РФФИ (проект № 20-54-53018 ГФЕН_а) и выполнено им в рамках этого проекта.

Список литературы

1. Russel S.J. Oncolytic virotherapy / S.J. Russel, K.W. Peng, J.C. Bell // Nat. Biotechnol. 2012. V. 30. P. 658-672.

2. Adderson E. Safety and immunogenicity of an intranasal Sendai virus-based human parainfluenza virus type 1 vaccine in 3- to 6-year-old children / E. Adderson, K. Branum, R.E. Sealy, B.G. Jones, S.L. Surman, R. Penkert, et al. // Clin. Vaccine. Immunol. 2015. V. 22. P. 298-303.

3. Schirrmacher V. Oncolytic Newcastle disease virus as a prospective anti-cancer therapy. A biological agent with potential to break therapy resistance / V. Schirrmacher // Expert. Opin. Biol. Ther. 2015. V. 15. P. 1757-1771.

4. Zamarin D. Oncolytic Newcastle disease virus for cancer therapy / D. Zamarin, P. Palese // Future Microbiol. 2012. V. 7. P. 347-367.

5. Nguyen L.-T. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies / L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh // Mol. Biol. Evol. 2015. V. 32. P. 268-274.

6. Hoang D.T. UFBoot2: Improving the ultrafast bootstrap approximation / D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh // Mol. Biol. Evol. 2018. V. 35. P. 518-522.

7. Saitou N. The neighbor-joining method: A new method for reconstructing phylogenetic trees / N. Saitou, M. Nei // Mol. Biol. Evol. 1987. V. 4. P. 406-425.

8. Tamura K. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 / K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar // Mol. Biol. Evol. 2013. V. 30. P. 2725-2729.

9. Sorensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content / T. Sorensen // Kongelige Danske Videnskabernes Selskab. Biol. krifter. 1948. V. 4. P. 1-34.

DOI: 10.34617/bche-gz73
УДК 578.81

РЕКОНСТРУКЦИИ ФИЛОГЕНЕТИЧЕСКИХ ОТНОШЕНИЙ МЕЖДУ SV1-РОДСТВЕННЫМИ АКТИНОФАГАМИ

Киселев Сергей Сергеевич[1], канд. биол. наук
Зимин Андрей Антонович[2], канд. биол. наук
Лу Иньхуа[3], PhD Biol. Sci.
Панюков Валерий Васильевич[4], канд. физ.-мат. наук
[1]*Институт биофизики клетки РАН – обособленное подразделение ФИЦ «Пущинский научный центр биологических исследований РАН», г. Пущино, Российская Федерация*
[2]*Институт биохимии и физиологии микроорганизмов им. Г.К. Скрябина РАН – обособленное подразделение ФИЦ «Пущинский научный центр биологических исследований РАН», г. Пущино, Российская Федерация*
[3]*Колледж наук о жизни, Педагогический университет Шанхая, г. Шанхай, Китай*
[4]*Институт математических проблем биологии РАН - филиал ФИЦ «Институт прикладной математики им. М.В. Келдыша РАН», г. Пущино, Российская Федерация*

В данной работе исследована таксономия десяти SV1-подобных актинофагов, хозяевами которых являются бактерии рода *Streptomyces*, с использованием методов

анализа полных геномов. С помощью метода полногеномного выравнивания было показано, что кластер актинофагов SV1, Ididsumtinwong, PapayaSalad, Bioscum и Austintatious представляет собой новую близкородственную группу и может быть предложен в таксономии вирусов как новый род актинофагов. Филогенетическое дерево, построенное при помощи подхода на основании попарного сравнения геномов без использования выравнивания, сопоставимо с деревом, полученным благодаря методу полногеномного выравнивания.

