

Санкт-Петербургский государственный университет

***ВАСИЛЬЕВ Артем Викторович***

**Выпускная квалификационная работа**

Эволюционные особенности структуры гена

*Nxf1* (nuclear export factor) у животных

Уровень образования: магистратура

Направление 06.04.01 “Биология”

Основная образовательная программа магистратуры “Биоинформатика”  
(шифр ВМ.5758.2023)

Работа выполнена на базе лаборатории генетики животных  
кафедры генетики и биотехнологии СПбГУ

Научный руководитель:

к.б.н., доцент, кафедра генетики и биотехнологии,

Голубкова Елена Валерьевна

Рецензент:

заведующая лабораторией, ведущий научный сотрудник,  
лаборатория эволюционной геномики и палеогеномики, ЗИН,

к.б.н., с.н.с.,

Абрамсон Наталья Иосифовна

Санкт-Петербург

2025

## Оглавление

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Введение .....</b>   | <b>3</b>  |
| 1.1      | Цель работы.....  | 4         |
| 1.2      | Задачи .....  | 4         |
| <b>2</b> | <b>Обзор литературы .....</b>                                       | <b>5</b>  |
| 2.1      | Механизмы усложнения организации генома .....                       | 5         |
| 2.2      | Значимость интронов .....   | 5         |
| 2.3      | Семейство генов <i>Nxf</i> .....                                    | 6         |
| 2.4      | Структура и функции гена <i>Nxf1</i> и его белковых продуктов ..... | 7         |
| <b>3</b> | <b>Материалы и методы .....</b>                                     | <b>9</b>  |
| 3.1      | Первичный анализ.....   | 9         |
| 3.2      | Загрузка данных.....  | 9         |
| 3.3      | Увеличение выборок .....  | 9         |
| 3.4      | Парсинг результатов .....   | 10        |
| 3.5      | Множественные выравнивания .....                                    | 10        |
| 3.6      | Поиск консервативных мотивов внутри “кассетного” интрона .....      | 10        |
| 3.7      | Построение и анализ вторичных структур РНК.....                     | 10        |
| 3.8      | Филогенетический анализ .....                                       | 11        |
| 3.9      | Настройки системы и доступность скриптов .....                      | 11        |
| <b>4</b> | <b>Результаты .....</b>   | <b>12</b> |
| 4.1      | Анализ всех найденных видов .....                                   | 12        |
| 4.2      | Подробный анализ <i>Actinopterygii</i> .....                        | 13        |
| <b>5</b> | <b>Обсуждение .....</b>   | <b>18</b> |
| 5.1      | Анализ всех найденных видов .....                                   | 18        |
| 5.2      | Подробный анализ <i>Actinopterygii</i> .....                        | 19        |
| <b>6</b> | <b>Выводы .....</b>   | <b>20</b> |
| <b>7</b> | <b>Список литературы .....</b>                                      | <b>21</b> |
| <b>8</b> | <b>Приложение .....</b>   | <b>25</b> |
| <b>9</b> | <b>Благодарности .....</b>  | <b>33</b> |

## Введение

Для большинства генов высших эукариот характерна мозаичная структура, в составе которой выделяют кодирующие участки — экзоны и некодирующие — интроны. В процессе созревания транскрипта интроны, как правило, вырезаются в процессе сплайсинга, и из ядра выходит мРНК, лишенная интронных последовательностей. Однако альтернативный сплайсинг позволяет получать несколько различных зрелых мРНК из одной пре-мРНК, что значительно расширяет протеом без увеличения числа генов.

Особый интерес представляют транскрипты, сохраняющие интрон (intron retention, IR). Как правило, в таком интроне присутствует преждевременный стоп-кодон (premature termination codon, PTC), поэтому существует специальный механизм для проверки качества транскриптов перед выходом из ядра — нонсенс-опосредованный распад мРНК (nonsense mediated mRNA decay, NMD), который препятствует выходу таких транскриптов в цитоплазму. Однако, несмотря на наличие специфического механизма, среди различных групп эволюционно далеких друг от друга, описаны случаи существования транскриптов с сохраненным интроном. Отдельно можно выделить дрозофилу и человека, для которых известно семейство генов *Nxf* (nuclear export factor), в котором нас заинтересовал ген *Nxf1*. Данный ген кодирует белок, являющийся основным транспортером мРНК из ядра в цитоплазму.

В состав последовательности гена *Nxf1* (nuclear export factor 1) входит так называемая “консервативная кассета”, которая включает два коротких экзона размером 110 и 37 нуклеотидов в каноническом варианте и “кассетный” интрон между ними. Названия сформулированы нашей научной группой и будут использоваться в дальнейшем повествовании. Эта структура сохраняется также и у представителей других филогенетических групп. Благодаря образованию специфической вторичной структуры или наличию в последовательности интрона специфических последовательностей, например конститутивного транспортного элемента (constitutive transport element, CTE), транскрипт, содержащий преждевременный стоп-кодон, избегает NMD и может кодировать укороченную форму белка.

Анализ подобных транскриптов показал, что консервативные элементы “касеты” *Nxf1* специфичны для разных клад организмов, а интрон-содержащие транскрипты формируют уникальные вторичные структуры, что подчеркивает эволюционную и функциональную значимость интронов.

Научная новизна работы заключается в сравнительном анализе структуры гена *Nxf1* у представителей различных филогенетических групп, данных по которым ранее не было, с целью выявления закономерностей эволюции нуклеотидной последовательности гена *Nxf1* и его белковых продуктов.

В моей бакалаврской работе было показано, что “консервативная кассета” сохраняет свойства внутри артропод, особенно внутри семейства Drosophilidae, однако вопрос о степени консервативности и специфике структурных элементов у более ши-

рокого круга организмов остается открытым. Помимо сравнительного анализа последовательностей, важной частью исследования является построение вторичных структур интрон-содержащих транскриптов и выявление консервативных мотивов внутри интрона, способствующих его сохранению и избеганию нонсенс-опосредованного распада.

### **Цель работы**

Изучить структуру гена *Nxf1* у представителей разных филогенетических групп животных для выявления эволюционных закономерностей и особенностей “кассетной” структуры, а также проанализировать вторичные структуры интрон-содержащих транскриптов.

### **Задачи**

1. Найти нуклеотидные и аминокислотные последовательности гена *Nxf1* у различных групп животных.
2. Произвести поиск “консервативной кассеты” в нуклеотидной последовательности гена у найденных организмов.
3. Выполнить анализ структуры “консервативной кассеты”, сравнить полученные последовательности между собой.
4. Выявить и охарактеризовать консервативные участки “кассетного” интрона и прилегающих экзонов у видов из исследуемых таксонов.
5. Провести анализ вторичной структуры интрон-содержащих транскриптов и оценить консервативные мотивы внутри интрона, потенциально способствующие его сохранению.

## Обзор литературы

### Механизмы усложнения организации генома

Увеличение разнообразия транскриптома и протеома у эукариот во многом достигается не только за счет классического эксцизионного сплайсинга [1], но и за счет ряда альтернативных механизмов обработки пре-мРНК. В частности, альтернативный сплайсинг позволяет из одного транскрипта формировать несколько зрелых мРНК, отличающихся включением или исключением отдельных экзонов и участков [2].

Одним из ключевых вариантов такого процесса является удержание интронов (intron retention, IR), когда интрон не удаляется и остается в составе зрелой мРНК [3]. Часто подобное сохранение интрона приводит к появлению в получившемся транскрипте преждевременных стоп-кодонах (premature termination codons, PTC), что запускает нонсенс-опосредованный распад (nonsense mediated mRNA decay, NMD) [4]. Тем не менее в ряде случаев, последовательность интрона может включать специфические, функционально-значимые последовательности, такие как, например, конститутивный транспортный элемент (constitutive transport element, CTE) [5]. Также интроны могут оказывать влияние на формирование устойчивой вторичной структуры, препятствующей связыванию факторов NMD, что позволяет транскриптам не подвергаться деградации [6]. Примечательно, что избегающие распада транскрипты способны даже участвовать в дальнейшем синтезе белка [7].

Кроме IR, альтернативный сплайсинг включает пропуск экзонов, использование альтернативных сайтов на 5'- и 3'-концах и кассетное включение/исключение блоков экзонов [4]. В совокупности эти механизмы значительно расширяют репертуар возможных транскриптов без необходимости увеличения числа генов. Например, у многих многоклеточных организмов до 95% генов подвергаются хотя бы одному типу альтернативного сплайсинга [7]. Для разных организмов продемонстрирована консервативность наличия транскрипта с сохраненным интроном, что подчеркивает эволюционную значимость интронов [6].

Таким образом, именно через комбинирование альтернативных способов сплайсинга, особенно удержания интронов, эукариоты получают мощный инструмент транскрипционной и белковой вариативности, что способствует адаптации и усложнению биологических процессов.

### Значимость интронов

Традиционно интроны воспринимались лишь как “ненужные” вставки, но современные исследования убедительно показывают, что их функции выходят далеко за рамки простой “пустоты”.

Во-первых, наличие интронных последовательностей может значительно усиливать уровень экспрессии генов. Эксперименты на клеточных системах SV40, дрожжах *Saccharomyces cerevisiae* и млекопитающих демонстрируют, что удаление ключевых интронов приводит к резкому снижению эффективности транскрипции и транс-

ляции [8, 9]. На этих же дрожжах более 10 лет назад наблюдали, как короткие открытые рамки считывания (open reading frame, ORF) эволюционировали в функциональные гены благодаря непрерывному эволюционному процессу [10].

Во-вторых, интроны влияют на чувствительность мРНК к нонсенс-опосредованному распаду. Если интрон попадает в 5'- или 3'-UTR, его присутствие может менять архитектуру сплайсосомного комплекса, корректируя доступность РТС и, соответственно, баланс между сохранением транскрипта и его деградацией через NMD [4, 11].

Третья важная роль интронов заключается в транспорте мРНК из ядра в цитоплазму. Долгое время считалось, что только полностью сплайсированные транскрипты эффективно экспортируются, однако при помощи флуоресцентной гибридизации *in situ* (fluorescent *in situ* hybridization, FISH) было показано, что РНК с сохраненными интронами также могут накапливаться в цитоплазме и функционировать там [11–13]. Это потребовало пересмотра классических представлений об экспорте мРНК.

Кроме регуляции экспрессии и транспорта, интроны участвуют в организации хроматиновой структуры. Концевые последовательности интронов образуют участки с пониженной плотностью нуклеосом, что способствует более четкому разделению экзонов и облегчает процесс транскрипции [14].

Наконец, интроны могут выполнять более специфические функции. Так, первый интрон гена *oskar* у *Drosophila* участвует в локализации мРНК в ооците [15], а длинные интронные вставки могут снижать интерференцию Хилла–Робертсона, улучшая кроссинговер в определенных регионах генома [16]. Результаты, полученные с помощью полногеномного поиска ассоциаций (genome-wide association study, GWAS) показывают, что однонуклеотидные варианты (single nucleotide variants, SNV) в интронных областях часто связаны с предрасположенностью к различным метаболическим и иммунным заболеваниям человека [17].

Таким образом, интроны выполняют сложные регуляторные функции — от контроля уровня экспрессии до обеспечения оптимальной архитектуры хроматина и тканеспецифической регуляции транскриптов.

### **Семейство генов *Nxf***

Перед описанием основного объекта данного исследования следует сказать несколько вводных слов про само семейство.

Семейство генов *Nxf* (nuclear export factor) названо по функции продукта их наиболее известного представителя — *Nxf1*, который обеспечивает экспорт большинства мРНК из ядра в цитоплазму. Распространение этих генов наблюдается у всех эукариот группы Opisthokonta, однако их число и структурные особенности заметно различаются между таксонами. У грибов обычно присутствует единственная копия *Nxf*, тогда как в геномах растений и некоторых протистов такие гены могут отсут-

ствовать полностью. У животных же часто встречается от двух до пяти паралогов, что свидетельствует об активных дупликационных процессах в эволюции этого семейства [6].

Характерной особенностью гена *Nxf1* является наличие в составе его структуры “кассетного” интрона, расположенного между двумя небольшими экзонами размером 110 и 37 пар нуклеотидов в каноническом варианте. При альтернативном сплайсинге этот интрон может сохраняться в зрелой мРНК, неся внутри себя преждевременный стоп-кодон, возникающий за счет особенностей размеров упомянутых ранее экзонов. Сохранению интрона может способствовать наличие определенных транспортных последовательностей, как у млекопитающих, или формирование устойчивой вторичной структуры, как у дрозофил. В итоге такие транскрипты избегают NMD и могут кодировать укороченные, но функционально активные белки [6, 18].

