

3. *Семак И. В., Зырянова Т. Н., Губич О. И.* Биохимия белков. Минск, 2008.
4. Камышников В. С. Справочник по клинико-биохимическим исследованиям и лабораторной диагностике М. 2009.
5. *Маршал В. Дж.* Клиническая биохимия М. 1999.
6. *Gonzales-Burgos E., Carretero M. E., Gomez-Serranillos M. P.* Sideritis spp.: uses, chemical composition and pharmacological activities // J. Ethnopharmacol. 2011. Vol. 135. № 2. P. 209–225.

SIGMOID: УДОБНАЯ КОМПЬЮТЕРНАЯ ПРОГРАММА ДЛЯ ИДЕНТИФИКАЦИИ И АННОТАЦИИ РЕГУЛЯТОРНЫХ УЧАСТКОВ БАКТЕРИАЛЬНЫХ ГЕНОМОВ

А. В. Доменикан

ВВЕДЕНИЕ

На сегодняшний день аннотация бактериальных геномов становится всё более автоматизированной. Принцип «от гена к гену», лежащий в основе аннотации с помощью компьютерных программ [1], является эффективным при идентификации таких функциональных элементов генома, как гены, кодирующие белки и РНК, сигнальные пептиды. Однако регуляторные последовательности (сайты связывания транскрипционных и сигма факторов, транскрипционные терминаторы) и структура оперонов редко принимаются во внимание, что снижает качество и информативность конечной аннотации.

В последние годы в открытом доступе накапливается всё больше информации, касающейся координированной регуляции экспрессии генов, особенно для модельных организмов. В частности, данные о сайтах связывания различных факторов транскрипции и сигма факторов бактерий представлены в таких базах данных, как RegPrecise [2] и RegulonDB [3]. Таким образом, предполагая, что функционально значимые регуляторные последовательности эволюционно консервативны, становится возможной идентификация таких сигналов при помощи построения математических/статистических моделей. Дальнейший поиск последовательностей в новосеквенированном штамме/виде бактерий производится на основе таких первичных моделей.

Основные биоинформатические подходы, которые решают проблему построения описательных моделей, являются позиционные весовые матрицы (ПВМ) [4] и профилированные скрытые Марковские модели (ПСММ) [5].

Существующие на момент написания данной статьи программные пакеты, которые осуществляют построение первичных моделей и поиск

схожих последовательностей в геномах (в частности HMMER [6] и MEME Suite [7]), не имеют графических интерфейсов и геномных обозревателей для визуализации результатов поиска. Более того, при поиске данные программы не учитывают геномное окружение найденных сайтов и не аннотируют их, что затрудняет последующий отбор ортологичных сайтов при построении вторичных моделей (специфичных для целевого генома).

Таким образом, целью данной работы было создание компьютерного приложения, которое бы сочетало в себе графический интерфейс к популярным консольным программам поиска регуляторных последовательностей; производило соответствующую фильтрацию и аннотацию найденных сайтов, исходя из их геномного окружения и желаний пользователя; было оснащено геномным обозревателем для визуализации результатов поиска.

МАТЕРИАЛЫ И РЕЗУЛЬТАТЫ

Графический интерфейс написан на языке программирования Xojo, что позволяет использовать созданную программу на трёх операционных системах (OS X, Linux и Windows). Sigmoid отображает сайты связывания в виде лого [8]. Построение моделей и поиск последовательностей производится программными пакетами HMMER (программы hmmbuild и nhmmer соответственно) и MEME Suite (программы MEME и MAST соответственно). Поиск транскрипционных терминаторов осуществляется программой TransTerm HP [9]. Соответствующая фильтрация и аннотация результатов поиска производится отдельными программными сценариями (скриптами): HmmGen.py, MastGen.py и TermGen.py для результатов соответствующих программ. Скрипты написаны на языке программирования Python с использованием библиотеки Biopython, и могут быть использованы отдельно от программы Sigmoid. Исходный код приложения и компилированные версии сопутствующих программ доступны на GitHub репозитории [10].

На рис. 1 представлена общая схема работы программы Sigmoid.

Поиск сайтов связывания транскрипционных и сигма факторов требует файла множественного выравнивания в формате FASTA и геномного файла в формате GenBank. Программа Sigmoid предоставляет онлайн доступ к базам данных RegPrecise и RegulonDB, что упрощает извлечение интересующих исследователя данных о сайтах связывания регуляторных белков в геномах родственных бактерий.

Поиск транскрипционных терминаторов требует лишь геномного файла в формате GenBank. Необходимые файловые конверсии для программы TransTerm HP выполняет скрипт TermGen.py.