Ключевые слова: *k*-меры; актинофаги; полногеномное выравнивание; методы без использования выравнивания

RECONSTRUCTIONS OF PHILOGENETIC RELATIONSHIPS BETWEEN SV1-RELATED ACTINOPHAGES

Kiselev Sergey Sergeevich[1], PhD Biol. Sci.
Zimin Andrey Antonovich[2], PhD Biol. Sci.
Lu Ynhua[3], PhD Biol. Sci.
Panyukov Valery Vasilievich[4], PhD Phys.-Math. Sci.
[1]*Institute of Cell Biophysics of the Russian Academy of Sciences, Pushchino, Russian Federation*
[2]*G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Russian Federation*
[3]*College of Life Sciences, Shanghai Normal University, Shanghai, China*
[4]*Institute of Mathematical Problems of Biology RAS - the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Russian Federation*

In this work, we studied the taxonomy of ten SV1-like actinophages of bacteria of the genus *Streptomyces*, based on methods for the analysis of complete genomes. Using the whole-genome alignment method, it was shown that the cluster of actinophages SV1, Ididsumtinwong, Papaya Salad, Bioscum and Austintatious is a new closely related group which can be proposed in the taxonomy of viruses as a new genus of actinophages. It is shown that a phylogenetic tree inferred with the approach based on alignment-free pairwise comparison of genomes is comparable to a tree obtained using the whole-genome alignment method.

Key words: *k*-mers; actinophages; whole-genome alignment; alignment-free methods

Бактериофаги являются наиболее распространенными и разнообразными генетическими объектами на планете Земля. Несмотря на это, было высказано предположение, что все фаги с двухцепочечной ДНК совместно используют один и тот же генофонд. Понимание механизмов, с помощью которых фаги генерируют такое огромное разнообразие, все еще остается проблемой, но факторы отбора и адаптации к росту у различных бактериальных хозяев явно оказывают сильное влияние. Актинобактерии, особенно из рода *Streptomyces*, обладают замечательной способностью синтезировать разнообразные вторичные метаболиты всех основных классов. Действительно, 70 % клинически полезных антибиотиков было выделено из *Streptomyces* spp. Генетические инструменты, полученные из фагов *Streptomyces*, были неоценимы в исследованиях, направленных на понимание фундаментальной биологии этих бактерий и на управление путями биосинтеза антибиотиков. Ряд фагов *Streptomyces* относится к семейству Siphoviridae, то есть обладают длинными, гибкими, не сокращающимися хвостами. Мы представляем доказательства того, что кластер актинофагов SV1, Ididsumtinwong, PapayaSalad, Bioscum и Austintatious представляет собой близкородственную группу семейства *Siphoviridae* и может предложен в таксономии вирусов как новый род актино-

фагов. Для бактериофага SV1, основного представителя этой группы актинофагов, показана способность к общей или генерализованной трансдукции [1]. Можно предположить, что все представители новой группы актинофагов обладают подобным свойством.

Методика исследований. Для поиска геномов бактериофагов, гомологичных SV1, использовалась программа BLASTn, при этом поиск осуществлялся только среди геномов хвостатых фагов (tailed phages). Полногеномное выравнивание было выполнено в программе Mauve (версия 2.4.0), оно в дальнейшем использовалось для построения филогенетического дерева с помощью метода максимального правдоподобия в программе IQ-TREE (версия 1.6.12). Оптимальная модель нуклеотидных замен была подобрана автоматически с помощью встроенного в IQ-TREE теста, ей оказалась GTR+G+I. Уровень поддержки ветвей определяли с помощью 2000 итераций «сверхбыстрого бутстреп-анализа» (ultrafast bootstrap). Для получения матрицы попарных расстояний между геномами бактериофагов по используемым *k*-мерам применяли программу собственной разработки GnmDistUsKm. Затем на основе данной матрицы расстояний было построено филогенетическое дерево в программе MEGA 6 с помощью метода присоединения ближайших соседей.

Результаты исследований и их обсуждение. С помощью BLASTn в базе данных NCBI GenBank было обнаружено 9 бактериофагов, гомологичных SV1 (табл. 1).