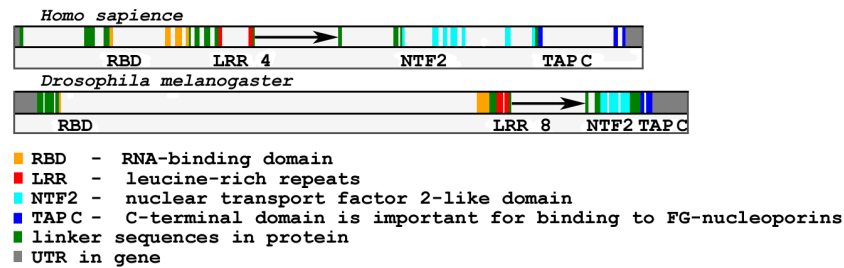
В одной из первых статей нашей научной группы [6] были определены следующие характеристики “кассетного” интрона у 3 таксономических групп животных:

- **Позвоночные.** У гена *Nxf1* интрон располагается между 10-м и 11-м экзонами. Вставка содержит несколько консервативных мотивов, включая фрагмент, похожий на СТЕ, необходимый для экспорта частично сплайсированных РНК.
- **Дрозофилиды.** У представителей данного семейства кассетный интрон локализуется между 5-м и 6-м экзонами. В нем отсутствуют длинные гомологичные вставки, но присутствуют два тракта, обогащенные аденином (А) и формирующие прочную вторичную структуру, которая предположительно защищает транскрипт от деградации в ядре.
- **Нематоды.** Интрон может располагаться между 5-м и 6-м или 6-м и 7-м экзонами, но он гораздо короче и богат тиминами (Т). Участки протяженной гомологии отсутствуют.

Перечисленные различия отражают долгую эволюцию семейства *Nxf*: от сравнительно простой короткой вставки у нематод до сложных СТЕ-подобных мотивов у позвоночных, подчеркивая ключевую роль “кассетного” интрона в посттранскрипционной регуляции [6, 18]. Перейдем к более подробному рассмотрению гена *Nxf1*.

### Структура и функции гена *Nxf1* и его белковых продуктов

Первоначально белок *Nxf1* у человека был охарактеризован как кофактор белка Тір герпеса *saimiri* и был назван ТАР – Tip-associated protein, или белок, ассоциированный с Тір [19]. Позже было выяснено, что он отвечает за экспорт несплайсированных и частично сплайсированных ретровирусных мРНК путем распознавания СТЕ-структуры в их последовательностях [5]. У дрозофилид этот ген также называют *sbr* от small bristles (маленькие щетинки), и мы предполагаем, что функцию СТЕ у них выполняет образуемая благодаря сохранению интрона вторичная структура.



**Рис. 1:** Интрон-экзонная структура для генов *Hs Nxf1* и *Dm Nxf1*. Стрелки обозначают “кассетный” интрон. Цвета экзонов отображают белковые домены [6].

Белок Nxf1 включает несколько функциональных доменов (рис. 1): RBD (домен связывания РНК), четыре лейцин-обогащенных повтора (LRR), NTF2-подобный домен, UBA-подобный домен и сигналы ядерной локализации в нетранслируемой области до экзонов. В совокупности эти домены обеспечивают узнавание мРНК и взаимодействие с компонентами ядерного порового комплекса, делая *Nxf1* основным экспортером мРНК [6, 20].

Кассетный интрон, встроенный между сегментами RBD+LRR и NTF2L+UBA, выступает в роли “переключателя”. При его сохранении происходит синтез укороченных изоформ белка, обладающих специфической активностью [6].

Кроме классического транспорта мРНК, у *Drosophila melanogaster* sbr выполняет органо- и тканеспецифические функции. В сперматогенных клетках ген продуцирует укороченную форму sbr, необходимую для нормального сперматогенеза: без нее наблюдается резкое снижение фертильности [21]. В центральной нервной системе sbr участвует в формировании границ между областями мозгового вещества зрительной системы, локализуясь в специфических нейронах и глиальных клетках и регулируя их ядерно-цитоплазматические комплексы [22].

У млекопитающих подобные семенниково-специфические транскрипты не выявлены, однако для мышей *Mus musculus* и человека известны паралоги из семейства генов *Nxf*, обладающие подобной функцией [7].

Аналогичные эволюционно значимые особенности кассетного интрона отмечены у Chiroptera (Летучие мыши). Сравнительный анализ продемонстрировал значимость кассетного интрона в эволюции гена *Nxf1* у данной группы организмов [23].

Также в моей бакалаврской работе было показано, что структура “консервативной кассеты” является специфической для таксонов более низкого ранга у всех взятых в анализ артропод (89 видов), а интрон-содержащие транскрипты проанализированных дрозофилид (37 видов) формируют специфические вторичные структуры, имеющие А-обогащенные участки.

Таким образом, Nxf1 представляет собой пример многофункционального белка, чья доменная организация и альтернативные формы позволяют выполнять как основную задачу — экспорт мРНК, так и специализированные функции в разных тканях у разных таксонов.



## Материалы и методы

### Первичный анализ

В качестве отправной точки был произведен поиск гена *Nxf1* внутри веб-сервиса NCBI [24]. Полученные данные были сохранены в текстовом формате и загружены в виде tsv-таблицы с помощью пакета pandas v2.2.3 [25] для языка программирования Python v3.12.6 [26]. Всего был найден 651 вид, содержащий анализируемый ген, большинство из которых относятся к Deuterostomia (Вторичноротые) – 436 видов. Таким образом, в качестве материалов выступали нуклеотидные и белковые последовательности, соответствующие гену *Nxf1*, из открытых баз данных NCBI [24].

Большинство этапов последующего анализа реализовано в виде отдельных скриптов, разработанных в рамках данной работы, если не указано другое. Для логического разделения на блоки был использован Jupyter Notebook v1.1.1 [27].

По данным из полученной таблицы в порядке поискового эксперимента было построено филогенетическое дерево по найденным видам для оценки количества видов в таксонах более низкого ранга. Для глубокого анализа было принято решение сфокусироваться на организмах, относящихся к группе Protostomia (Первичноротые), Cnidaria (Стрекающие), а также на всех группах из Deuterostomia в таблице за исключением Mammalia (Млекопитающие).

### Загрузка данных

Для найденных организмов с помощью пакета NCBI E-utilities из BioPython v1.85 [28] и NCBI Datasets Command-Line Interface (CLI) v18.0.2 [29] были загружены нуклеотидные последовательности гена, кодирующих участков и мРНК, а также аминокислотные последовательности белка в формате FASTA и аннотации для гена в GenBank-формате, необходимые для получения нуклеотидных последовательностей экзонов и поиска “консервативной кассеты”. Затем были получены и проанализированы интересующие нас участки экзон-интрон-экзонной структуры и созданы файлы со всеми экзонами и “кассетным” интроном для всех организмов, у которых получилось найти “кассету”. Данные файлы будут необходимы для последующего анализа.

### Увеличение выборок

Учитывая очень маленькие выборки во многих анализируемых группах (например, Cnidaria – 4 вида, Spiralia – 9 видов), было принято решение по увеличению их количества. Для этой цели, учитывая разнообразия полученных генов даже внутри одной таксономической группы, самым эффективным вариантом оказалось использование PSI-BLAST [30]. В качестве запроса (Query), или референса, использовались белковые последовательности тех организмов, у которых была найдена “кассета”. Для проведения PSI-BLAST были выбраны настройки по-умолчанию за исключением параметра Organism: поиск проводился внутри таксономической группы, к которой принадлежал референс, также референс был исключен из поиска.

## **Парсинг результатов**

Парсинг результатов BLAST также осуществлялся с помощью пакета BioPython [28] и специально разработанных скриптов. Он включал в себя фильтрацию данных по параметрам процента покрытия (Query Coverage, QC), длине и сходству (Per. Ident) найденных последовательностей (Subject), а также загрузку нуклеотидных и белковых последовательностей, однако реализация отличалась из-за особенностей баз данных NCBI [24]. Получение “кассеты” было произведено по тому же принципу, но, опять же, с отличиями. Благодаря данному шагу удалось увеличить выборки суммарно на 117 видов. К сожалению, для некоторых таксономических групп увеличение выборки оказалось невозможным в связи с отсутствием у некоторых организмов интересующего нас участка.

## **Множественные выравнивания**

Множественные выравнивания осуществлялись с помощью алгоритма MAFFT [31], 10 итераций, остальные настройки по-умолчанию, в программе Unipro UGENE v52.0 [32].

## **Поиск консервативных мотивов внутри “кассетного” интрона**

Анализ видов из *Deuterostomia* изначально шел более благоприятно за счет большого сходства последовательностей, в том числе интронных, и большего количества видов в группах. Для них также были загружены все необходимые файлы и произведен поиск и анализ “консервативной кассеты”. Мы решили сосредоточить свое внимание на организмах из *Actinopterygii* (Лучеперые рыбы), 72 вида, так как данных по ним ранее получено не было.

Учитывая большую степень сходства интронных последовательностей, с помощью пакета инструментов MEME Suite v5.5.8 [33] локально был произведен поиск консервативных мотивов внутри “кассетного” интрона. Найденные мотивы, у которых  $E\text{-value} < 0.05$  также локально были проанализированы с помощью Tomtom [34] из того же пакета. Для описанного шага была взята база данных JASPAR2024 CORE (NON-REDUNDANT) DNA.

## **Построение и анализ вторичных структур РНК**

С помощью инструмента RNAfold v2.7.0 из пакета ViennaRNA [35] были построены вторичные структуры РНК для нуклеотидных последовательностей в двух вариантах (MFE и Centroid), содержащих экзоны и “кассетный” интрон, так как мы предполагаем, что избегание интроном сплайсинга может быть опосредовано образованной им специфической вторичной структурой. Учитывая данное предположение, разумным шагом также являлся анализ “силы сайтов сплайсинга”, проведенный с помощью MaxEntScan [36].

Также с помощью скриптов цветом были выделены интронные последовательности внутри вторичной структуры и найденный мотив у Actinopterygii, который предположительно является СТЕ.

### **Филогенетический анализ**

Для Actinopterygii также был проведен филогенетический анализ, включающий построение и визуализацию деревьев. Для данной цели использовались самые популярные и проверенные временем инструменты. Построение деревьев осуществлялось с помощью IQ-TREE v2.4.0 [37], визуализация – с помощью Figtree v1.4.4 [38].

### **Настройки системы и доступность скриптов**

Работа проводилась в виртуальном окружении Mamba v1.5.5 [39], использованные пакеты и примеры анализа в Jupyter Notebooks можно найти в GitHub [40] репозитории автора: <https://github.com/ArtemVaska/Diploma>.

Для написания ВКР была использована система верстки LaTeX v4.76 [41], таблицы генерировались в веб-сервисе TablesGenerator [42]. Большинство рисунков обработано с помощью веб-сервиса draw.io [43]. Построение графиков осуществлялось с помощью пакета для Python matplotlib v3.10.3 [44].

Все шаги анализа проводились на базе операционной системы Linux Ubuntu 22.04 [45].

## Результаты

### Анализ всех найденных видов

Были проанализированы 413 нуклеотидных последовательностей гена *Nxf1* у представителей различных филогенетических групп из клад Cnidaria (Стрекающие) и Bilateria (Двусторонне-симметричные). Организмы, относящиеся к Mammalia, в анализ не были взяты в связи с уже имеющимися для них данными.

