

ОТЗЫВ официального оппонента

на диссертацию МУСТАФИНА ЗАХАРА СЕРГЕЕВИЧА на тему «РАЗРАБОТКА КОМПЛЕКСА ПРОГРАММ ДЛЯ АНАЛИЗА ЭВОЛЮЦИОННЫХ ХАРАКТЕРИСТИК ГЕННЫХ СЕТЕЙ», представленную к защите на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика

27 сентября 2021г.

Актуальность

Исследование механизмов эволюции живых систем необходимо для решения многих вопросов современной биологии. Все больше методов создается в настоящее время для компьютерного анализа эволюционных процессов. Одним из важных направлений эволюционных исследований стал филостратиграфический анализ — подход к исследованию эволюционных преобразований генов на основе сравнения ортологов и определения наиболее вероятного времени возникновения гена. За счет комбинирования с методами, основанными на сравнении последовательностей близкородственных генов и поиске отношения частот несинонимичных и синонимичных замен, эволюционный анализ позволяет оценить возраст гена и степень его эволюционной изменчивости. Большая часть приложений, осуществляющих подобный анализ, распространяется в виде пакетов к языку программирования R и работают они со списками генов. Существует большая потребность в программ, работающих в других средах и расширяющих методологию анализа, что и обуславливает актуальность рецензируемой работы.

Диссертация Мустафина Захара Сергеевича посвящена разработке методов анализа эволюционных характеристик генных сетей, а также функционально связанных групп генов, не объединенных в сеть. В диссертации представлены приложения Orthoscape и Orthoweb для анализа эволюционных характеристик генных сетей и групп функционально связанных генов, не объединенных в сеть. С помощью разработанных методов было проведено исследование генных сетей заболеваний человека, а также генов, ассоциированных с различными типами стресса у *Arabidopsis thaliana*.

Научная новизна

С использованием современных технологий и языков программирования 3.C. Мустафиным разработаны программа Orthoscape, являющаяся плагином для востребованного в биоинформатике программного комплекса Cytoscape и самостоятельное приложение Orthoweb. Orthoscape и Orthoweb позволяют проводить анализ генов с целью выявления эволюционного этапа их возникновения, основываясь на исследовании групп ортологических генов. Приложения позволяют проводить анализ степени давления отбора на гены на основе сравнения отношения долей несинонимичных и синонимичных замен организмов и определения отношения долей несинонимичных и синонимичных замен в них. Научная новизна Orthoscape состоит в том, что он сконцентрирован на анализе генных сетей и позволяет визуализировать результаты анализа с помощью средств, предоставленных в Cytoscape. На данный момент Orthoscape является единственным доступным в Cytoscape плагином, проводящим эволюционный анализ генов, входящих в сеть. Orthoweb является самостоятельным веб-приложением, выполняющим схожие с Orthoscape функции, однако менее требовательным к вычислительным ресурсам и работающим быстрее, за счет клиент-серверной архитектуры и заранее подготовленной базы данных. Orthoweb не имеет таких же возможностей визуализации сетей, как Orthoscape, и ориентирован на анализ списков генов, не объединенных в сеть.

С помощью разработанных средств были проанализированы гены, ассоциированные с заболеваниями человека. С использованием Orthoscape было показано, что в функционировании заболеваний, связанных с иммунной системой, вовлечены эволюционно молодые гены, в то время как в функционировании заболеваний, ответственных за

формирование зависимостей человека от веществ, вызывающих привыкание, вовлечены эволюционно древние гены. Также высокая доля эволюционно молодых генов отмечена среди генов, задействованных в инфекционных заболеваниях, вызванных паразитами, а высокая доля эволюционно древних генов отмечена среди генов, ассоциированных со специфическими типами рака. Была показана высокая достоверная корреляция эволюционного возраста и эволюционной изменчивости гена – чем эволюционно моложе ген, тем сильнее он подвержен генетической изменчивости.

С помощью Orthoweb были проанализированы гены, ассоциированные с абиотическим стрессом у *A. thaliana* и показано, что среди таких генов достоверно перепредставлены эволюционно древние и консервативные гены, по сравнению с полным набором генов *A. thaliana*, кодирующих белки.

Структура диссертационной работы

Диссертация состоит из введения, пяти глав, заключения, выводов, списка литературы из 139 наименований. Содержание основного текста диссертации изложено на 116 страницах, содержит 42 рисунка и 6 таблиц.

Первая глава является обзором литературы, в котором описаны эволюционные характеристики генов и все связанные с ними понятия, генные сети, а также базы биологических данных, с которыми проводилась работа. Приведены формулы для расчета ранее созданных эволюционных индексов и результаты, ранее полученные с помощью филогенетического анализа.

Во второй главе приведены материалы и методы, использованные в работе. В главе представлено краткое описание технологий Spring, Webix и MongoDB, использованных для написания приложения Orthoweb. Представлено описание программного комплекса Cytoscape, необходимого для функционирования плагина Orthoscape.

