисунок 12В). Два кластера включали только по одной фазе жизненного цикла: редию *P. simillimum* (второй кластер) и метацеркарию *F. gigantica* (четвертый кластер). Первый кластер объединял 14 фаз, все из которых, за исключением яиц фасциол, принадлежат амфимиктическому поколению. Третий кластер содержал две фазы партеногенетического поколения *F. gigantica* (мирацидия и редии) и ювениль *F. hepatica*, покинувшую цисту за 24 часа до момента забора образца транскриптома.

У «спороцистоидных» видов дигеней обнаружено только два кластера при сравнении как молекулярных подписей фаз целиком (Рисунок 12С), так и результатов анализа дифференциальной экспрессии (Рисунок 12D). В первом случае церкарии *T. regenti* и *T. szidati* объединились в отдельный кластер от других фаз сложного жизненного цикла, тогда как шистосомулы *Trichobilharzia* были включены в один кластер с церкарией, шистосомулой и маритой *S. mansoni* (Рисунок 12С). Сравнительный анализ повышенной экспрессии ортологов продемонстрировал совместную кластеризацию шистосомул *Trichobilharzia* отдельно от церкарий *T. regenti* и *T. szidati*, которые сформировали единый кластер со всеми рассмотренными фазами *S. mansoni* (Рисунок 12D).

Наличие крупного кластера, включавшего большинство исследованных фаз сложных жизненных циклов шести видов дигеней, являлось общим результатом как при сравнении молекулярных подписей фаз целиком (Рисунок 12Е), так и при сопоставлении результатов анализа дифференциальной экспрессии ортологов (Рисунок 12F). В обоих случаях кластеры содержали фазы как партеногенетического, так и амфимиктического поколения. В отдельный кластер были объединены яйца фасциол, 70-дневная ювениль *F. gigantica*, а также мариты

F. gigantea и *P. simillimum* (Рисунок 12E). Результаты кластеризации фаз *Trichobilharzia* в обоих случаях совпадали с результатами, полученными при анализе «спороцистоидных» видов: церкарии (Рисунок 12E) или шистосомулы (Рисунок 12F) располагались отдельно от других фаз сложных жизненных циклов дигеней. В то же время, при рассмотрении повышенной экспрессии ортологов, редия *P. simillimum* объединилась вместе с шистосомулами *Trichobilharzia* (Рисунок 12F).

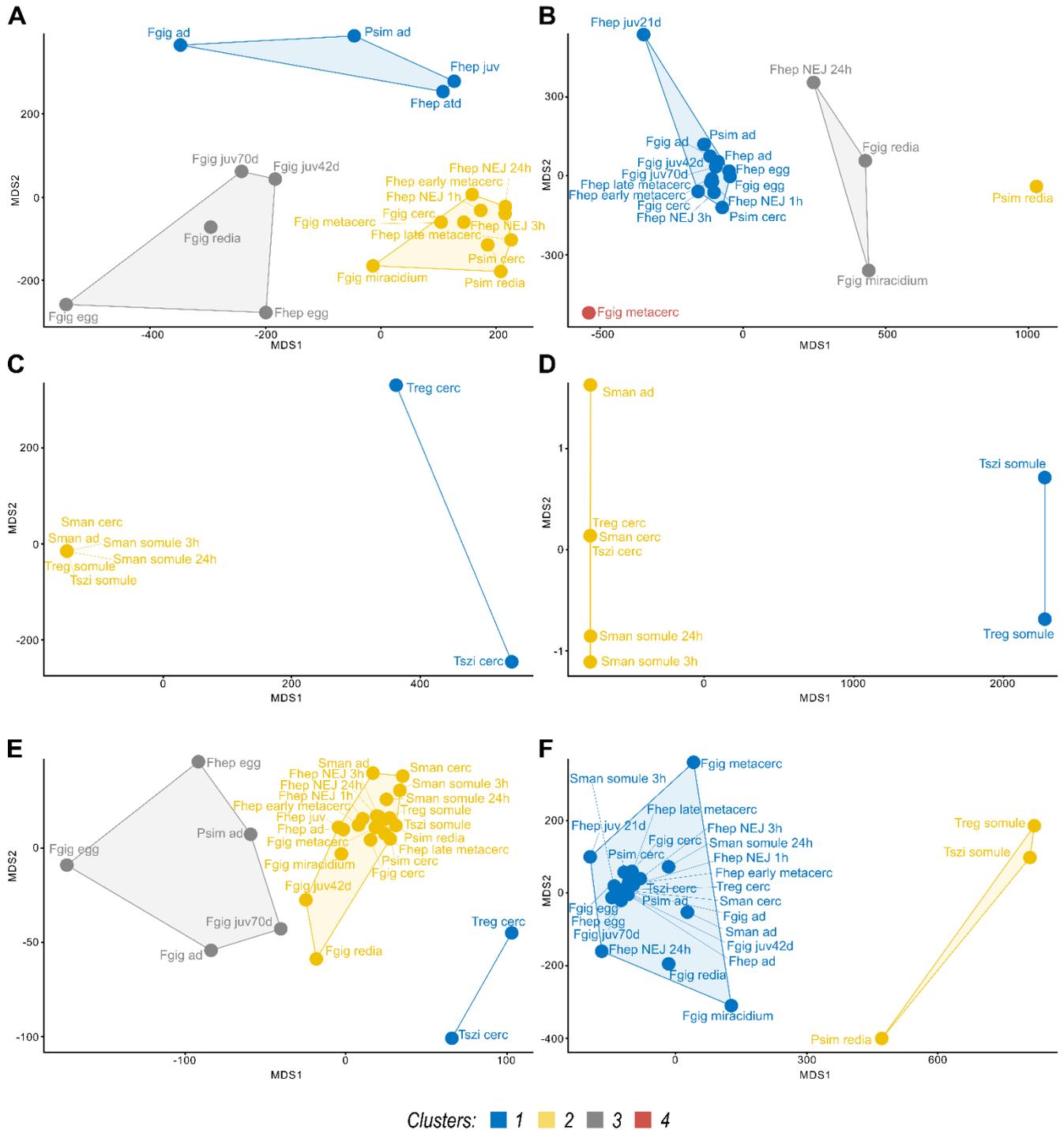


Рисунок 12 – Результаты межвидового многомерного шкалирования молекулярных подписей фаз. Были рассмотрены «редииоидные» (А, В), «спороцистоидные» (С, D), а также все виды дигеней одновременно (Е, F). Многомерное шкалирование было проведено на матрицах присутствия/отсутствия ортолога в молекулярной подписи фазы жизненного цикла (А, С, Е) и повышенной экспрессии ортолога на рассматриваемой фазе (В, 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 цветовые идентификаторы кластеров представлены в нижней части рисунка. Обозначения: *serc* – церкария, *metacerc* – метацеркария, *juv* – ювениль, *juv42/70d* – 42- и 70-дневные ювенили, соответственно; *NEJ 1/3/24h* – ювенили 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 генной онтологии

С целью выявления биологических процессов, в которых принимают участие дифференциально экспрессирующиеся гены, нами был проведен анализ обогащения терминов генной 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. Полный перечень обогащенных «родительских» терминов приведен в Приложении 4 к публикации Nesterenko et al., 2022. Далее будут представлены обобщенные результаты сравнения списков «родительских» биологических процессов для схожих фаз сложных жизненных циклов различных видов дигеней.

Транскриптомные данные по яйцам дигеней доступны только для фасциOLID. Из 61 биологического процесса, обогащенных как минимум на одном из образцов транскриптомов яиц *Fasciola*, 14 оказались общими для *F. gigantica* и *F. hepatica*. В числе общих процессов обнаружены «регуляция клеточной смерти» (regulation of cell death, GO:0010941) и «пигментация в ходе развития» (developmental pigmentation, GO:0048066). Данные по мирацидиям доступны только для *F. gigantica* и для личинки материнской спороцисты выявлена активность таких биологических процессов, как «клеточное дыхание» (cellular respiration, GO:0045333), «транспорт токсинов» (toxin transport, GO:1901998), «организация митотического веретена» (mitotic spindle organization, GO:0007052), «регенерация печени» (liver regeneration, GO:0097421), а также процессов, ассоциированных с метаболизмом и процессингом РНК.

Секвенирование транскриптомов редий было проведено только для *F. gigantea* и *P. simillimum*. Транскриптом *P. simillimum* получен в ходе нашего исследования. Для редий *F. gigantea* и *P. simillimum* обнаружено 7 общих обогащенных биологических процессов: «биогенез рибонуклеопротеинового комплекса» (ribonucleoprotein complex biogenesis, GO:0022613), «метаболический процесс клеточных соединений азота» (cellular nitrogen compound metabolic process, GO:0034641), «процессинг РНК» (RNA processing, GO:0006396), «сборка рибонуклеопротеинового комплекса» (ribonucleoprotein complex assembly, GO:0022618), «ядерно-транскрибируемый катаболический процесс мРНК» (nuclear-transcribed mRNA catabolic process, GO:0000956), «трансляция в митохондриях» (mitochondrial translation, GO:0032543), «метаболический процесс ДНК» (DNA metabolic process, GO:0006259). К числу «специфичных» для редии *F. gigantea* могут быть отнесены такие биологические процессы, как «регенерация печени» (liver regeneration, GO:0097421), «реакция на повреждение ДНК, обнаружение повреждения ДНК» (DNA damage response, detection of DNA damage, GO:0042769), «регуляция внутреннего апоптотического сигнального пути с помощью медиатора класса p53» (regulation of intrinsic apoptotic signaling pathway by p53 class mediator, GO:1902253). У редий *P. simillimum* удалось обнаружить активность биологических процессов «ДНК-зависимая репликация ДНК» (DNA-dependent DNA replication, GO:0006261), «G1/S переход в митотическом клеточном цикле» (G1/S transition of mitotic cell cycle, GO:0000082), «определение пола» (sex determination, GO:0007530), «поддержание клеточного состава» (maintenance of cell number, GO:0098727), «определение судьбы эндодермальных клеток» (endodermal cell fate commitment, GO:0001711), «дифференцировка дофаминергических нейронов среднего мозга» (midbrain dopaminergic neuron differentiation, GO:1904948).

Данные по церкариям доступны для 5 видов дигеней: *F. gigantea*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. Мы рассматривали в качестве «общих» те биологические процессы, которые были классифицированы как «обогащенные» как минимум у церкарий 3 видов. 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: «сокращение мышц» (muscle contraction, GO:0006936) (*F. gigantea*), «регуляция мышечного сокращения» (regulation of muscle contraction, GO:0006937) (*P. simillimum*, *S. mansoni*) и «сокращение скелетных мышц» (skeletal muscle contraction, GO:0003009) (*P. simillimum*).

Среди шести исследуемых видов дигеней транскриптомы метацеркарий доступны для фасциолид: ранняя и поздняя метацеркария *F. hepatica* и метацеркария *F. gigantea*. Для метацеркарий *F. gigantea* и *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» (response to abiotic stimulus, GO:0009628), «проведение сигнала» (signaling, GO:0023052), «локомоция» (locomotion, GO:0040011), «межклеточная передача сигнала» (cell-cell signaling, GO:0007267), «развитие тканей» (tissue development, GO:0009888), «миграция эпителия» (epithelium migration, GO:0090132), «поведение» (behavior, GO:0007610), «ответ на внешний раздражитель» (response to external stimulus, GO:0009605), «защитная реакция на другой организм» (defense response to other organism, GO:0098542), «постэмбриональное развитие органов животного» (post-embryonic animal organ development, GO:0048569). Единственным общим биологическим процессом для ранней метацеркарии *F. hepatica* и метацеркарии *F. gigantica* оказался «межвидовое взаимодействие между организмами» (interspecies interaction between organisms, GO:0044419).

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» (oxidation-reduction process, GO:0055114), «регуляция клеточной смерти» (regulation of cell death, GO:0010941), «локомоция» (locomotion, GO:0040011), «негативная регуляция биологических процессов» (negative regulation of biological process, GO:0048519), «гликозилирование» (glycosylation, GO:0070085) и «проведение сигнала через рецептор на клеточной поверхности» (cell surface receptor signaling pathway, GO:0007166). Среди 22 биологических процессов, активных у ювенили через один час после покидания цисты, наиболее репрезентативными оказались «процессы развития» (developmental process, GO:0032502), «межвидовое взаимодействие между организмами» (interspecies interaction between organisms, GO:0044419), «ответ на внешний раздражитель» (response to external stimulus, GO:0009605), «биологическая адгезия» (biological adhesion, GO:0022610), «иммунитет, опосредованный нейтрофилами» (neutrophil mediated immunity, GO:0002446). К числу специфичных для ювенили через 3 часа после покидания цисты, были отнесены 27 биологических процессов, в том числе «регуляция уровней гормонов» (regulation of hormone levels, GO:0010817), «негативная регуляция секреции» (negative regulation of secretion, GO:0051048), «регуляция клеточной дифференцировки» (regulation of cell differentiation, GO:0045595). Для особей паразита, покинувших оболочку цисты сутки назад, отмечена активность 26 специфичных биологических процессов, в числе которых «метаболический

процесс производных жирных кислот» (fatty acid derivative metabolic process, GO:1901568), «организация саркомеров» (sarcomere organization, GO:0045214), «процесс биосинтеза ацилкоэнзима А» (acyl-CoA biosynthetic process, GO:0071616), «процесс биосинтеза углеводов» (carbohydrate biosynthetic process, GO:0016051), «позитивная регуляция адгезии между клеткой и субстратом» (positive regulation of cell-substrate adhesion, GO:0010811), «ответ на питательные вещества» (response to nutrient, GO:0007584).

Для двух исследованных видов фасциол были также доступны транскриптомные данные ювенилий с более поздних этапов миграции в организме окончательного хозяина: транскриптомы 21-дневной ювенили *F. hepatica* и 42- и 70-дневных ювенилей *F. gigantica*. Сравнительный анализ списков «родительских» терминов геномной онтологии обнаружил, что из 52 биологических процессов большинство «обогащено» только на одной из рассмотренных ювенилей паразита. Среди них обнаружены такие процессы как «организация ресничек» (cilium organization, GO:0044782) (21-дневная ювениль *F. hepatica*), «подвижность клеток, зависящая от ресничек или жгутиков» (cilium or flagellum-dependent cell motility, GO:0001539) (21-дневная ювениль *F. hepatica*), «материнский процесс, связанный с женской беременностью» (maternal process involved in female pregnancy, GO:0060135) (42-дневная ювениль *F. gigantica*), «женская беременность» (female pregnancy, GO:0007565) (70-дневная ювениль *F. gigantica*). В числе 14 общих для различных ювенилей биологических процессов, за исключением разнообразных метаболических процессов, присутствовали «ответ на глюкокортикоиды» (response to glucocorticoid, GO:0051384) (42- и 70-дневные ювенили *F. gigantica*), «пролиферация эпителиальных клеток» (epithelial cell proliferation, GO:0050673) (42- и 70-дневные ювенили *F. gigantica*), и «организация внеклеточного матрикса» (extracellular matrix organization, GO:0030198) (42- и 70-дневные ювенили *F. gigantica*). Ни одного общего биологического процесса для всех трех рассматриваемых ювенилей фасциол не обнаружено.

Транскриптомы шистосомул опубликованы для *S. mansoni* (шистосомула через 3 и 24 часа после заражения), *T. regenti* и *T. szidati*. Сравнительный анализ списков обогащенных терминов геномной онтологии показал, что 40 биологических процессов являлись общими для различных шистосомул. Большинство биологических процессов (33 из 40) оказались общими для двух рассмотренных видов *Trichobilharzia*. В их числе были «клеточное общение» (cell communication, GO:0007154), «миграция клеток» (cell migration, GO:0016477), «определение клеточной судьбы» (cell fate commitment, GO:0045165), «пролиферация клеточной популяции» (cell population proliferation, GO:0008283), «эмбриональное развитие» (embryo development, GO:0009790), «процессы иммунной системы» (immune system process, GO:0002376), «размножение» (reproduction, GO:0000003), «нейрогенез» (neurogenesis, GO:0022008). Три биологических процесса являлись общими для одного вида *Trichobilharzia* и одной из рассмотренных

шистосомул *S. mansoni*: «постэмбриональное развитие органов животного» (post-embryonic animal organ development, GO:0048569) (шистосомула *S. mansoni* через 24 часа после заражения, шистосомула *T. szidati*), «шаперон-опосредованный фолдинг белка» (chaperone-mediated protein folding, GO:0061077) (шистосомула *S. mansoni* через 24 часа после заражения, шистосомула *T. regenti*), «полиаденилирование РНК» (RNA polyadenylation, GO:0043631) (шистосомула *S. mansoni* через 3 часа после заражения, шистосомула *T. szidati*). Общими между двумя проанализированными видами *Trichobilharzia* и одной из шистосомул *S. mansoni* оказались такие биологические процессы, как «регуляция клеточного процесса» (regulation of cellular process, GO:0050794) (шистосомулы *Trichobilharzia* и шистосомула *S. mansoni* через 3 часа после заражения), «негативная регуляция клеточных процессов» (negative regulation of cellular process, GO:0048523) (шистосомулы *Trichobilharzia* и шистосомула *S. mansoni* через 3 часа после заражения), «морфогенез клеточных компонентов» (cellular component morphogenesis, GO:0032989) (шистосомулы *Trichobilharzia* и шистосомула *S. mansoni* через 24 часа после заражения), «регуляция ответа на раздражитель» (regulation of response to stimulus, GO:0048583) (шистосомулы *Trichobilharzia* и шистосомула *S. mansoni* через 3 часа после заражения). Для шистосомул *S. mansoni*, собранных через 3 и 24 часа после заражения, не найдено ни одного общего биологического процесса.

Наибольшее количество транскриптомов доступно для марит дигеней и в нашем исследовании было проведено сравнение между *F. gigantica*, *F. hepatica*, *P. simillimum* и *S. mansoni*. Среди 11 биологических процессов, общих для марит различных видов, обнаружены такие, как «организация реснички» (cilium organization, GO:0044782) (*F. gigantica*, *P. simillimum*), «движение реснички» (cilium movement, GO:0003341) (*F. gigantica*, *P. simillimum*), «подвижность сперматозоидов» (sperm motility, GO:0097722) (*F. gigantica*, *P. simillimum*) и «ремоделирование тканей» (tissue remodeling, GO:0048771) (*F. gigantica*, *F. hepatica*). Напротив, 77 биологических процессов классифицированы как «обогащённые» только в одном из рассмотренных видов: «развитие кожи» (skin development, GO:0043588) (*F. gigantica*), «клеточный ответ на тиреоидные гормоны» (cellular response to thyroid hormone stimulus, GO:0097067) (*F. gigantica*), «развитие первичных мужских половых признаков» (development of primary male sexual characteristics, GO:0046546) (*F. hepatica*), «определение левой/правой симметрии» (determination of left/right symmetry, GO:0007368) (*P. simillimum*), «движение клетки или субклеточного компонента» (movement of cell or subcellular component, GO:0006928) (*P. simillimum*), «вителлогенез» (vitellogenesis, GO:0007296) (*S. mansoni*), «ответ на эстроген» (response to estrogen, GO:0043627) (*S. mansoni*) и другие.

3.12 Определение и анализ наборов потенциальных секреторных белков

Принимая во внимание паразитический образ жизни дигеней, мы провели *in silico* определение наборов потенциальных секреторных белков для 11 видов дигеней: *C. sinensis*, *F. gigantica*, *F. hepatica*, *O. felineus*, *O. viverrini*, *P. simillimum*, *S. haematobium*, *S. japonicum*, *S. mansoni*, *T. regenti* и *T. szidati*. Все потенциальные секреторные белки были разделены на две группы на основании присутствия («классические») или отсутствия («неклассические») в последовательности классического N-концевого сигнального пептида.

Для каждого из 11 рассматриваемых видов дигеней удалось обнаружить сотни секреторных белков: от 327 (239 «классических» + 88 «неклассических» секреторных белков) у *T. szidati* до 1899 (974 «классических» + 925 «неклассических» секреторных белков) у *P. simillimum*. Согласно результатам проведенных анализов обогащения терминов генной онтологии, потенциальные «классические» секреторные белки вовлечены в различные биологические процессы, в числе которых «организация внеклеточного матрикса» (extracellular structure organization, GO:0043062) (*F. gigantica*, *O. felineus*, *S. mansoni*), «ремоделинг тканей» (tissue remodeling, GO:0048771) (*F. gigantica*, *F. hepatica*, *P. simillimum*), «регуляция хемотаксиса» (regulation of chemotaxis, GO:0050920) (*F. hepatica*, *P. simillimum*, *S. japonicum*, *S. mansoni*). Среди обогащенных биологических процессов четыре являлись общими для более чем половины исследованных видов дигеней: «протеолиз» (proteolysis, GO:0006508), «ответ на стресс эндоплазматического ретикулума» (response to endoplasmic reticulum stress, GO:0034976), «ответ на раздражитель» (response to stimulus, GO:0050896) и «дегрануляция лейкоцитов» (leukocyte degranulation, GO:0043299).

Потенциальные «неклассические» секреторные белки принимали участие в 23 биологических процессах, в том числе «реакция на органическое циклическое соединение» (response to organic cyclic compound, GO:0014070) (*C. sinensis*), «регуляция пролиферации нервных клеток-предшественников» (regulation of neural precursor cell proliferation, GO:2000177) (*O. felineus*), «эмбриональное развитие органов» (embryonic organ development, GO:0048568) (*O. felineus*), «дифференцировка клеток» (cell differentiation, GO:0030154) (*O. viverrini*), «клеточный ответ на повреждение ДНК» (cellular response to DNA damage stimulus, GO:0006974) (*S. japonicum*), «модификация макромолекул» (macromolecule modification, GO:0043412) (*T. szidati*). 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

Реконструкция моделей геномов последних общих предков Platyhelminthes и Digenea выполнена по результатам анализа ранее выявленных иерархических групп ортологов. Первичные модели геномов последних общих предков включали 5952 (Platyhelminthes) и 10372 (Digenea) ортогруппы. После исключения иерархичных ортогрупп, включавших менее 75% рассматриваемых видов плоских червей, обновленные модели последних общих предков Platyhelminthes и Digenea состояли из 2579 и 4622 групп ортологов, соответственно. Результаты сравнительного анализа построенных моделей указали на то, что в модели Digenea 2258 генов осталось от Platyhelminthes, 64 гена были дублированы, а 1850 генов появились у последнего общего предка дигеней.

В реконструированных моделях были представлены ортогруппы, а не нуклеотидные последовательности генов. Для дальнейшего анализа было необходимо подготовить наборы последовательностей генов и кодируемых ими белков на основе информации о составе ортогрупп, включенных в модели. Для реализации поставленной цели мы выбрали геном *S. mansoni* в качестве «справочного», принимая во внимание высокое качество доступных геномных данных и многочисленные результаты исследований молекулярной биологии этого вида дигеней. Общее количество генов *S. mansoni* в реконструированных моделях составило 3687 (Platyhelminthes) и 6431 (Digenea). Среди генов *S. mansoni*, вошедших в модель генома Digenea, мы отобрали 2633 и 116 генов, которые, предположительно, появились и дублировались у последнего общего предка дигеней, соответственно.

Согласно результатам анализа обогащения терминов геной онтологии, гены, появившиеся у последнего общего предка дигеней, принимают участие в 705 биологических процессах, а гены, дублированные у предка дигеней, вовлечены в 177 биологических процесса. Полный перечень обогащенных биологических процессов приведен в Приложении 5 к публикации Nesterenko et al., 2022. Сравнительный анализ между списками обогащенных биологических процессов обнаружил, что 62 наименования являлись общими, в числе которых «развитие органов животного» (animal organ development, GO:0048513), «развитие репродуктивной системы» (reproductive system development, GO:0061458), «размножение» (reproduction, GO:0000003), «развитие эмбриона, заканчивающееся рождением или вылуплением яиц» (embryo development ending in birth or egg hatching, GO:0009792), «миграция клеток» (cell migration, GO:0016477), «дифференцировка клеток» (cell differentiation, GO:0030154).