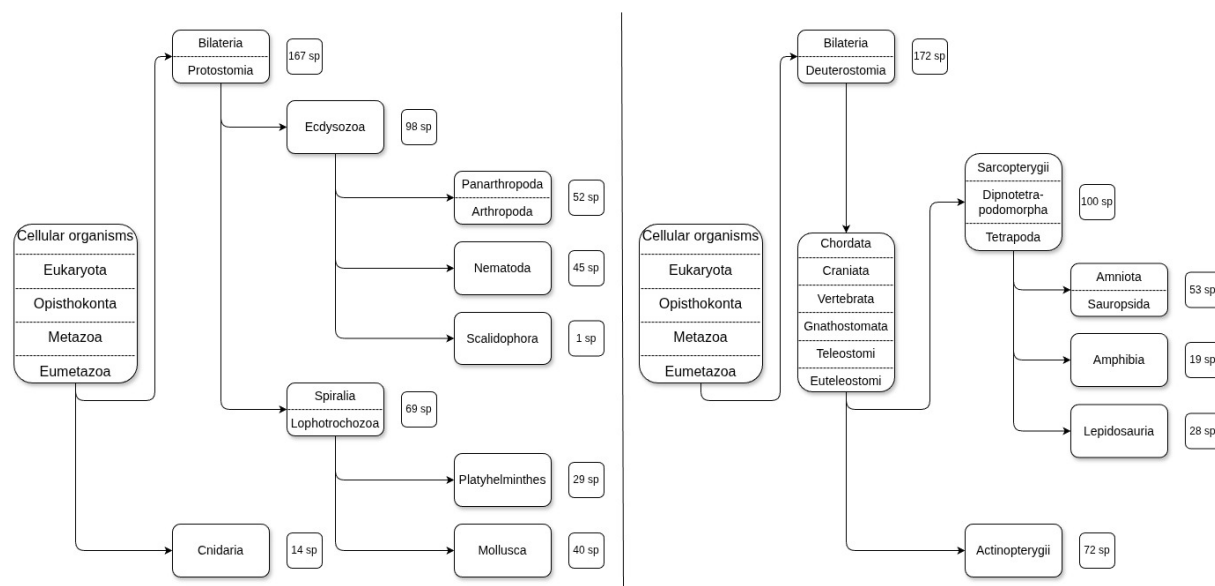
Для таксономических групп более низкого ранга с небольшим количеством видов в них с помощью PSI-BLAST были увеличены выборки, где это оказалось возможным, результат продемонстрирован на таблице 1.

**Таблица 1: Результат увеличения выборки с помощью PSI-BLAST.**

| Филогенетическая группа | Таксон высокого ранга | Видов до PSI-BLAST | Видов добавлено | Итого видов |
|-------------------------|-----------------------|--------------------|-----------------|-------------|
| Bilateria→Protostomia   | Ecdysozoa             | 56                 | 42              | 98          |
|                         | Spiralia              | 6                  | 63              | 69          |
| Cnidaria                | Anthozoa              | 2                  | 12              | 14          |

В итоге для 353 видов удалось найти “консервативную кассету” и продолжить дальнейший анализ.

На рисунке 2 отображено распределение исследованных видов по таксонам высокого ранга. Последовательно идущие таксоны объединены в один блок и разделены пунктиром.



**Рис. 2: Количество видов, взятых в анализ, для разных таксономических групп.**

В таблице 2 можно ознакомиться с некоторыми характеристиками “консервативной кассеты” для всех исследованных групп.

Для всех видов, имеющих “кассету”, были построены вторичные структуры для интрон-содержащего транскрипта с выделением цветом “кассетного” интрона.

**Таблица 2: Характеристики “консервативной кассеты” для всех исследованных групп.**

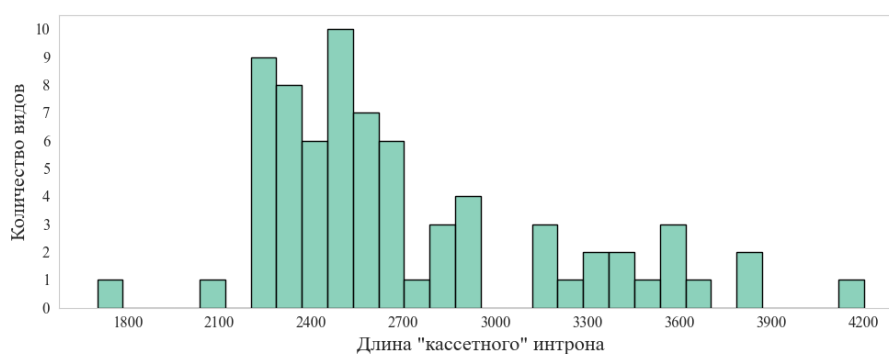
| Название таксона           | Стат. метрики        | Кол-во нуклеотидов до стоп-кодона в интроне | Длина 1-го экзона в кассете | Длина кассетного интрона | Длина 2-го экзона в кассете |
|----------------------------|----------------------|---|-----------------------------|--------------------------|-----------------------------|
| Ecdysozoa,<br>98 видов     | Min<br>Median<br>Max | 1<br>22<br>1345                             | 83<br>110<br>307            | 47<br>375<br>9821        | 37                          |
| Spiralia,<br>69 видов      | Min<br>Median<br>Max | 1<br>1<br>19                                | 56<br>221<br>242            | 222<br>1706<br>18219     | 37                          |
| Cnidaria,<br>14 видов      | Min<br>Median<br>Max | 10<br>40<br>103                             | 86<br>116<br>116            | 116<br>531<br>991        | 37                          |
| Sauropsida,<br>53 вида     | Min<br>Median<br>Max | 1<br>142<br>1519                            | 107<br>110<br>110           | 443<br>1393<br>3942      | 37                          |
| Amphibia,<br>19 видов      | Min<br>Median<br>Max | 1<br>25<br>187                              | 110                         | 2004<br>3029<br>10340    | 37                          |
| Lepidosauria,<br>28 видов  | Min<br>Median<br>Max | 1   | 110                         | 2293<br>2579<br>4667     | 37                          |
| Actinopterygii,<br>72 вида | Min<br>Median<br>Max | 1<br>22<br>283                              | 110                         | 1702<br>2549<br>4202     | 37                          |

### Подробный анализ Actinopterygii

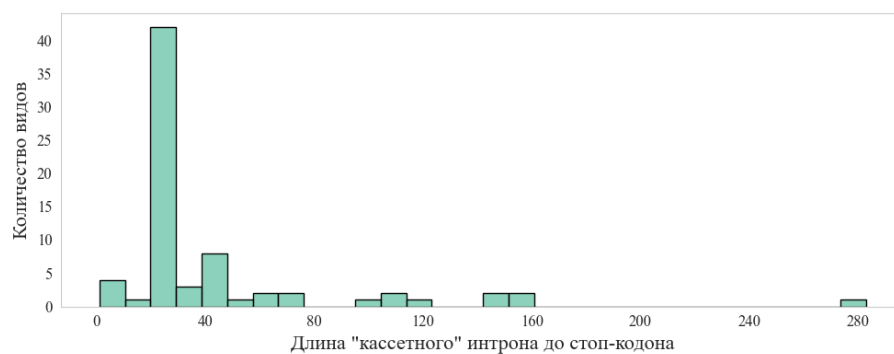
Для таксономической группы Actinopterygii проводился более углубленный анализ, так как на текущий момент данных по гену *Nxf1* для них не было. Были взяты все найденные нуклеотидные последовательности гена у представителя данной филогенетической группы – 72 вида.

В таблице 9 из приложения показана характеристика “консервативной кассеты” исследуемой группы. Результаты по другим группам можно также найти в приложении, таблицы 3–8.

На рисунках 3 и 4 показано распределение длин “кассетного” интрона и длин части “кассетного” интрона до стоп-кодона, соответственно.

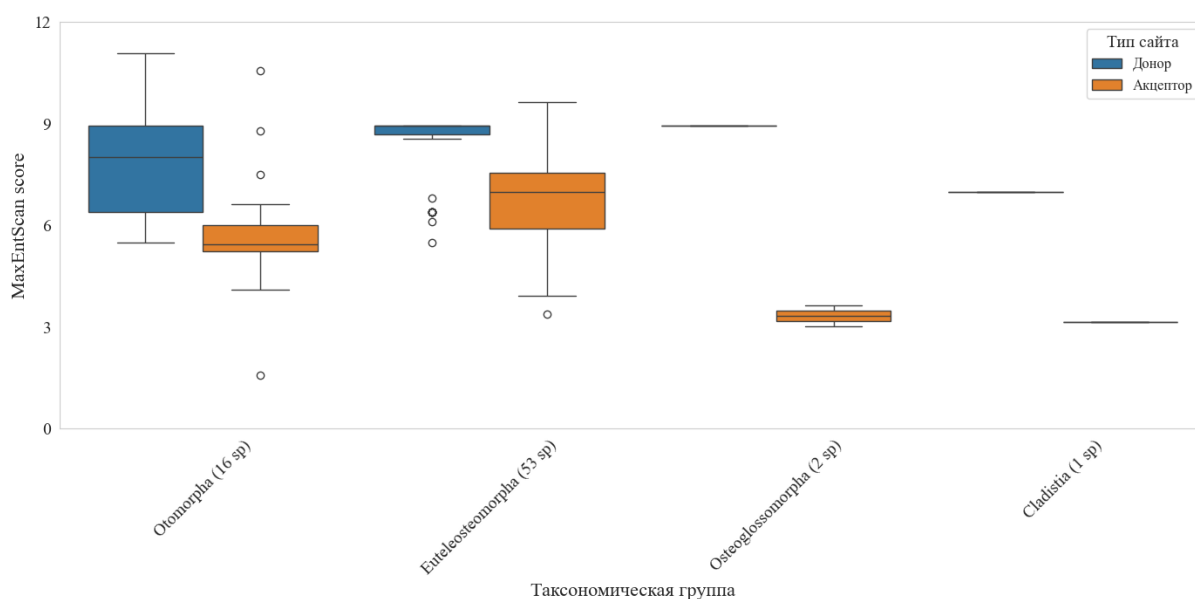


**Рис. 3: Распределение длин “кассетного” интрона у Actinopterygii**



**Рис. 4:** Распределение длин части “кассетного” интрона до стоп-кодона у *Actinopterygii*

На рисунке 5 представлены результаты оценки “силы сайтов сплайсинга” – “ящички с усами”, отображающие распределение MaxEntScan score для таксонов более низкого ранга внутри группы *Actinopterygii*. Разбиение на подгруппы основано на их удаленности друг от друга. Порядок групп на графике не несет смысловой нагрузки.



**Рис. 5:** Результаты проведения MaxEntScan для *Actinopterygii*.

Рисунок 6 демонстрирует результаты, полученные с помощью MEME Suite.

Найденные мотивы присутствуют не у всех видов, взятых в анализ изначально, их количество отображено в столбце Sites. Нас заинтересовал 2-й найденный мотив, так как его начало схоже с предложенной авторами [46] консенсусной последовательностью для СТЕ из рисунка 7. К сожалению, использование Tomtom для сравнения найденных консервативных мотивов из “кассетного” интрона с базой данных не дало статистически значимых результатов.

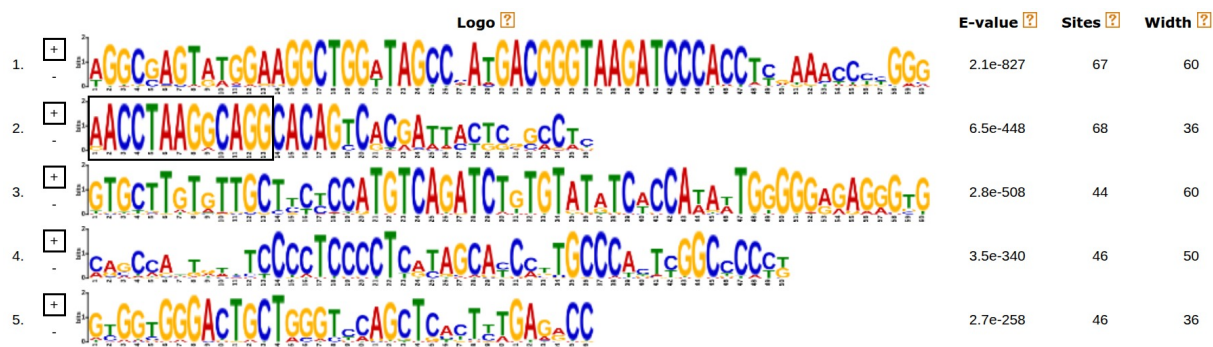


Рис. 6: Результат поиска мотивов внутри “кассетного” интрона с помощью MEME Suite для Actinopterygii. Черным прямоугольником выделен участок, похожий на консенсусную последовательность СТЕ (рис. 7) [46]

TBE    CACC-TA-GAGCTGTG  
 СТЕ { AACCTAA-GA-CAG-G consensus  
       AGCC-AATGA-C-GGG

Рис. 7: Консенсусный конститутивный транспортный элемент (СТЕ) [46].

Репрезентация вторичной структуры интрон-содержащего транскрипта с выделенным кассетным интроном и найденным мотивом показана на рисунке 8. У всех видов были сходные структуры и в качестве иллюстрации представлен один из проанализированных видов.