Третья глава содержит описание самих приложений Orthoscape и Orthoweb. В разделе 3.1 отмечены общие черты приложений Orthoscape и Orthoweb. Показано, на основании каких данных приложениями проводится поиск гомологичных генов, описано, каким параметрами можно контролировать этот отбор. В этом же разделе описаны эволюционные характеристики, используемые в приложениях Orthoscape и Orthoweb, а именно PAI (филогенетический индекс гена) и DI (индекс изменчивости гена), и способы их расчета. Показаны способы визуализации результатов, доступные в обоих средствах. Это генерируемые приложениями графики изменения филогенетического индекса, в зависимости от входных параметров, а также гистограммы распределения этого индекса по значениям и графики, на которых эти гистограммы объединены в виде диаграмм размаха, позволяющих сравнить входные наборы генов между собой.

В разделе 3.2 перечислены отличия Orthoscape и Orthoweb. В первую очередь это касается времени работы приложения. Orthoweb является веб-приложением и имеет заранее подготовленную и периодически обновляемую базу данных на сервере, что делает его работу достаточно быстрой, в то время как Orthoscape работает на компьютере пользователя. Orthoscape также позволяет создать локальную базу данных для ускорения дальнейшей работы, однако первый запуск приложения на определенных данных всегда будет более длительным. В этом же разделе описаны типы генных сетей, с которыми может работать Orthoscape, а также необходимые условия для того, чтобы импортировать исходный набор генов в Cytoscape и привести к формату, совместимому с Orthoscape. Приведен пример визуализации сетей, проводимой с помощью цветовых схем, реализованных в Orthoscape. В разделе 3.3 показан пример эволюционного анализа и интерпретации его результатов на основе анализа сетей биосинтеза стероидов и биосинтеза стероидных гормонов, взятых из базы данных KEGG Pathway. И визуальный анализ окрашенных в соответствии с посчитанными характеристиками сетей, и результаты сравнения долей эволюционно молодых и эволюционно древних генов с помощью точного теста Фишера, позволяют сделать вывод, что в сети биосинтеза стероидов достоверно больше эволюционно древних генов и меньше эволюционно молодых, чем в сети биосинтеза стероидных гормонов. Полученные результаты согласуются

с текущими знаниями об эволюции эукариот.

В четвертой главе приведены результаты использования Orthoscape для анализа генных сетей заболеваний человека, взятых из базы данных KEGG Pathway, раздела Human Diseases. В работе рассмотрено 80 генных сетей, разбитых на 11 категорий заболеваний – различные типы рака, сердечно-сосудистые заболевания, инфекционные заболевания и т.д. Представлены данные по количеству сетей в категориях и среднему количеству генов в сетях. Приведены результаты эволюционных характеристик PAI и DI, для категорий, полученные на основании усреднения значений характеристик для сетей. Анализ показал, что сети с максимальным значением PAI и DI, т.е. в которых наибольшая доля эволюционно молодых и эволюционно изменчивых генов, как правило, принадлежат к группе заболеваний, связанных с иммунной системой. Указано, что высокая доля эволюционно молодых и изменчивых генов в этих сетях приходится на позвоночных, что соответствует современным представлениям о развитии специфического иммунитета. В то же время сети с минимальным значением PAI и DI, т.е. в которых наибольшая доля эволюционно древних и эволюционно консервативных генов, как правило, принадлежат к сетям, описывающим зависимость от веществ, вызывающих привыкание. Отмечено, что, как правило, такие гены описывают основополагающие для организма процессы. С чем и может быть связано их древнее происхождение и низкая изменчивость. Кроме того, отмечено, что между значениями индексов PAI и DI наблюдается высокая достоверная корреляция – эволюционно молодые гены, как правило, являются изменчивыми, а эволюционно древние – консервативными.

Отдельными разделами приведены результаты анализа генной сети болезни Паркинсона и Диабета 1-го и 2-го типов, как заболеваний, актуальных в настоящее время. Исследование эволюционных характеристик генов, вовлеченных в генную сеть болезни Паркинсона, позволило выявить несколько генов, выделяющихся по показателям эволюционных индексов. Большинство из них относятся к генам, мутации в которых приводят к возникновению определенного типа болезни Паркинсона, однако есть и гены, за которыми не была отмечена такая важная роль и возможно эти гены могли бы стать целью более подробного исследования. По результатам анализа генной сети диабета первого типа были выделены два кластера – эволюционно молодых и изменчивых генов и эволюционно древних и консервативных генов. Однако анализ генной сети диабета второго типа позволил выделить наименьшее число генов для дальнейшего анализа.

