



DOI: <https://doi.org/10.15688/jvolsu11.2018.2.12>

UDC 004:577.2:579.852.11

LBC 28.0

A META-ANALYSIS OF GENOMIC AND INTERACTOMIC INTERACTION DATA BETWEEN HOST AND *B. ANTHRACIS*

Gregory A. Pechkovsky

Stavropol Anti-plague Institute, Stavropol, Russian Federation

Alena V. Abramovich

Stavropol Anti-plague Institute, Stavropol, Russian Federation

Elena A. Koteneva

Stavropol Anti-plague Institute, Stavropol, Russian Federation

Abstract. A variety of techniques are used to obtain omix data, and the output of these experiments is heterogeneous in type data, which, however, can be reduced to sequences (proteins or genes) for their further meta-analysis. The proposed algorithm was implemented in Python, and it is based on the use of the program BLAST with the request of a certain assembly, the subsequent multiple alignment of the initial and grouped sequences. With the help of this algorithm, we analyzed the sequences of proteins of *B. anthracis* or *H. sapiens* from interactome (IMEx IM-13779) obtained from Uniprot database. Were used genomic data 36 RefSeq complete genomes of representatives of the *B. cereus* group (*B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. mycoides*) for this analysis.

Key words: meta-analysis, genome, interactive data, *B. Anthracis*.

УДК 004:577.2:579.852.11

ББК 28.0

МЕТААНАЛИЗ ГЕНОМНЫХ И ИНТЕРАКТОМНЫХ ДАННЫХ ВЗАИМОДЕЙСТВИЯ МЕЖДУ ХОЗЯИНОМ И *B. ANTHRACIS*

Григорий Александрович Печковский

Ставропольский противочумный институт Роспотребнадзора, г. Ставрополь, Российская Федерация

Алёна Владимировна Абрамович

Ставропольский противочумный институт Роспотребнадзора, г. Ставрополь, Российская Федерация

Елена Анатольевна Котенева

Ставропольский противочумный институт Роспотребнадзора, г. Ставрополь, Российская Федерация

Аннотация. Для получения омиксных данных используются разнообразные методики, на выходе этих экспериментов получаются разнородные по типу данные, которые, тем не менее, могут быть сведены к последовательностям (белков или генов) для их дальнейшего метаанализа. Предлагаемый алгоритм был реализован на языке python, и основан на использовании программы BLAST с запросом определенной сборки, последующего множественного выравнивания начальных и сгруппированных последовательностей. С помощью данного алгоритма были проанализированы последовательности белков *B. anthracis* или *H. sapiens* из интерактома (IMEx IM-13779), полученные из базы данных Uniprot. В анализе использовались геномные данные 36 полных RefSeq геномов представителей *B. cereus* group (*B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. mycoides*).

Ключевые слова: метаанализ, геном, интерактомные данные, *B. Anthracis*.

Возбудитель сибирской язвы, *Bacillus anthracis*, это граммположительная, спорообразующая, неподвижная палочка, факультативный анаэроб. Сибирской язве подвержены преимущественно травоядные животные, как дикие как и сельскохозяйственные, а так же человек. В организм хозяина *B. anthracis* попадает, как правило, в покоящемся состоянии – в виде спор, которые, попав в благоприятные условия, прорастают, переходя в фазу вегетативного роста. В ходе этих процессов наблюдается значительное изменение уровня экспрессии многих генов *B. anthracis* [3]. Так же известно, что течение сибиреязвенной инфекции у животных разных эволюционных групп отличается [1]. При этом выполнение экспериментов *in vivo* ограничено в виду сложности, подобных экспериментов, в том числе, соблюдения требований биологической безопасности. В настоящее время собран большой объем данных, но остается много загадок о жизненном цикле *B. anthracis* в организме хозяина. В том числе требуются экстраполяция знаний с помощью биоинформатических подходов. Нами разработан алгоритм (пайлайн) для анализа мультиомиксных данных и его применения для анализа геномных, протеомных и интерактомных данных.

Для получения омиксных данных используются разнообразные методики такие как микрочипы, секвенирование масс-спектрометрии, двугибридный анализ и т. д. Соответственно на выходе этих экспериментов получаются разнородные по типу данные, которые, тем не менее, могут быть сведены к последовательностям (белков или генов) для их дальнейшего метаанализа.

Предлагаемый нами алгоритм был реализован на языке *python*. На вход пайлайну подается сборка последовательностей из омиксных данных (например, белки из геномов) и таблица к каким группам или кластерам принадлежит каждая из сборок. На первом этапе используется программа BLAST с запросом определенной сборки или всех сборок, отбираются последовательности с определенной степенью идентичности в отдельные начальные *fasta* файлы. Далее происходит группировка начальных файлов в соответствии с их кластером в отдельные группиро-

ванные *fasta* файлы. И производится множественное выравнивание как начальных, так и сгруппированных последовательностей. Это дает возможность сравнения множеств последовательностей и параметров множественных выравниваний между группами. За параметр множественного выравнивания бралась информационная энтропия, рассчитанная по формуле как отрицательная сумма вероятности, умноженная на двоичный логарифм от вероятности присутствия буквы в последовательности. Чем больше энтропия, тем больше неопределенность той или иной буквы, или в биологическом смысле, изменчивость.