Учитывая тот факт, что мотив с интересующим нас участком, был найден у 68 видов, именно для них был проведен последующий анализ.

Рисунок 9 отображает результаты множественного выравнивания, а на рисунке 10 представлено филогенетическое дерево, построенное по результатам этого выравнивания.

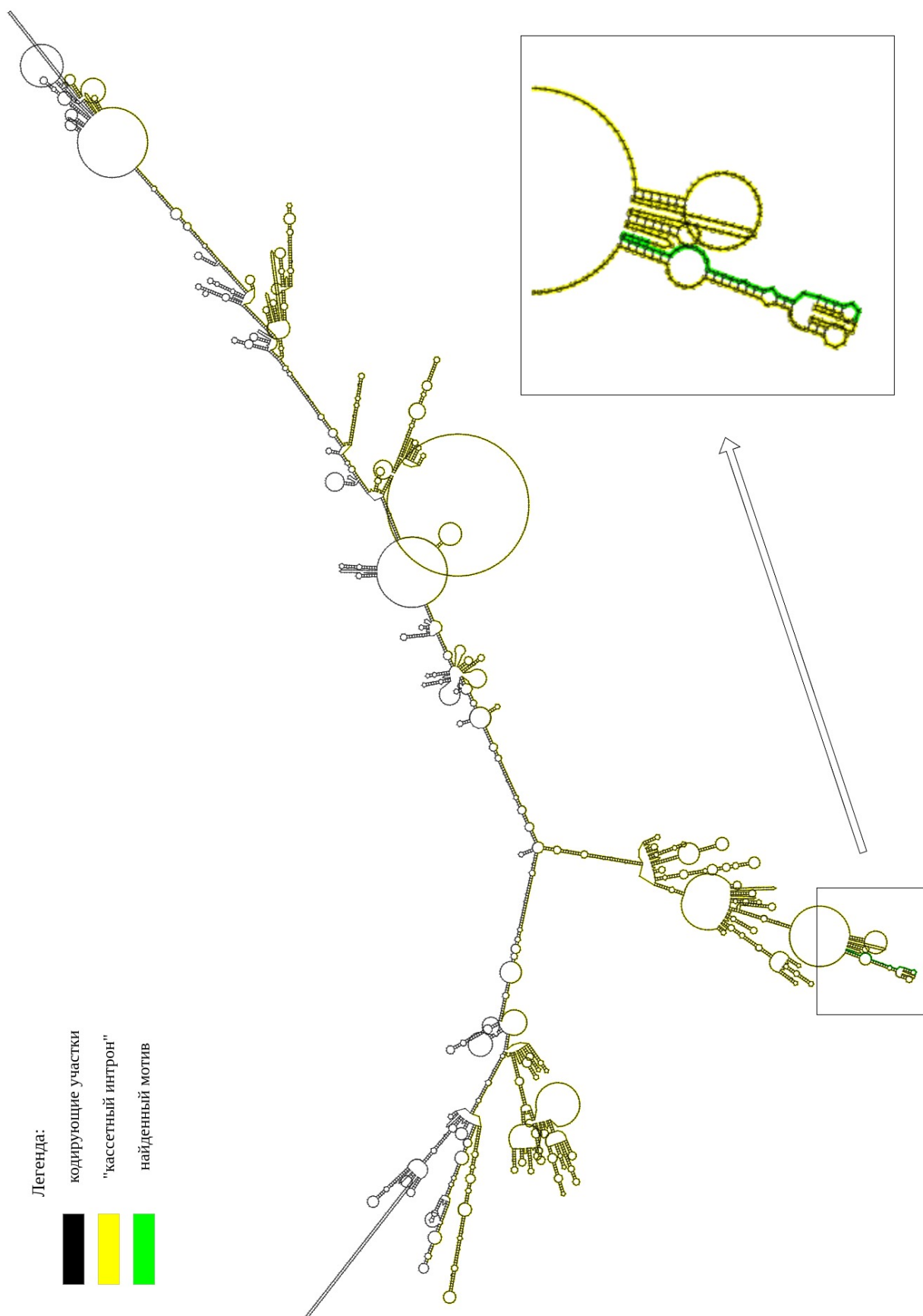


Рис. 8: Вторичная структура РНК-транскрипта для *Chanos chanos* из *Otomorpha*, содержащая "каскадный" интрон.





Рис. 9: Фрагмент множественного выравнивания для Actinopterygii. Участок, соответствующий 1-му экзону из “консервативной кассеты”.

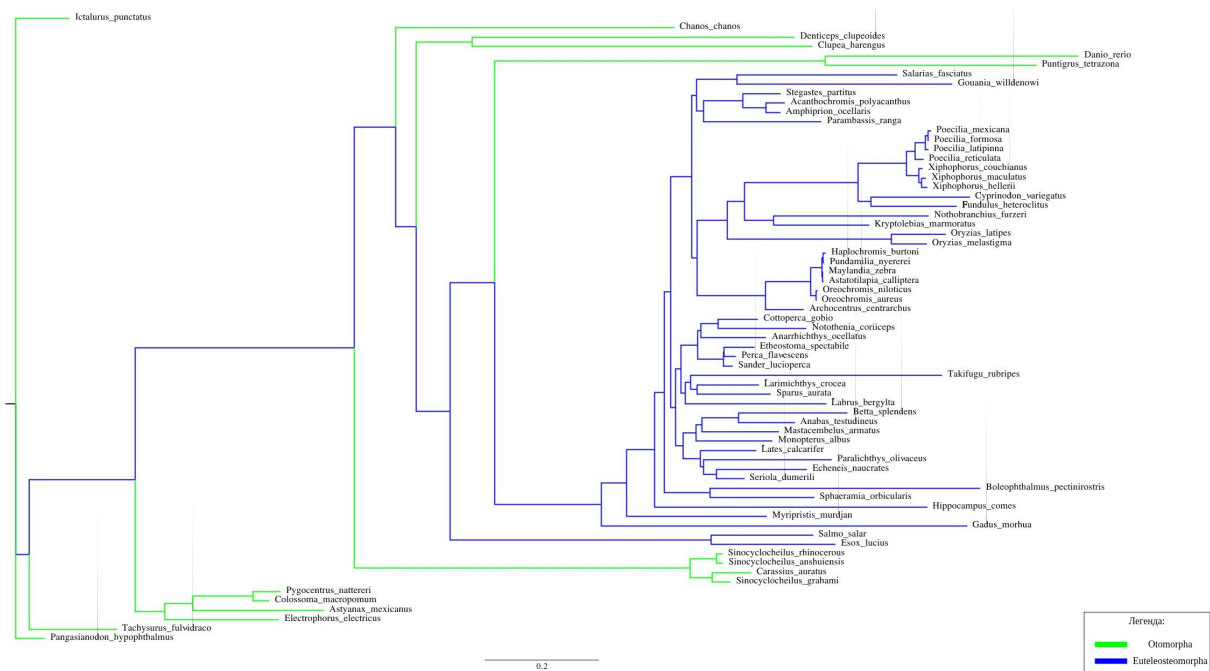
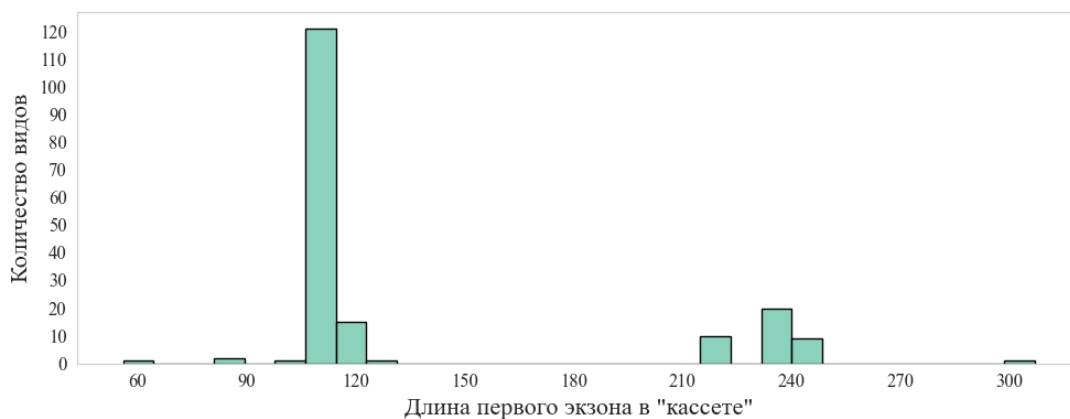


Рис. 10: Филогенетическое дерево для Actinopterygii.

## Обсуждение

### Анализ всех найденных видов

У всех проанализированных видов размер второго экзона из “консервативной кассеты” равен 37 нуклеотидам, в то время как размер первого экзона варьирует в различных группах (см. таблицу 2). На рисунке 11 показано распределение длины первого экзона из “кассеты” для Protostomia.



**Рис. 11: Распределение длины первого экзона из “консервативной кассеты” для Protostomia.**

Для Ecdysozoa и Cnidaria первый экзон как правило размером 110 нуклеотидов, но встречаются и исключения. У Spiralia размер этого экзона гораздо больше и чаще всего составляет 239 нуклеотидов. По данному отличию и встречающимся уникальным вариантам размера экзона требуется углубленное исследование.

У Deuterostomia размер первого экзона в абсолютном большинстве случаев (171 из 172 исследованных видов) составляет 110 нуклеотидов, что также характерно и для млекопитающих.

Длина участка внутри интрона до стоп-кодона, как и длина самого интрона, варьирует в более широких пределах в разных группах. Тем не менее внутри отдельных групп, например Lepidosauria (таблица 8 в приложении), наблюдается высокая степень консервативности обоих параметров.

Также встречаются виды, у которых происходит частичная или даже полная трансляция “кассетного” интрона, потому что в нем не встречается преждевременный стоп-кодон. Например, таким видом является давно изученная в данном контексте *Caenorhabditis elegans*, у которой преждевременный стоп-кодон встречается в одном из экзонов после “кассетного” интрона. В данном исследовании были найдены еще 2 вида, у которых интрон полностью считывается: Aves — *Vidua chalybeata*, Paraneoptera — *Rhopalosiphum maidis*. Упомянутые виды также требуют тщательного изучения.

## Подробный анализ *Actinopterygii*

Данная группа организмов была исследована более подробно по перечисленным ранее причинам. Внутри группы размеры первого и второго экзона из “консервативной кассеты” для всех исследованных видов составляют 110 и 37 нуклеотидов, соответственно. Длина участка внутри интрона до стоп-кодона у большинства видов составляет 22 нуклеотида (39 из 72 исследованных в работе). Размер “кассетного” интрона варьирует от 1702 до 4202 нуклеотидов (медиана – 2549).

Анализ “сайтов силы сплайсинга” (рис. 5) говорит о том, что у всех исследованных видов в работе данный интрон успешно вырезается сплайсосомой. Учитывая большую выборку видов, взятую для анализа, было принято решение ориентироваться на эмпирическую интерпретацию результатов, которая выглядит следующим образом:

- 0–3: слабый сайт сплайсинга
- 3–6: умеренный сайт сплайсинга
- >6: сильный сайт сплайсинга

Так как у большинства видов значение MaxEntScan score больше или около 6, был сделан вывод, высказанный выше. Соответственно, удержание интрона является не ошибкой сплайсинга, а закономерным событием варианта альтернативного сплайсинга. В связи с этим и было принято решение о поиске консервативных мотивов внутри “кассетного” интрона.

Несмотря на то, что на рисунке 6 представлено 5 найденных мотивов, их количество может быть больше, потому что данное значение мотивов было ограничением запуска MEME Suite. Учитывая высокую степень сходства начала 2-го найденного мотива с консенсусной последовательностью СТЕ из рисунка 7, можно предположить сохранение интрона благодаря этой и возможно другим структурам внутри интрон-содержащего транскрипта (рис. 8). Было бы интересно узнать, как именно СТЕ-подобная последовательность оказалось в данном интроне.

Проведенное множественное выравнивание на рисунке 9 говорит о высокой степени консервативности как кодирующих участков – левый и правый крайние части диаграммы под выравниванием, так и некоторых участков внутри интрона – центр диаграммы под выравниванием.

Филогенетическое древо (рис. 10), построенное по результатам выравнивания, несмотря на наличие “кассетного” интрона в последовательности, использованной для его построения, успешно разделяет виды на таксоны более высокого ранга – Otomorpha и Euteleosteomorpha.