В пятой главе описана работа по исследованию генов *A. thaliana*, ассоциированных с семью различными типами абиотического стресса. В подразделах описаны все этапы работы: отбор генов, эволюционный анализ, статистический анализ результатов, анализ функций исследуемых генов на основе ассоциаций с терминами генной онтологии, реконструкция генных сетей и анализ их структуры, обсуждение результатов. Результаты анализа позволили предположить, что в процессе эволюции новые функции могли выполнять эволюционно молодые гены, но при этом в основе сетей лежат кластеры эволюционно древних генов. В целом, среди генов, ассоциированных с типами абиотического стресса, достоверно перепредставлены эволюционно древние и консервативные гены.

В заключении собраны воедино и сформулированы результаты. Выводы диссертации носят конкретный характер и адекватно отражают полученные результаты.

Автореферат

Текст автореферата соответствует содержанию диссертационной работы. Содержание работы детально и качественно отражено в публикациях. По теме работы опубликовано 3 статьи в изданиях, входящих в список ВАК, 12 тезисов конференций и одно авторское свидетельство. Результаты работы были доложены на международных и российских конференциях.

Замечания

Должен отметить ряд недостатков диссертации З.С. Мустафина

1. Литературный обзор носит технический характер, оставляя в стороне обсуждение таких важных понятий, например – «возраст гена» .
2. Представляется необходимым перечисление в явном виде (например, в виде списка) набора предположений, сделанных при разработке программ. Это особенно необходимо, поскольку эволюционный анализ геновых сетей - весьма новое направление, методология которого существенно отличается от методологии, в частности, молекулярно-филогенетического анализа. Это необходимо для адекватного использования разработанных Захаром Сергеевичем программ;
3. В тексте встречается «телеграфный стиль», который порой даже затрудняет понимание. При описании Orthoweb упомянут некий «стандартный браузер», который нужен для работы программы.
4. Защищаемые положения 1 и 2, на мой взгляд, сформулированы слишком категорично. Положение 2: «У человека эволюционно молодыми генами обогащены геновые сети, связанные с заболеваниями иммунной системы, а эволюционно древними – с зависимостью от веществ...» следовало бы заменить на «На основании ... сформулирована гипотеза о том, что у человека эволюционно молодыми генами обогащены геновые сети, связанные с заболеваниями иммунной системы, а эволюционно древними – с зависимостью от веществ, вызывающих привыкание.», поскольку иначе единственный обнаруженный случай противоположного качества геновой сети опровергает это утверждение. Такое же замечание относится и к 3-му положению.

Высказанные замечания носят в основном редакционный характер и не умаляют научного значения представленной работы.

ЗАКЛЮЧЕНИЕ

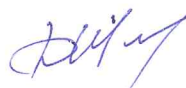
Диссертационная работа Мустафина Захара Сергеевича «Разработка комплекса программ для анализа эволюционных характеристик геновых сетей», представленная на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика, является завершенной научно-исследовательской работой по разработке и применению средств для компьютерного анализа биологических данных.

В рамках работы было создано два приложения для анализа эволюционных характеристик генов. Приложение Orthoscape, являющееся плагином к программному комплексу Cytoscape, позволяет провести эволюционный анализ генов, объединенных в геновую сеть, визуализировать результаты, воспользовавшись всеми преимуществами Cytoscape. Менее требовательное к ресурсам компьютера веб-приложение Orthoweb, для доступа к которому достаточно интернет браузера, позволяет провести более быстрый анализ эволюционных характеристик генов, не объединенных в сеть, за счет заранее подготовленной базы данных. Данные приложения позволяют определить такие эволюционные характеристики, как возраст гена и степень его изменчивости.

Представленные в работе результаты по анализу заболеваний человека и генов, ассоциированных со стрессом у *A. thaliana*, также могут послужить основой для более подробных исследований генов, выделенных с помощью разработанных приложений. Результаты работы опубликованы в рецензируемых журналах, входящих в перечень ВАК. Результат разработки программы Orthoscape опубликован в статье в журнале BMC Bioinformatics, а результаты анализа генов, ассоциированных с абиотическими стрессами у *A. thaliana* и геновых сетей заболеваний человека, в журналах Genes и Вавиловском журнале Генетики и Селекции соответственно. Тема диссертации, публикации по работе и положения, выносимые на защиту, полностью соответствуют специальности 03.01.09 – математическая биология, биоинформатика.

Представленная работа полностью соответствует критериям пп. 9-14 «Положения о присуждении научных степеней», утвержденного постановлением Правительства Российской Федерации № 842 от 24 сентября 2013 г. (в редакции с изменениями, утвержденными Постановлением Правительства РФ от 11 сентября 2021 года №1539), а ее автор Мустафин Захар Сергеевич заслуживает присуждения ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

д.б.н., заведующий лабораторией геносистематики
Лимнологического института СО РАН,
664033 Иркутск, ул. Улан-Баторская 3,
тел. (3952) 42-29-23,
Электронная почта: sherb@lin.irk.ru

 Д.Ю.Щербаков

Подпись заведующего лабораторией, д.б.н. Щербакова Д.Ю. ЗАВЕРЯЮ
Ученый секретарь ЛИИ СО РАН к.б.н. Максимова Н.В.