На рисунках 13 и 14 представлено разнообразие «родительских» биологических процессов, обогащенных появившимися и дублированными генами последнего общего предка дигеней, соответственно. Гены, появившиеся у последнего общего предка дигеней, принимают участие в процессах, связанных с проведением сигналов, функционированием клеток, развитием, размножением и регуляцией экспрессии генов (Рисунок 13). 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 (Рисунок 14).

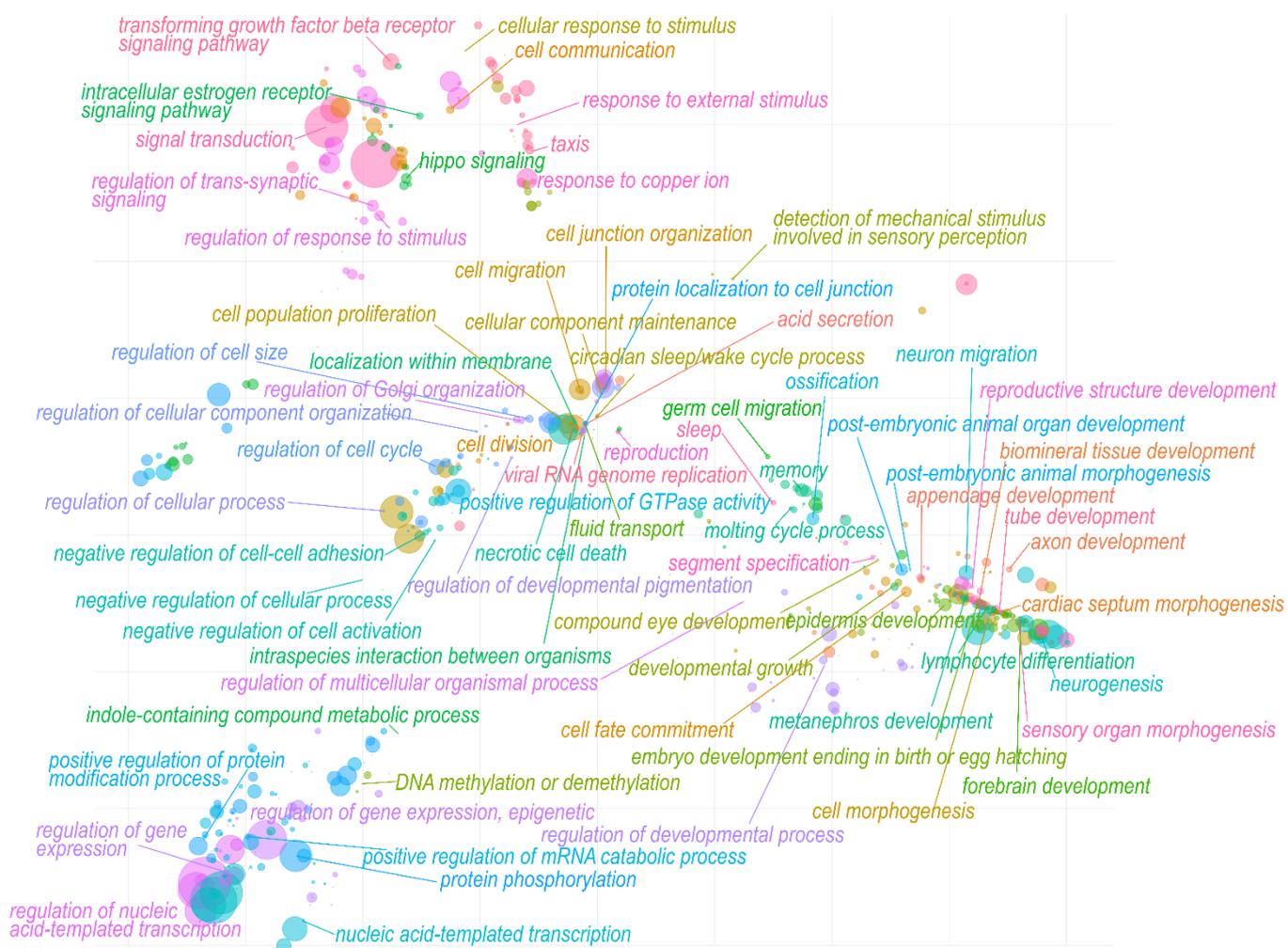


Рисунок 13 – Диаграмма рассеивания, построенная для «родительских» биологических процессов, обогащенных генами, появившимися у последнего общего предка дигеней. Каждая отдельная точка представляет собой «родительский» биологический процесс, а расстояние между точками обратно пропорционально сходству между процессами. На двумерной плоскости точки расположены тем ближе, чем больше схожи между собой «родительские» биологические процессы. Размер точек отражает отрицательную величину десятичного логарифма Р-значения точного теста Фишера: чем больше статистическая значимость, тем больше размер точки.



Рисунок 14 – Диаграмма рассеивания, построенная для «родительских» биологических процессов, обогащенных генами, дублированными у последнего общего предка дигеней. Каждая отдельная точка представляет собой «родительский» биологический процесс, а расстояние между точками обратно пропорционально сходству между процессами. На двумерной плоскости точки расположены тем ближе, чем больше схожи между собой «родительские» биологические процессы. Размер точек отражает отрицательную величину десятичного логарифма Р-значения точного теста Фишера: чем больше статистическая значимость, тем больше размер точки.

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

В нашем исследовании был проведен филостратиграфический анализ наборов генов 11 видов дигеней (*C. sinensis*, *F. hepatica*, *F. gigantica*, *O. viverrini*, *O. felineus*, *P. simillimum*, *S. mansoni*, *S. haematobium*, *S. japonicum*, *T. regenti* и *T. szidati*) и 3 (*S. mediterranea*, *M. lignano*, *P. vittatus*) видов свободноживущих плоских червей. Все гены исследованных видов успешно

Рисунок 15 – Филостратиграфический состав наборов генов, кодирующих белки длиной большей или равной 100 аминокислот у исследованных видов плоских червей. Каждая филострата представлена собственным цветом, а числовые значения соответствуют процентам от общего числа рассмотренных для вида белок-кодирующих генов.

Полученные результаты филостратиграфии были использованы нами для определения состава различных групп генов. Проведенный анализ продемонстрировал, что в модели генома последнего общего предка дигеней процент генов, отнесённых к филострате «Cellular organisms», снизился в сравнении с моделью генома последнего общего предка Platyhelminthes (Рисунок 16А). В модели Platyhelminthes к филострате «Cellular organisms» отнесено 53,19% генов, тогда как в модели Digenea доля таких генов составила 42,51%. По сравнению с моделью генома последнего общего предка Platyhelminthes, в модели генома последнего общего предка Digenea увеличилась доля таких филострат, как «Opisthokonta» (5,8% против 7,54% генов в моделях геномов Platyhelminthes и Digenea, соответственно), «Metazoa» (3,55% против 4,45%), «Eumetazoa» (2,17% против 4,49%), «Bilateria» (0,62% против 2,1%), «Protostomia» (0,46% против 0,75%), «Spiralia» (0,03% против 0,06%), «Lophotrochozoa» (0,24% против 0,75%), «Platyhelminthes» (0,62% против 4,37%). Процент генов, отнесенных к филострате «Digenea» («Class»), в реконструированной модели генома последнего общего предка дигеней равен 0,9%. Сравнительный анализ построенных моделей показал, что размер каждой филостраты в модели Digenea превышал размер этой же филостраты в модели Platyhelminthes (Рисунок 16А).

Мы также проанализировали филостратиграфический состав генов, имеющих заметную экспрессию на всех рассмотренных фазах сложных жизненных циклов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. У каждого из шести исследованных видов дигеней больше 55% генов с указанным профилем экспрессии принадлежали филостратам «Cellular organisms» и «Eukaryota» (Рисунок 16В). В то же время среди генов с заметной экспрессией на протяжении сложного жизненного цикла количество видоспецифичных у большинства видов не превышало 2%. Единственным исключением стал вид *P. simillimum*, у которого доля таких генов составила приблизительно 6,5%.

Сложный филостратиграфический состав обнаружен и у выявленных кластеров ко-экспрессии *F. gigantica*, *F. hepatica*, *P. simillimum* и *S. mansoni*. Каждый кластер ко-экспрессии включал гены из различных филострат, но в большинстве случаев (48 из 53 кластеров) наибольшее количество ко-экспрессирующихся генов относилось к филострате «Cellular organisms». Вклад видоспецифичной филостраты варьировал в широких пределах: от 0 до 15,38% у *F. gigantica*, от 0,38% до 2,31% у *F. hepatica*, от 7,88% до 18,17% у *P. simillimum*, и от 0 до 1,21% у *S. mansoni*.

Филостратиграфический состав наборов генов, кодирующих как «классические» (Рисунок 16С), так и «неклассические» (Рисунок 16D) секреторные белки, различался у всех 11 рассмотренных видов дигеней. Все филостраты были разделены на две группы: 1) от «Cellular organisms» до «Platyhelminthes», 2) от «Class» до «Species». У большинства видов первая группа включала большинство «классических» (11 из 11 видов) и «неклассических» (8 из 11 видов) секреторных белков. Исключением стали *C. sinensis*, *O. viverrini* и *P. simillimum*, у которых большинство предполагаемых «неклассических» секреторных белков принадлежали второй группе филострат. В целом, процент «классических» и «неклассических» секреторных белков во второй группе варьировал от 17,58% (*T. szidati*) до 49,18% (*P. simillimum*) и от 27,27% (*T. szidati*) до 73,13% (*C. sinensis*), соответственно.

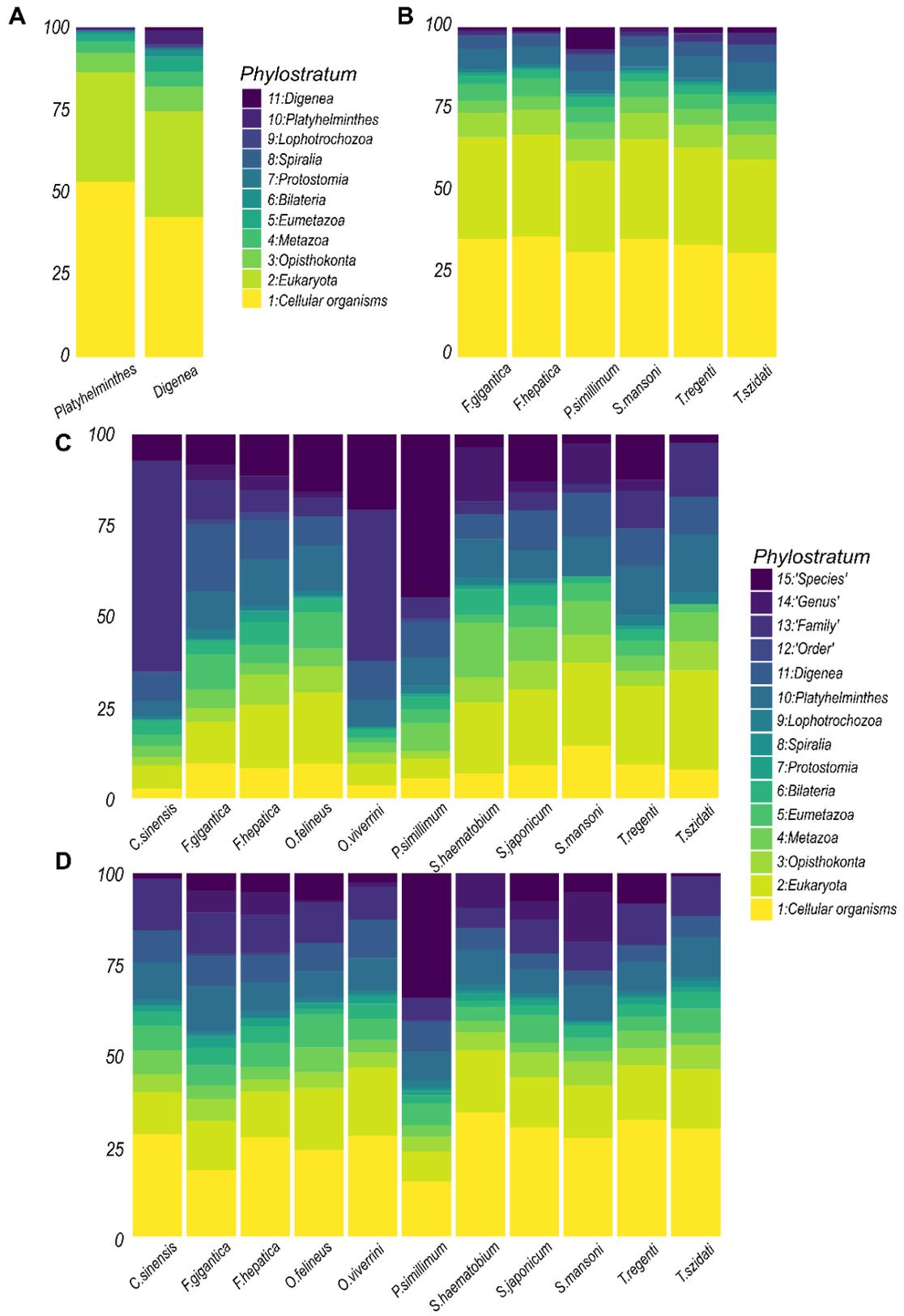


Рисунок 16 – Филостратиграфический состав реконструированных моделей геномов предков (А), наборов генов с заметным (≥ 2 транскрипта-на-миллион) уровнем экспрессии на всех рассмотренных фазах сложных жизненных циклов дигеней (В) и наборов генов, кодирующих потенциальные «классические» (С) и «неклассические» (D) секреторные белки. Каждая филострата представлена собственным цветом, а числовые значения соответствуют процентам.

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

Основываясь на результатах определения филогенетического состава кластеров ко-экспрессии, мы предположили, что молекулярные подписи фаз жизненных циклов дигеней также будут различаться по вкладам различных филогенетических групп. С целью проверить предположение мы провели расчет возрастных индексов для молекулярных подписей, выявленных нами для фаз сложных жизненных циклов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. Для проверки статистической значимости вариации возрастных индексов мы использовали тест, оценивающий отклонение филогенетического профиля от плоской линии. Любое значимое отклонение филогенетического профиля от плоской линии рассматривалось в качестве эволюционного сигнала. У *F. gigantica*, *F. hepatica*, *S. mansoni*, *T. regenti* и *T. szidati* статистически значимые отклонения выявлены при одновременном рассмотрении всех обнаруженных филогенетических групп.

Мы предположили, что отсутствие эволюционного сигнала у *P. simillimum* могло быть связано с увеличенным размером видоспецифичной филогенетической группы. При исключении из анализа всех генов, отнесенных к филогенетической группе «Species», у *P. simillimum* изменились величины возрастных индексов молекулярных подписей фаз сложного жизненного цикла (Рисунок 17А, В) и поменялся общий филогенетический профиль (Рисунок 17А, В). Обновленный профиль изменения возрастных индексов молекулярных подписей фаз жизненного цикла *P. simillimum* демонстрировал статистически значимое отклонение от плоской линии (Рисунок 17В). В дальнейшем анализе для *P. simillimum* мы рассматривали только результаты, полученные после исключения влияния видоспецифичной филогенетической группы.

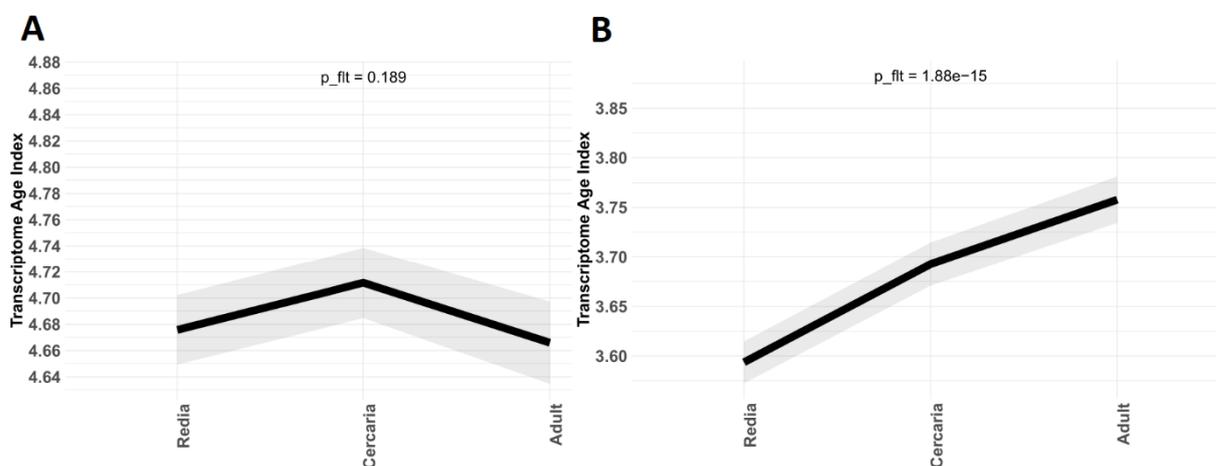
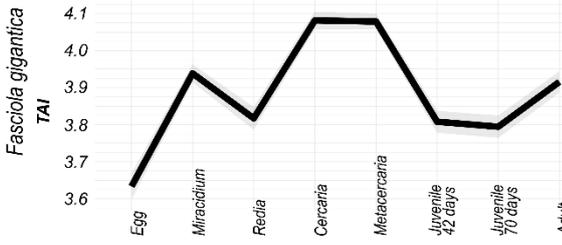
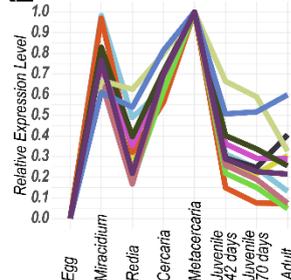


Рисунок 17 – Возрастные индексы транскриптомов (ТАИ) фаз сложного жизненного цикла *P. simillimum* до (А) и после (В) исключения генов, относящихся к видоспецифичной филострате. В верхних частях профилей вариации возрастных индексов приведены Р-значения (p_{flt}), рассчитанные при анализе статистической значимости отклонения филотранскриптомного профиля от плоской линии. Серая зона вокруг профиля соответствует стандартному отклонению. Обозначения: Transcriptome Age Index – возрастной индекс транскриптома; Redia – редия, Cercaria – церкария, Adult – марита.

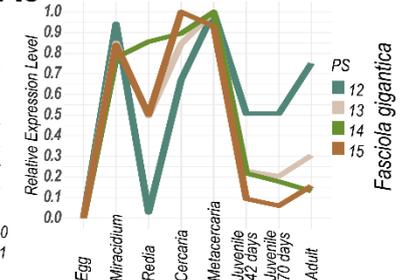
A1



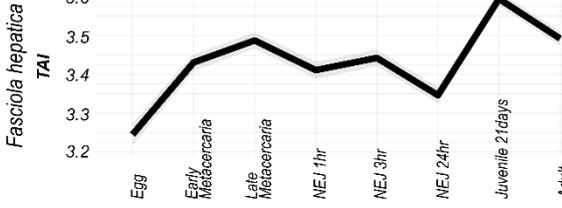
A2



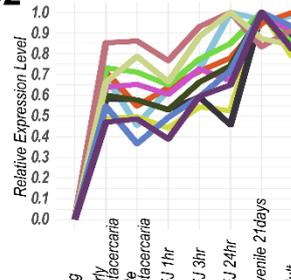
A3



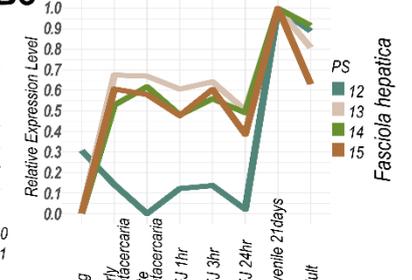
B1



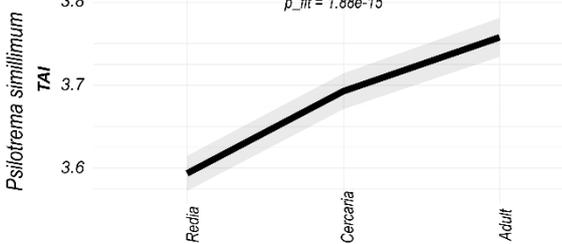
B2



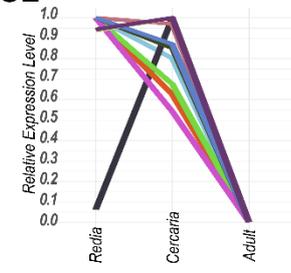
B3



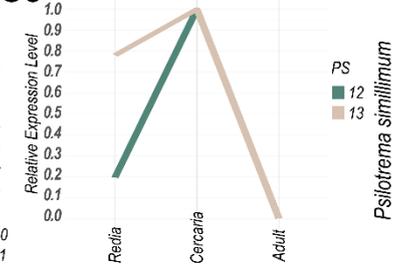
C1



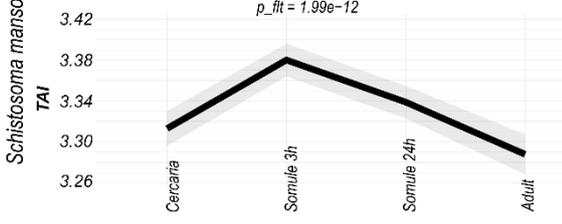
C2



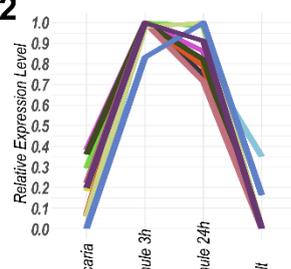
C3



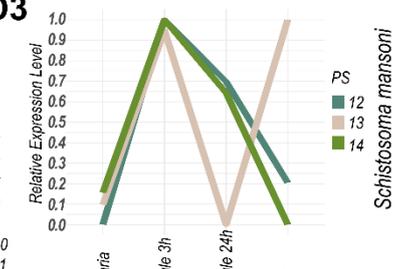
D1



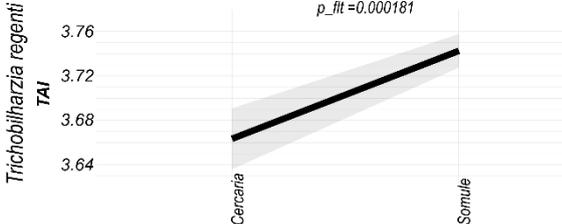
D2



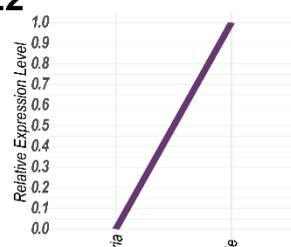
D3



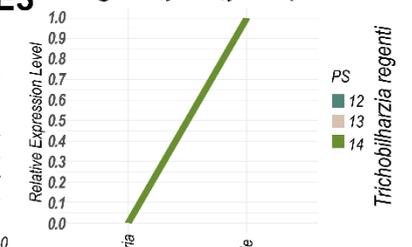
E1



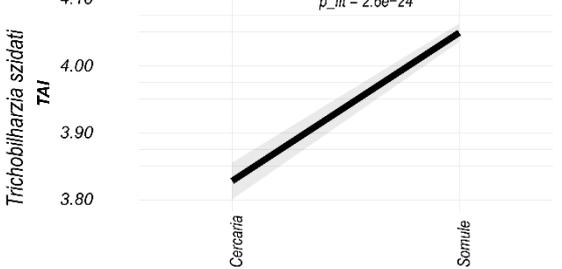
E2



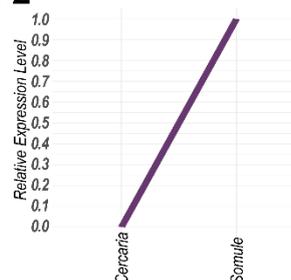
E3



F1



F2



F3

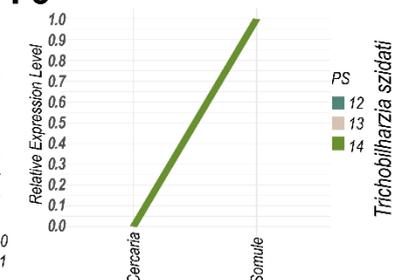


Рисунок 18 – Возрастные индексы транскриптомов фаз сложных жизненных циклов дигеней и относительные уровни экспрессии филострат. Представлены результаты анализов для *F. gigantica* (A1-3), *F. hepatica* (B1-3), *P. simillimum* (C1-3), *S. mansoni* (D1-3), *T. regenti* (E1-3) и *T. szidati* (F1-3). Значения 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 Р-значения (p_{flt}), 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) метацеркария; 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.