Остальные группы, не включенные в подробный анализ, требуют его проведения.

## Выводы

1. “Консервативная кассета” имеет определенные характеристики, которые меняются в пределах разных таксономических ветвей. Вариативность характеристик больше у тех групп, которые дальше эволюционно друг от друга.
2. Внутри “кассетного” интрона существуют участки, которые способствуют формированию особой вторичной структуры интрон-содержащего транскрипта. За счет их наличия возможно сохранение такого транскрипта и последующая трансляция с синтезом укороченной формы белка.

## Список литературы

1. *Jurica M. S., Roybal G. A.* RNA splicing // Encyclopedia of Biological Chemistry: Second Edition. — Elsevier, 2013. — P. 185–190. — DOI: [10.1016/B978-0-12-378630-2.00674-5](https://doi.org/10.1016/B978-0-12-378630-2.00674-5).
2. *Krebs J. E., Goldstein E. S., Kilpatrick S. T.* Lewin's GENES XII. — 12th ed. — Burlington, MA : Jones & Bartlett Learning, 2017. — ISBN 9781284104493.
3. Intron retention enhances gene regulatory complexity in vertebrates / U. Schmitz [et al.] // Genome Biology. — 2017. — Vol. 18, no. 1. — P. 216. — DOI: [10.1186/s13059-017-1339-3](https://doi.org/10.1186/s13059-017-1339-3).
4. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis / M. Kalyna [et al.] // Nucleic Acids Research. — 2012. — Vol. 40, no. 6. — P. 2454–2469. — DOI: [10.1093/nar/gkr932](https://doi.org/10.1093/nar/gkr932).
5. Retroviral Constitutive Transport Element Evolved from Cellular TAP(NXF1)-Binding Sequences / A. S. Zolotukhin [et al.] // Journal of Virology. — 2001. — Vol. 75, no. 12. — P. 5567–5575. — DOI: [10.1128/jvi.75.12.5567-5575.2001](https://doi.org/10.1128/jvi.75.12.5567-5575.2001).
6. *Mamon L. A., Kliver S. F., Golubkova E. V.* Evolutionarily conserved features of the retained intron in alternative transcripts of the nxfl (nuclear export factor) genes in different organisms // Open Journal of Genetics. — 2013. — Vol. 3, no. 3. — P. 159–170. — DOI: [10.4236/ojgen.2013.33018](https://doi.org/10.4236/ojgen.2013.33018).
7. *Mamon L. A., Ginanova V., Golubkova E. V.* Organ-specific transcripts as a source of gene multifunctionality: Lessons learned from the *Drosophila melanogaster* sbr (Dm nxfl) gene // Biological Communications. — 2019. — Vol. 64, no. 2. — P. 146–157. — DOI: [10.21638/spbu03.2019.206](https://doi.org/10.21638/spbu03.2019.206).
8. Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40 / P. Gruss [et al.] // Proceedings of the National Academy of Sciences of the United States of America. — 1979. — Vol. 76, no. 9. — P. 4317–4321. — DOI: [10.1073/pnas.76.9.4317](https://doi.org/10.1073/pnas.76.9.4317).
9. *Juneau K., Nislow C., Davis R. W.* Introns regulate RNA and protein abundance in yeast // Genetics. — 2006. — Vol. 174, no. 1. — P. 511–518. — DOI: [10.1534/genetics.106.058560](https://doi.org/10.1534/genetics.106.058560).
10. Proto-gene and de novo gene birth / A.-R. Carvunis [et al.] // Nature. — 2012. — Vol. 487, no. 7407. — P. 370–374. — DOI: [10.1038/nature11184](https://doi.org/10.1038/nature11184).
11. *Jo B.-S., Choi S. S.* Introns: The Functional Benefits of Introns in Genomes // Genomics & Informatics. — 2015. — Vol. 13, no. 4. — P. 112. — DOI: [10.5808/gi.2015.13.4.112](https://doi.org/10.5808/gi.2015.13.4.112).

12. *Valencia P., Dias A. P., Reed R.* Splicing promotes rapid and efficient mRNA export in mammalian cells // *Proceedings of the National Academy of Sciences of the United States of America*. — 2008. — Vol. 105, no. 9. — P. 3386–3391. — DOI: [10.1073/pnas.0800250105](https://doi.org/10.1073/pnas.0800250105).
13. *Roy S. W., Gilbert W.* The evolution of spliceosomal introns: Patterns, puzzles and progress // *Nature Reviews Genetics*. — 2006. — Vol. 7, no. 3. — P. 211–221. — DOI: [10.1038/nrg1807](https://doi.org/10.1038/nrg1807).
14. *Schwartz S., Meshorer E., Ast G.* Chromatin organization marks exon-intron structure // *Nature Structural and Molecular Biology*. — 2009. — Vol. 16, no. 9. — P. 990–995. — DOI: [10.1038/nsmb.1659](https://doi.org/10.1038/nsmb.1659).
15. *Cadherin 23 Is a component of the tip link in hair-cell stereocilia / J. Siemens [et al.]* // *Nature*. — 2004. — Vol. 428, no. 6986. — P. 950–955. — DOI: [10.1038/nature02483](https://doi.org/10.1038/nature02483).
16. *Comeron J. M., Williford A., Kliman R. M.* The Hill-Robertson effect: Evolutionary consequences of weak selection and linkage in finite populations // *Heredity*. — 2008. — Vol. 100, no. 1. — P. 19–31. — DOI: [10.1038/sj.hdy.6801059](https://doi.org/10.1038/sj.hdy.6801059).
17. *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations / D. Welter [et al.]* // *Nucleic Acids Research*. — 2014. — Vol. 42, no. D1. — P. 1001–1006. — DOI: [10.1093/nar/gkt1229](https://doi.org/10.1093/nar/gkt1229).
18. *Golubkova E., Shidlovskii Y., Schedl P.* The evolutionarily conserved family of nuclear export factor (NXF) in *Drosophila melanogaster* // *Drosophila Melanogaster: Life Cycle, Genetics and Development*. — 2012. — P. 63–82. — Published December 2015.
19. *Tap: A novel cellular protein that interacts with tip of herpesvirus saimiri and induces lymphocyte aggregation / D. W. Yoon [et al.]* // *Immunity*. — 1997. — Vol. 6, no. 5. — P. 571–582. — DOI: [10.1016/S1074-7613\(00\)80345-3](https://doi.org/10.1016/S1074-7613(00)80345-3).
20. *Herold A., Klymenko T., Izaurralde E.* TAP (NXF1) Belongs to a Multigene Family of Putative RNA Export Factors with a Conserved Modular Architecture // *Molecular and Cellular Biology*. — 2000. — Vol. 20, no. 23. — P. 8996–9008. — DOI: [10.1128/mcb.20.23.8996-9008.2000](https://doi.org/10.1128/mcb.20.23.8996-9008.2000).
21. *Testis-specific products of the *Drosophila melanogaster* sbr gene, encoding nuclear export factor 1, are necessary for male fertility / V. Ginanova [et al.]* // *Gene*. — 2016. — Vol. 577, no. 2. — P. 153–160. — DOI: [10.1016/j.gene.2015.11.030](https://doi.org/10.1016/j.gene.2015.11.030).
22. *The RNA-Binding Protein SBR (Dm NXF1) Is Required for the Constitution of Medulla Boundaries in *Drosophila melanogaster* Optic Lobes / L. Mamon [et al.]* // *Cells*. — 2021. — Vol. 10, no. 5. — P. 1144. — DOI: [10.3390/cells10051144](https://doi.org/10.3390/cells10051144).



23. *Bondaruk D. D., Golubkova E. V., Mamon L. A.* Contribution of the intron retained in the Nxf1 gene transcript to the phylogeny of the order Chiroptera // *Ecological Genetics*. — 2022. — Vol. 20, no. 2. — P. 73–88. — DOI: [10.17816/ecogen90940](https://doi.org/10.17816/ecogen90940).
24. Database resources of the National Center for Biotechnology Information / E. W. Sayers, E. E. Bolton, J. R. Brister, [et al.] // *Nucleic Acids Research*. — 2022. — Vol. 50, no. D1. — P. D20–D26. — DOI: [10.1093/nar/gkab1112](https://doi.org/10.1093/nar/gkab1112).
25. *McKinney W.* Data Structures for Statistical Computing in Python. — 2010.
26. *Python Software Foundation.* Python, Version 3.12. — 2023. — <https://www.python.org/downloads/release/python-3120/>.
27. Jupyter Notebooks – a publishing format for reproducible computational workflows / T. Kluyver [et al.]. — 2016. — DOI: [10.3233/978-1-61499-649-1-87](https://doi.org/10.3233/978-1-61499-649-1-87).
28. Biopython: Freely available Python tools for computational molecular biology and bioinformatics / P. J. A. Cock [et al.] // *Bioinformatics*. — 2009. — Vol. 25, no. 11. — P. 1422–1423. — DOI: [10.1093/bioinformatics/btp163](https://doi.org/10.1093/bioinformatics/btp163).
29. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets / N. A. O’Leary [et al.] // *Scientific Data*. — 2024. — Vol. 11, no. 1. — P. 732. — DOI: [10.1038/s41597-024-03571-y](https://doi.org/10.1038/s41597-024-03571-y).
30. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs / S. F. Altschul [et al.] // *Nucleic Acids Research*. — 1997. — Vol. 25, no. 17. — P. 3389–3402. — DOI: [10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389).
31. *Katoh K., Standley D. M.* MAFFT multiple sequence alignment software version 7: improvements in performance and usability // *Molecular Biology and Evolution*. — 2013. — Vol. 30, no. 4. — P. 772–780. — DOI: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
32. Unipro UGENE: a unified bioinformatics toolkit / K. Okonechnikov [et al.] // *Bioinformatics*. — 2012. — Vol. 28, no. 8. — P. 1166–1167. — DOI: [10.1093/bioinformatics/bts091](https://doi.org/10.1093/bioinformatics/bts091).
33. The MEME Suite / T. L. Bailey [et al.] // *Nucleic Acids Research*. — 2015. — Vol. 43, W1. — W39–W49. — DOI: [10.1093/nar/gkv416](https://doi.org/10.1093/nar/gkv416).
34. Quantifying similarity between motifs / S. Gupta [et al.] // *Genome Biology*. — 2007. — Vol. 8, no. 2. — R24. — DOI: [10.1186/gb-2007-8-2-r24](https://doi.org/10.1186/gb-2007-8-2-r24).
35. ViennaRNA Package 2.0 / R. Lorenz [et al.] // *Algorithms for Molecular Biology*. — 2011. — Vol. 6, no. 1. — P. 26. — DOI: [10.1186/1748-7188-6-26](https://doi.org/10.1186/1748-7188-6-26).
36. *Yeo G., Burge C. B.* Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals // *Bioinformatics*. — 2004. — Vol. 20, no. 3. — P. 327–335. — DOI: [10.1093/bioinformatics/btg005](https://doi.org/10.1093/bioinformatics/btg005).

37. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era / B. Q. Minh [et al.] // *Molecular Biology and Evolution*. — 2020. — Vol. 37, no. 5. — P. 1530–1534. — DOI: [10.1093/molbev/msaa015](https://doi.org/10.1093/molbev/msaa015).
38. *Rambaut A.* FigTree v1.4.4. — 2018. — Institute of Evolutionary Biology, University of Edinburgh. <http://tree.bio.ed.ac.uk/software/figtree/>.
39. *QuantStack, contributors mamba.* Mamba: The Fast Cross-Platform Package Manager. — 2024. — <https://github.com/mamba-org/mamba>.
40. *GitHub, Inc.* GitHub. — 2008. — URL: <https://github.com>.
41. *Lamport L.* LaTeX: A Document Preparation System. — 2nd ed. — Reading, Massachusetts : Addison-Wesley, 1994.
42. *TablesGenerator.com.* Tables Generator – LaTeX Tables Editor. — 2025. — URL: <https://www.tablesgenerator.com>.
43. *diagrams.net.* draw.io - Online Diagram Software. — 2025. — URL: <https://www.diagrams.net/>.
44. *Hunter J. D.* Matplotlib: A 2D Graphics Environment // *Computing in Science & Engineering*. — 2007. — Vol. 9, no. 3. — P. 90–95. — DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
45. *Canonical Ltd.* Ubuntu 22.04 LTS (Jammy Jellyfish). — 2022. — <https://releases.ubuntu.com/22.04/>.
46. Replication of Human Herpesvirus 6A and 6B Is Associated with Distinct Nuclear Domains / F. Tajima [et al.] // *Journal of Virology*. — 2001. — Vol. 75, no. 12. — P. 5567–5575. — DOI: [10.1128/JVI.75.12.5567-5575.2001](https://doi.org/10.1128/JVI.75.12.5567-5575.2001).