Геномный обозреватель программы Sigmoid позволяет открывать геномные файлы в формате GenBank как после произведённого поиска регуляторных сигналов (с возможностью переключения от одного найденного сайта к другому), так и независимо от остальных функций программы (рис. 3). Геномный обозреватель предоставляет также быстрый доступ к базам данных белковых и нуклеотидных последовательностей, а также позволяет быстро редактировать аннотацию функциональных элементов генома, что в совокупности делает обозреватель удобным инструментом ручной аннотации.

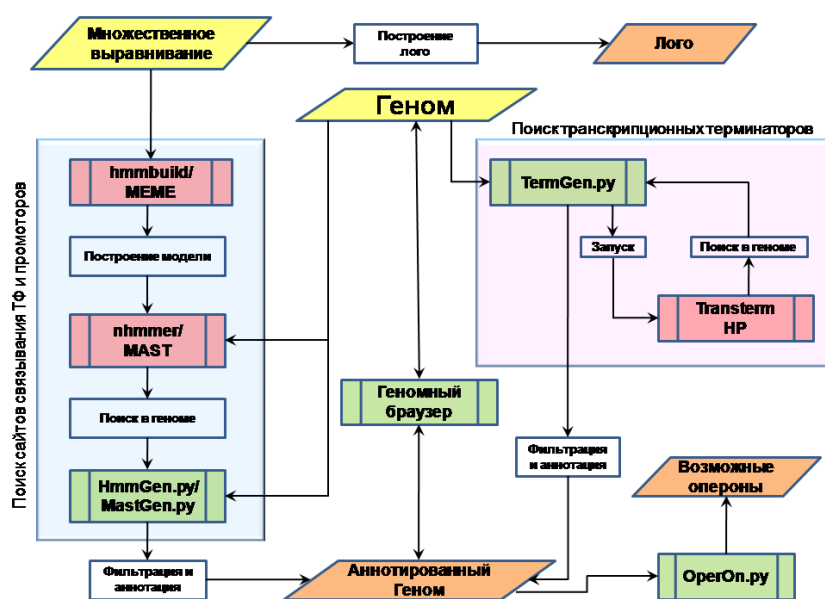


Рис. 1. Общая блок-схема работы программы Sigmoid.

В параллелограммах представлены данные: жёлтые – входные данные, оранжевые – выходные. В прямоугольниках с вертикальными линиями – программы или их элементы: красные – сторонние, зелёные – элементы программы Sigmoid. В не закрашенных прямоугольниках отображены исполняемые процедуры. Направления стрелок указывает поток информации между блоками. Синяя область отображает контур поиска сайтов связывания регуляторных белков, розовая – поиска транскрипционных терминаторов.

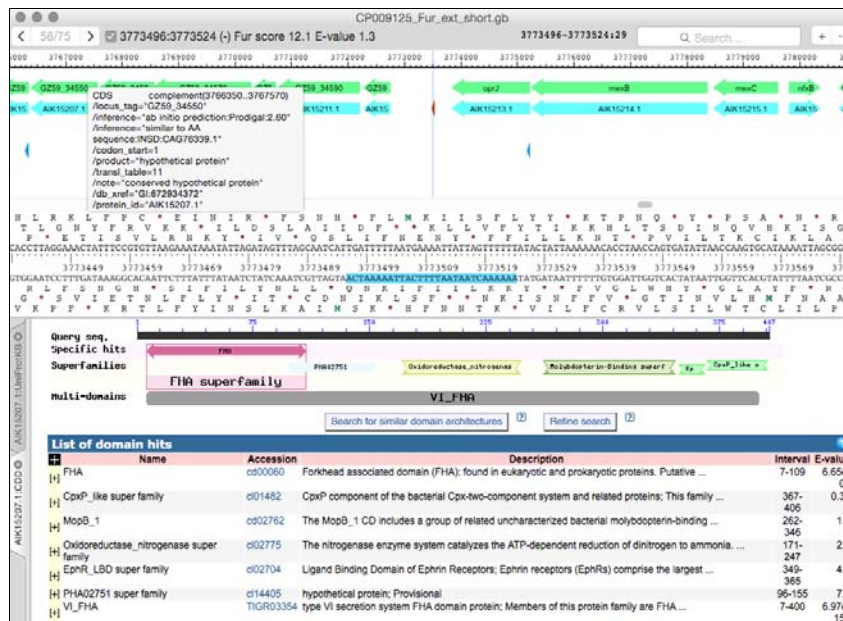


Рис. 2. Окно геномного обозревателя программы SigmodID.