Наименьшие значения возрастных индексов (приведены в скобках) получены для яиц *F. gigantica* (3.63) (Рисунок 18A1) и *F. hepatica* (3.24) (Рисунок 18B1), редии *P. simillimum* (3.59) (Рисунок 18C1), мариты *S. mansoni* (3.28) (Рисунок 18D1), а также церкарий *T. regenti* (3.66) (Рисунок 18E1) и *T. szidati* (3.82) (Рисунок 18F1). Напротив, наибольшими значениями возрастных индексов обладали церкария *F. gigantica* (4.08) (Рисунок 18A1), 21-дневная ювениль *F. hepatica* (3.59) (Рисунок 18B1), марита *P. simillimum* (3.75) (Рисунок 18C1), шистосомула *S. mansoni* через 3 часа после заражения (3.37) (Рисунок 18D1), шистосомулы *T. regenti* (3.74) (Рисунок 18E1) и *T. szidati* (4.04) (Рисунок 18F1).

Для всех 12 перечисленных фаз сложных жизненных циклов мы определили вклад каждого гена в возрастную индекс молекулярной подписи. Среди генов с наибольшими вкладами мы отобрали 500, обладающих функциональной аннотацией по базе данных геной онтологии, с целью определить биологические процессы, в которых принимают участие такие гены. Согласно полученным результатам анализа, отобранные гены вовлечены в 125 «родительских» биологических процессов. Общими для большинства рассмотренных фаз с наибольшими и наименьшими возрастными индексами молекулярных подписей оказались три процесса, связанные с синтезом белка в митохондриях: «митохондриальная трансляция» (mitochondrial translation, GO:0032543), «митохондриальная трансляционная элонгация» (mitochondrial

translational elongation, GO:0070125), «митохондриальная трансляционная терминация» (mitochondrial translational termination, GO:0070126). В числе 30 биологических процессов, общих для нескольких фаз сложных жизненных циклов, были такие, как: «развитие глиальных клеток» (glial cell development, GO:0021782) (21-дневная ювениль *F. hepatica*, марита *P. simillimum*, марита *S. mansoni*, шистосомула *S. mansoni* через 3 часа после заражения, шистосомула *T. szidati*), «регуляция клеточного ответа на стимул фактора роста» (regulation of cellular response to growth factor stimulus, GO:0090287) (редия *P. simillimum*, марита *S. mansoni*, шистосомулы *T. regenti* и *T. szidati*), «морфогенез мышечной ткани» (muscle tissue morphogenesis, GO:0060415) (редия *P. simillimum*, марита и шистосомула *S. mansoni* через 3 часа после заражения, шистосомула *T. regenti*), «митохондриальный синтез АТФ, связанный с транспортом электронов» (mitochondrial ATP synthesis coupled electron transport, GO:0042775) (церкария *F. gigantica*, яйцо и 21-дневная ювениль *F. hepatica*, марита *P. simillimum*), «транsepителиальный транспорт» (transepithelial transport, GO:0070633) (марита и шистосомула *S. mansoni* через 3 часа после заражения, шистосомула *T. regenti*).

Подавляющее большинство биологических процессов (92 из 125) активно только на одной из рассмотренных фаз сложных жизненных циклов дигеней. Список таких «специфичных» биологических процессов включал: «регуляция пролиферации нейробластов» (regulation of neuroblast proliferation, GO:1902692) (яйцо *F. gigantica*), «процесс на основе актиновых филаментов» (actin filament-based process, GO:0030029) (церкария *F. gigantica*), «созревание в развитии» (developmental maturation, GO:0021700) (яйцо *F. hepatica*), «движение клетки или субклеточного компонента» (movement of cell or subcellular component, GO:0006928) (21-дневная ювениль *F. hepatica*), «регуляция морфогенеза органов животного» (regulation of animal organ morphogenesis, GO:2000027) (редия *P. simillimum*), «трансляционная элонгация» (translational elongation, GO:0006414) (марита *P. simillimum*), «женская беременность» (female pregnancy, GO:0007565) (марита *S. mansoni*), «развитие мышечных структур» (muscle structure development, GO:0061061) (шистосомула *S. mansoni* через 3 часа после заражения), «дыхательная электрон-транспортная цепь» (respiratory electron transport chain, GO:0022904) (церкария *T. regenti*), «положительная регуляция развития нервной системы» (positive regulation of nervous system development, GO:0051962) (шистосомула *T. szidati*), «митохондриальный транспорт электронов, НАДН в убихинон» (mitochondrial electron transport, NADH to ubiquinone, GO:0006120) (церкария *T. szidati*), «регуляция развития нервной системы» (regulation of nervous system development, GO:0051960) (шистосомула *T. szidati*). Обогащение биологических процессов, связанных с развитием, выявлено у яйца и церкарии *F. gigantica*, яйца *F. hepatica*, редии *P. simillimum*, мариты и шистосомулы *S. mansoni* через 3 часа после заражения, шистосомул *T. regenti* и *T. szidati*.

Далее с целью описать и сравнить профили активности филострат в ходе реализации жизненных циклов дигеней, вычисленные значения уровней экспрессии филострат были преобразованы в относительные величины. Наибольший уровень экспрессии был взят за «1», наименьший – за «0», а все остальные значения экспрессии были преобразованы в этом диапазоне. Большинство филострат демонстрировали наименьшие относительные уровни экспрессии на фазах яиц *F. gigantica* (15 из 15 филострат) (Рисунок 18A2,3) и *F. hepatica* (14 из 15) (Рисунок 18B2,3), марит *P. simillimum* (13 из 13) (Рисунок 18C2,3) и *S. mansoni* (10 из 14) (Рисунок 18D2,3), а также церкарий *T. regenti* (14 из 14) (Рисунок 18E2,3) и *T. szidati* (14 из 14) (Рисунок 18F2,3). В то же время, наибольший относительный уровень экспрессии у большинства филострат обнаружен на фазах метацеркарии *F. gigantica* (14 из 15 филострат) (Рисунок 18A2,3), ювенили *F. hepatica* (11 из 15) (Рисунок 18B2,3), редии *P. simillimum* (8 из 13) (Рисунок 18C2,3), шистосомулы *S. mansoni* через 3 часа после заражения (11 из 14) (Рисунок 18D2,3), шистосомул *T. regenti* (14 из 14) (Рисунок 18E2,3) и *T. szidati* (14 из 14) (Рисунок 18F2,3). Для проверки статистической значимости различий в экспрессии различных филострат на рассмотренных фазах сложного жизненного цикла, все филостраты были разделены на две группы относительно того, сформировались они «до» (Рисунок 18A2, B2, C2, D2, E2, F2) или «после» (Рисунок 18A3, B3, C3, D3, E3, F3) расхождения различных видов дигеней. Статистически значимые различия в уровнях экспрессии между выделенными группами филострат обнаружены у ювенили *F. hepatica* через 24 часа после выхода из цисты, редии *P. simillimum* и шистосомулы *S. mansoni* через 24 часа после заражения хозяина.

Глава 4 Обсуждение

Дигенетические сосальщики или дигенеи представляют собой одну из крупнейших групп паразитических плоских червей. Сложный жизненный цикл дигеней с последовательным и закономерным чередованием половых поколений является ярким примером формирования и функционирования множества фенотипов путем регуляции работы одного генома. Тем не менее несмотря на долгую историю изучения молекулярных основ реализации сложных жизненных циклов дигеней, вопрос разделения единого генома между разными поколениями и отдельными фазами со специфичным клеточным составом, гистологическим и морфологическим строением, физиологическими особенностями, а также характерными чертами поведения и образа жизни, остается открытым.

Определить набор генов, активных на различных этапах жизненного цикла, позволяет транскриптомный анализ, успешно применяемый в исследовании дигеней на протяжении последних 20 лет. 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, который может быть использован в качестве маркера рассматриваемого фенотипа, используют термин «молекулярная подпись». В зависимости от выбранных исследователями 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 анализу транскриптомов дигеней, в качестве молекулярной подписи фазы сложного жизненного цикла рассматривается именно набор дифференциально экспрессирующихся генов (например, Zhang X.-X. et al., 2019; Leontovyc et al., 2016, 2019). Для определения дифференциальной экспрессии генов проводят попарное сравнение фаз одного жизненного цикла. Как следствие, характеристики выявленной молекулярной подписи рассматриваемой фазы жизненного цикла зависят от того, с какой фазой было проведено 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 и одной из сложностей для определения эволюционно консервативных компонентов молекулярных подписей фаз различных видов дигеней.

Исследование молекулярных основ реализации сложного жизненного цикла, как общего признака для целого таксона Digenea, осложнено также тем, что высокая эволюционная

пластичность дигеней нашла отражение в разнообразии их жизненных циклов. Многочисленные виды дигенетических сосальщиков различаются не только по морфологическим и физиологическим признакам фаз жизненных циклов, но и условиями обитания, количеством и особенностями их хозяев. 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 представители Psilostomatidae, *P. simillimum* и *S. pseudoglobulus* (Nesterenko et al., 2020). Во-первых, оба вида используют пресноводных переднежаберных моллюсков *B. tentaculata* в качестве промежуточных хозяев, что делает возможным содержание паразитов в лабораторных условиях. Во-вторых, дочерние партеногенетические поколения у *P. simillimum* и *S. pseudoglobulus* представлены редиями. Обитая в первом промежуточном хозяине и размножаясь партеногенезом, редия представляет собой одну из ключевых фаз сложного жизненного цикла дигеней. Исследование транскриптома редии важно как для изучения взаимодействия паразита с хозяином, так и для определения молекулярных основ усложнения жизненного цикла последнего общего предка дигеней. Тем не менее до момента проведения исследования транскриптом редии был опубликован только для *F. gigantea*. В-третьих, мариты *P. simillimum* и *S. pseudoglobulus* способны достигать половой зрелости не только в водоплавающих птицах, но и в цыплятах, что способствует постановке жизненных циклов обоих видов в лабораторных условиях. В-четвертых, укороченные матки марит *P. simillimum* и *S. pseudoglobulus* облегчают транскриптомный анализ половозрелых особей, так как значительно снижают контаминацию образцов транскриптомов марит материалом развивающихся мирацидиев. В-пятых, общие схемы реализации жизненных циклов обоих рассмотренных видов различаются. В отличие от *P. simillimum* с диксенным жизненным циклом, жизненный цикл *S. pseudoglobulus* является триксенным. Мы предполагаем, что изменение числа хозяев в жизненном цикле требует от паразита перераспределения процессов метаморфоза и развития между хозяевами, что, вероятно, должно отразиться и на экспрессии генов, связанных с указанными биологическими процессами. Принимая во внимание перечисленные особенности *P. simillimum* и *S. pseudoglobulus*, мы можем утверждать, что результаты анализа транскриптомов фаз псилостоматид являются важным дополнением в исследовании молекулярных основ реализации сложных жизненных циклов дигеней (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. В каждой из библиотек больше 90 процентов

коротких парных прочтений успешно прошло контроль качества. На основании полученных результатов мы можем утверждать, что транскриптомы редий, церкарий и марит *P. simillimum* и *S. pseudoglobulus*, полученные в нашем исследовании, характеризуются высоким качеством (Nesterenko et al., 2020). Более того, представленные транскриптомы являлись первыми не только для указанных видов, но и для всего таксона Psilostomatidae (Nesterenko et al., 2020).

Принимая во внимание отсутствие «справочных» геномов нами было принято решение проводить сборку протяженных последовательностей *de novo* с одновременным использованием библиотек коротких парных прочтений с трех фаз жизненного цикла. В отличие от стратегии, в которой для каждого из рассматриваемых образцов производится независимая сборка транскриптома, выбранный нами подход позволяет компенсировать возможный недостаток информации в одних образцах данными из других. Объединение библиотек коротких прочтений также не препятствует успешной сборке транскриптов, специфичных только для одной из исследуемых фаз жизненного цикла, ввиду особенностей алгоритмом построения и прохождения графов де Брюйна, используемых в анализе. В то же время нельзя исключать вероятность того, что при увеличении количества данных возрастает шанс объединения в единую последовательность изначально разных фрагментов. Принимая это во внимание, мы ввели дополнительный этап контроля качества и полноты сборки протяженных последовательностей в наш протокол биоинформатического анализа данных. Контроль был осуществлен с помощью программы TransRate и, согласно полученным данным, как у *P. simillimum*, так и у *S. pseudoglobulus* более 95% ассемблированных протяженных последовательностей характеризовались высокими показателями качества и полноты сборки (Nesterenko et al., 2020). Из этого следует, что 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. Ввиду отсутствия в открытом доступе геномных данных *P. simillimum* и *S. pseudoglobulus*, для определения вероятных наборов генов исследуемых видов дигеней мы использовали таблицу, созданную программой Trinity во время *de novo* сборки транскриптомов. В таблице содержалась информация о соответствии собранных последовательностей «генам». Программа Trinity использует термин «ген» для обозначения совокупности коротких парных прочтений, объединённых на основании сходства их последовательностей. Сборка протяженных последовательностей для каждого «гена» производилась отдельно от других выявленных объединений. Исходя из того, что сравнительный анализ последовательностей, проведенный

Trinity на начальных этапах сборки транскриптома, успешно определил достаточное количество различий для разделения последовательностей по «генам», мы считаем возможным использовать полученную информацию для определения вероятного набора генов исследуемого организма.

С целью уточнить списки предположительных генов *P. simillimum* и *S. pseudoglobulus* мы ввели два дополнительных критерия: 1) наличие заметного уровня экспрессии и 2) кодирование протяженного белка. Согласно первому критерию, как минимум на одной из рассмотренных фаз жизненных циклов ген должен был иметь уровень экспрессии не меньше 1 транскрипта на миллион проанализированных. При расчете значения уровня экспрессии в транскриптах-на-миллион сначала учитывается длина гена, а затем производится нормализация на глубину секвенирования (Wagner et al., 2012). Полученные нормализованные значения позволяют проводить корректное сравнение уровней экспрессии генов между образцами, что активно применяется в современных исследованиях. Выбор порога минимальной экспрессии для включения гена в анализ определяется только целью исследования. Для нас наиболее важно было охватить весь набор активных генов, даже если небольшая часть генов будет ошибочно классифицирована, как имеющие биологически заметную экспрессию, поэтому мы использовали в качестве порога минимальной экспрессии значение равное 1 транскрипту-на-миллион (Nesterenko et al., 2020). В то же время мы ввели второй критерий, согласно которому, как минимум один из собранных транскриптов отобранного гена должен был кодировать белок длиной не меньше 100 аминокислот. При выборе минимального размера кодируемого белка мы приняли во внимание, что размеры белков животных варьируют в широких пределах и средний размер белка у Eumetazoa составляет 486 аминокислот (Tiessen et al., 2012), а белковые домены в среднем имеют длину 100 аминокислот (Lin, Zewail, 2012). Следовательно, использование 100 аминокислот в качестве порога минимальной длины кодируемого белка, предположительно, позволяет охватить большую часть набора белков двух исследуемых видов дигеней.

Применение критериев минимальной экспрессии гена и длины кодируемого белка позволило нам подготовить «справочные» наборы генов *P. simillimum* и *S. pseudoglobulus*, которые могут служить основой для дальнейшего анализа молекулярных основ реализации сложных жизненных циклов дигеней. Обновленные наборы включали 21433 (*P. simillimum*) и 46424 (*S. pseudoglobulus*) генов. В обоих случаях количество генов, кодирующих белки длиной равной или большей 100 аминокислот, превышало число подобных генов у 10 рассмотренных видов дигеней (*C. sinensis*, *F. gigantea*, *F. hepatica*, *O. felineus*, *O. viverrini*, *S. haematobium*, *S. japonicum*, *S. mansoni*, *T. regenti* и *T. szidati*). Выявленные различия могут быть объяснены избыточностью транскриптомных данных, на основе которых и было проведено определение вероятного набора генов двух исследованных Psilostomatidae. В то же время мы не можем полностью исключать возможное отсутствие части активных генов в полученных данных. Без

высококачественных сборок геномов *P. simillimum* и *S. pseudoglobulus* определить точное количество белок-кодирующих генов у обоих видов не представляется возможным. Тем не менее учитывая высокие показатели качества полученных сборок, мы можем предположить, что в подготовленных «справочных» наборах представлено подавляющее большинство генов двух исследованных видов псилостоматид (Nesterenko et al., 2020).

С целью определить полноту подготовленных нами наборов генов псилостоматид, мы провели поиск в «справочных» наборах белков *P. simillimum* и *S. pseudoglobulus* последовательностей ортологов, присутствующих в геномах Metazoa в количестве одной копии. В анализе была использована девятая версия базы данных Metazoa odb, содержащая последовательности ортологов, необходимых для работы различных биологических процессов в клетках, в том числе гены рибосомальных и трансмембранных белков. Результаты проведенного сравнения с базой данных продемонстрировали, что количество успешно собранных однокопийных ортологов у *P. simillimum* и *S. pseudoglobulus* сопоставимо с количеством ортологов, обнаруженных у 10 видов дигеней (Рисунок 3) (Nesterenko et al., 2020). Часть ортологов, отсутствующих у исследованных псилостоматид, не найдены и у других рассмотренных видов дигеней, что вероятно связано с паразитическим образом жизни Digenea (Nesterenko et al., 2020). Возможно, потеря данных генов произошла у последнего общего предка дигеней в связи с возложением утраченных паразитом функций на хозяина. Несмотря на всю важность полученных результатов, подходить к их интерпретации следует с осторожностью. С одной стороны, мы можем сделать вывод, что выбранный нами подход к анализу данных позволил успешно собрать последовательности ортологов, важных для Metazoa. Однокопийные ортологи из базы данных Metazoa odb, вероятнее всего, характеризуются стабильной экспрессией на всем протяжении жизненного цикла. Объединение библиотек коротких прочтений транскриптомов с трех различных фаз жизненного цикла дигеней, предположительно, способствовало восстановлению полноразмерных последовательностей обнаруженных ортологов. С другой стороны, результаты сравнения с базой данных указывают только на наличие, количество и полноту сборки последовательностей ортологов в исследуемом наборе, но не предоставляют никакой информации о качестве сборки других последовательностей. Тем не менее такой подход позволят сравнивать разные наборы геномных, транскриптомных и протеомных данных. Более того, в отличие от технических характеристик сборок, таких как N50 (наименьшая из длин последовательностей, включивших 50% нуклеотидов, рассмотренных при сборке) или минимального размера собранной последовательности, результаты сравнения с базой данных однокопийных ортологов предоставляют также функциональную информацию об анализируемых последовательностях, в том числе отсутствующих или дублированных.

В роли дополнительных критериев оценки качества сборки транскриптома и определения наборов белок-кодирующих генов могут быть рассмотрены результаты сравнения с базами данных нуклеотидных и/или аминокислотных последовательностей. В этом случае двумя основными характеристиками являются количество последовательностей, имеющих значимые совпадения с базой данных, а также наличие в рассматриваемом наборе данных консервативных последовательностей. Как у *P. simillimum*, так и у *S. pseudoglobulus* более 73% генов успешно аннотировано, то есть для большинства генов из «справочных» наборов определена функциональная роль и/или установлена доменная архитектура кодируемого белка. Отсутствие аннотации у оставшихся генов означает лишь то, что для их последовательностей не удалось найти совпадений с данными в рассмотренных базах NCBI, SwissProt, eggNOG и PfamA. Включение большого числа баз данных в анализ или добавление близких последовательностей в перечисленные базы должно улучшить результаты аннотации. Принимая во внимание, что в настоящем исследовании были использованы одни из самых крупных баз данных, а при поиске гомологов были выставлены «чувствительные» параметры, мы можем утверждать, что неаннотированные последовательности могут быть рассмотрены в качестве специфичных или ограниченных таксонами *P. simillimum* и *S. pseudoglobulus* или Psilostomatidae.

При интерпретации результатов поиска консервативных последовательностей в рассматриваемых наборах данных важно учитывать особенности биологии исследуемого организма. В качестве яркого примера может быть рассмотрен пул генов, связанных с биосинтезом жирных кислот. Результаты проведенного нами анализа показали, что как у *P. simillimum*, так и у *S. pseudoglobulus*, присутствуют 7 ферментов, участвующих в биосинтезе жирных кислот. Одним из обнаруженных ферментов являлась синтаза жирных кислот FasN. Несмотря на то, что для последовательностей обоих псилостоматид удалось обнаружить совпадения с последовательностью FasN, доменная архитектура гомологов у *P. simillimum* и *S. pseudoglobulus* различалась. Наличие лишь одного из ключевых белковых доменов синтазы жирных кислот у обоих Psilostomatidae не позволяет рассматривать последовательности FasN из базы данных eggNOG и у *P. simillimum*, и *S. pseudoglobulus* в качестве полноценных гомологов (Nesterenko et al., 2020). Основываясь на полученных результатах, можно сделать вывод о том, что у исследованных псилостоматид, как и у других видов дигеней, нарушен биосинтез жирных кислот (Nesterenko et al., 2020). Отсутствие в «справочных» наборах *P. simillimum* и *S. pseudoglobulus* генов ферментов, необходимых для полного биосинтеза жирных кислот, вероятнее всего, связано с паразитическим образом жизни дигеней и не является следствием выбранной стратегии сборки транскриптома и определения белок-кодирующих генов. Обобщая полученные результаты поиска однокопийных ортологов и сравнения с базами данных, мы утверждаем, что подготовленные наборы 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.

В отличие от *P. simillimum*, прямое сравнение *S. pseudoglobulus* с другими видами плоских червей осложнено возможным наличием материала криптических видов *Sphaeridiotrema* в анализируемых образцах. Предположение о наличии смеси криптических видов *Sphaeridiotrema* в наших данных основано на обнаружении двухкратного различия в числе выявленных генов у *P. simillimum* и *S. pseudoglobulus* и необычно высокого количества дубликатов однокопийных ортологов Metazoa (Рисунок 3). Наличие дубликатов 53% однокопийных ортологов выявлено и в данных по *F. gigantica* (Рисунок 3), но о возможном присутствии контаминации в оригинальной публикации Чжана с соавторами (Zhang X.-X. et al., 2019) не упоминается. С целью проверить предположение о смеси видов *Sphaeridiotrema* в образцах мы провели поиск и анализ последовательностей основной субъединицы цитохром с-оксидазы. Исследование последовательности указанного гена широко используется в проектах, направленных на определение видовой принадлежности рассматриваемого организма. В полученном нами «справочном» наборе последовательностей *S. pseudoglobulus* удалось обнаружить две различающиеся копии гена цитохром с-оксидазы. Реконструкция филогенетического дерева доказала, что одна из копий гена отличается от последовательностей гена цитохром с-оксидазы как *S. pseudoglobulus*, так и *S. globulus*, но кластеризуется совместно с ними (Рисунок 4) (Nesterenko et al., 2020). Основываясь на том, что вид *S. pseudoglobulus* изначально описан, как крайне схожий с космополитичным видов *S. globulus* (McLaughlin et al., 1993), мы считаем, что вторая выявленная копия гена цитохром с-оксидазы принадлежит близкому к *S. globulus* виду дигеней, информация о котором отсутствует в базе данных GenBank (Nesterenko et al., 2020). Преобладание количества генов с редия- и марито-специфичной экспрессией, выявленное в результате сравнительного анализа молекулярных подписей фаз сложного жизненного цикла *S. pseudoglobulus* (Рисунок 9B), дает основания предполагать, что контаминация образцов произошла при сборе материала паразитических фаз (Nesterenko et al., 2020). Принимая во внимание все полученные результаты, мы пришли к выводу, что данные по *S. pseudoglobulus* могут быть использованы в дальнейшем анализе, но только для выявления общих тенденций.