## Приложение

**Таблица 3: Сводная таблица с характеристикой кассетного интрона для Ecdysozoa. Сортировка по возрастанию количества нуклеотидов до стоп-кодона в кассетной интроне.**

| Название организма                | Кол-во нуклеотидов до стоп-кодона в интроне | Длина 1-го экзона в кассете | Длина кассетного интрона | Длина 2-го экзона в кассете |
|-----------------------------------|---|-----------------------------|--------------------------|-----------------------------|
| <i>Trichinella spiralis</i>       | 1   | 83                          | 417                      | 37                          |
| <i>Priapulus caudatus</i>         | 1   | 110                         | 2114                     | 37                          |
| <i>Galendromus occidentalis</i>   | 1   | 110                         | 1491                     | 37                          |
| <i>Ixodes scapularis</i>          | 1   | 110                         | 3567                     | 37                          |
| <i>Limulus polyphemus</i>         | 1   | 110                         | 915                      | 37                          |
| <i>Parasteatoda tepidariorum</i>  | 1   | 110                         | 1725                     | 37                          |
| <i>Cryptotermes secundus</i>      | 1   | 110                         | 4335                     | 37                          |
| <i>Maniola hyperantus</i>         | 1   | 110                         | 920                      | 37                          |
| <i>Cimex lectularius</i>          | 1   | 110                         | 4437                     | 37                          |
| <i>Vespa mandarinia</i>           | 1   | 113                         | 379                      | 37                          |
| <i>Zerene cesonia</i>             | 1   | 110                         | 1162                     | 37                          |
| <i>Pararge aegeria</i>            | 1   | 110                         | 2657                     | 37                          |
| <i>Myzus persicae</i>             | 1   | 107                         | 772                      | 37                          |
| <i>Halyomorpha halys</i>          | 1   | 110                         | 7270                     | 37                          |
| <i>Diuraphis noxia</i>            | 1   | 107                         | 742                      | 37                          |
| <i>Sipha flava</i>                | 1   | 107                         | 58                       | 37                          |
| <i>Manduca sexta</i>              | 1   | 110                         | 1796                     | 37                          |
| <i>Apis laboriosa</i>             | 1   | 113                         | 1254                     | 37                          |
| <i>Orussus abietinus</i>          | 1   | 113                         | 74                       | 37                          |
| <i>Danarus plexippus</i>          | 1   | 110                         | 1009                     | 37                          |
| <i>Colletes gigas</i>             | 1   | 113                         | 379                      | 37                          |
| <i>Ostrinia furnacalis</i>        | 1   | 110                         | 1946                     | 37                          |
| <i>Vespa crabro</i>               | 1   | 113                         | 381                      | 37                          |
| <i>Venturia canescens</i>         | 1   | 113                         | 621                      | 37                          |
| <i>Papilio polytes</i>            | 1   | 110                         | 1674                     | 37                          |
| <i>Vespa velutina</i>             | 1   | 113                         | 377                      | 37                          |
| <i>Cephus cinctus</i>             | 1   | 113                         | 75                       | 37                          |
| <i>Bombus pyrosoma</i>            | 1   | 113                         | 244                      | 37                          |
| <i>Papilio xuthus</i>             | 1   | 110                         | 999                      | 37                          |
| <i>Vanessa tameamea</i>           | 1   | 110                         | 2352                     | 37                          |
| <i>Megalopta genalis</i>          | 1   | 113                         | 373                      | 37                          |
| <i>Vespula pennsylvanica</i>      | 1   | 113                         | 363                      | 37                          |
| <i>Leptopilina heterotoma</i>     | 1   | 113                         | 921                      | 37                          |
| <i>Acromyrmex echinator</i>       | 1   | 113                         | 438                      | 37                          |
| <i>Aphidius gifuensis</i>         | 1   | 113                         | 240                      | 37                          |
| <i>Polistes fuscatus</i>          | 1   | 113                         | 400                      | 37                          |
| <i>Dirofilaria immitis</i>        | 7   | 98                          | 248                      | 37                          |
| <i>Odontomachus brunneus</i>      | 10  | 113                         | 498                      | 37                          |
| <i>Diploscapter pachys</i>        | 10  | 110                         | 662                      | 37                          |
| <i>Bactrocera dorsalis</i>        | 13  | 110                         | 1808                     | 37                          |
| <i>Drosophila melanogaster</i>    | 13  | 110                         | 1602                     | 37                          |
| <i>Ceratitis capitata</i>         | 19  | 110                         | 2023                     | 37                          |
| <i>Pediculus humanus corporis</i> | 19  | 110                         | 631                      | 37                          |
| <i>Aphelenchoides avenae</i>      | 19  | 110                         | 441                      | 37                          |
| <i>Litomosoides sigmodontis</i>   | 19  | 110                         | 242                      | 37                          |
| <i>Acanthocheilonema viteae</i>   | 19  | 110                         | 225                      | 37                          |
| <i>Aethina tumida</i>             | 19  | 110                         | 1729                     | 37                          |

|   |      |     |      |    |
|---|------|-----|------|----|
| <i>Lepeophtheirus salmonis</i>          | 22   | 110 | 1555 | 37 |
| <i>Anoplophora glabripennis</i>         | 22   | 110 | 3664 | 37 |
| <i>Varroa jacobsoni</i>                 | 22   | 110 | 3077 | 37 |
| <i>Varroa destructor</i>                | 22   | 110 | 3077 | 37 |
| <i>Thelazia callipaeda</i>              | 25   | 110 | 209  | 37 |
| <i>Bursaphelenchus xylophilus</i>       | 25   | 110 | 638  | 37 |
| <i>Acyrtosiphon pisum</i>               | 28   | 107 | 68   | 37 |
| <i>Anisakis simplex</i>                 | 30   | 219 | 665  | 37 |
| <i>Tetranychus urticae</i>              | 31   | 122 | 648  | 37 |
| <i>Homarus americanus</i>               | 31   | 110 | 9821 | 37 |
| <i>Bursaphelenchus okinawaensis</i>     | 37   | 110 | 593  | 37 |
| <i>Globodera pallida</i>                | 43   | 113 | 47   | 37 |
| <i>Amphibalanus amphitrite</i>          | 73   | 110 | 369  | 37 |
| <i>Cotesia glomerata</i>                | 73   | 116 | 236  | 37 |
| <i>Caenorhabditis angaria</i>           | 79   | 110 | 96   | 37 |
| <i>Onchocerca ochengi</i>               | 88   | 110 | 243  | 37 |
| <i>Brugia pahangi</i>                   | 91   | 110 | 232  | 37 |
| <i>Ditylenchus destructor</i>           | 97   | 307 | 1167 | 37 |
| <i>Mesorhabditis belari</i>             | 97   | 110 | 147  | 37 |
| <i>Melanaphis sacchari</i>              | 97   | 107 | 71   | 37 |
| <i>Enterobius vermicularis</i>          | 100  | 110 | 195  | 37 |
| <i>Pristionchus mayeri</i>              | 103  | 110 | 131  | 37 |
| <i>Cercopithifilaria johnstoni</i>      | 103  | 110 | 238  | 37 |
| <i>Steinernema carpocapsae</i>          | 106  | 110 | 131  | 37 |
| <i>Wuchereria bancrofti</i>             | 106  | 125 | 242  | 37 |
| <i>Parelaphostrongylus tenuis</i>       | 112  | 110 | 228  | 37 |
| <i>Toxocara canis</i>                   | 115  | 110 | 1062 | 37 |
| <i>Necator americanus</i>               | 136  | 110 | 243  | 37 |
| <i>Brugia malayi</i>                    | 139  | 110 | 243  | 37 |
| <i>Caenorhabditis auriculariae</i>      | 145  | 110 | 156  | 37 |
| <i>Auanema sp. JU1783</i>               | 145  | 110 | 80   | 37 |
| <i>Pristionchus entomophagus</i>        | 151  | 110 | 154  | 37 |
| <i>Steinernema hermaphroditum</i>       | 157  | 110 | 131  | 37 |
| <i>Caenorhabditis brenneri</i>          | 175  | 110 | 130  | 37 |
| <i>Angiostrongylus cantonensis</i>      | 181  | 110 | 213  | 37 |
| <i>Dictyocaulus viviparus</i>           | 190  | 110 | 832  | 37 |
| <i>Caenorhabditis elegans</i>           | 193  | 110 | 106  | 37 |
| <i>Cooperia oncophora</i>               | 205  | 110 | 215  | 37 |
| <i>Caenorhabditis sp. 36 PRJEB53466</i> | 205  | 110 | 133  | 37 |
| <i>Caenorhabditis nigoni</i>            | 214  | 110 | 142  | 37 |
| <i>Pristionchus pacificus</i>           | 214  | 110 | 251  | 37 |
| <i>Trichostrongylus colubriformis</i>   | 214  | 110 | 224  | 37 |
| <i>Caenorhabditis briggsae</i>          | 217  | 110 | 145  | 37 |
| <i>Cylicocyclus nassatus</i>            | 229  | 110 | 239  | 37 |
| <i>Haemonchus contortus</i>             | 304  | 110 | 220  | 37 |
| <i>Caenorhabditis bovis</i>             | 316  | 110 | 235  | 37 |
| <i>Nippostrongylus brasiliensis</i>     | 316  | 110 | 235  | 37 |
| <i>Dracunculus medinensis</i>           | 334  | 110 | 122  | 37 |
| <i>Mesorhabditis spiculigera</i>        | 376  | 110 | 173  | 37 |
| <i>Pollicipes pollicipes</i>            | 436  | 110 | 367  | 37 |
| <i>Rhopalosiphum maidis</i>             | 1345 | 107 | 69   | 37 |

Таблица 4: Сводная таблица с характеристикой кассетного интрона для *Spiralia*. Сортировка по возрастанию количества нуклеотидов до стоп-кодона в “кассетном” интроне.