С помощью данного алгоритма были проанализированы последовательности белков *B. anthracis* или *H. sapiens* из интерактома (IMEx IM-13779) [2]. Последовательности были получены с помощью программы *Uniprot id* из базы данных *Uniprot*. Сравнение последовательностей производился с запросом белков из интерактома отдельно *B. anthracis* и *H. sapiens* против геномов бацилл, относящихся к *Bacillus cereus group* и геномов восприимчивых животных соответственно. В анализе использовались геномные данные 36 полных RefSeq геномов представителей *B. cereus group* (*B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. mycoides*), 12 геномов хорошо (модельные животные и жвачные животные) и 7 (отряда хищных *Carnivora*) слабо восприимчивых к сибирской язве животных [1]. Группировка производилась в соответствии со штаммами и восприимчивостью.

Были вычислены параметры множественных выравниваний (энтропия, матрица расстояний) и получены множества гомологичных белков, принадлежащих определенным группам и пересекающиеся между группами. Общих белков (белки, присутствующие во всех множествах) в *B. cereus group* 450 из 936 со средней энтропией 37,16 (SD 50,7), общих белков среди штаммов *B. anthracis* 854 из 935 со средней энтропией 4,25 (SD 25,5). Множество пересекающихся белков между хорошо и слабо восприимчивыми животными 925 из 1638 со средней энтропией 353,9 (SD 404) в группе с хорошо и 261,3 (SD 311,5) в группе животных со слабой восприимчивостью.

Выводы. Таким образом, полученные данные говорят, что достаточно большое коли-

чество белков представителей *B. cereus group* гомологично к белкам участвующим во взаимодействии между хозяином и *B. anthracis*, но с достаточно большим количеством замен в сравнении со штаммами *B. anthracis*. Возможно, эти белки подверглись эволюционным изменениям для поддержания патогенности. При этом среди самих штаммов разнообразие белков и их изменчивость значительно ниже по сравнению с другими видами *B. cereus group*, что может говорить о большей стабильности генома *B. anthracis*. Так же значение между хорошо и слабо восприимчивыми к сибиреязвенной инфекции животными говорят о сильной вариации в белках. Возможно, такие данные будут полезны для понимания наследственной предрасположенности к инфекции и смогут помочь в поиске белков влияющих на течение сибирской язвы. Дальнейшее улучшение алгоритма по-

зволят дать более детальное описание данных и выявить новые закономерности. Полученные данные и скрипты доступны через запрос по электронной почте.

СПИСОК ЛИТЕРАТУРЫ

1. Beyer, W. Anthrax in animals / W. Beyer, P. C. B. Turnbull // *Molecular aspects of medicine*. – 2009. – Vol. 30, № 6. – P. 481–489.
2. The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis* / M.D. Dyer, C. Neff, M. Dufford, et al. // *PloS One*. – 2010. – Vol. 5, № 8. – e12089.
3. Transcriptional profiling of *Bacillus anthracis* during infection of host macrophages / N. H. Bergman, E. C. Anderson, E. E. Swenson [et al.] // *Infection and immunity*. – 2007. – Vol. 75, № 7. – P. 3434–3444.

Information about the Authors

Gregory A. Pechkovsky, Junior Researcher, Laboratory of Postgenomic Technologies, Stavropol Anti-plague Institute, Sovetskaya St., 13-15, 355035 Stavropol, Russian Federation, stavnipchi@mail.ru.

Alena V. Abramovich, Junior Researcher, Laboratory of Postgenomic Technologies, Stavropol Anti-plague Institute, Sovetskaya St., 13-15, 355035 Stavropol, Russian Federation, alena.abramovich.1992@yandex.ru.

Elena A. Koteneva, Researcher, Laboratory of Postgenomic Technologies, Stavropol Anti-plague Institute, Sovetskaya St., 13-15, 355035 Stavropol, Russian Federation, stavnipchi@mail.ru.

Информация об авторах

Григорий Александрович Печковский, младший научный сотрудник лаборатории постгеномных технологий, Ставропольский противочумный институт Роспотребнадзора, ул. Советская 13-15, 355012 г. Ставрополь, Российская Федерация, stavnipchi@mail.ru.

Алёна Владимировна Абрамович, младший научный сотрудник лаборатории постгеномных технологий, Ставропольский противочумный институт Роспотребнадзора, ул. Советская 13-15, 355035 г. Ставрополь, Российская Федерация, alena.abramovich.1992@yandex.ru.

Елена Анатольевна Котенева, научный сотрудник лаборатории постгеномных технологий, Ставропольский противочумный институт Роспотребнадзора, ул. Советская 13-15, 355035 г. Ставрополь, Российская Федерация, stavnipchi@mail.ru.