С целью определить сходства и различия в наборах белок-кодирующих генов как паразитических, так и свободноживущих видов плоских червей, мы провели поиск и построение групп ортологов. Мы использовали два разных подхода, OrthoFinder и OMA standalone, каждый из которых определил наличие большого числа общих ортогрупп для представителей Psilostomatidae и *Fasciola*. Полученные результаты указали на то, 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 дигеней. В основном все рассмотренные виды плоских

червей объединялись в кластеры в соответствии с современными представлениями об их филогенетическом родстве, вне зависимости от того, были ли рассмотрены ортогруппы, построенные OrthoFinder (Рисунок 5) (Nesterenko et al., 2020) или OMA standalone (Рисунок 8) (Nesterenko et al., 2022). Выделение программой OrthoFinder вида *S. mansoni* в качестве сестринского таксона ко всем исследованным дигенейам (Рисунок 5) являлось следствием использования устаревшей версии аннотации генома шистосомы. Обновление данных позволило улучшить результаты поиска ортогрупп и, как следствие, расположить *S. mansoni* совместно с другими представителями Schistosomatidae (Рисунок 8).

На результаты поиска ортогрупп с помощью OrthoFinder предположительно повлияло и наличие смеси видов *Sphaeridiotrema* в анализируемых данных. Результаты сравнения с базой данных однокопийных ортологов продемонстрировали, что многие гены из базы данных Metazoa_odb присутствуют в числе нескольких копий в «справочном» наборе *S. pseudoglobulus* (Рисунок 3). Мы можем предположить, что копии одного гена, схожие по своим нуклеотидным последовательностям, но принадлежащие криптическим видам, вероятнее всего, войдут в состав одной группы ортологов. Проведенный анализ показал, что во многих ортогруппах присутствуют несколько генов, которые могут быть как паралогами, так и генами криптических видов. В настоящее время разделить схожие последовательности не представляется возможным, так как мы не обладаем достаточной информацией и необходимыми данными для определения видовой принадлежности генов, вошедших в одну ортогруппу. Мы предполагаем, что наиболее явно присутствие видоспецифичных последовательностей могло проявиться в увеличении числа ортогрупп, уникальных для *S. pseudoglobulus*. Для сравнения, у *P. simillimum* обнаружено 13 видоспецифичных групп ортологов, а у *S. pseudoglobulus* – 43 ортогрупп (Рисунок 6). Количество вошедших в состав специфичных ортогрупп генов у двух рассмотренных псилостоматид различалось практически в 13 раз: 34 гена у *P. simillimum* и 435 у *S. pseudoglobulus*. С одной стороны, выявленные отличия могут быть связаны с особенностями схем реализации сложных жизненных циклов *P. simillimum* и *S. pseudoglobulus*, в том числе и с различиями в числе промежуточных хозяев. С другой стороны, часть генов, включенных в состав специфичных ортогрупп, вероятно, является молекулярными различиями между криптическими видами *Sphaeridiotrema*. Ввиду отсутствия функциональной аннотации у большинства генов, входящих в состав видоспецифичных ортогрупп, установить в каких биологических процессах они принимают участие в настоящее время затруднительно. Следует также учитывать, что часть видоспецифичных генов, вероятно, не включена в состав ни одной из построенных ортогрупп, если для их последовательностей не выявлено сходства с последовательностями других генов. В то же время на тепловой карте (Рисунок 5) отчетливо видно, что *S. pseudoglobulus* кластеризуется совместно с *P. simillimum* и для них обнаружено наибольшее число общих ортогрупп.

Полученный результат может быть связан как с филогенетическим родством *P. simillimum* и представителями *Sphaeridiotrema*, так и с тем, что в сравнительном анализе учитывалось количество общих ортогрупп, но не их состав. Принимая во внимание, что исследование состава ортогрупп важно не только для детального сравнения разных видов, но и для построения моделей геномов предков дигеней, мы приняли решение не рассматривать *S. pseudoglobulus* при повторном поиске ортогрупп с помощью OMA standalone. Тем не менее представленные данные по *S. pseudoglobulus* станут ценным дополнением в исследовании разнообразия Psilostomatidae.

Независимо от того входит или нет ген в одну из выявленных ортогрупп, для оказания заметного влияния на организм ему необходимы специфичное окружение, в том числе подходящие физико-химические условия, формируемые в клетке, и совместное действие множества других генов (Orgogozo et al., 2015). Мы можем предположить, что степень перекрытия между списками активных генов для фаз сложного жизненного цикла будет коррелировать со сходством фаз на физиологическом и морфологическом уровне (Nesterenko et al., 2020). Количество «общих» активных генов будет тем больше, чем сильнее схожи сравниваемые фазы жизненного цикла между собой (Nesterenko et al., 2020). Основную часть пересечений между множествами будут, вероятно, составлять последовательности, необходимые для функционирования организма, такие как гены домашнего хозяйства, сигнальных каскадов и факторов транскрипции (Nesterenko et al., 2020). В случае, если сравниваемые фазы принадлежат одному поколению жизненного цикла, и, как следствие, их клеточный состав схож, это также должно отразиться в увеличении числа «общих» генов (Nesterenko et al., 2020). В то же время каждая фаза жизненного цикла характеризуется наличием специфичных морфологических структур и адаптаций, способствующих её обитанию в определенных условиях. Связанные с такими признаками гены будут составлять уникальную часть списка активных генов (Nesterenko et al., 2020).

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 и 2 транскриптам-на-миллион. Первый из них, как уже было указано выше, позволяет наиболее широко охватить набор активных генов и был использован нами для предварительного внутривидового анализа молекулярных подписей *P. simillimum* и *S. pseudoglobulus*. Второй же порог, согласно математической модели (Wagner et al., 2013), предоставляет возможность отобрать наиболее активно транскрибируемые гены. Порог минимальной экспрессии в 2 транскрипта-на-миллион был использован нами для уточнения результатов определения молекулярных подписей фаз *P. simillimum* и для их сравнительного анализа с молекулярными подписями фаз двух «редииоидных» (*F. gigantea* и *F. hepatica*) и трех «спороцистоидных» (*S. mansoni*, *T. regenti* и *T. szidati*) видов дигеней.

Построение диаграмм Венна для молекулярных подписей редий, церкарий и марит *P. simillimum* (Рисунок 9А) и *S. pseudoglobulus* (Рисунок 9В) позволило визуализировать взаимоотношения между множествами активных генов. Удалось показать, что молекулярная подпись любой из рассмотренных фаз жизненного цикла дигеней может быть поделена на три части (Nesterenko et al., 2020). Первую часть молекулярной подписи составляли гены, активные на трех контрастных фазах жизненного цикла. Ко второй части молекулярной подписи мы отнесли гены, экспрессирующиеся на двух фазах жизненного цикла. Пересечения между молекулярными подписями двух фаз жизненного цикла, вероятно, вызвано наличием эмбрионов церкарий в редиях, паразитическим образом жизни редий и марит, а также принадлежностью церкарий и марит амфимиктического поколению (Nesterenko et al., 2020). Гены с фазо-специфичной экспрессией могут быть рассмотрены в качестве третьей части молекулярной подписи фазы жизненного цикла. Количество генов, включенных в состав выделенных частей молекулярных подписей фаз, значительно различалось у двух исследованных видов псилостоматид. В отличие от *S. pseudoglobulus*, у *P. simillimum* большая часть генов из «справочного» набора вошла одновременно в состав молекулярных подписей как редии, так и церкарии с маритой. У выявленных межвидовых различий может быть несколько причин. С одной стороны, молекулярная подпись, как и весь транскриптом, частью которого она является, отражает состояние фазы жизненного цикла в конкретный момент времени и в определенных условиях. Это означает, что у схожих фаз жизненных циклов филогенетически близких видов может различаться набор активных генов, если условия, воздействующие на сравниваемые фазы, отличаются. С другой стороны, наличие материала криптических видов *Sphaeridiotrema* в образцах может оказывать влияния на результаты сравнения молекулярных подписей фаз жизненного цикла. В случае, если ортологичные последовательности по результатам анализа экспрессии вошли в состав не всех молекулярных подписей, мы будем наблюдать увеличение числа генов, специфичных для одной фазы и/или общих для двух из трех рассмотренных фаз сложного жизненного цикла. Мы не исключаем, что в основе выявленных различий между *P. simillimum* и *S. pseudoglobulus* лежат как пластичность молекулярных подписей фаз, так и наличие смеси близких видов *Sphaeridiotrema* в образцах (Nesterenko et al., 2020). Тем не менее важно учитывать, что сбор материала и выделение образцов транскриптомов у обоих исследованных видов дигеней производились в одних и тех же лабораторных условиях, а биоинформатический анализ всех полученных данных был проведен по единому протоколу. Принимая это во внимание, мы считаем, что определяющее влияние на определение молекулярных подписей фаз жизненного цикла *S. pseudoglobulus* оказало именно наличие смеси видов. Преобладание количества «специфичных» генов в молекулярных подписях редий и марит *S. pseudoglobulus* в сравнении с церкарией дают основания предполагать, что образцы

паразитических фаз содержали материал криптических видов. В завершение обсуждения результатов первичного анализа молекулярных подписей фаз сложных жизненных циклов псилостоматид, важно сделать несколько обобщений. Во-первых, в ходе реализации сложных жизненных циклов *P. simillimum* и *S. pseudoglobulus* происходит изменение экспрессии множества генов, что приводит к формированию фаза-специфичных молекулярных подписей (Nesterenko et al., 2020). Во-вторых, большая часть генов, входящих в состав молекулярных подписей редий, церкарий и марит, активна на нескольких фазах одного жизненного цикла (Nesterenko et al., 2020).

В качестве яркого примера изменения активности генов в ходе реализации сложных жизненных циклов дигеней могут быть рассмотрены профили экспрессии гомеобоксных генов (Рисунок 10). Результаты проведенного анализа доказали, что большая часть гомеобокс-содержащих генов демонстрируют повышение экспрессии на фазах редии и церкарии, а также её снижение на фазе мариты. Повышенная экспрессия гомеобокс-содержащих генов на фазе редии жизненных циклов *P. simillimum* и *S. pseudoglobulus* может быть связана с эмбриональным развитием потомства, протекающим в половозрелых особях партеногенетического поколения (Nesterenko et al., 2020). Объяснить повышенную экспрессию гомеобоксных генов на фазе церкарии у обоих рассмотренных видов дигеней несколько сложнее. С одной стороны, выделение образцов транскриптома производилось из активно плавающих церкарий, покинувших организм моллюска. Мы можем ожидать, что у таких личинок уже сформированы все провизорные признаки, имеющие адаптивные функции, то есть активность гомеобоксных генов не может быть связана с их развитием. С другой стороны, церкария является личинкой мариты и содержит зачатки систем органов взрослого организма. Мы можем предположить, что повышенная экспрессия гомеобокс-содержащих генов, вовлеченных в регуляцию развития, является проявлением гетерохронии, т. е. задержки в развитии одних систем органов при ускоренном формировании других. В то же время важно учитывать, что в отличие от *P. simillimum*, жизненный цикл *S. pseudoglobulus* является триксенным. Наличие второго промежуточного хозяина в жизненном цикле позволяет перенести формирование дефинитивной организации систем органов на фазу метацеркарии, где указанный процесс может протекать за счет энергетических ресурсов хозяина. Отсутствие транскриптомных данных адолескарии *P. simillimum* и метацеркарии *S. pseudoglobulus* препятствует проверке предположения о наличии различий в экспрессии генов, связанных с развитием систем органов марит, у филогенетически близких видов с ди- и триксенным сложным жизненным циклом. Тем не менее мы можем предположить, что на момент выделения образцов транскриптомов из половозрелых особей амфимиктического поколения большая часть процессов развития завершена ввиду отсутствия и слабой экспрессии большинства идентифицированных гомеобоксных генов на фазе мариты

P. simillimum и *S. pseudoglobulus*. На тепловой карте (Рисунок 10) отчетливо видна также согласованность экспрессии гомеобокс-содержащих генов у *P. simillimum* и *S. pseudoglobulus*. Мы предполагаем, что совпадение профилей экспрессии гомеобоксных генов является следствием эволюционной консервативности системы регуляции формирования (редия) и переключения (церкария) молекулярных подписей фаз псилостоматид (Nesterenko et al., 2020).

С целью определить сходства и различия между молекулярными подписями фаз не только филогенетически близких, но и отдаленных видов дигеней, мы включили в анализ транскриптомные данные *F. gigantica*, *F. hepatica*, *S. mansoni*, *T. regenti* и *T. szidati*. На момент проведения анализа качественные транскриптомы партеногенетических фаз сложных жизненных циклов дигеней были опубликованы только для представителей *Fasciola*, тогда как у трех оставшихся видов, относящихся к Schistosomatidae, библиотеки коротких парных прочтений были доступны только для фаз амфимиктического поколения. Повышенный интерес к анализу транскриптомов фаз амфимиктического поколения может быть объяснен тем, что полученные результаты предоставляют возможность не только определить молекулярные основы взаимодействия паразита с окончательным хозяином, но и обнаружить перспективные мишени для лекарственных препаратов. В то же время опубликованные транскриптомы представляют большой интерес и для исследования молекулярных основ реализации сложных жизненных циклов дигеней. Для исключения возможного влияния небологических факторов на результаты анализа мы провели подготовку библиотек прочтений транскриптомов фаз *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati* с помощью единого протокола. Проведенный контроль качества обнаружил, что образцы различались по характеристикам не только между исследованиями, но даже в пределах одного вида. Тем не менее большая часть коротких парных прочтений соответствовали установленным требованиям по минимальной длине и точности определения нуклеотидов. Использование в анализе обширной базы данных, включающей в том числе и геномы хозяев исследуемых видов, позволило найти и устранить предполагаемую контаминацию из библиотеках прочтений (Nesterenko et al., 2022). На основании полученных результатов мы можем утверждать, что подготовленные данные могут быть использованы для определения и сравнения молекулярных подписей фаз сложных жизненных циклов дигеней.

Для сравнительного межвидового анализа мы повысили порог минимальной экспрессии гена для включения в молекулярную подпись с 1 до 2 транскриптов-на-миллион. Согласно регрессионной модели, построенной Вангером, Кином и Линчем (Wagner et al., 2013), гены с экспрессией больше 2 транскриптов-на-миллион наиболее вероятно являются активно транскрибируемыми. Результаты проведенного нами анализа продемонстрировали, что в состав молекулярной подписи каждой из рассмотренных фаз сложных жизненных циклов дигеней

вошло не меньше шести тысяч генов. Более того, у 5 из 6 исследованных видов дигеней большая часть белок-кодирующих генов демонстрировала заметную экспрессию на всех рассмотренных фазах сложных жизненных циклов (Nesterenko et al., 2022). 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.

Наименьшими размерами молекулярных подписей обладали яйца (*F. gigantica*, *F. hepatica*), церкарии (*T. regenti*, *T. szidati*) и мариты (*P. simillimum*, *S. mansoni*) (Таблица 2). Наибольшее количество генов вошло в молекулярные подписи редии *P. simillimum*, метацеркарии *F. gigantica*, ювенили *F. hepatica*, а также шистосомул *S. mansoni*, *T. regenti* и *T. szidati* (Таблица 2). Выявленные различия в размерах молекулярных подписей фаз отражают изменения транскрипционной активности геномов дигеней в ходе реализации сложных жизненных циклов паразитов. Начать обсуждение следует с яиц *Fasciola* и церкарий *Trichobilharzia*, которые после покидания организма хозяина попадают в окружающую среду, где вынуждены эффективно использовать накопленные ранее ограниченные ресурсы. Находящиеся в яйцах *F. gigantica* и *F. hepatica* запасы идут на клеточный рост и эмбриональное развитие мирацидиев, которые протекают под защитой оболочек яйца в окружающей среде. Исследованные транскриптомы выделены из яиц, собранных в желчном пузыре окончательного хозяина, из чего мы можем сделать вывод, что в них представлен сигнал преимущественно от яйцеклеток и желточных клеток. Церкарии *T. regenti* и *T. szidati* используют запасенные вещества для активного плавания в воде, поиска окончательного хозяина и проникновения в него. Мы предполагаем, что относительно небольшие размеры молекулярных подписей яиц и церкарий связаны с подавлением экспрессии генов, не вовлеченных в описанные биологические процессы. Напротив, увеличение размеров молекулярных подписей редии *P. simillimum*, метацеркарии *F. gigantica*, ювенили *F. hepatica*, а также шистосомул *S. mansoni*, *T. regenti* и *T. szidati*, вероятно, вызвано активностью биологических процессов, участвующих в эмбриональном развитии потомства (редия) и метаморфозе (метацеркария, шистосомула, ювениль). Принимая во внимание, что процессы развития требуют не только экспрессии множества генов, но и большого количества энергии, мы можем ожидать, что на указанных фазах жизненного цикла происходит усиление метаболизма, что, предположительно, влияет на размеры их молекулярных подписей. В завершении обсуждения размеров молекулярных подписей важно обратить внимание на то, что при изменении параметров анализа или набора исследуемых фаз, результаты сравнения количества генов, входящих в молекулярные подписи, могут значительно измениться. Классификация молекулярных подписей по размерам на «наименьшие» и «наибольшие»

относительная и зависит от того, с какими фазами проводится сравнение. Так, к примеру, в настоящем исследовании наименьшими размерами молекулярных подписей у *P. simillimum* и *S. mansoni* обладали мариты, несмотря на то что в них вошло больше генов, чем в наибольшие из молекулярных подписей *Fasciola* и *Trichobilharzia*. Необходимо также отметить, что, как уже было показано ранее, молекулярные подписи контрастных фаз пересекаются и ключевое значение имеют уровни экспрессии генов (Nesterenko et al., 2020, 2022).

Для определения статистически значимых различий в уровнях экспрессии генов мы использовали новый подход, имплементированный в библиотеке RNentropy. Большинство популярных методов при поиске дифференциальной экспрессии проводит попарное сравнение образцов. Это означает, что будет ли ген классифицирован, как «дифференциально экспрессирующийся», зависит от того, какой образец использован в качестве «контроля». Как следствие, сравнение результатов разных исследований будет осложнено, если при проведении анализов были использованы разные «контроли» или параметры анализа. Принимая во внимание, что вывод о понижении или повышении уровня экспрессии всегда основан на результатах сравнения с контрольным значением, исключить полностью влияние «контроля» не представляется возможным. В отличие от широко используемых методов, выбранный нами подход рассматривает одновременно все образцы и определяет, отличается ли экспрессия гена в любом количестве из них от построенного фонового распределения. В этом случае образцы не сравниваются между собой напрямую, и мы можем ожидать, что применение такого подхода в исследовании позволит получить более полную картину вариации экспрессии генов в ходе реализации сложных жизненных циклов 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). За исключением *S. mansoni*, наибольшее количество генов с повышенной экспрессией обнаружено у фаз жизненных циклов с наибольшим размером молекулярной подписи: метацеркария *F. gigantica*, ювенили *F. hepatica*, редия *P. simillimum*, шистосомулы *T. regenti* и *T. szidati*. Мы предполагаем, что это может быть связано с формированием специфичных признаков в ходе эмбрионального развития или метаморфоза, протекающем на перечисленных фазах жизненного цикла. 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 геной 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. Следовательно, мы можем сделать вывод, что наборы дифференциально экспрессирующихся генов могут быть использованы в качестве репрезентативных частей молекулярных подписей фаз сложного жизненного цикла дигеней. В нашем исследовании удалось также установить, что многие гены у *F. gigantica*, *F. hepatica*, *P. simillimum* и *S. mansoni* меняют свою экспрессию согласованно в ходе реализации сложного жизненного цикла. Мы предполагаем, что совместная активация генов на разных фазах жизненного цикла указывает на наличие сложной системы регуляции работы генома у дигеней (Nesterenko et al., 2020, 2022).

При обсуждении результатов определения дифференциальной экспрессии необходимо также обратить внимание на то, что в анализе были рассмотрены фазы сложного жизненного цикла, различающиеся между собой по множеству признаков. К числу основных отличий мы можем отнести не только морфологическое строение и физиологические особенности, но и клеточный состав фаз. При выделении образцов мРНК фазы жизненного цикла мы частично или полностью охватываем транскриптомы всех клеточных типов, составляющих организм паразита. Транскрипционный сигнал от различных клеточных типов будет отличаться в зависимости от того, в каком количестве представлены клетки разных типов в составе организма. Так, к примеру, у церкарии с сильно развитой мускулатурой транскрипционный сигнал от мышечных клеток, вероятно, будет перекрывать транскрипционный сигнал от других клеточных типов. Рассмотрим гипотетический пример, представленный на рисунке 19.

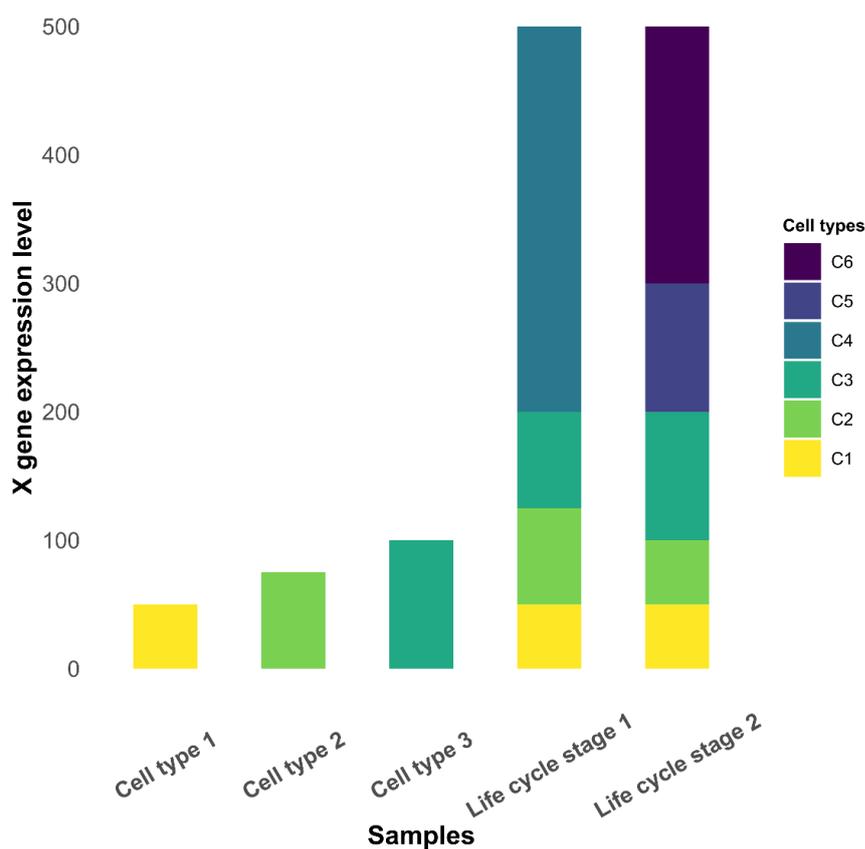


Рисунок 19 - Уровни экспрессии гипотетического гена X в различных клеточных типах и на разных фазах жизненного цикла. На вертикальной оси отмечены уровни экспрессии, а на горизонтальной оси расположены образцы. Разными цветами обозначены уровни экспрессии гена X в различных клеточных типах. Обозначения: Cell type 1, 2, 3 – клеточный тип 1, 2, 3; C1-6 – уровень экспрессии гена X в клеточных типах 1-6; Life cycle stage 1, 2 – первая и вторая фаза жизненного цикла.

Представим, что было произведено секвенирование транскриптомов трех клеточных типов (Cell type 1–3) и двух фаз жизненного цикла (Life cycle stage 1, 2), и для каждого из образцов был вычислен уровень экспрессии гена X (Рисунок 19). Мы видим, что уровни экспрессии гипотетического гена различаются между рассмотренными клеточными типами, тогда как у двух фаз жизненного цикла они совпадают и равны 500 (Рисунок 19). Вычисленный уровень экспрессии гена X у двух фаз жизненного цикла соответствует сумме уровней экспрессии этого же гена в различных клеточных типах, формируемых организмом фазы. Клеточный тип может быть как общим (клеточный тип 1–3), так и специфичным (клеточные типы 4–6) для нескольких фаз жизненного цикла. Определить какое количество данных от какого клеточного типа вошло в состав библиотеки коротких парных прочтений, анализируя только её состав без дополнительной информации, не представляется возможным. Более того, уровень экспрессии гипотетического гена в каждом из клеточных типов может меняться в зависимости от влияния внешних и внутренних факторов. 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).