| Название организма               | Кол-во нуклеотидов до стоп-кодона в интроне | Длина 1-го экзона в кассете | Длина кассетного интрона | Длина 2-го экзона в кассете |
|----------------------------------|---|-----------------------------|--------------------------|-----------------------------|
| <i>Schistosoma haematobium</i>   | 1   | 239                         | 652                      | 37                          |
| <i>Magallana gigas</i>           | 1   | 110                         | 1537                     | 37                          |
| <i>Mya arenaria</i>              | 1   | 110                         | 1727                     | 37                          |
| <i>Crassostrea virginica</i>     | 1   | 110                         | 1613                     | 37                          |
| <i>Aplysia californica</i>       | 1   | 221                         | 4146                     | 37                          |
| <i>Gigantopelta aegis</i>        | 1   | 110                         | 1869                     | 37                          |
| <i>Mercenaria mercenaria</i>     | 1   | 110                         | 1690                     | 37                          |
| <i>Dreissena polymorpha</i>      | 1   | 110                         | 2207                     | 37                          |
| <i>Ruditapes philippinarum</i>   | 1   | 110                         | 1646                     | 37                          |
| <i>Mactra antiquata</i>          | 1   | 110                         | 2319                     | 37                          |
| <i>Mytilus coruscus</i>          | 1   | 110                         | 1234                     | 37                          |
| <i>Potamilus streckersoni</i>    | 1   | 110                         | 4567                     | 37                          |
| <i>Saccostrea echinata</i>       | 1   | 110                         | 1556                     | 37                          |
| <i>Mytilus edulis</i>            | 1   | 110                         | 1360                     | 37                          |
| <i>Mytilus trossulus</i>         | 1   | 110                         | 1357                     | 37                          |
| <i>Pecten maximus</i>            | 1   | 110                         | 5000                     | 37                          |
| <i>Ostrea edulis</i>             | 1   | 110                         | 1643                     | 37                          |
| <i>Mizuhopecten yessoensis</i>   | 1   | 110                         | 4836                     | 37                          |
| <i>Saccostrea cucullata</i>      | 1   | 110                         | 1706                     | 37                          |
| <i>Ylistrum balloti</i>          | 1   | 110                         | 4649                     | 37                          |
| <i>Argopecten irradians</i>      | 1   | 110                         | 5057                     | 37                          |
| <i>Magallana angulata</i>        | 1   | 110                         | 1534                     | 37                          |
| <i>Mytilus californianus</i>     | 1   | 110                         | 1248                     | 37                          |
| <i>Pinctada imbricata</i>        | 1   | 110                         | 4144                     | 37                          |
| <i>Haliotis asinina</i>          | 1   | 110                         | 2375                     | 37                          |
| <i>Sinanodonta woodiana</i>      | 1   | 110                         | 4580                     | 37                          |
| <i>Haliotis cracherodii</i>      | 1   | 110                         | 2506                     | 37                          |
| <i>Haliotis rufescens</i>        | 1   | 110                         | 2505                     | 37                          |
| <i>Patella caerulea</i>          | 1   | 110                         | 1362                     | 37                          |
| <i>Patella vulgata</i>           | 1   | 110                         | 1384                     | 37                          |
| <i>Lymnaea stagnalis</i>         | 1   | 221                         | 2705                     | 37                          |
| <i>Batillaria attramentaria</i>  | 1   | 110                         | 8614                     | 37                          |
| <i>Schistosoma turkestanicum</i> | 1   | 239                         | 905                      | 37                          |
| <i>Paragonimus westermani</i>    | 1   | 239                         | 13971                    | 37                          |
| <i>Pomacea canaliculata</i>      | 1   | 56                          | 255                      | 37                          |
| <i>Bradybaena similaris</i>      | 1   | 221                         | 3811                     | 37                          |
| <i>Elysia crispata</i>           | 1   | 221                         | 8063                     | 37                          |
| <i>Elysia chlorotica</i>         | 1   | 221                         | 7182                     | 37                          |
| <i>Bulinus truncatus</i>         | 1   | 221                         | 1873                     | 37                          |
| <i>Biomphalaria pfeifferi</i>    | 1   | 221                         | 1885                     | 37                          |
| <i>Biomphalaria glabrata</i>     | 1   | 221                         | 1889                     | 37                          |
| <i>Schistosoma guineensis</i>    | 1   | 239                         | 652                      | 37                          |
| <i>Schistosoma curassoni</i>     | 1   | 239                         | 652                      | 37                          |
| <i>Schistosoma bovis</i>         | 1   | 239                         | 652                      | 37                          |
| <i>Schistosoma margrebowiei</i>  | 1   | 239                         | 650                      | 37                          |
| <i>Schistosoma intercalatum</i>  | 1   | 239                         | 652                      | 37                          |
| <i>Schistosoma rodhaini</i>      | 1   | 239                         | 671                      | 37                          |
| <i>Schistosoma japonicum</i>     | 1   | 239                         | 847                      | 37                          |
| <i>Clonorchis sinensis</i>       | 1   | 242                         | 6006                     | 37                          |
| <i>Hydatigera taeniaeformis</i>  | 1   | 242                         | 375                      | 37                          |

|                                  |    |     |       |    |
|----------------------------------|----|-----|-------|----|
| <i>Taenia crassiceps</i>         | 1  | 242 | 278   | 37 |
| <i>Taenia asiatica</i>           | 1  | 242 | 480   | 37 |
| <i>Heterobilharzia americana</i> | 1  | 239 | 2163  | 37 |
| <i>Trichobilharzia szidati</i>   | 1  | 239 | 1336  | 37 |
| <i>Trichobilharzia regenti</i>   | 1  | 239 | 996   | 37 |
| <i>Opisthorchis felineus</i>     | 1  | 242 | 14603 | 37 |
| <i>Rodentolepis nana</i>         | 1  | 242 | 222   | 37 |
| <i>Calicophoron daubneyi</i>     | 1  | 239 | 4214  | 37 |
| <i>Taenia solium</i>             | 1  | 242 | 480   | 37 |
| <i>Echinococcus granulosus</i>   | 1  | 242 | 521   | 37 |
| <i>Fasciola hepatica</i>         | 1  | 239 | 2631  | 37 |
| <i>Fasciola gigantica</i>        | 1  | 239 | 2581  | 37 |
| <i>Schistosoma mattheei</i>      | 1  | 239 | 649   | 37 |
| <i>Fasciolopsis buskii</i>       | 1  | 239 | 1303  | 37 |
| <i>Dicrocoelium dendriticum</i>  | 1  | 239 | 2612  | 37 |
| <i>Paragonimus heterotremus</i>  | 1  | 239 | 18219 | 37 |
| <i>Hymenolepis diminuta</i>      | 1  | 242 | 224   | 37 |
| <i>Solemya velum</i>             | 4  | 110 | 2071  | 37 |
| <i>Littorina saxatilis</i>       | 19 | 218 | 6746  | 37 |

**Таблица 5: Сводная таблица с характеристикой кассетного интрона для Cnidaria. Сортировка по возрастанию количества нуклеотидов до стоп-кодона в “кассетном” интроне.**

| Название организма             | Кол-во нуклеотидов до стоп-кодона в интроне | Длина 1-го экзона в кассете | Длина кассетного интрона | Длина 2-го экзона в кассете |
|--------------------------------|---|-----------------------------|--------------------------|-----------------------------|
| <i>Actinia tenebrosa</i>       | 10  | 116                         | 173                      | 37                          |
| <i>Dendronephthya gigantea</i> | 10  | 116                         | 328                      | 37                          |
| <i>Nematostella vectensis</i>  | 25  | 116                         | 991                      | 37                          |
| <i>Montipora foliosa</i>       | 31  | 116                         | 907                      | 37                          |
| <i>Pocillopora verrucosa</i>   | 34  | 116                         | 390                      | 37                          |
| <i>Acropora digitifera</i>     | 40  | 116                         | 670                      | 37                          |
| <i>Acropora millepora</i>      | 40  | 116                         | 682                      | 37                          |
| <i>Acropora muricata</i>       | 40  | 116                         | 679                      | 37                          |
| <i>Pocillopora damicornis</i>  | 46  | 116                         | 392                      | 37                          |
| <i>Pocillopora meandrina</i>   | 46  | 116                         | 392                      | 37                          |
| <i>Porites lutea</i>           | 61  | 116                         | 711                      | 37                          |
| <i>Porites evermanni</i>       | 61  | 116                         | 711                      | 37                          |
| <i>Exaiptasia diaphana</i>     | 76  | 86                          | 227                      | 37                          |
| <i>Xenia sp. Carnegie-2017</i> | 103   | 116                         | 116                      | 37                          |

**Таблица 6: Сводная таблица с характеристикой кассетного интрона для Sauropsida. Сортировка по возрастанию количества нуклеотидов до стоп-кодона в “кассетном” интроне.**

| Название организма         | Кол-во нуклеотидов до стоп-кодона в интроне | Длина 1-го экзона в кассете | Длина кассетного интрона | Длина 2-го экзона в кассете |
|----------------------------|---|-----------------------------|--------------------------|-----------------------------|
| <i>Molothrus aeneus</i>    | 1   | 110                         | 745                      | 37                          |
| <i>Taeniopygia guttata</i> | 1   | 110                         | 443                      | 37                          |
| <i>Lonchura striata</i>    | 1   | 110                         | 629                      | 37                          |
| <i>Gallus gallus</i>       | 7   | 110                         | 1616                     | 37                          |

|                                     |      |     |      |    |
|-------------------------------------|------|-----|------|----|
| <i>Cygnus atratus</i>               | 25   | 110 | 1257 | 37 |
| <i>Haliaeetus leucocephalus</i>     | 25   | 110 | 1375 | 37 |
| <i>Phalacrocorax carbo</i>          | 25   | 110 | 1345 | 37 |
| <i>Grus americana</i>               | 25   | 110 | 1659 | 37 |
| <i>Haliaeetus albicilla</i>         | 25   | 110 | 1378 | 37 |
| <i>Oxyura jamaicensis</i>           | 25   | 110 | 1246 | 37 |
| <i>Anser cygnoides</i>              | 25   | 110 | 1279 | 37 |
| <i>Ciconia boyciana</i>             | 25   | 107 | 1459 | 37 |
| <i>Anas acuta</i>                   | 25   | 110 | 1346 | 37 |
| <i>Astur gentilis</i>               | 25   | 110 | 1393 | 37 |
| <i>Aquila chrysaetos chrysaetos</i> | 25   | 110 | 1375 | 37 |
| <i>Aythya fuligula</i>              | 25   | 110 | 1227 | 37 |
| <i>Struthio camelus</i>             | 64   | 110 | 1405 | 37 |
| <i>Chelonia mydas</i>               | 79   | 110 | 1674 | 37 |
| <i>Dermochelys coriacea</i>         | 79   | 110 | 1661 | 37 |
| <i>Caretta caretta</i>              | 79   | 110 | 1656 | 37 |
| <i>Ammospiza caudacuta</i>          | 82   | 110 | 3942 | 37 |
| <i>Aphelocoma coerulescens</i>      | 85   | 110 | 3626 | 37 |
| <i>Gopherus flavomarginatus</i>     | 142  | 110 | 1655 | 37 |
| <i>Chelonoidis abingdonii</i>       | 142  | 110 | 1645 | 37 |
| <i>Malaclemys terrapin pileata</i>  | 142  | 110 | 1652 | 37 |
| <i>Mauremys mutica</i>              | 142  | 110 | 1662 | 37 |
| <i>Mauremys reevesii</i>            | 142  | 110 | 1661 | 37 |
| <i>Trachemys scripta elegans</i>    | 142  | 110 | 1661 | 37 |
| <i>Chrysemys picta bellii</i>       | 142  | 110 | 1662 | 37 |
| <i>Emys orbicularis</i>             | 142  | 110 | 1650 | 37 |
| <i>Alligator sinensis</i>           | 148  | 110 | 1497 | 37 |
| <i>Alligator mississippiensis</i>   | 148  | 110 | 1618 | 37 |
| <i>Caloenas nicobarica</i>          | 184  | 110 | 1245 | 37 |
| <i>Rissa tridactyla</i>             | 205  | 110 | 1388 | 37 |
| <i>Terrapene triunguis</i>          | 211  | 110 | 1662 | 37 |
| <i>Emydura macquarii macquarii</i>  | 223  | 110 | 1647 | 37 |
| <i>Catharus ustulatus</i>           | 241  | 110 | 3252 | 37 |
| <i>Gopherus evgoodei</i>            | 301  | 110 | 1639 | 37 |
| <i>Strigops habroptila</i>          | 457  | 110 | 1317 | 37 |
| <i>Neopsephotus bourkii</i>         | 502  | 110 | 1245 | 37 |
| <i>Melopsittacus undulatus</i>      | 517  | 110 | 1257 | 37 |
| <i>Apteryx rowi</i>                 | 541  | 110 | 1359 | 37 |
| <i>Apteryx mantelli</i>             | 541  | 110 | 1359 | 37 |
| <i>Dromaius novaehollandiae</i>     | 553  | 110 | 1365 | 37 |
| <i>Chroicocephalus ridibundus</i>   | 562  | 110 | 1373 | 37 |
| <i>Pezoporus wallicus</i>           | 568  | 110 | 1328 | 37 |
| <i>Pezoporus flaviventris</i>       | 568  | 110 | 1328 | 37 |
| <i>Rhea pennata</i>                 | 568  | 110 | 1348 | 37 |
| <i>Pezoporus occidentalis</i>       | 568  | 110 | 1319 | 37 |
| <i>Pelodiscus sinensis</i>          | 640  | 110 | 1643 | 37 |
| <i>Phaenicophaeus curvirostris</i>  | 892  | 110 | 2155 | 37 |
| <i>Camarhynchus parvulus</i>        | 1360 | 110 | 2456 | 37 |
| <i>Vidua chalybeata</i>             | 1519 | 110 | 678  | 37 |