Окно разделено на три панели. Верхняя панель непосредственно отображает функциональные элементы генома в виде стрелок различного цвета. Средняя панель отображает нуклеотидную/аминокислотную последовательность выбранного участка генома. Нижняя панель – интегрированный интернет обозреватель, оснащённый возможностью быстрого поиска в базах данных нуклеотидных и белковых последовательностей.

После поиска и аннотации регуляторных последовательностей возможно перечисление теоретических транскрипционных единиц (ТЕ), которые находятся под контролем интересующих белков. Извлечение данной информации производится скриптом OperOn.py. Уточнение границы теоретических ТЕ возможно с помощью отображения в геномном обозревателе графиков данных по РНК - секвенированию генома (рис. 3).

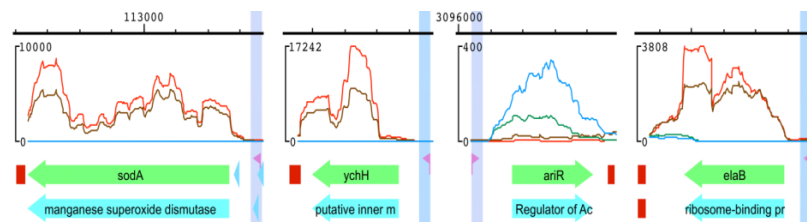


Рис. 3. Верификация транскрипционных единиц в геномном обозревателе.

Промоторные последовательности обозначены флажками, сайты связывания транскрипционных факторов – синими стрелками, транскрипционные терминаторы – красными прямоугольниками, гены – зелёными стрелками, белок - кодирующие последовательности –

бирюзовыми стрелками. Линии над элементами аннотации отображают глубину покрытия данных РНК - секвенирования на геном.

Таким образом, созданная программа Sigmoid снабжена всеми необходимыми функциями и свойствами для упрощения биоинформатической идентификации и аннотации регуляторных последовательностей в бактериальных геномах, что делает её удобным инструментом для изучения транскрипционной регуляции метаболизма бактерий.

Литература

1. *Seemann T.* Prokka: rapid prokaryotic genome annotation // *Bioinformatics*. 2014. Vol. 30. № 14. P. 2068-2069.
2. *P.S. Novichkov* RegPrecise web services interface: programmatic access to the transcriptional regulatory interactions in bacteria reconstructed by comparative genomics // *Nucleic Acids Research*. 2012. Vol. 40. № W1. P. W604-W608.
3. *S. Gama-Castro* RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units) // *Nucleic Acids Research*. 2011. Vol. 39. № Database. P. D98-D105.
4. *Stormo G.D.* DNA binding sites: representation and discovery // *Bioinformatics* (Oxford, England). 2000. Vol. 16. DNA binding sites. № 1. P. 16-23.
5. *Yoon B.-J.* Hidden Markov Models and their Applications in Biological Sequence Analysis // *Current Genomics*. 2009. Vol. 10. № 6. P. 402-415.
6. *Eddy S.R.* Accelerated Profile HMM Searches // *PLoS Computational Biology*. 2011. Vol. 7. № 10. P. e1002195.
7. *T.L. Bailey* The MEME Suite // *Nucleic Acids Research*. 2015. Vol. 43. № W1. P. W39-W49.
8. *T.D. Schneider, R.M. Stephens* Sequence logos: a new way to display consensus sequences. // *Nucleic Acids Research*. 1990. Vol. 18. № 20. P. 6097-6100.
9. *C.L. Kingsford, K. Ayanbule, S.L. Salzberg* Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake // *Genome Biology*. 2007. Vol. 8. № 2. P. R22.
10. <https://github.com/nikolaichik/Sigmoid>.

МОЛЕКУЛЯРНЫЕ МЕХАНИЗМЫ ШТАММ-СПЕЦИФИЧЕСКОЙ ИНДУКЦИИ РЕАКЦИИ СВЕРХЧУВСТВИТЕЛЬНОСТИ РАСТЕНИЙ *NICOTIANA TABACUM* ПРИ ЗАРАЖЕНИИ *PECTOBACTERIUM ATROSEPTICUM*

Ю. В. Дюбо

Pectobacterium atrosepticum – грамотрицательные энтеробактерии, способные вызывать заболевания картофеля – "черную ножку" и мягкую гниль. Штамм 21А, в отличие от большинства других штаммов *P. atrosepticum*, индуцирует реакцию сверхчувствительности у растений табака *Nicotiana tabacum*, которые не являются для него естественным