Основываясь на указанных особенностях молекулярных подписей, 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, так и между видами, выявленная по результатам многомерного шкалирования молекулярных подписей. В качестве ярких примеров могут быть рассмотрены результаты кластеризации фаз в жизненных циклах *F. gigantea* (Рисунок 11А), *F. hepatica* (Рисунок 11В) и *P. simillimum* (Рисунок 11С). У *F. gigantea* в единый кластер были объединены церкарии и метацеркарии (Рисунок 11А), которые являются близкими фазами амфимиктического поколения. Схожий результат получен и для вида *F. hepatica*, у которого ранние и поздние метацеркарии кластеризовались вместе с различными ювенилями, покинувшими оболочки цисты (Рисунок

11В). Для обоих рассмотренных видов *Fasciola* (Рисунок 11А, В) также можно отметить объединение в общий кластер ювенилей и марит, т. е. неполовозрелых и половозрелых особей амфимиктического поколения. Представляя собой сменяющие друг друга фазы сложного жизненного цикла, метацеркария, ювениль и марита обладают схожим клеточным составом и отличаются между собой, преимущественно, степенью развития различных систем органов. Для *P. simillimum* выявлена совместная кластеризация редий и церкарий (Рисунок 11С), различающихся по множеству признаков, в том числе и по принадлежности к разным половым поколениям. Мы предполагаем, что совместная кластеризация молекулярных подписей редий и церкарий *P. simillimum* может быть следствием присутствия развивающихся эмбрионов церкарий внутри редий (Nesterenko et al., 2020, 2022).

Отдельно следует рассмотреть результаты кластеризации молекулярных подписей фаз сложного жизненного цикла *S. mansoni*. В настоящем исследовании мы использовали транскриптомные данные церкарий, шистосомул и марит, принадлежащие одной пуэрто-риканской линии (Protasio et al., 2012). Несмотря на то, что все образцы содержали материал особей обоих полов, мы не исключаем, что биологические повторности могут различаться между собой по величине транскрипционного сигнала от мужских и женских особей, даже если количество разнополых особей совпадало. Наше предположение основано на том, что значительные молекулярные различия между особями разных полов проявляются уже на фазе церкарии, однако фенотипический диморфизм становится заметен только после заражения позвоночного хозяина (Picard et al., 2016). Если наше предположение верно, то преобладание транскрипционного сигнала от мужских или женских особей могло повлиять на результаты кластеризации молекулярных подписей. На рисунке 11D отчетливо видно, что две биологические повторности транскриптомов шистосомул, собранные через 3 часа после заражения хозяина, вошли в состав двух разных кластеров, содержащие разные биологические повторности транскриптомов церкарий. Мы предполагаем, что формирование кластеров молекулярных подписей шистосомул и церкарий, основано не только на сходстве клеточного состава двух фаз жизненного цикла, но и на молекулярных различиях между особями разных полов.

Результаты межвидового сравнения между близкими видами дигеней также подтверждают сходство молекулярных подписей похожих фаз сложных жизненных циклов (Nesterenko et al., 2022). При рассмотрении молекулярных подписей целиком мы обнаружили совместную кластеризацию марит *F. gigantica*, *F. hepatica*, *P. simillimum* и ювенилей *F. hepatica* (Рисунок 12А), что может быть связано со сходством их клеточного состава. В единый кластер были также объединены мирацидий *F. gigantica*, церкарии *F. gigantica* и *P. simillimum*, редия *P. simillimum*, метацеркарии *F. gigantica* и *F. hepatica*, а также ювенили *F. hepatica*, покинувшие оболочку цисты (Рисунок 12А). Совместная кластеризация молекулярных подписей указанных

фаз жизненных циклов может быть вызвана несколькими причинами. Во-первых, мигриций и церкария являются свободноживущими фазами жизненного цикла и обитание в схожих условиях приводит к активации похожих наборов биологических процессов. Во-вторых, молекулярная подпись редии *P. simillimum* содержит в себе сигнал от эмбрионов церкарий, формирующихся в партените, что и могло привести к её сближению с молекулярными подписями церкарии *P. simillimum* и *F. gigantica*, филогенетически близкого к псилотреме вида. В-третьих, мы можем ожидать, что клеточный состав и, как следствие, наборы активных генов у церкарий, метацеркарий и ювенилей через 1, 3 и 24 часа после покидания оболочки цисты у филогенетически близких видов также будут перекрываться, что и выразилось в совместной кластеризации их молекулярных подписей.

При многомерном шкалировании, использованном в анализе, каждый ген задает свою плоскость, а его экспрессия или отсутствие в образце рассматривается, как отдельный признак. Из этого следует, что даже частичное совпадение между молекулярными подписями по набору активных генов может приводить к их сближению на двумерной плоскости и формированию таких крупных кластеров со сложным составом. Если в молекулярных подписях много «специфичных» генов или профиль экспрессии генов, входящих в их состав, сильно отличается от всех рассмотренных молекулярных подписей, то это будет приводить к формированию более компактных и однородных кластеров. В качестве ярких примеров можно рассмотреть результаты кластеризации молекулярных подписей, полученных для трех исследованных представителей Schistomatidae (Рисунок 12C, D). На рисунке 12C видно, что в единый кластер были объединены молекулярные подписи церкарии и мариты *S. mansoni* совместно с шистосомулами *S. mansoni*, *T. regenti* и *T. szidati*. Мы предполагаем, что к сближению их молекулярных подписей привело сходство клеточного состава перечисленных фаз. Выделение молекулярных подписей шистосомул *T. regenti* и *T. szidati* в отдельный кластер при рассмотрении списков генов с повышенной экспрессией (Рисунок 12D) вероятно связано с различиями в условиях обитания внутри хозяев. Оба исследованных вида *Trichobilharzia* в качестве окончательного хозяина используют водоплавающих птиц, тогда как *S. mansoni* заражает млекопитающих. Напротив, обитание в похожих условиях, вероятно, привело к повышению экспрессии схожих наборов генов, что и могло привести к объединению в единый кластер церкарий *S. mansoni*, *T. regenti* и *T. szidati* (Рисунок 12D). Наличие в этом же кластере шистосомул и мариты *S. mansoni* предположительно связано с повышением экспрессии генов, связанных как с особенностями строения амфимиктического поколения, так и с паразитическим образом жизни.

Особый интерес для исследования молекулярных основ реализации сложных жизненных циклов представляют результаты совместной кластеризации молекулярных подписей фаз *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. При рассмотрении

молекулярных подписей целиком (Рисунок 12E) и сравнении списков генов с повышенной экспрессией (Рисунок 12F), мы наблюдали объединение в единый кластер большинства фаз сложных жизненных циклов шести видов дигеней. В обоих случаях, как и при исследовании «редиидных» и «спороцистоидных» видов отдельно, в анализ были включены только ортогруппы, включающие белки всех сравниваемых видов. Как следствие, при увеличении числа анализируемых видов с трех до шести, количество групп ортологов, соответствующих выставленному требованию, уменьшилось. Тем не менее мы можем предположить, что оставшиеся ортогруппы важны для реализации сложных жизненных циклов дигеней, так как присутствуют у шести видов, различающихся между собой по множеству признаков. Результаты проведенного анализа указывают, что гены из рассмотренных ортогрупп не только входят в молекулярные подписи фаз как партеногенетического, так и амфимиктического поколений (Рисунок 12E), но и демонстрируют статистически значимое повышение экспрессии на свободноживущих и паразитических фазах жизненного цикла (Рисунок 12F). 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.

Определить, в каких биологических процессах принимают участие гены, входящие в состав молекулярных подписей, позволяет анализ обогащения терминов геной онтологии. В рамках настоящего исследования мы сфокусировали внимание только на генах со статистически значимым повышением экспрессии, как на репрезентативных частях молекулярных подписей. Результаты проведенного анализа указали, что не менее 80 биологических процессов активно на каждой из рассмотренных фаз жизненного цикла. Для корректного обсуждения полученных списков биологических процессов, необходимо учитывать особенности биологии как вида, так и конкретной фазы сложного жизненного цикла (Nesterenko et al., 2020, 2022). У Fasciolidae, как и у многих семейств дигеней с активно заражающим мирацидием, деление оплодотворенного ооцита происходит после выхода яйца во внешнюю среду, но самые ранние этапы могут протекать в матке мариты (Galaktionov, Dobrovolskij, 2003). Исследованные нами транскриптомы яиц *Fasciola* получены из материала, собранного в желчном пузыре окончательного хозяина. Мы предполагаем, что выявленная для яиц *F. gigantica* и *F. hepatica* активность биологических процессов, участвующих в регуляции клеточной смерти и пигментации в ходе развития, может быть связана с начальными этапами эмбрионального развития мирацидиев. Для мирацидия, покинувшего оболочку яйца и активно плавающего в воде, обнаружена работа биологических процессов, вовлеченных в высвобождение энергии (клеточное дыхание) и, возможно, в митотическое деление клеток герминальной массы (организация митотического веретена). Несмотря на то, что наименование биологического процесса «регенерация печени» приведено в

соответствии с информацией о *H. sapiens*, мы можем предполагать, что указанный процесс описывает митоз клеток в составе герминальной массы.

Биологические процессы, вероятно связанные с активностью герминальной массы и эмбриональным развитием потомства, обнаружены у редии *F. gigantea* и *P. simillimum*. Как и у мирацидия *F. gigantea*, у редии этого же вида выявлена активность биологического процесса «регенерация печени», который мы связываем с митотическими делениями клеток герминальной массы. Согласно материалам публикации Чжана с соавторами (Zhang X.-X. et al., 2019), для выделения транскриптома были использованы редии *F. gigantea*, собранные на 15 день после заражения моллюска в лабораторных условиях. В публикации Огамбо-Онгома (Ogambo-Ongoma) и Гудмана (Goodman) (1976) указано, что на 12 день после заражения моллюска материнская редия содержит эмбрионы дочерних редий, а на 14 день их количество увеличивается. К 16-му дню начинается миграция материнской редии в гепатопанкреас моллюска, а у дочерних редий наблюдается развитие глотки (Ogambo-Ongoma, Goodman, 1976). Основываясь на указанной информации, мы можем сделать предположение, что рассматриваемый транскриптом принадлежит именно материнской редии *F. gigantea*, которая готовится или начинает миграцию в промежуточном хозяине. Обнаруженные у редии фасциолы реакция на повреждения ДНК и регуляция апоптотического сигнала вероятно связаны с ответом паразита на агрессивное воздействие окружающей среды. В отличие от Чжана с соавторами, для получения транскриптома партениты мы собирали материал дочерних редий *P. simillimum* из природного заражения. В редиях присутствовали церкарии на разных этапах развития, поэтому мы можем утверждать, что выявленное для партениты обогащение биологических процессов, ассоциированных с митотическим циклом и развитием, вызвано транскрипционным сигналом от формирующихся личинок амфимиктического поколения.

Движение церкарий в толще воды является энергозатратным процессом и происходит с помощью сокращения мышц, что подтверждается результатами анализа обогащения терминов генной онтологии. С представлениями о том, что на фазе адолескарии/метацеркарии происходит изменение тканей и постэмбриональное развитие органов, согласуется активность таких биологических процессов, как «проведение сигнала», «развитие тканей», «миграция эпителия», «постэмбриональное развитие органов животного», обнаруженная нами для метацеркарий *Fasciola*. На взаимодействие метацеркарий с хозяином или с внешней окружающей средой указывает обогащение следующих терминов генной онтологии: «ответ на абиотический раздражитель», «проведение сигнала», «межклеточная передача сигнала», «ответ на внешний раздражитель», «защитная реакция на другой организм», «межвидовое взаимодействие между организмами».

Активность биологических процессов, связанных с постэмбриональным развитием («процессы развития») и взаимодействием с хозяином («межвидовое взаимодействие между организмами», «ответ на внешний раздражитель», «иммунитет, опосредованный нейтрофилами»), обнаружена и у ювенилей *F. hepatica* на первом часу после покидания оболочки цисты внутри окончательного хозяина. Гены, повышающие экспрессию на третьем часу после выхода паразита из цисты, принимают участие в регуляции клеточной дифференцировки, уровней гормонов и секреции, что, вероятно, указывает на адаптацию паразита к условиям обитания внутри хозяина. Мигрируя в организме хозяина, паразит активно использует его ресурсы для получения энергии, что подтверждается усилением на фазе ювенили через 24 часа после покидания оболочки цисты экспрессии генов, участвующих в метаболизме производных жирных кислот, биосинтезе ацилкоэнзима А и углеводов, а также в ответе на питательные вещества. Наблюдаемую на этом же этапе активацию биологических процессов «организация саркомеров» и «позитивная регуляция адгезии между клеткой и субстратом» мы связываем с формированием мужской половой системы. Известно, что зрелые семенники окружены базальной пластинкой и толстым слоем внеклеточного матрикса с прикрепленными к нему отростками мышечных клеток (Galaktionov, Dobrovolskij, 2003). Отростки образуют вокруг семенника мышечную оболочку, которая обеспечивает пульсирующее движение, выталкивающее сперму в семявыносящий поток (Galaktionov, Dobrovolskij, 2003). Образование мужских гамет происходит, вероятно, на более поздних этапах миграции, на что указывает активность таких биологических процессов, как «организация ресничек» и «подвижность клеток, зависящая от ресничек или жгутиков», выявленная для 21-дневной ювенили *F. hepatica*. Для 42- и 70-дневных ювенилей *F. gigantica* обнаружена повышенная экспрессия генов, принимающих участие в беременности, пролиферации эпителиальных клеток, организации внеклеточного матрикса, что, вероятно, связано с развитием женской половой системы.

Результаты анализа списков обогащенных биологических процессов доказали, что постэмбриональное развитие различных систем органов происходит и на фазе шистосомулы *S. mansoni*, *T. regenti* и *T. szidati*. Нам не удалось обнаружить ни одного общего биологического процесса для шистосомул *S. mansoni*, собранных через 3 и 24 часа после заражения хозяина. Принимая во внимание, что и при сравнительном анализе молекулярных подписей шистосомулы *S. mansoni* разошлись по разным кластерам (Рисунок 11D), мы можем сделать вывод, что для адаптаций к изменяющимся условиям среды в ходе миграции в организме хозяина, паразит регулирует экспрессию многих генов.

У большинства дигеней онтогенез мартиты разделен на два этапа: подготовительный (мартитогония), во время которого происходит завершение развития дефинитивных систем органов и/или физиологическое созревание, и функциональный – период репродукционной

активности половозрелых особей (Galaktionov, Dobrovolskij, 2003). В проанализированных нами молекулярных подписях марит *F. gigantea*, *F. hepatica*, *P. simillimum* и *S. mansoni*, гены с повышенной экспрессией принимали участие как в процессах развития, так и в формировании половых продуктов. Активность таких биологических процессов таких, как «развитие кожи» (*F. gigantea*), «ремоделирование тканей» (*F. gigantea*, *F. hepatica*), «развитие первичных мужских половых признаков» (*F. hepatica*), «определение левой/правой симметрии» (*P. simillimum*) и «ответ на эстроген» (*S. mansoni*), мы связываем с маритогонией. В то время как «подвижность сперматозоидов» (*F. gigantea*, *P. simillimum*) и, вероятно, «организация реснички» (*F. gigantea*, *P. simillimum*) и «движение реснички» (*F. gigantea*, *P. simillimum*), указывают на наличие и подвижность мужских гамет, а «вителлогенез» (*S. mansoni*) – на образование и накопление желтка в желточных клетках в составе яйца. Одновременная активность биологических процессов, связанных как с развитием, так и размножением, в молекулярных подписях марит, вероятнее всего, связана с тем, что выделение транскрипта производилось из пулированных образцов, которые могли содержать особей на разных этапах онтогенеза.

Суммируя результаты анализа обогащения терминов генной онтологии, мы можем утверждать, что исследование наборов дифференциально экспрессирующихся генов позволяет определить молекулярные основы фаза-специфичных признаков. Тем не менее гены со статистически значимым повышением экспрессии, как уже было показано ранее, составляют важную, но относительно небольшую часть молекулярных подписей фаз сложных жизненных циклов дигенетических сосальщиков. Результаты проведенных внутри- и межвидовых сравнительных анализов продемонстрировали, что молекулярные подписи схожих фаз жизненных циклов дигеней значимо перекрываются, но также обладают и набором «специфичных» генов (Nesterenko et al., 2020, 2022). Совместная работа «общих» и «специфичных» генов позволяют каждой фазе жизненного цикла успешно адаптироваться к особым условиям обитания. Принимая во внимание, что в настоящее время нет никаких сомнений, что предком дигеней являлся свободноживущий плоский червь, возникает три закономерных вопроса. Во-первых, каким набором белок-кодирующих генов обладал последний общий предок дигеней? Во-вторых, какие изменения в геноме предка дигеней, относительно генома общего предка дигеней и свободноживущих плоских червей, сделали возможным эволюционное усложнение жизненного цикла с чередованием поколений? В-третьих, различаются ли между собой фазы одного жизненного цикла по вкладам в их молекулярные подписи групп генов с разным филогенетическим происхождением?

Первая успешная реконструкция модели генома последнего общего предка дигеней выполнена Заяк с соавторами (Zajac et al., 2021). В исследовании авторы ставили перед собой цель определить, какие изменения в геноме, а именно расширение семейств генов или

приобретение новых кодирующих последовательностей, потенциально участвовали в адаптации к паразитизму. На основе полученных результатов анализа авторы предположили, что гены, появившиеся у предка дигеней, принимают участие в биологических процессах, связанных с функционированием тегумента. На примере *A. winterbourni* авторам удалось показать, что гены, присутствующие во множестве копий по сравнению с набором генов предка *Opisthorchiata/Xiphidiata*, участвуют в проникновении в ткани хозяина, избегании иммунного ответа и манипулировании поведением хозяина. Описанные результаты предоставляют дополнительные доказательства в пользу того, что одним из определяющих направлений эволюции дигеней является формирование и усложнение взаимоотношений с хозяевами.

Для взаимодействия с хозяевами гельминты используют обширный и разнообразный набор экскреторных/секреторных белков (Zarowiecki, Bergman, 2015). Мета-анализ геномов различных видов нематод и Platyhelminthes, показал, что семейства генов, кодирующих экскреторные/секреторные белки, стремительно меняются в ходе эволюции, что может быть рассмотрено в качестве адаптации к специфичным условиям обитания (Zarowiecki, Bergman, 2015). В рамках нашего исследования установлено, что у каждого из 11 рассмотренных видов дигеней не менее трех сотен генов кодируют потенциальные секреторные белки. Все потенциальные секреторные белки разделены на две группы на основании присутствия («классические») или отсутствия («неклассические») в последовательности классического N-концевого сигнального пептида. Наличие сигнальных N-концевых пептидов направляют белки к эндоплазматической сети, откуда они посредством везикулярного транспорта направляются через аппарат Гольджи на поверхность клетки (Nickel, 2003). Высвобождение белков во внеклеточное пространство происходит путем слияния секреторных везикул с плазматической мембраной. Описанный путь экспорта белков из эукариотических клеток известен в качестве «классического» или ЭПС/Гольджи-зависимого (Nickel, 2003). Альтернативные пути выделения белков, проходящие без ЭПС и аппарата Гольджи рассматриваются в качестве «неклассических». К числу «неклассических» путей секреции относятся эндосомальная рециркуляция, использование транспортеров плазматической мембраны, трансмембранное движение (флип-флоп-переход) и выпячивание плазматической мембраны (Nickel, 2003). В нашем исследовании разделение потенциальных секреторных белков на «классические» и «неклассические» было осуществлено только на основании результатов анализа их последовательностей. Несмотря на то, что дифференцировать пути их выделения возможно только с помощью экспериментальных методов исследования, мы считаем, что обнаруженные различия позволяют рассматривать две группы секреторных белков отдельно.

Основываясь на результатах проведенных анализов, мы предполагаем, что часть «классических» секреторных белков дигеней принимает участие в биологических процессах,

связанных с формированием тегумента («организация внеклеточного матрикса», «ремоделинг тканей»), изменением поведения («регуляция хемотаксиса», «ответ на раздражитель»), разрушением белков хозяина («протеолиз») и избеганием его иммунитета («дегрануляция лейкоцитов»). В то же время, необходимо учитывать, что часть белков может оставаться во внутриклеточных везикулах или внутри полостей тела паразита и не участвовать во взаимодействии паразита с хозяином. Так, например, часть «неклассических» секреторных белков, вероятно, участвует в регуляции развития («регуляция пролиферации нервных клеток-предшественников», «эмбриональное развитие органов», дифференцировка клеток»). Отсутствие между исследованными видами дигеней «общих» биологических процессов, обогащенных потенциальными «неклассическими» секреторными белками, указывает на разнообразие и эволюционную пластичность набора генов, кодирующих такие белки.

Исследование Заяк с соавторами (Zajac et al., 2021) наглядно показало, что реконструкция и сравнение моделей геномов предков дигеней позволяют описать молекулярные основы формирования таксон-специфичного признака. С целью ответить на вопрос, какие изменения в геноме предка дигеней сделали возможным эволюционное усложнение жизненного цикла, мы провели собственный анализ реконструированных моделей генов предков. Основным изменением в дизайне исследования, относительно описанного протокола в публикации Заяк с соавторами (Zajac et al., 2021), стало то, что мы не включили в анализ данные по нематодам, цестодам или моногенеям и рассматривали только наборы белок-кодирующих генов дигеней и свободноживущих плоских червей. Расширенный набор данных позволяет построить большее число моделей геномов предков по сравнению с набором данных, ограниченным двумя интересующими нас таксонами. В то же время для определения потерянных, оставшихся, дублированных и появившихся генов используют попарное сравнение моделей геномов предков. Это означает, что вывод об изменениях генома делается относительно набора генов, характерного для выбранного «контроля». Для более полной реконструкции эволюционной истории изменения генома последнего общего предка дигеней, требуется провести минимум три попарных сравнения реконструированных моделей геномов предков: Platyhelminthes против Neodermata, Neodermata против Cestoda/Trematoda, Cestoda/Trematoda против Digenea. Результаты определения биологических процессов, обогащенных появившимися или дублированными генами, будут отличаться в зависимости от того, используется ли в анализе результаты сравнения моделей Platyhelminthes и Digenea или Cestoda/Trematoda и Digenea. Различия между наборами генов предка Digenea и предка Cestoda/Trematoda будут намного меньше, чем между предком Digenea и предком Platyhelminthes. Исходя из того, что нас интересовали молекулярные основы «эволюционного скачка» от свободноживущих плоских

червей с простым жизненным циклом до дигеней с гетерогонией, мы ограничились построением и сравнительным анализом двух моделей геномов предков.