**Таблица 7: Сводная таблица с характеристикой кассетного интрона для Amphibia. Сортировка по возрастанию количества нуклеотидов до стоп-кодона в “кассетном” интроне.**

| Название организма      | Кол-во нуклеотидов до стоп-кодона в интроне | Длина 1-го экзона в кассете | Длина кассетного интрона | Длина 2-го экзона в кассете |
|-------------------------|---|-----------------------------|--------------------------|-----------------------------|
| Ambystoma mexicanum     | 1   | 110                         | 10340                    | 37                          |
| Pelobates fuscus        | 1   | 110                         | 2424                     | 37                          |
| Bufo bufo               | 7   | 110                         | 3002                     | 37                          |
| Bufo gargarizans        | 7   | 110                         | 2879                     | 37                          |
| Hyperolius riggenbachi  | 10  | 110                         | 3902                     | 37                          |
| Rana temporaria         | 10  | 110                         | 3036                     | 37                          |
| Pseudophryne corroboree | 19  | 110                         | 3561                     | 37                          |
| Spea bombifrons         | 25  | 110                         | 2840                     | 37                          |
| Engystomops pustulosus  | 25  | 110                         | 2004                     | 37                          |
| Nanorana parkeri        | 25  | 110                         | 3038                     | 37                          |
| Hyla sarda              | 25  | 110                         | 3029                     | 37                          |
| Pyxicephalus adspersus  | 25  | 110                         | 2917                     | 37                          |
| Ranitomeya imitator     | 37  | 110                         | 2650                     | 37                          |
| Xenopus tropicalis      | 46  | 110                         | 2596                     | 37                          |
| Xenopus laevis          | 52  | 110                         | 3791                     | 37                          |
| Geotrypetes seraphini   | 55  | 110                         | 3065                     | 37                          |
| Rhinatrema bivittatum   | 103   | 110                         | 4053                     | 37                          |
| Pleurodeles waltl       | 151   | 110                         | 3245                     | 37                          |
| Microcaecilia unicolor  | 187   | 110                         | 2784                     | 37                          |

**Таблица 8: Сводная таблица с характеристикой кассетного интрона для Lepidosauria. Сортировка по возрастанию количества нуклеотидов до стоп-кодона в “кассетном” интроне.**

| Название организма                   | Кол-во нуклеотидов до стоп-кодона в интроне | Длина 1-го экзона в кассете | Длина кассетного интрона | Длина 2-го экзона в кассете |
|--------------------------------------|---|-----------------------------|--------------------------|-----------------------------|
| <i>Python bivittatus</i>             | 1   | 110                         | 2374                     | 37                          |
| <i>Notechis scutatus</i>             | 1   | 110                         | 2507                     | 37                          |
| <i>Pseudonaja textilis</i>           | 1   | 110                         | 2519                     | 37                          |
| <i>Anolis sagrei</i>                 | 1   | 110                         | 4667                     | 37                          |
| <i>Pituophis catenifer annectens</i> | 1   | 110                         | 2420                     | 37                          |
| <i>Lacerta agilis</i>                | 1   | 110                         | 2499                     | 37                          |
| <i>Candoia aspera</i>                | 1   | 110                         | 2293                     | 37                          |
| <i>Sphaerodactylus townsendi</i>     | 1   | 110                         | 2825                     | 37                          |
| <i>Thamnophis elegans</i>            | 1   | 110                         | 2426                     | 37                          |
| <i>Ahaetulla prasina</i>             | 1   | 110                         | 2432                     | 37                          |
| <i>Gekko japonicus</i>               | 1   | 110                         | 2924                     | 37                          |
| <i>Crotalus tigris</i>               | 1   | 110                         | 3091                     | 37                          |
| <i>Pogona vitticeps</i>              | 1   | 110                         | 2746                     | 37                          |
| <i>Podarcis raffonei</i>             | 1   | 110                         | 2495                     | 37                          |
| <i>Protobothrops mucrosquamatus</i>  | 1   | 110                         | 3264                     | 37                          |
| <i>Varanus komodoensis</i>           | 1   | 110                         | 2658                     | 37                          |
| <i>Pantherophis guttatus</i>         | 1   | 110                         | 2411                     | 37                          |
| <i>Elgaria multicarinata webbiai</i> | 1   | 110                         | 2800                     | 37                          |
| <i>Rhineura floridana</i>            | 1   | 110                         | 2581                     | 37                          |
| <i>Podarcis muralis</i>              | 1   | 110                         | 2506                     | 37                          |
| <i>Heteronotia binoei</i>            | 1   | 110                         | 3002                     | 37                          |

|                               |   |     |      |    |
|-------------------------------|---|-----|------|----|
| <i>Anolis carolinensis</i>    | 1 | 110 | 4026 | 37 |
| <i>Erythrolamprus reginae</i> | 1 | 110 | 2638 | 37 |
| <i>Sceloporus undulatus</i>   | 1 | 110 | 2380 | 37 |
| <i>Eublepharis macularius</i> | 1 | 110 | 2577 | 37 |
| <i>Euleptes europaea</i>      | 1 | 110 | 2901 | 37 |
| <i>Hemicordylus capensis</i>  | 1 | 110 | 2830 | 37 |
| <i>Zootoca vivipara</i>       | 1 | 110 | 2516 | 37 |

**Таблица 9: Сводная таблица с характеристикой кассетного интрона для таксономической группы Actinopterygii. Сортировка по возрастанию количества нуклеотидов до стоп-кодона в “кассетном” интроне.**

| Название организма                   | Кол-во нуклеотидов до стоп-кодона в интроне | Длина 1-го экзона в кассете | Длина кассетного интрона | Длина 2-го экзона в кассете |
|--------------------------------------|---|-----------------------------|--------------------------|-----------------------------|
| <i>Chanos chanos</i>                 | 1   | 110                         | 3568                     | 37                          |
| <i>Danio rerio</i>                   | 1   | 110                         | 3580                     | 37                          |
| <i>Denticeps clupeioides</i>         | 7   | 110                         | 2629                     | 37                          |
| <i>Labrus bergylta</i>               | 10  | 110                         | 2684                     | 37                          |
| <i>Cottoperca gobio</i>              | 16  | 110                         | 2388                     | 37                          |
| <i>Xiphophorus couchianus</i>        | 22  | 110                         | 2227                     | 37                          |
| <i>Larimichthys crocea</i>           | 22  | 110                         | 2340                     | 37                          |
| <i>Lates calcarifer</i>              | 22  | 110                         | 2434                     | 37                          |
| <i>Notothenia coriiceps</i>          | 22  | 110                         | 2886                     | 37                          |
| <i>Betta splendens</i>               | 22  | 110                         | 2274                     | 37                          |
| <i>Poecilia reticulata</i>           | 22  | 110                         | 2262                     | 37                          |
| <i>Takifugu rubripes</i>             | 22  | 110                         | 2114                     | 37                          |
| <i>Salarias fasciatus</i>            | 22  | 110                         | 3855                     | 37                          |
| <i>Poecilia mexicana</i>             | 22  | 110                         | 2247                     | 37                          |
| <i>Stegastes partitus</i>            | 22  | 110                         | 2900                     | 37                          |
| <i>Clupea harengus</i>               | 22  | 110                         | 3219                     | 37                          |
| <i>Archocentrus centrarchus</i>      | 22  | 110                         | 2644                     | 37                          |
| <i>Esox lucius</i>                   | 22  | 110                         | 2848                     | 37                          |
| <i>Monopterus albus</i>              | 22  | 110                         | 2353                     | 37                          |
| <i>Echeneis naucrates</i>            | 22  | 110                         | 2314                     | 37                          |
| <i>Paralichthys olivaceus</i>        | 22  | 110                         | 3148                     | 37                          |
| <i>Maylandia zebra</i>               | 22  | 110                         | 2565                     | 37                          |
| <i>Parambassis ranga</i>             | 22  | 110                         | 2484                     | 37                          |
| <i>Sander lucioperca</i>             | 22  | 110                         | 2494                     | 37                          |
| <i>Xiphophorus maculatus</i>         | 22  | 110                         | 2231                     | 37                          |
| <i>Nothobranchius furzeri</i>        | 22  | 110                         | 2290                     | 37                          |
| <i>Anabas testudineus</i>            | 22  | 110                         | 2352                     | 37                          |
| <i>Acanthochromis polyacanthus</i>   | 22  | 110                         | 2797                     | 37                          |
| <i>Anarrhichthys ocellatus</i>       | 22  | 110                         | 2355                     | 37                          |
| <i>Boleophthalmus pectinirostris</i> | 22  | 110                         | 1702                     | 37                          |
| <i>Sparus aurata</i>                 | 22  | 110                         | 2361                     | 37                          |
| <i>Oryzias melastigma</i>            | 22  | 110                         | 2212                     | 37                          |
| <i>Seriola dumerili</i>              | 22  | 110                         | 2494                     | 37                          |
| <i>Poecilia formosa</i>              | 22  | 110                         | 2259                     | 37                          |
| <i>Oreochromis niloticus</i>         | 22  | 110                         | 2580                     | 37                          |
| <i>Kryptolebias marmoratus</i>       | 22  | 110                         | 2556                     | 37                          |
| <i>Xiphophorus hellerii</i>          | 22  | 110                         | 2240                     | 37                          |
| <i>Poecilia latipinna</i>            | 22  | 110                         | 2261                     | 37                          |
| <i>Pundamilia nyererei</i>           | 22  | 110                         | 2527                     | 37                          |
| <i>Hippocampus comes</i>             | 22  | 110                         | 2622                     | 37                          |
| <i>Oreochromis aureus</i>            | 22  | 110                         | 2579                     | 37                          |

|                                     |     |     |      |    |
|-------------------------------------|-----|-----|------|----|
| <i>Amphiprion ocellaris</i>         | 22  | 110 | 2752 | 37 |
| <i>Seriola lalandi dorsalis</i>     | 22  | 110 | 2481 | 37 |
| <i>Austrofundulus limnaeus</i>      | 22  | 110 | 2541 | 37 |
| <i>Puntigrus tetrazona</i>          | 25  | 110 | 2440 | 37 |
| <i>Fundulus heteroclitus</i>        | 25  | 110 | 2476 | 37 |
| <i>Cyprinodon variegatus</i>        | 28  | 110 | 2533 | 37 |
| <i>Haplochromis burtoni</i>         | 31  | 110 | 2535 | 37 |
| <i>Astatotilapia calliptera</i>     | 31  | 110 | 2571 | 37 |
| <i>Gouania willdenowi</i>           | 37  | 110 | 2616 | 37 |
| <i>Oryzias latipes</i>              | 40  | 110 | 2331 | 37 |
| <i>Sphaeramia orbicularis</i>       | 43  | 110 | 2376 | 37 |
| <i>Pygocentrus nattereri</i>        | 46  | 110 | 2649 | 37 |
| <i>Astyanax mexicanus</i>           | 46  | 110 | 2791 | 37 |
| <i>Colossoma macropomum</i>         | 46  | 110 | 2644 | 37 |
| <i>Ictalurus punctatus</i>          | 46  | 110 | 3166 | 37 |
| <i>Tachysurus fulvidraco</i>        | 46  | 110 | 3493 | 37 |
| <i>Pangasianodon hypophthalmus</i>  | 46  | 110 | 3348 | 37 |
| <i>Erpetoichthys calabaricus</i>    | 55  | 110 | 3662 | 37 |
| <i>Perca flavescens</i>             | 58  | 110 | 2378 | 37 |
| <i>Mastacembelus armatus</i>        | 64  | 110 | 2371 | 37 |
| <i>Salmo salar</i>                  | 67  | 110 | 3553 | 37 |
| <i>Gadus morhua</i>                 | 67  | 110 | 3151 | 37 |
| <i>Etheostoma spectabile</i>        | 97  | 110 | 2457 | 37 |
| <i>Scleropages formosus</i>         | 112 | 110 | 3412 | 37 |
| <i>Myripristis murdjan</i>          | 112 | 110 | 2492 | 37 |
| <i>Paramormyrops kingsleyae</i>     | 121 | 110 | 2929 | 37 |
| <i>Carassius auratus</i>            | 148 | 110 | 3854 | 37 |
| <i>Sinocyclocheilus grahami</i>     | 148 | 110 | 3330 | 37 |
| <i>Sinocyclocheilus rhinoceros</i>  | 154 | 110 | 3449 | 37 |
| <i>Sinocyclocheilus anshuiensis</i> | 154 | 110 | 4202 | 37 |
| <i>Electrophorus electricus</i>     | 283 | 110 | 2874 | 37 |



## Благодарности

Я хотел бы поблагодарить моего научного руководителя, Голубкову Елену Валерьевну, и моего куратора, Бондарука Дмитрия Денисовича, за постоянную поддержку и помощь в обсуждении результатов работы.

Отдельно я хотел бы поблагодарить Абрамсон Наталью Иосифовну за повторное рецензирование работы моего авторства.

Также хочу выразить благодарность преподавателям программы “Биоинформатика” и кафедры генетики и биотехнологии СПбГУ, и коллективу преподавателей и ассистентов Института Биоинформатики за полученные знания в процессе обучения, благодаря которым стало возможным осуществление данной работы.