

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА ГЕНЕТИКИ И БИОТЕХНОЛОГИИ

Васильев Артем Викторович

Эволюционные особенности структуры гена *Nxf1*
(nuclear export factor) у *Opisthokonta*

Работа выполнена в
лаборатории генетики животных СПбГУ

Научный руководитель:
к.б.н., доцент, кафедра генетики и биотехнологии,
Голубкова Елена Валерьевна

Санкт-Петербург
2024

Оглавление

Введение	3
Материалы и методы	5
Результаты и обсуждение	6
Заключение	7

Введение

Данная работа посвящена анализу нуклеотидных последовательностей генов семейства *Nxf* у представителей различных филогенетических групп. Она является прямым продолжением бакалаврской работы, выполненной на кафедре генетики и биотехнологии СПбГУ в лаборатории генетики животных.

Ген *Nxf1* является нашим объектом интереса в связи с тем, что возможно образование транскрипта с сохраненным интроном в нем, что само по себе является нечастым явлением. Более того, в данном так называемом "кассетном интроне" присутствует стоп-кодон, и несмотря на это, показано, что данный транскрипт необходим для правильного формирования нервной ткани у *Drosophila melanogaster*.

В предыдущей работе были проанализированы нуклеотидные последовательности данного гена у видов из группы Arthropoda, почти половина из которых являлись представителями семейства Drosophilidae. Сейчас мы хотим сосредоточить наше внимание на группу Stenophora, в связи с тем, что их систематическое положение на данный момент под вопросом. Мы предполагаем, что использование гена *Nxf1* в качестве маркерного поможет разрешить данный вопрос.

Следовательно, актуальность работы заключается в потенциальном разрешении гипотезы о дважды независимом возникновении нервной системы у гребневилов, а также разрешении их систематического положения с использованием *Nxf1* в качестве гена-маркера. Кроме того, подробное изучение нуклеотидной последовательности гена и "кассетного" интрона в частности, позволит проследить его эволюцию в лучшем случае, начиная от прокариот.

Целью данной работы является изучение структуры гена *Nxf1* у представителей разных филогенетических групп для выявления закономерностей эволюции генов семейства *Nxf*.

В рамках цели сформулированы следующие **задачи**:

1. Разработка пайплайна для автоматизации обработки и анализа данных
 - использование нуклеотидных последовательностей представителей семейства Drosophilidae с целью разработки метода автоматизированного анализа результатов BLAST
2. Поиск
 - "консервативной кассеты" в нуклеотидной последовательности гена *Nxf1* у представителей Opisthokonta в геномных базах данных
 - полной нуклеотидной последовательности гена *Nxf1* у представителей Stenophora
 - эволюционного предшественника гена *Nxf1* у простейших
 - РНК-связывающих участков у прокариот

3. Сравнение найденных нуклеотидных последовательностей с

- полной нуклеотидной последовательностью гена *Nxf1*
- "консервативной кассетой" в составе гена *Nxf1*
- "кассетным" интроном в составе "консервативной кассеты" гена *Nxf1* между представителями разных филогенетических групп

На данный момент ведется активная работа над разработкой пайплайна для обработки результатов BLAST.

Материалы и методы

В качестве **материалов** выступают нуклеотидные последовательности представителей разных филогенетических групп, взятые из открытых и закрытых баз данных.

Для выполнения вышестоящих задач планируется использование следующих **методов**:

1. Поиск ортологов внутри таксономических групп с помощью BLAST или Diamond
2. Анализ и обработка полученных результатов путем выполнения скриптов, написанных на языке программирования Python, с использованием различных библиотек (например, Biopython)
3. Использование командной строки Linux и различных программ обработки и визуализации данных
 - для построения пайплайнов – SnakeMake
 - для выравнивания последовательностей на локальном компьютере – BWA (Burrows-Wheeler Alignment)
 - для визуализации полученных выравниваний – IGV (Integrative Genomics Viewer)
4. Множественное выравнивание нуклеотидных последовательностей с помощью алгоритма MUSCLE в программах UNIPRO UGENE и MEGA-X

Пайплайн для обработки результатов BLAST включает следующие этапы:

1. Получить и прочесть результаты BLAST
2. Посчитать Query Coverage (QC) в каждой найденной последовательности
3. Отфильтровать по заданному порогу для QC TODO
4. Сгруппировать оставшиеся с помощью алгоритма кластеризации TODO
5. Загрузить найденные последовательности в соответствующие папки
6. Провести множественные выравнивания TODO
7. Проанализировать итоговые результаты TODO

Пункты, помеченные "TODO" находятся на этапе разработки. С реализацией остальных стадий анализа можно ознакомиться в GitHub репозитории:

https://github.com/ArtemVaska/Diploma/tree/bachelor_repeat
в ноутбуке bachelor_repeat.ipynb в папке Scripts.

Результаты и обсуждение

На данной стадии написания дипломной работы:

- был проведен анализ литературы по теме
- изучены методы проведения BLAST в разных вариантах (первые 2 являются предпочтительными вариантами):
 - NCBI веб-инструмент
 - командная строка локальная БД
 - Python пакет Bio.Blast
 - командная строка удаленная БД
- разработан подход для парсинга результатов .xml2 файлов
- частично разработан пайплайн для обработки результатов BLAST

Заключение

Продолжается работа по автоматизации процесса запуска BLAST и анализа его результатов. В ближайшее время планируется закончить пайплайн и перейти к автоматизации анализа результатов множественных выравниваний. В последующем будет произведен переход к решению поставленных ранее задач, а именно поиску нуклеотидных последовательностей гена *Nxf1* у представителей различных филогенетических групп, а также сравнению консервативных участков гена и "консервативной кассеты" у найденных последовательностей. Особое внимание будет уделено интрону из "консервативной кассеты".