Сравнительный анализ обнаружил, что количество появившихся генов в модели последнего общего предка Digenea превышает количество дублицированных последовательностей практически в 29 раз. В качестве одного из возможных объяснений выявленных различий может выступать то, что «появившиеся» гены — это гетерогенный набор генов, объединенных по признаку отсутствия схожих последовательностей в построенной модели генома последнего общего предка Platyhelminthes. Например, если «новый» ген возник путем сдвига рамки считывания или горизонтального переноса от другого организма уже после расхождения последних общих предков свободноживущих плоских червей и Neodermata, то его последовательность будет отсутствовать в реконструированной модели последнего общего предка Platyhelminthes. В указанную модель также не войдут последовательности, специфичные для последних общих предков Neodermata, Cestoda/Trematoda и Digenea, возникшие путем дубликации предкового гена и стремительного накопления мутаций одной из копий или фактического появления гена *de novo* в ранее некодирующих областях генома. Согласно одной из моделей, белок-кодирующий ген может сформироваться *de novo* из белок-некодирующего протогена, если РНК протогена способны формировать ассоциации с другими молекулярными комплексами (Tautz et al., 2013). Исследования генома *S. mansoni* доказали присутствие в нем тысячи генов длинных белок-некодирующих РНК (Vasconcelos et al., 2017, 2018; Oliveira et al., 2018), формирующих на каждой фазе сложного жизненного цикла как минимум один модуль ко-экспрессии с белок-кодирующими генами (Maciel et al., 2019). В исследовании Масиель (Maciel) с соавторами (Maciel et al., 2020) показано, что длинные белок-некодирующие РНК и гены, кодирующие белки, у самцов и самок *S. japonicum* демонстрируют как согласованную экспрессию на протяжении всего полового созревания, так и дифференциальную экспрессию между особями разных полов. Интересно отметить, что гены, кодирующие белки и входящие в такие модули ко-экспрессии, связаны с развитием нервной системы, липидным обменом и общей выживаемостью паразита (Maciel et al., 2020). Результаты исследований транскриптомов спороцисты и мариты *S. mansoni* также продемонстрировали, что паразитирующие и размножающиеся фазы жизненного цикла различаются по профилям экспрессии генов длинных белок-некодирующих РНК, вероятно принимающих участие в регуляции развития (Kim et al., 2020). Присутствие большого числа белок-некодирующих генов в геномах современных видов дигеней дает нам основания сделать три предположения. Во-первых, белок-некодирующие гены могли присутствовать в геномах предков на разных этапах эволюции дигенетических сосальщиков. Во-вторых, часть белок-кодирующих генов в построенной модели генома предка Digenea могли произойти от белок-некодирующих генов предков. В-третьих, гены, возникшие от

белок-некодирующих последовательностей, могли сыграть важную роль в формировании разнообразия фаз жизненного цикла последнего общего предка *Digenea* путем усложнения системы генетической регуляции эмбрионального развития.

Результаты проведенного анализа обогащения терминов геной онтологии указали на большой вклад как появившихся (Рисунок 13), так и дублированных (Рисунок 14) генов во многие биологические процессы последнего общего предка дигеней, в том числе и в его эмбриональное развитие и/или метаморфоз (Nesterenko et al., 2022). Для лучшего понимания полученных результатов требуется вновь обратить внимание на то, что мы сравнивали только две реконструированные модели геномов последних общих предков *Platyhelminthes* и *Digenea*. Это означает, что расширение молекулярных основ биологических процессов, представленных на рисунках 13 и 14, происходило не одномоментно, а последовательно от предка к предку. 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). Идея «закрепления» полифенизма в ходе эволюции, приводящая к сосуществованию контрастных генетически контролируемых фенотипов в качестве частей одной системы, впервые была высказана Захваткиным (1949) при обсуждении формирования множества клеточных типов в многоклеточном организме. Позже идея о «закреплении» полифенизма в контексте усложнения жизненного цикла была развита Минелли и Фуско (Minelli, Fusco, 2010), которые высказали два ключевых предположения. Согласно первому предположению, из полифенизма мог развиваться согласованный во времени, а не только в пространстве, предсказуемый набор фенотипов в виде структурно различающихся стадий в жизненном цикле. Согласно второму предположению, усложнение жизненного цикла, вероятно, происходило, преимущественно, в менее стабильной, возможно и сезонно меняющейся среде (Minelli, Fusco, 2010). Результаты проведенного нами анализа указывают на то, что приобретенные гены вовлечены как в регуляцию ответа на различные стимулы, так и в контроль экспрессии генов, в том числе и с помощью эпигенетических механизмов (Рисунок 13) (Nesterenko et al., 2022). Вместе с тем, в настоящий момент затруднительно определить, связано ли выявленное обогащение указанных биологических процессов с «закреплением» полифенизма в жизненном цикле последнего общего предка дигеней. Тем не менее мы можем утверждать, что последовательное добавление «новых» генов в различные биологические процессы в ходе эволюции *Digenea* сформировало основу для увеличения пластичности как регуляции, так и реализации закодированной в геноме информации.

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». При обсуждении результатов филостратиграфического анализа необходимо учитывать, что выделение филострат происходит на основе анализа топологии филогенетического дерева и таксономической аффилиации анализируемых видов. Из этого следует, что, если филогенетическое дерево неполное и в нем нет данных для определения более общей филостраты, часть генов будет отнесена к видоспецифичной филострате. Так, к примеру, в отличие от всех остальных рассмотренных видов плоских червей, у *P. simillimum* и *P. vittatus* филострата «Species» оказалась наиболее крупной (Рисунок 15). Вероятно, что из-за отсутствия данных по большему числу филогенетически близких видов, гены, общие для семейства и рода, также были включены в видоспецифичные филостраты двух указанных видов (Nesterenko et al., 2022).

Несмотря на то, что видоспецифичная филострата, предположительно, искусственно расширена, мы можем определить и общие черты между разными видами плоских червей в филостратиграфическом составе их наборов генов. В качестве одного из ярких примеров может быть рассмотрен тот факт, что филострата «Spiralia» оказалась наименьшей у всех исследованных видов (Nesterenko et al., 2022). Исследования показывают, что, скорее всего, предковыми признаками Platyhelminthes является наличие эндолецитального яйца, квартетное спиральное дробление, эпиболическая гастрюляция и детерминированное развитие (Martín-Durán, Egger, 2012). У дигеней яйца эктолецитальные, а спиральное дробление полностью отсутствует (Egger et al., 2015). Мы предполагаем, что небольшой размер филостраты «Spiralia» у рассмотренных плоских червей может быть отчасти связан с видоизменением дробления, что также могло внести вклад в усложнение жизненного цикла последнего общего предка дигеней.

Полученные результаты филостратиграфического анализа были использованы нами в исследовании состава различных групп генов с целью определить, как менялся их состав в ходе эволюции. Первыми мы рассмотрели реконструированные модели геномов последних общих предков Platyhelminthes и Digenea. Построенные модели различались между собой по количеству генов, принадлежащих разным филостратам (Рисунок 16А). Размер каждой филостраты в модели генома предка Digenea превышал размер этой же филостраты в модели генома Platyhelminthes. Принимая это во внимание, мы можем сделать вывод, что гены, появившиеся или дублированные у последнего общего предка дигеней, имеют разную филогенетическую историю. Сложный филостратиграфический состав обнаружен как для кластеров ко-экспрессии,

так и для наборов генов с заметным уровнем экспрессии на всех рассмотренных фазах сложного жизненного цикла (Рисунок 16B). В обоих случаях большинство генов принадлежало первым филостратам, и, в первую очередь, филострате «Cellular organisms». Вклад более «молодых» филострат в состав рассмотренных групп генов сильно варьировал. Основываясь на полученных результатах, мы предполагаем, что усложнение жизненного цикла последнего общего предка дигеней потребовало не только добавление «новых» генов в уже имеющиеся генетические сети, но и согласование экспрессии генов из разных филострат. Формирование сложного жизненного цикла последнего общего предка дигеней тесно связано с эволюцией взаимоотношений паразита с хозяевами. Результаты проведенного анализа показали, что филостратиграфический состав наборов генов, кодирующих как «классические» (Рисунок 16C), так и «неклассические» (Рисунок 16D) секреторные белки, различается у всех 11 рассмотренных видов дигеней. Выявленные различия между видами, использующими разные виды в качестве промежуточных и окончательных хозяев, согласуются с выводами Заровецки и Берримана (Zarowiecki, Berriman, 2015) о том, что наборы экскреторных/секреторных белков стремительно меняются в ходе эволюции для адаптации к специфичным условиям обитания.

Возможность разделить набор белок-кодирующих генов исследуемого вида на неперекрывающиеся группы с разным филогенетическим происхождением имеет большой потенциал для исследования молекулярных подписей фаз сложного жизненного цикла. Используя информацию о филостратиграфической принадлежности генов, возможно как исследовать профили экспрессии отдельных филострат, так и измерить их вклад в молекулярные подписи фаз. На первом этапе анализа мы провели расчет возрастных индексов молекулярных подписей фаз сложных жизненных циклов *F. gigantica*, *F. hepatica*, *P. simillimum*, *S. mansoni*, *T. regenti* и *T. szidati*. Статистически значимые вариации возрастных индексов при одновременном рассмотрении всех выделенных филострат получены для *F. gigantica*, *F. hepatica*, *S. mansoni*, *T. regenti* и *T. szidati*. Общий филотранскриптомный профиль у *P. simillimum* «сглажен» видоспецифичной филостратой (Рисунок 17A), которая, как уже было указано ранее, вероятнее всего искусственно расширена и включает в себя также гены, общий для Psilostomatidae и *Psilotrema*. При исключении филостраты «Species» мы наблюдали уменьшение значений возрастных индексов молекулярных подписей всех рассмотренных фаз сложного жизненного цикла *P. simillimum*. Выявленные изменения указывают на то, что гены, как специфичные для вида, так и, предположительно, общие для семейства и рода, вносят существенный вклад в молекулярные подписи фаз и, вероятно, играют важную роль в формировании фаза-специфичных признаков.

Согласно результатам внутривидовых сравнений возрастных индексов молекулярных подписей фаз, у обоих исследованных видов *Fasciola* наиболее «старой» фазой являлось яйцо, у

P. simillimum – редия, а наиболее «молодой» фазой у всех трех рассмотренных представителей Schistosomatidae оказалась шистосомула (Nesterenko et al., 2022). Так, к примеру, возрастной индекс молекулярной подписи материнской редии *F. gigantea* больше значений, полученных для яйца и ювенилей, но меньше, чем у мирацидия, церкарии, метацеркарии и мариты этого же вида (Рисунок 18A1). Означает ли это, что материнская редия появилась в жизненном цикле последнего общего предка позже яйца и ювенилей, но раньше мирацидия, церкарии, метацеркарии и мариты? Вероятнее всего нет. Выявленные различия в возрастных индексах не могут быть напрямую использованы для реконструкции последовательности включения фаз в жизненный цикл по ряду причин (Nesterenko et al., 2022). Во-первых, молекулярная подпись любой из фазы жизненного цикла современного вида, вероятно, значительно отличается от молекулярной подписи соответствующей фазы жизненного цикла предка. Формирование новой фазы жизненного цикла, по всей видимости, происходит преимущественно путем преобразования молекулярной подписи фазы, уже присутствующей в жизненном цикле. Вероятно, что преобразование молекулярной подписи происходит как путем включения «новых» генов, так и изменением уровней экспрессии генов, входящих в её состав. Как следствие, молекулярная подпись предковой фазы будет иметь определенную степень сходства по списку генов и/или уровнях их экспрессии с молекулярными подписями одной или даже нескольких фаз жизненных циклов современных видов. Во-вторых, наборы доступных транскриптомных данных у рассмотренных видов дигеней сильно различаются. В-третьих, молекулярные подписи фаз жизненных циклов пластичны. Экспрессия генов, входящих в молекулярную подпись фазы жизненного цикла, может меняться в зависимости от внутренних и внешних факторов. Пластичность молекулярной подписи находит отражение в вариации возрастных индексов одной и той же фазы жизненного цикла, если рассматриваются транскриптомные данные с нескольких временных точек. В качестве примеров могут быть рассмотрены молекулярные подписи ювенилей *F. hepatica* (Рисунок 18B1), покинувших оболочки цисты, и шистосомул *S. mansoni* (Рисунок 18D1). Вероятно, что определяющее влияние на молекулярные подписи указанных фаз жизненных циклов оказывает их паразитический образ жизни, в том числе и необходимость адаптироваться к изменяющимся условиям в ходе миграции в организме хозяина. В-четвертых, каждая фаза жизненного цикла характеризуется множеством признаков, которые могут формироваться и изменяться на всем протяжении эволюции вида. Рассматривать «появление» фазы в жизненном цикле как одномоментное событие – некорректно (Nesterenko et al., 2022). Например, у *S. mansoni*, *T. regenti* и *T. szidati* наиболее высоким значением возрастного индекса молекулярной подписи обладала шистосомула. Основываясь на том, что шистосомула представляет собой мигрирующую фазу жизненного цикла, характерную только для Schistosomatidae, мы можем предположить, что высокое значение возрастного индекса её

молекулярной подписи указывает на позднее «появление» этой фазы в жизненном цикле последнего общего предка шистосоматид. Возникает закономерный вопрос, как может «появиться» такая специализированная фаза в жизненном цикле? Вероятнее всего шистосомула соответствует метацеркарии других видов дигеней (Galaktionov, Dobrovolskij, 2003). В таком случае высокий возрастной индекс её молекулярной подписи может быть связан с адаптацией к паразитированию и завершению жизненного цикла в промежуточном хозяине при выпадении исходного окончательного хозяина. Таким образом, мы можем заключить, что 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.

Принимая во внимание, что молекулярная подпись представляет собой «снимок» транскриптомов клеточных типов, формирующих организм рассматриваемой фазы жизненного цикла, выявленные различия по возрастным индексам могут указывать на эволюционную трансформацию отдельной ткани или органа (Nesterenko et al., 2022). Мы предположили, что обнаружить признаки фазы, претерпевшие наиболее заметные изменения в ходе эволюции возможно путем определения биологических процессов, в которых принимают участие гены с наибольшими вкладами в возрастной индекс её молекулярной подписи. Общими для большинства рассмотренных фаз с наибольшими и наименьшими возрастными индексами оказались три биологических процесса, связанных с синтезом белка в митохондриях. Большой вклад митохондрий в молекулярные подписи обусловлен присутствием митохондрий во всех клеточных типах исследованных фаз и, следовательно, их сильным транскрипционным сигналом. Биологические процессы, имеющие отношение к регуляции развития организма, обнаружены среди специфичных, как для одной, так и нескольких фаз. Часть биологических процессов связана с развитием нервной («развитие глиальных клеток», «регуляция пролиферации нейробластов», «положительная регуляция развития нервной системы», «регуляция развития нервной системы») и мышечной системы («морфогенез мышечной ткани», «развитие мышечных структур»). 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. Вероятно, что

включение «новых» генов в регуляторные сети, связанные с развитием или метаморфозом, способствовало преобразованию метацеркарии в шистосомулу в жизненном цикле последнего общего предка Schistosomatidae.

Проведенный анализ относительных уровней экспрессии филострат доказал, что группы генов с разным филогенетическим происхождением изменяют активность на разных фазах сложного жизненного цикла. Большинство филострат демонстрировали наименьшие относительные уровни экспрессии на фазах с наименьшими размерами молекулярных подписей (яйца *Fasciola*, мартиты *P. simillimum* и *S. mansoni*, церкарии *Trichobilharzia*) и наибольшие относительные уровни экспрессии на фазах с наибольшими размерами молекулярных подписей (редии *P. simillimum*, метацеркарии *F. gigantica*, ювенили *F. hepatica*, шистосомулы *S. mansoni*, *T. regenti* и *T. szidati*). Мы предполагаем, что выявленный профиль экспрессии большинства филострат вызван подавлением (яйца *Fasciola*, мартиты *P. simillimum* и *S. mansoni*, церкарии *Trichobilharzia*) и усилением (редии *P. simillimum*, метацеркарии *F. gigantica*, ювенили *F. hepatica*, шистосомулы *S. mansoni*, *T. regenti* и *T. szidati*) морфогенетических процессов на отдельных фазах 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 ортогрупп доказали, 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 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. Мы уверены, что 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 филогенетики 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 филогенетической аффилиации генов позволило нам продемонстрировать сложный филогенетический 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 фазо-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 литературы

1. Атаев Г. Л. Размножение партенит трематод: Обзор основных теорий. / Г. Л. Атаев – СПб.: Наука. – 2017. – 87 с.
2. Белякова Ю. В. Жизненный цикл *Psilotrema simillimum* (Muhling, 1898) (Trematoda: Psilostomidae) / Ю. В. Белякова // Паразитология. – 1978. – Т. 12. – №. 1. – С. 62–67.
3. Галактионов К. В. Происхождение и эволюция жизненных циклов трематод. / К. В. Галактионов, А. А. Добровольский – СПб.: Наука. – 1998. – 404 с.
4. Гинецинская Т. А. О природе жизненных циклов трематод / Т. А. Гинецинская // Вестник ЛГУ, сер. биол. – 1965. – Т. 21. – №. 4. – С. 5–14.
5. Гинецинская Т. А. Трематоды их жизненные циклы, биология и эволюция. / Т. А. Гинецинская – Наука. Ленингр. отд-ние – 1968. – 410 с.
6. Гранович А. И. Морфопроеесс и жизненные циклы организмов / А. И. Гранович, А. Н. Островский, А. А. Добровольский // Журнал общей биологии. – 2010. – Т. 71. – №. 6. – С. 514–522.
7. Добровольский А. А. Партеногенетические поколения трематод. / А. А. Добровольский, К. В. Галактионов, Г. К. Мухамедов, Б. К. Синха, И. А. Тихомиров // Труды Ленингр. Об-ва естествоиспытателей. – 1983. – Т. 82. – №. 4. – С. 1–108.
8. Догель В. А. Курс общей паразитологии. / В. А. Догель – Гос. учебно-педагогическое изд-во Министерства просвещения СССР. – 1947. – 372 с.
9. Захваткин А. А. Сравнительная эмбриология низших беспозвоночных: Источники и пути формирования индивидуального развития многоклеточных. Учебное пособие. / Захваткин А. А. – Советская наука. – 1949. – 396 с.
10. Abu-Jamous B. Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data / B. Abu-Jamous, S. Kelly // Genome biology. – 2018. – Vol. 19. – №. 1. – P. 1–11.
11. Alié A. The ancestral gene repertoire of animal stem cells / A. Alié, T. Hayashi, I. Sugimura, M. Manuel, W. Sugano, A. Mano, N. Satoh, K. Agata, N. Funayama // Proceedings of the National Academy of Sciences. – 2015. – Vol. 112. – №. 51. – P. E7093-E7100.
12. Allam A. Karect: accurate correction of substitution, insertion and deletion errors for next-generation sequencing data / A. Allam, P. Kalnis, V. Solovyev // Bioinformatics. – 2015. – Vol. 31. – №. 21. – P. 3421-3428.
13. Almagro Armenteros J. J. SignalP 5.0 improves signal peptide predictions using deep neural networks / J. J. Almagro Armenteros, K. D. Tsirigos, C. K. Sønderby, T. N. Petersen, O. Winther,

- S. Brunak, G. von Heijne, H. Nielsen // *Nature biotechnology*. – 2019. – Vol. 37. – №. 4. – P. 420-423.
14. Altenhoff A. M. OMA standalone: orthology inference among public and custom genomes and transcriptomes / A. M. Altenhoff, J. Levy, M. Zarowiecki, B. Tomiczek, A. W. Vesztrocy, D. A. Dalquen, S. Müller, M. J. Telford, N. M. Glover, D. Dylus, C. Dessimoz // *Genome research*. – 2019. – Vol. 29. – №. 7. – P. 1152-1163.
 15. Arendsee Z. Phylostratr: a framework for phylostratigraphy / Z. Arendsee, J. Li, U. Singh, A. Seetharam, K. Dorman, E. S. Wurtele // *Bioinformatics*. – 2019. – Vol. 35. – №. 19. – P. 3617-3627.
 16. Bankers L. *De novo* transcriptome characterization of a sterilizing trematode parasite (*Microphallus* sp.) from two species of New Zealand snails / L. Bankers, M. Neiman // *G3: Genes, Genomes, Genetics*. – 2017. – Vol. 7. – №. 3. – P. 871-880.
 17. Bendtsen J. D. Feature-based prediction of non-classical and leaderless protein secretion / J. D. Bendtsen, L. J. Jensen, N. Blom, G. von Heijne, S. Brunak // *Protein Engineering Design and Selection*. – 2004. – Vol. 17. – №. 4. – P. 349-356.
 18. Berriman M. The genome of the blood fluke *Schistosoma mansoni* / M. Berriman, B. J. Haas, P. T. LoVerde, R. A. Wilson, G. P. Dillon, G. C. Cerqueira, S. T. Mashiyama, B. Al-Lazikani, L. F. Andrade, P. D. Ashton, M. A. Aslett, D. C. Bartholomeu, G. Blandin, C. R. Caffrey, A. Coghlan, R. Coulson, T. A. Day, A. Delcher, R. DeMarco, A. Djikeng, T. Eyre, J. A. Gamble, E. Ghedin, Y. Gu, C. Hertz-Fowler, H. Hirai, Y. Hirai, R. Houston, A. Ivens, D. A. Johnston, D. Lacerda, C. D. Macedo, P. McVeigh, Z. Ning, G. Oliveira, J. P. Overington, J. Parkhill, M. Pertea, R. J. Pierce, A. V. Protasio, M. A. Quail, M.-A. Rajandream, J. Rogers, M. Sajid, S. L. Salzberg, M. Stanke, A. R. Tivey, O. White, D. L. Williams, J. Wortman, W. Wu, M. Zamanian, A. Zerlotini, C. M. Fraser-Liggett, B. G. Barrell, N. M. El-Sayed // *Nature*. – 2009. – Vol. 460. – №. 7253. – P. 352-358.
 19. Bolger A. M. Trimmomatic: a flexible trimmer for Illumina sequence data / A. M. Bolger, M. Lohse, B. Usadel // *Bioinformatics*. – 2014. – Vol. 30. – №. 15. – P. 2114-2120.
 20. Brandl H. PlanMine—a mineable resource of planarian biology and biodiversity / H. Brandl, HK. Moon, M. Vila-Farré, S.-Y. Liu, I. Henry, J. C. Rink // *Nucleic acids research*. – 2016. – Vol. 44. – №. D1. – P. D764-D773.
 21. Brooks F. G. Studies on the Germ Cell Cycle of Trematodes / F. G. Brooks // *American Journal of Hygiene*. – 1930. – Vol. 12. – P. 299-340.
 22. Buchfink B. Fast and sensitive protein alignment using DIAMOND / B. Buchfink, C. Xie, D. H. Huson // *Nature methods*. – 2015. – Vol. 12. – №. 1. – P. 59-60.

23. Bürglin T. R. Homeodomain proteins: an update / T. R. Bürglin, M. Affolter // *Chromosoma*. – 2016. – Vol. 125. – №. 3. – P. 497-521.
24. Bychovskaja-Pavlovskaja I. E. Cercariae of bithyniid snails (*Bithynia tentaculata* and *B. leachi*) of Kurshskii gulf / I. E. Bychovskaja-Pavlovskaja, A. P. Kulakova // *Parazitologiya*. – 1971. – Vol. 5. – P. 222-32.
25. Camacho C. BLAST+: architecture and applications / C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden // *BMC bioinformatics*. – 2009. – Vol. 10. – №. 1. – P. 1-9.
26. Cantacessi C. A deep exploration of the transcriptome and “excretory/secretory” proteome of adult *Fascioloides magna* / C. Cantacessi, J. Mulvenna, N. D. Young, M. Kasny, P. Horak, A. Aziz, A. Hofmann, A. Loukas, R. B. Gasser // *Molecular & Cellular Proteomics*. – 2012. – Vol. 11. – №. 11. – P. 1340-1353.
27. Cantacessi C. Coming out of the shell: building the molecular infrastructure for research on parasite-harboring snails / C. Cantacessi, S. Prasopdee, J. Sotillo, J. Mulvenna, S. Tesana, A. Loukas // *PLoS neglected tropical diseases*. – 2013. – Vol. 7. – №. 9. – P. e2284.
28. Cantalapiedra C. P. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale / C. P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas // *Molecular biology and evolution*. – 2021. – Vol. 38. – №. 12. – P. 5825-5829.
29. Capella-Gutiérrez S. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses / S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón // *Bioinformatics*. – 2009. – Vol. 25. – №. 15. – P. 1972-1973.
30. Chen S. fastp: an ultra-fast all-in-one FASTQ preprocessor / S. Chen, Y. Zhou, Y. Chen, J. Gu // *Bioinformatics*. – 2018. – Vol. 34. – №. 17. – P. i884-i890.
31. Cherezov R. O. The Phenomenon of Evolutionary “*De Novo* Generation” of Genes / R. O. Cherezov, J. E. Vorontsova, O. B. Simonova // *Russian Journal of Developmental Biology*. – 2021. – Vol. 52. – №. 6. – P. 390-400.
32. Choi I. RNA-Seq of plant-parasitic nematode *Meloidogyne incognita* at various stages of its development / I. Choi, P. Subramanian, D. Shim, B.-J. Oh, B.-S. Hahn // *Frontiers in Genetics*. – 2017. – Vol. 8. – P. 190.
33. Choudhary V. Transcriptome analysis of the adult rumen fluke *Paramphistomum cervi* following next generation sequencing / V. Choudhary, S. Garg, R. Chourasia, J. J. Hasnani, P. V. Patel, T. M. Shah, V. D. Bhatt, A. Mohapatra, D. P. Blake, C. G. Joshi // *Gene*. – 2015. – Vol. 570. – №. 1. – P. 64-70.

34. Clark W. C. Interpretation of life history pattern in the Digenea / W. C. Clark // International Journal for Parasitology. – 1974. – Vol. 4. – №. 2. – P. 115-123.
35. Cwiklinski K. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution / K. Cwiklinski, J. P. Dalton, P. J. Dufresne, J. L. Course, D. J. L. Williams, J. Hodgkinson, S. Paterson // Genome biology. – 2015. – Vol. 16. – №. 1. – P. 1-13.
36. Darriba D. ProtTest 3: fast selection of best-fit models of protein evolution / D. Darriba, G. L. Taboada, R. Doallo, D. Posada // Bioinformatics. – 2011. – Vol. 27. – №. 8. – P. 1164-1165.
37. De León G. P.-P. Testing the higher-level phylogenetic classification of Digenea (Platyhelminthes, Trematoda) based on nuclear rDNA sequences before entering the age of the 'next-generation' Tree of Life / G. P.-P. de León, D. I. Hernández-Mena // Journal of Helminthology. – 2019. – Vol. 93. – №. 3. – P. 260-276.
38. Dinnik J. A. Observations on the succession of redial generations of *Fasciola gigantica* Cobbold in a snail host / J. A. Dinnik, N. N. Dinnik // Zeitschrift fur Tropenmedizin und Parasitologie. – 1956. – Vol. 7. – №. 4. – P. 397-419.
39. Dinnik J. A. The influence of temperature on the succession of redial and cercarial generations of *Fasciola gigantica* in a snail host / J. A. Dinnik, N. N. Dinnik // Parasitology. – 1964. – Vol. 54. – №. 1. – P. 59-65.
40. Dobrovolskij A. A. The nature of reproduction of digenea rediae and sporocysts / A. A. Dobrovolskij, G. L. Ataev // Taxonomy, ecology and evolution of metazoan parasites. – 2003. – P. 249-272.
41. Domazet-Lošo T. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns / T. Domazet-Lošo, D. Tautz // Nature. – 2010. – Vol. 468. – №. 7325. – P. 815-818.
42. Domazet-Lošo T. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages / T. Domazet-Lošo, J. Brajković, D. Tautz // Trends in Genetics. – 2007. – Vol. 23. – №. 11. – P. 533-539.
43. Domazet-Lošo T. No evidence for phylostratigraphic bias impacting inferences on patterns of gene emergence and evolution / T. Domazet-Lošo, A.-R. Carvunis, M. Mar Albà, M. S. Šestak, R. Bakarić, R. Neme, D. Tautz // Molecular biology and evolution. – 2017. – Vol. 34. – №. 4. – P. 843-856.
44. Drost H.-G. myTAI: evolutionary transcriptomics with R / H.-G. Drost, A. Gabel, J. Liu, M. Quint, I. Grosse // Bioinformatics. – 2018. – Vol. 34. – №. 9. – P. 1589-1590.
45. Dujon B. The yeast genome project: what did we learn? / B. Dujon // Trends in Genetics. – 1996. – Vol. 12. – №. 7. – P. 263-270.

46. Dylus D. How to build phylogenetic species trees with OMA / D. Dylus, Y. Nevers, A. M. Altenhoff, A. Gürtler, C. Dessimoz, N. M. Glover // *F1000Research*. – 2020. – Vol. 9. – №. 511. – P. 511.
47. Edgar R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput / R. C. Edgar // *Nucleic acids research*. – 2004. – Vol. 32. – №. 5. – P. 1792-1797.
48. Egger B. A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms / B. Egger, F. Lapraz, B. Tomiczek, S. Müller, C. Dessimoz, J. Girstmair, N. Škunca, K. A. Rawlinson, C. B. Cameron, E. Beli, M. A. Todaro, M. Gammoudi, C. Noreña, M. J. Telford // *Current Biology*. – 2015. – Vol. 25. – №. 10. – P. 1347-1353.
49. El-Gebali S. The Pfam protein families database in 2019 / S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto, R. D. Finn // *Nucleic acids research*. – 2019. – Vol. 47. – №. D1. – P. D427-D432.
50. Emanuelsson O. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence / O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne // *Journal of molecular biology*. – 2000. – Vol. 300. – №. 4. – P. 1005-1016.
51. Emms D. M. OrthoFinder: phylogenetic orthology inference for comparative genomics / D. M. Emms, S. Kelly // *Genome biology*. – 2019. – Vol. 20. – №. 1. – P. 1-14.
52. Emms D. M. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy / D. M. Emms, S. Kelly // *Genome biology*. – 2015. – Vol. 16. – №. 1. – P. 1-14.
53. Ershov N. I. New insights from *Opisthorchis felineus* genome: update on genomics of the epidemiologically important liver flukes / N. I. Ershov, V. A. Mordvinov, E. B. Prokhortchouk, M. Y. Pakharukova, K. V. Gunbin, K. Ustyantsev, M. A. Genaev, A. G. Blinov, A. Mazur, E. Boulygina, S. Tsygankova, E. Khrameeva, N. Chekanov, G. Fan, A. Xiao, H. Zhang, X. Xu, H. Yang, V. Solovyev, S. M.-Y. Lee, X. Liu, D. A. Afonnikov, K. G. Skryabin // *BMC genomics*. – 2019. – Vol. 20. – №. 1. – P. 1-22.
54. Franco G. R. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library / G. R. Franco, M. D. Adams, M. B. Soares, A. J.G. Simpson, J. C. Venter, S. D. J. Pena // *Gene*. – 1995. – Vol. 152. – №. 2. – P. 141-147.
55. Fried B. *Advances in trematode biology*. / B. Fried, T. K. Graczyk (ed) – CRC Press. – 1997. – 476 p.
56. Fu L. CD-HIT: accelerated for clustering the next-generation sequencing data / L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li // *Bioinformatics*. – 2012. – Vol. 28. – №. 23. – P. 3150-3152.

57. Fuchs B. Regulation of polyp-to-jellyfish transition in *Aurelia aurita* / B. Fuchs, W. Wang, S. Graspeuntner, Y. Li, S. Insua, E.-M. Herbst, P. Dirksen, A.-M. Böhm, G. Hemmrich, F. Sommer, T. Domazet-Lošo, U. C. Klostermeier, F. Anton-Erxleben, P. Rosenstiel, T. C.G. Bosch, K. Khalturin // *Current Biology*. – 2014. – Vol. 24. – №. 3. – P. 263-273.
58. Galaktionov K. V. Self-sustaining infrapopulation or colony? Radial clonal groups of *Himasthla elongata* (Mehlis, 1831) (Trematoda: Echinostomatidae) in *Littorina littorea* (Linnaeus) (Gastropoda: Littorinidae) do not support the concept of eusocial colonies in trematodes / K. V. Galaktionov, I. M. Podvyaznaya, K. E. Nikolaev, I. A. Levakin // *Folia parasitologica*. – 2015. – Vol. 62. – №. 067. – P. 1-14.
59. Galaktionov K. V. The Biology and Evolution of Trematodes: An Essay on the Biology, Morphology, Life Cycles, Transmissions, and Evolution of Digenetic Trematodes. / K. V. Galaktionov, A. A. Dobrovolskij – Springer Science & Business Media. – 2003. – 592 p.
60. Gao J.-F. *De novo* assembly and functional annotations of the transcriptome of *Metorchis orientalis* (trematoda: Opisthorchiidae) / J.-F. Gao, Y. Gao, J.-H. Qiu, Q.-c. Chang, Y. Zhang, M. Fang, C.-R. Wang // *Experimental parasitology*. – 2018. – Vol. 184. – P. 90-96.
61. Garg G. *In silico* secretome analysis approach for next generation sequencing transcriptomic data / G. Garg, S. Ranganathan // *BMC Genomics*. – BioMed Central. – 2011. – Vol. 12. – №. 3. – P. 1-10.
62. Garg G. The transcriptome of *Echinostoma caproni* adults: further characterization of the secretome and identification of new potential drug targets / G. Garg, D. Bernal, M. Trelis, J. Forment, J. Ortiz, M. L. Valero, L. Pedrola, J. Martinez-Blanch, J. G. Esteban, S. Ranganathan, R. Toledo, A. Marcilla // *Journal of proteomics*. – 2013. – Vol. 89. – P. 202-214.
63. Gouy M. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building / M. Gouy, S. Guindon, O. Gascuel // *Molecular biology and evolution*. – 2010. – Vol. 27. – №. 2. – P. 221-224.
64. Grabherr M. G. Full-length transcriptome assembly from RNA-Seq data without a reference genome / M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev // *Nature biotechnology*. – 2011. – Vol. 29. – №. 7. – P. 644-652.
65. Heberle H. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams / H. Heberle, G. Vaz Meirelles, F. R. da Silva, G. P. Telles, R. Minghim // *BMC bioinformatics*. – 2015. – Vol. 16. – №. 1. – P. 1-7.

66. Howe K. L. WormBase ParaSite – a comprehensive resource for helminth genomics / K. L. Howe, B. J. Bolt, M. Shafie, P. Kersey, M. Berriman // *Molecular and biochemical parasitology*. – 2017. – Vol. 215. – P. 2-10.
67. Hu W. Evolutionary and biomedical implications of a *Schistosoma japonicum* complementary DNA resource / W. Hu, Q. Yan, D.-K. Shen, F. Liu, Z.-D. Zhu, H.-D. Song, X.-R. Xu, Z.-J. Wang, Y.-P. Rong, L.-C. Zeng, J. Wu, X. Zhang, J.-J. Wang, X.-N. Xu, S.-Y. Wang, G. Fu, X.-L. Zhang, Z.-Q. Wang, P. J. Brindley, D. P. McManus, C.-L. Xue, Z. Feng, Z. Chen, Z.-G. Han // *Nature genetics*. – 2003. – Vol. 35. – №. 2. – P. 139-147.
68. Huerta-Cepas J. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper / J. Huerta-Cepas, K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen, C. von Mering, P. Bork // *Molecular biology and evolution*. – 2017. – Vol. 34. – №. 8. – P. 2115-2122.
69. Husnik F. Functional horizontal gene transfer from bacteria to eukaryotes / F. Husnik, J. P. McCutcheon // *Nature Reviews Microbiology*. – 2018. – Vol. 16. – №. 2. – P. 67-79.
70. International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms / A. Coghlan, R. Tyagi, J. A. Cotton, N. Holroyd, B. A. Rosa, I. J. Tsai, D. R. Laetsch, R. N. Beech, T. A. Day, K. Hallsworth-Pepin, H.-M. Ke, T.-H. Kuo, T. J. Lee, J. Martin, R. M. Maizels, P. Mutowo, P. Ozersky, J. Parkinson, A. J. Reid, N. D. Rawlings, D. M. Ribeiro, L. S. Swapna, E. Stanley, D. W. Taylor, N. J. Wheeler, M. Zamanian, X. Zhang, F. Allan, J. E. Allen, K. Asano, S. A. Babayan, G. Bah, H. Beasley, H. M. Bennett, S. A. Bisset, E. Castillo, J. Cook, P. J. Cooper, T. Cruz-Bustos, C. Cuéllar, E. Devaney, S. R. Doyle, M. L. Eberhard, A. Emery, K. S. Eom, J. S. Gilleard, D. Gordon, Y. Harcus, B. Harsha, J. M. Hawdon, D. E. Hill, J. Hodgkinson, P. Horák, K. L. Howe, T. Huckvale, M. Kalbe, G. Kaur, T. Kikuchi, G. Koutsovoulos, S. Kumar, A. R. Leach, J. Lomax, B. Makepeace, J. B. Matthews, A. Muro, N. M. O'Boyle, P. D. Olson, A. Osuna, F. Partono, K. Pfarr, G. Rinaldi, P. Foronda, D. Rollinson, M. G. Samblas, H. Sato, M. Schnyder, T. Scholz, M. Shafie, V. N. Tanya, R. Toledo, A. Tracey, J. F. Urban, L.-C. Wang, D. Zarlenga, M. L. Blaxter, M. Mitreva, M. Berriman // *Nature genetics*. – 2019. – Vol. 51. – №. 1. – P. 163.
71. Kanehisa M. KEGG for linking genomes to life and the environment / M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi // *Nucleic acids research*. – 2007. – Vol. 36. – №. suppl_1. – P. D480-D484.
72. Katoh K. MAFFT multiple sequence alignment software version 7: improvements in performance and usability / K. Katoh, D. M. Standley // *Molecular biology and evolution*. – 2013. – Vol. 30. – №. 4. – P. 772-780.

73. Khalturin K. More than just orphans: are taxonomically-restricted genes important in evolution? / K. Khalturin, G. Hemmrich, S. Fraune, R. Augustin, T. C. G. Bosch // Trends in Genetics. – 2009. – Vol. 25. – №. 9. – P. 404-413.
74. Kim H. C. LncRNAs in molluscan and mammalian stages of parasitic schistosomes are developmentally-regulated and coordinately expressed with protein-coding genes / H. C. Kim, A. M. Khalil, E. R. Jolly // RNA biology. – 2020. – Vol. 17. – №. 6. – P. 805-815.
75. Krogh A. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes / A. Krogh, B. Larsson, G. von Heijne, E. L. L. Sonnhammer // Journal of molecular biology. – 2001. – Vol. 305. – №. 3. – P. 567-580.
76. Laha T. Gene discovery for the carcinogenic human liver fluke, *Opisthorchis viverrini* / T. Laha, P. Pinlaor, J. Mulvenna, B. Sripa, M. Sripa, M. J. Smout, R. B. Gasser, P. J. Brindley, A. Loukas // BMC genomics. – 2007. – Vol. 8. – №. 1. – P. 1-15.
77. Lee J.-S. Analysis of the genes expressed in *Clonorchis sinensis* adults using the expressed sequence tag approach / J.-S. Lee, J. Lee, S.-J. Park, T.-S. Yong // Parasitology research. – 2003. – Vol. 91. – №. 4. – P. 283-289.
78. Leontovyč R. Comparative transcriptomic exploration reveals unique molecular adaptations of neuropathogenic *Trichobilharzia* to invade and parasitize its avian definitive host / R. Leontovyč, N. D. Young, P. K. Korhonen, R. S. Hall, P. Tan, L. Mikeš, M. Kašný, P. Horák, R. B. Gasser // PLoS neglected tropical diseases. – 2016. – Vol. 10. – №. 2. – P. e0004406.
79. Leontovyč R. Molecular evidence for distinct modes of nutrient acquisition between visceral and neurotropic schistosomes of birds / R. Leontovyč, N. D. Young, P. K. Korhonen, R. S. Hall, J. Bulantová, V. Jeřábková, M. Kašný, R. B. Gasser, P. Horák // Scientific reports. – 2019. – Vol. 9. – №. 1. – P. 1-12.
80. Lepitzki D. A. W. A plug in the cyst wall of metacercariae of *Sphaeridiotrema pseudoglobulus* (Digenea: Psilostomidae) and a possible novel mode of transmission / D. A. W. Lepitzki, B. M. Bunn // International journal for parasitology. – 1994. – Vol. 24. – №. 2. – P. 273-275.
81. Li B. Conservation and diversification of the transcriptomes of adult *Paragonimus westermani* and *P. skrjabini* / B.-w. Li, S. N. McNulty, B. A. Rosa, R. Tyagi, Q. R. Zeng, K.-z. Gu, G. J. Weil, M. Mitreva // Parasites & Vectors. – 2016. – Vol. 9. – №. 1. – P. 1-11.
82. Lin M. M. Hydrophobic forces and the length limit of foldable protein domains / M. M. Lin, A. H. Zewail // Proceedings of the National Academy of Sciences. – 2012. – Vol. 109. – №. 25. – P. 9851-9856.
83. Liu F. MolluscDB: an integrated functional and evolutionary genomics database for the hyper-diverse animal phylum Mollusca / F. Liu, Y. Li, H. Yu, L. Zhang, J. Hu, Z. Bao, S. Wang // Nucleic acids research. – 2021. – Vol. 49. – №. D1. – P. D988-D997.

84. Liu G.-H. *De novo* assembly and characterization of the transcriptome of the pancreatic fluke *Eurytrema pancreaticum* (trematoda: Dicrocoeliidae) using Illumina paired-end sequencing / G.-H. Liu, M.-J. Xu, H.-Q. Song, C.-R. Wang, X.-Q. Zhu // *Gene*. – 2016. – Vol. 576. – №. 1. – P. 333-338.
85. Maciel L. F. Dynamic Expression of Long Non-Coding RNAs Throughout Parasite Sexual and Neural Maturation in *Schistosoma japonicum* / L. F. Maciel, D. A. Morales-Vicente, S. Verjovski-Almeida // *Non-coding RNA*. – 2020. – Vol. 6. – №. 2. – P. 15.
86. Maciel L. F. Weighted Gene Co-Expression Analyses Point to Long Non-Coding RNA Hub Genes at Different *Schistosoma mansoni* Life-Cycle Stages / L. F. Maciel, D. A. Morales-Vicente, G. O. Silveira, R. O. Ribeiro, G. G. O. Olberg, D. S. Pires, M. S. Amaral, S. Verjovski-Almeida // *Frontiers in genetics*. – 2019. – Vol. 10. – P. 823.
87. Martín-Durán J. M. Developmental diversity in free-living flatworms / J. M. Martín-Durán, B. Egger // *EvoDevo*. – 2012. – Vol. 3. – №. 1. – P. 1-23.
88. Martín-Durán J. M. Increased taxon sampling reveals thousands of hidden orthologs in flatworms / J. M. Martín-Durán, J. F. Ryan, B. C. Vellutini, K. Pang, A. Hejnlol // *Genome research*. – 2017. – Vol. 27. – №. 7. – P. 1263-1272.
89. McLaughlin J. D. *Sphaeridiotrema globulus* (Rudolphi, 1814) (Digenea): evidence for two species known under a single name and a description of *Sphaeridiotrema pseudoglobulus* n. sp / J. D. McLaughlin, M. E. Scott, J. E. Huffman // *Canadian Journal of Zoology*. – 1993. – Vol. 71. – №. 4. – P. 700-707.
90. McNulty S. N. Genomes of *Fasciola hepatica* from the Americas reveal colonization with *Neorickettsia* endobacteria related to the agents of Potomac horse and human Sennetsu fevers / S. N. McNulty, J. F. Tort, G. Rinaldi, K. Fischer, B. A. Rosa, P. Smircich, S. Fontenla, Y.-J. Choi, R. Tyagi, K. Hallsworth-Pepin, V. H. Mann, L. Kammili, P. S. Latham, N. Dell’Oca, F. Dominguez, C. Carmona, P. U. Fischer, P. J. Brindley, M. Mitreva // *PLoS genetics*. – 2017. – Vol. 13. – №. 1. – P. e1006537.
91. Miller M. A. Creating the CIPRES Science Gateway for inference of large phylogenetic trees / M. A. Miller, W. Pfeiffer, T. Schwartz // 2010 gateway computing environments workshop (GCE). – 2010. – P. 1-8.
92. Minelli A. Developmental plasticity and the evolution of animal complex life cycles / A. Minelli, G. Fusco // *Philosophical Transactions of the Royal Society B: Biological Sciences*. – 2010. – Vol. 365. – №. 1540. – P. 631-640.
93. Minh B. Q. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era / B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A.

- von Haeseler, R. Lanfear // *Molecular biology and evolution*. – 2020. – Vol. 37. – №. 5. – P. 1530-1534.
94. Mirdita M. MMseqs2 desktop and local web server app for fast, interactive sequence searches / M. Mirdita, M. Steinegger, J. Söding // *Bioinformatics*. – 2019. – Vol. 35. – №. 16. – P. 2856-2858.
95. Mistry J. Pfam: The protein families database in 2021 / J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, A. Bateman // *Nucleic acids research*. – 2021. – Vol. 49. – №. D1. – P. D412-D419.
96. Moazeni M. Controversial aspects of the life cycle of *Fasciola hepatica* / M. Moazeni, A. Ahmadi // *Experimental parasitology*. – 2016. – Vol. 169. – P. 81-89.
97. Mustafin Z. S. Phylostratigraphic analysis shows the earliest origination of the abiotic stress associated genes in *A. thaliana* / Z.S. Mustafin, V.I. Zamyatin, D. K. Konstantinov, A. V. Doroshkov, S. A. Lashin, D.A. Afonnikov // *Genes*. – 2019. – Vol. 10. – №. 12. – P. 963.
98. Neme R. Phylogenetic patterns of emergence of new genes support a model of frequent *de novo* evolution / R. Neme, D. Tautz // *BMC genomics*. – 2013. – Vol. 14. – №. 1. – P. 1-13.
99. 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.
100. 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.
101. Nesterenko M. A. The phylostratigraphy analysis of the digenean molecular signatures [Электронный ресурс] / M. A. Nesterenko // GitHub – 2022. – Режим доступа: https://github.com/maxnest/The_phylostratigraphy_analysis_of_the_digenean_molecular_signatures/blob/main/pyHAM_flatworm_ancestor_genomes.py (дата обращения: 22.08.2022).
102. Nguyen L.-T. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies / L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh // *Molecular biology and evolution*. – 2015. – Vol. 32. – №. 1. – P. 268-274.
103. Nickel W. The mystery of nonclassical protein secretion: A current view on cargo proteins and potential export routes / W. Nickel // *European Journal of Biochemistry*. – 2003. – Vol. 270. – №. 10. – P. 2109-2119.
104. Odening K. Verwandtschaft, System und zyko-ontogenetische Besonderheiten der Trematoden / K. Odening // *Zoologischer Jahrbucher* – 1974. – Vol. 101. – №. 3. – P. 345–396.

105. Ogambo-Ongoma A. H. *Fasciola gigantica* Cobbold 1856 in the snail / A. H. Ogambo-Ongoma, J. D. Goodman // *The Journal of Parasitology*. – 1976. – Vol. 62. – №. 1. – P. 33-38.
106. Oliveira V. F. Identification of 170 New Long Noncoding RNAs in *Schistosoma mansoni* / V. F. Oliveira, L. A. G. Moares, E. A. Mota, L. K. Jannotti-Passos, P. M. Z. Coelho, A. C. A. Mattos, F. F. B. Couto, B. E. Caffrey, A. Marsico, R. Guerra-Sá // *BioMed research international*. – 2018. – Vol. 2018.
107. Orgogozo V. The differential view of genotype – phenotype relationships / V. Orgogozo, B. Morizot, A. Martin // *Frontiers in genetics*. – 2015. – Vol. 6. – P. 179.
108. Paradis E. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R / E. Paradis, K. Schliep // *Bioinformatics*. – 2019. – Vol. 35. – №. 3. – P. 526-528.
109. Parkinson J. Expressed Sequence Tags: An Overview. In: Parkinson J. (eds) *Expressed Sequence Tags (ESTs). Methods in Molecular Biology (Methods and Protocols)* / J. Parkinson, M. Blaxter – Humana Press. – 2009. – Vol. 533.
110. Patro R. Salmon provides fast and bias-aware quantification of transcript expression / R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford // *Nature methods*. – 2017. – Vol. 14. – №. 4. – P. 417-419.
111. Pearson J. C. A phylogeny of life-cycle patterns of the Digenea / J. C. Pearson // *Advances in parasitology*. – 1972. – Vol. 10. – P. 153-189.
112. Picard M. A. L. Sex-biased transcriptome of *Schistosoma mansoni*: host-parasite interaction, genetic determinants and epigenetic regulators are associated with sexual differentiation / M. A. L. Picard, J. Boissier, D. Roquis, C. Grunau, J.-F. Allienne, D. Duval, E. Toulza, N. Arancibia, C. R. Caffrey, T. Long, S. Nidelet, M. Rohmer, C. Cosseau // *PLoS neglected tropical diseases*. – 2016. – Vol. 10. – №. 9. – P. e0004930.
113. Pomaznoy M. Y. Whole transcriptome profiling of adult and infective stages of the trematode *Opisthorchis felineus* / M. Y. Pomaznoy, M. D. Logacheva, N. D. Young, A. A. Penin, N. I. Ershov, A. V. Katokhin, V. A. Mordvinov // *Parasitology international*. – 2016. – Vol. 65. – №. 1. – P. 12-19.
114. Protasio A. V. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni* / A. V. Protasio, I. J. Tsai, A. Babbage, S. Nichol, M. Hunt, M. A. Aslett, N. De Silva, G. S. Velarde, T. J. C. Anderson, R. C. Clark, C. Davidson, G. P. Dillon, N. E. Holroyd, P. T. LoVerde, C. Lloyd, J. McQuillan, G. Oliveira, T. D. Otto, S. J. Parker-Manuel, M. A. Quail, R. A. Wilson, A. Zerlotini, D. W. Dunne, M. Berriman // *PLoS neglected tropical diseases*. – 2012. – Vol. 6. – №. 1. – P. e1455.
115. Reinhard E. G. Landmarks of parasitology I. The discovery of the life cycle of the liver fluke / E. G. Reinhard // *Experimental Parasitology*. – 1957. – Vol. 6. – №. 2. – P. 208-232.

116. Ronquist F. MrBayes 3: Bayesian phylogenetic inference under mixed models / F. Ronquist, J. P. Huelsenbeck // *Bioinformatics*. – 2003. – Vol. 19. – №. 12. – P. 1572-1574.
117. Sahoo P. K. *De novo* whole transcriptome analysis of the fish louse, *Argulus siamensis*: first molecular insights into characterization of Toll downstream signalling molecules of crustaceans / P. K. Sahoo, B. Kar, A. Mohapatra, J. Mohanty // *Experimental parasitology*. – 2013. – Vol. 135. – №. 3. – P. 629-641.
118. Sandland G. J. Patterns of *Sphaeridiotrema pseudoglobulus* infection in sympatric and allopatric hosts (*Bithynia tentaculata*) originating from widely separated sites across the USA / G. J. Sandland, J. P. Peirce // *Parasitology Research*. – 2021. – Vol. 120. – №. 1. – P. 187-195.
119. Schlötterer C. Genes from scratch—the evolutionary fate of *de novo* genes / C. Schlötterer // *Trends in Genetics*. – 2015. – Vol. 31. – №. 4. – P. 215-219.
120. Shi L. Evolutionary analysis of the *Bacillus subtilis* genome reveals new genes involved in sporulation / L. Shi, A. Derouiche, S. Pandit, S. Rahimi, A. Kalantari, M. Futo, V. Ravikumar, C. Jers, V. R. S. S. Mokkalapati, K. Vlahoviček, I. Mijakovic // *Molecular biology and evolution*. – 2020. – Vol. 37. – №. 6. – P. 1667-1678.
121. Shoop W. L. Trematode transmission patterns / W. L. Shoop // *The Journal of Parasitology*. – 1988. – Vol. 74. – №. 1. – P. 46-59.
122. Simão F. A. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs / F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov // *Bioinformatics*. – 2015. – Vol. 31. – №. 19. – P. 3210-3212.
123. Smith-Unna R. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies / R. Smith-Unna, C. Boursnell, R. Patro, J. M. Hibberd, S. Kelly // *Genome research*. – 2016. – Vol. 26. – №. 8. – P. 1134-1144.
124. Steenstrup J. J. S. On the Alternation of Generations; Or the Propagation and Development of Animals Through Alternate Generations: A Peculiar Form of Fostering the Young in the Lower Classes of Animals / J. J. S. Steenstrup – Ray society – 1845.
125. Tautz D. Evolutionary Origin of Orphan Genes. In eLS, (Ed.) / D. Tautz, R. Neme, T. Domazet-Lošo – Wiley & Sons – 2013.
126. Tautz D. The discovery of *de novo* gene evolution / D. Tautz // *Perspectives in biology and medicine*. – 2014. – Vol. 57. – №. 1. – P. 149-161.
127. Tautz D. The evolutionary origin of orphan genes / D. Tautz, T. Domazet-Lošo // *Nature Reviews Genetics*. – 2011. – Vol. 12. – №. 10. – P. 692-702.
128. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021 / The UniProt Consortium // *Nucleic acids research*. – 2021. – Vol. 49. – №. D1. – P. D480-D489.

129. Tiessen A. Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes / A. Tiessen, P. Pérez-Rodríguez, L. J. Delaye-Arredondo // *BMC Research notes*. – 2012. – Vol. 5. – №. 1. – P. 1-23.
130. Toll-Riera M. Origin of primate orphan genes: a comparative genomics approach / M. Toll-Riera, N. Bosch, N. Bellora, R. Castelo, L. Armengol, X. Estivill, M. M. Albà // *Molecular biology and evolution*. – 2009. – Vol. 26. – №. 3. – P. 603-612.
131. Train C.-M. iHam and pyHam: visualizing and processing hierarchical orthologous groups / C.-M. Train, M. Pignatelli, A. Altenhoff, C. Dessimoz // *Bioinformatics*. – 2019. – Vol. 35. – №. 14. – P. 2504-2506.
132. Vasconcelos E. J. R. Atlas of *Schistosoma mansoni* long non-coding RNAs and their expression correlation to protein-coding genes / E. J. R. Vasconcelos, V. C. Mesel, L. F. daSilva, D. S. Pires, G. M. Lavezzo, A. S. A. Pereira, M. S. Amaral, S. Verjovski-Almeida // *Database*. – 2018. – Vol. 2018. – P. bay068.
133. Vasconcelos E. J. R. The *Schistosoma mansoni* genome encodes thousands of long non-coding RNAs predicted to be functional at different parasite life-cycle stages / E. J. R. Vasconcelos, L. F. daSilva, D. S. Pires, G. M. Lavezzo, A. S. A. Pereira, M. S. Amaral, S. Verjovski-Almeida // *Scientific Reports*. – 2017. – Vol. 7. – P. 10508.
134. Verjovski-Almeida S. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni* / S. Verjovski-Almeida, R. DeMarco, E. A. L. Martins, P. E. M. Guimarães, E. P. B. Ojopi, A. C. M. Paquola, J. P. Piazza, M. Y. Nishiyama Jr., J. P. Kitajima, R. E. Adamson, P. D. Ashton, M. F. Bonaldo, P. S. Coulson, G. P. Dillon, L. P. Farias, S. P. Gregorio, P. L. Ho, R. A. Leite, L. C. C. Malaquias, R. C. P. Marques, P. A. Miyasato, A. L. T. O. Nascimento, F. P. Ohlweiler, E. M. Reis, M. A. Ribeiro, R. G. Sá, G. C. Stukart, M. B. Soares, C. Gargioni, T. Kawano, V. Rodrigues, A. M. B. N. Madeira, R. A. Wilson, C. F. M. Menck, J. C. Setubal, L. C. C. Leite, E. Dias-Neto // *Nature genetics*. – 2003. – Vol. 35. – №. 2. – P. 148-157.
135. Wagner G. P. A model based criterion for gene expression calls using RNA-seq data / G. P. Wagner, K. Kin, V. J. Lynch // *Theory in Biosciences*. – 2013. – Vol. 132. – №. 3. – P. 159-164.
136. Wagner G. P. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples / G. P. Wagner, K. Kin, V. J. Lynch // *Theory in biosciences*. – 2012. – Vol. 131. – №. 4. – P. 281-285.
137. Wang B. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni* / B. Wang, J. J. Collins III, P. A. Newmark // *eLife*. – 2013. – Vol. 2. – P. e00768.

138. Wang B. Stem cell heterogeneity drives the parasitic life cycle of *Schistosoma mansoni* / B. Wang, J. Lee, P. Li, A. Saberi, H. Yang, C. Liu, M. Zhao, P. A. Newmark // eLife. – 2018. – Vol. 7. – P. e35449.
139. Wang J. Evolutionary transcriptomics of metazoan biphasic life cycle supports a single intercalation origin of metazoan larvae / J. Wang, L. Zhang, S. Lian, Z. Qin, X. Zhu, X. Dai, Z. Huang, C. Ke, Z. Zhou, J. Wei, P. Liu, N. Hu, Q. Zeng, B. Dong, Y. Dong, D. Kong, Z. Zhang, S. Liu, Y. Xia, Y. Li, L. Zhao, Q. Xing, X. Huang, X. Hu, Z. Bao, S. Wang // Nature Ecology & Evolution. – 2020. – Vol. 4. – №. 5. – P. 725-736.
140. Wang X. The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis* / X. Wang, W. Chen, Y. Huang, J. Sun, J. Men, H. Liu, F. Luo, L. Guo, X. Lv, C. Deng, C. Zhou, Y. Fan, X. Li, L. Huang, Y. Hu, C. Liang, X. Hu, J. Xu, X. Yu // Genome biology. – 2011. – Vol. 12. – №. 10. – P. 1-14.
141. Wang Z. RNA-Seq: a revolutionary tool for transcriptomics / Z. Wang, M. Gerstein, M. Snyder // Nature reviews genetics. – 2009. – Vol. 10. – №. 1. – P. 57-63.
142. Waterhouse R. M. BUSCO applications from quality assessments to gene prediction and phylogenomics / R. M. Waterhouse, M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, E. M. Zdobnov // Molecular biology and evolution. – 2018. – Vol. 35. – №. 3. – P. 543-548.
143. Wilson G. A. Orphans as taxonomically restricted and ecologically important genes / G. A. Wilson, N. Bertrand, Y. Patel, J. B. Hughes, E. J. Feil, D. Field // Microbiology. – 2005. – Vol. 151. – №. 8. – P. 2499-2501.
144. Wollenberg Valero K. C. Transcriptomic and macroevolutionary evidence for phenotypic uncoupling between frog life history phases / K. C. Wollenberg Valero, J. Garcia-Porta, A. Rodríguez, M. Arias, A. Shah, R. D. Randrianiaina, J. L. Brown, F. Glaw, F. Amat, S. Künzel, D. Metzler, R. D. Isokpehi, M. Vences // Nature communications. – 2017. – Vol. 8. – №. 1. – P. 1-9.
145. Wood D. E. Improved metagenomic analysis with Kraken 2 / D. E. Wood, J. Lu, B. Langmead // Genome biology. – 2019. – Vol. 20. – №. 1. – P. 1-13.
146. Wood D. E. Kraken: ultrafast metagenomic sequence classification using exact alignments / D. E. Wood, S. L. Salzberg // Genome biology. – 2014. – Vol. 15. – №. 3. – P. 1-12.
147. Xu F. High expression of new genes in trochophore enlightening the ontogeny and evolution of trochozoans / F. Xu, T. Domazet-Lošo, D. Fan, T. L. Dunwell, L. Li, X. Fang, G. Zhang // Scientific reports. – 2016. – Vol. 6. – №. 1. – P. 1-10.
148. Yates A. D. Ensembl 2020 / A. D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C.

- Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. McDowall, A. McMahon, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, G. Iisley, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino, P. Flicek // *Nucleic acids research*. – 2020. – Vol. 48. – №. D1. – P. D682-D688.
149. Young N. D. A portrait of the transcriptome of the neglected trematode, *Fasciola gigantica*— biological and biotechnological implications / N. D. Young, A. R. Jex, C. Cantacessi, R. S. Hall, B. E. Campbell, T. W. Spithill, S. Tangkawattana, P. Tangkawattana, T. Laha, R. B. Gasser // *PLoS neglected tropical diseases*. – 2011. – Vol. 5. – №. 2. – P. e1004.
150. Young N. D. The *Opisthorchis viverrini* genome provides insights into life in the bile duct / N. D. Young, N. Nagarajan, S. J. Lin, P. K. Korhonen, A. R. Jex, R. S. Hall, H. Safavi-Hemami, W. Kaewkong, D. Bertrand, S. Gao, Q. Seet, S. Wongkham, B. T. Teh, C. Wongkham, P. M. Intapan, W. Maleewong, X. Yang, M. Hu, Z. Wang, A. Hofmann, P. W. Sternberg, P. Tan, J. Wang, R. B. Gasser // *Nature communications*. – 2014. – Vol. 5. – №. 1. – P. 1-11.
151. Young N. D. Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini* / N. D. Young, B. E. Campbell, R. S. Hall, A. R. Jex, C. Cantacessi, T. Laha, W.-M. Sohn, B. Sripa, A. Loukas, P. J. Brindley, R. B. Gasser // *PLoS neglected tropical diseases*. – 2010. – Vol. 4. – №. 6. – P. e719.
152. Young N. D. Whole-genome sequence of *Schistosoma haematobium* / N. D. Young, A. R. Jex, B. Li, S. Liu, L. Yang, Z. Xiong, Y. Li, C. Cantacessi, R. S. Hall, X. Xu, F. Chen, X. Wu, A. Zerlotini, G. Oliveira, A. Hofmann, G. Zhang, X. Fang, Y. Kang, B. E. Campbell, A. Loukas, S. Ranganathan, D. Rollinson, G. Rinaldi, P. J. Brindley, H. Yang, J. Wang, J. Wang, R. B. Gasser // *Nature genetics*. – 2012. – Vol. 44. – №. 2. – P. 221-225.
153. Zajac N. Gene duplication and gain in the trematode *Atriophallophorus winterbourni* contributes to adaptation to parasitism / N. Zajac, S. Zoller, K. Seppälä, D. Moi, C. Dessimoz, J. Jokela, H. Hartikainen, N. Glover // *Genome biology and evolution*. – 2021. – Vol. 13. – №. 3. – P. evab010.
154. Zambelli F. RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments / F. Zambelli, F. Mastropasqua, E. Picardi, A.

- M. D'Erchia, G. Pesole, G. Pavesi // *Nucleic acids research*. – 2018. – Vol. 46. – №. 8. – P. e46-e46.
155. Zarowiecki M. What helminth genomes have taught us about parasite evolution / M. Zarowiecki, M. Berriman // *Parasitology*. – 2015. – Vol. 142. – №. S1. – P. S85-S97.
156. Zerlotini A. SchistoDB: an updated genome resource for the three key schistosomes of humans / A. Zerlotini, E. R. G. R. Aguiar, F. Yu, H. Xu, Y. Li, N. D. Young, R. B. Gasser, A. V. Protasio, M. Berriman, D. S. Roos, J. C. Kissinger, G. Oliveira // *Nucleic Acids Research*. – 2012. – Vol. 41. – №. D1. – P. D728-D731.
157. Zhang X.-X. Complex and dynamic transcriptional changes allow the helminth *Fasciola gigantica* to adjust to its intermediate snail and definitive mammalian hosts / X.-X. Zhang, K. Cwiklinski, R.-S. Hu, W.-B. Zheng, Z.-A. Sheng, F.-K. Zhang, H. M. Elsheikha, J. P. Dalton, X.-Q. Zhu // *BMC genomics*. – 2019. – Vol. 20. – №. 1. – P. 1-18.
158. Zhang X.-X. *De novo* transcriptome sequencing and analysis of the juvenile and adult stages of *Fasciola gigantica* / X.-X. Zhang, W. Cong, H. M. Elsheikha, G.-H. Liu, J.-G. Ma, W.-Y. Huang, Q. Zhao, X.-Q. Zhu // *Infection, Genetics and Evolution*. – 2017. – Vol. 51. – P. 33-40.
159. Zhang Y. ComBat-seq: batch effect adjustment for RNA-seq count data / Y. Zhang, G. Parmigiani, W. E. Johnson // *NAR genomics and bioinformatics*. – 2020. – Vol. 2. – №. 3. – P. lqaa078.
160. Zhou Y. The *Schistosoma japonicum* genome reveals features of host-parasite interplay / Y. Zhou, H. Zheng, X. Chen, L. Zhang, K. Wang, J. Guo, Z. Huang, B. Zhang, W. Huang, K. Jin, D. Tonghai, M. Hasegawa, L. Wang, Y. Zhang, J. Zhou, L. Tao, Z. Cao, Y. Li, T. Vinar, B. Brejova, D. Brown, M. Li, D. J. Miller, D. Blair, Y. Zhong, Z. Chen, F. Liu, W. Hu, Z.-Q. Wang, Q.-H. Zhang, H.-D. Song, S. Chen, X. Xu, B. Xu, Z. Ju, Y. Cheng, P. J. Brindley, D. P. McManus, Z. Feng, Z.-G. Han, G. Lu, S. Ren, Y. Wang, W. Gu, H. Kang, J. Chen, X. Chen, S. Chen, L. Wang, J. Yan, B. Wang, X. Lv, L. Jin, B. Wang, S. Pu, X. Zhang, W. Zhang, Q. Hu, G. Zhu, J. Wang, J. Yu, J. Wang, H. Yang, Z. Ning, M. Beriman, C.-L. Wei, Y. Ruan, G. Zhao, S. Wang // *Nature*. – 2009. – Vol. 460. – №. 7253. – P. 345.

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 phylostratigraphy 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).

ПРИЛОЖЕНИЕ 1. Список библиотек коротких парных прочтений, использованных в исследовании (страницы 42, 50)

Вид	ID в базе данных SRA	Фаза жизненного цикла	Повторность	DOI
<i>F. gigantea</i>	SRR13334789	Яйцо	1	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334765	Яйцо	2	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334766	Яйцо	3	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334767	Мирацидий	1	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334768	Мирацидий	2	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334769	Мирацидий	3	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334770	Редия	1	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334771	Редия	2	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334772	Редия	3	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334773	42-дневная ювениль	1	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334774	42-дневная ювениль	2	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334775	42-дневная ювениль	3	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334776	Церкария	1	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334777	Церкария	2	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334778	Церкария	3	https://doi.org/10.1186/s12864-019-6103-5

<i>F. gigantea</i>	SRR13334779	Метацеркария	1	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334780	Метацеркария	2	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334781	Метацеркария	3	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334782	70-дневная ювениль	1	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334783	70-дневная ювениль	2	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334784	70-дневная ювениль	3	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334785	Марита	1	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334786	Марита	2	https://doi.org/10.1186/s12864-019-6103-5
<i>F. gigantea</i>	SRR13334787	Марита	3	https://doi.org/10.1186/s12864-019-6103-5
<i>F. hepatica</i>	ERR576952	Марита	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576953	Марита	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576954	Яйцо	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576955	Яйцо	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576956	21-дневная ювениль	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576957	21-дневная ювениль	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576958	Метацеркария	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576959	Метацеркария	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576964	Метацеркария	2	https://doi.org/10.1186/s13059-015-0632-2

<i>F. hepatica</i>	ERR576965	Метацеркария	3	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576960	Ювениль через 1 час после покидания цисты	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576961	Ювениль через 1 час после покидания цисты	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576966	Ювениль через 1 час после покидания цисты	2	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576962	Ювениль через 3 часа после покидания цисты	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576963	Ювениль через 3 часа после покидания цисты	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576969	Ювениль через 3 часа после покидания цисты	2	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576967	Ювениль через 24 часа после покидания цисты	1	https://doi.org/10.1186/s13059-015-0632-2
<i>F. hepatica</i>	ERR576968	Ювениль через 24 часа после покидания цисты	2	https://doi.org/10.1186/s13059-015-0632-2

<i>F. hepatica</i>	SRR2038743	Яйцо	1	https://doi.org/10.1371/journal.pgen.1006537
<i>F. hepatica</i>	SRR2039050	Метацеркария	1	https://doi.org/10.1371/journal.pgen.1006537
<i>F. hepatica</i>	SRR2039051	Метацеркария	2	https://doi.org/10.1371/journal.pgen.1006537
<i>F. hepatica</i>	SRR2038730	Марита	1	https://doi.org/10.1371/journal.pgen.1006537
<i>F. hepatica</i>	SRR2038734	Марита	2	https://doi.org/10.1371/journal.pgen.1006537
<i>P. simillimum</i>	SRR8466830	Редия	1	https://doi.org/10.1186/s13071-020-04424-4
<i>P. simillimum</i>	SRR8466836	Редия	2	https://doi.org/10.1186/s13071-020-04424-4
<i>P. simillimum</i>	SRR8466809	Церкария	1	https://doi.org/10.1186/s13071-020-04424-4
<i>P. simillimum</i>	SRR8466817	Церкария	2	https://doi.org/10.1186/s13071-020-04424-4
<i>P. simillimum</i>	SRR8466919	Марита	1	https://doi.org/10.1186/s13071-020-04424-4
<i>P. simillimum</i>	SRR8466955	Марита	2	https://doi.org/10.1186/s13071-020-04424-4
<i>S. pseudoglobulus</i>	SRR8466753	Редия	1	https://doi.org/10.1186/s13071-020-04424-4
<i>S. pseudoglobulus</i>	SRR8466759	Редия	2	https://doi.org/10.1186/s13071-020-04424-4
<i>S. pseudoglobulus</i>	SRR8466356	Церкария	1	https://doi.org/10.1186/s13071-020-04424-4
<i>S. pseudoglobulus</i>	SRR8466748	Церкария	2	https://doi.org/10.1186/s13071-020-04424-4
<i>S. pseudoglobulus</i>	SRR8466776	Марита	1	https://doi.org/10.1186/s13071-020-04424-4
<i>S. pseudoglobulus</i>	SRR8466792	Марита	2	https://doi.org/10.1186/s13071-020-04424-4
<i>S. mansoni</i>	ERR022872	Церкария	1	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022877	Церкария	2	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022878	Церкария	3	https://doi.org/10.1371/journal.pntd.0001455

<i>S. mansoni</i>	ERR022876	Шистосомула через 3 часа после заражения	1	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022874	Шистосомула через 3 часа после заражения	1	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022879	Шистосомула через 3 часа после заражения	2	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022880	Шистосомула через 24 часа после заражения	1	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022881	Шистосомула через 24 часа после заражения	2	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022882	Шистосомула через 24 часа после заражения	3	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022883	Шистосомула через 24 часа после заражения	4	https://doi.org/10.1371/journal.pntd.0001455
<i>S. mansoni</i>	ERR022873	Марита	1	https://doi.org/10.1371/journal.pntd.0001455
<i>T. regenti</i>	SRR2170925	Церкария	1	https://doi.org/10.1371/journal.pntd.0004406
<i>T. regenti</i>	SRR2170926	Церкария	2	https://doi.org/10.1371/journal.pntd.0004406
<i>T. regenti</i>	SRR2170927	Церкария	3	https://doi.org/10.1371/journal.pntd.0004406
<i>T. regenti</i>	SRR2170928	Церкария	4	https://doi.org/10.1371/journal.pntd.0004406
<i>T. regenti</i>	SRR2170929	Шистосомула	1	https://doi.org/10.1371/journal.pntd.0004406

<i>T. regenti</i>	SRR2170930	Шистосомула	2	https://doi.org/10.1371/journal.pntd.0004406
<i>T. regenti</i>	SRR2170931	Шистосомула	3	https://doi.org/10.1371/journal.pntd.0004406
<i>T. regenti</i>	SRR2170935	Шистосомула	4	https://doi.org/10.1371/journal.pntd.0004406
<i>T. szidati</i>	SRR7687402	Церкария	1	https://doi.org/10.1038/s41598-018-37669-2
<i>T. szidati</i>	SRR7687403	Церкария	2	https://doi.org/10.1038/s41598-018-37669-2
<i>T. szidati</i>	SRR7687405	Церкария	3	https://doi.org/10.1038/s41598-018-37669-2
<i>T. szidati</i>	SRR7687404	Церкария	4	https://doi.org/10.1038/s41598-018-37669-2
<i>T. szidati</i>	SRR7687401	Шистосомула	1	https://doi.org/10.1038/s41598-018-37669-2
<i>T. szidati</i>	SRR7687400	Шистосомула	2	https://doi.org/10.1038/s41598-018-37669-2
<i>T. szidati</i>	SRR7687399	Шистосомула	3	https://doi.org/10.1038/s41598-018-37669-2
<i>T. szidati</i>	SRR7687398	Шистосомула	4	https://doi.org/10.1038/s41598-018-37669-2