Таблица 1 – Геномы бактериофагов *Streptomyces*, гомологичные SV1

№ доступа в GenBank	Название	Длина генома, н.п.	GC-состав, %
JX182371.1	SV1	37612	72,69
KY092479.1	Ididsumtinwong	37817	72,37
KY092480.1	Picard	39522	72,6
KY092481.1	PapayaSalad	38411	72,64
KY092482.1	Mojorita	38496	72,54
KY092483.1	Bioscum	37830	72,34
KY092484.1	Raleigh	40785	71,8
KY676784.1	ToastyFinz	39693	72,51
MH825699.1	Darolandstone	40725	72,31
MK305888.1	Austintatious	36213	72,6

Все найденные геномы бактериофагов имеют следующее таксономическое положение: Viruses; царство – dsDNA viruses, класс – no RNA stage; отряд – Caudovirales; семейство – Siphoviridae. Минимальная степень покрытия при этом была равна 54 %. Хозяевами всех фагов были бактерии, являющиеся представителями рода *Streptomyces*. Длина геномов варьировалась в диапазоне от 36213 до 40785 нуклеотидных пар, степень идентичности с геномом фага SV1 по результатам BLASTn составляла от 71,6 % до 89,7 %. Все геномы являются GC-богатыми (GC-состав от 71,8 % до 72,69 %).

Для 10 обнаруженных геномов бактериофагов было сделано полногеномное выравнивание в программе Mauve. Данное выравнивание использовалось для построения филогенетического дерева в программе IQ-TREE с помощью метода максимального правдоподобия [2]. При этом было обнаружено, что оптимальной моделью нуклеотидных замен являлась GTR+G+I с 5 категориями для учёта разных скоростей эволюции и инвариантными сайтами. Полученное филогенетическое дерево показано на рис. 1.

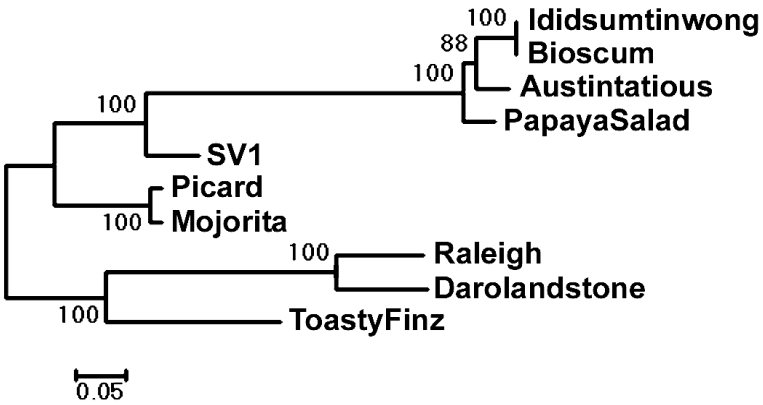

Рисунок 1 – Филогенетическое дерево, построенное методом максимального правдоподобия на основе выравнивания из программы Mauve. Хозяевами всех фагов являются бактерии рода *Streptomyces*. Значения бутстреп-поддержки в процентах от 2000 реплик «ultrafast bootstrap» [3] приведены рядом с узлами ветвления. На масштабной линейке отражено число нуклеотидных замен на позицию.

Второе дерево было построено с помощью подхода, свободного от выравнивания (alignment-free). При этом учитывались все варианты используемых в геномах *k*-меров, т.е. коротких нуклеотидов длиной *k* оснований. Анализировались все *k*-меры со сдвигом в 1 нуклеотид; прямая последовательность и инвертированная комплементарная считались идентичными. Оптимальный размер *k* для данных геномов был равен 9. Для каждой пары геномов было определено сходство по формуле Сёренсена [4] с помощью программы собственной разработки GnmDistUsKm. Затем матрица сходства переводилась в матрицу расстояний по формуле

$$dist = 1 - sim,$$

где *dist* — расстояние между двумя геномами, а *sim* — сходство по Сёренсену. Эта матрица расстояний использовалась на входе алгоритма присоединения ближайших соседей [5], реализованного в MEGA 6 [6]. На рис. 2 представлено полученное филогенетическое дерево.

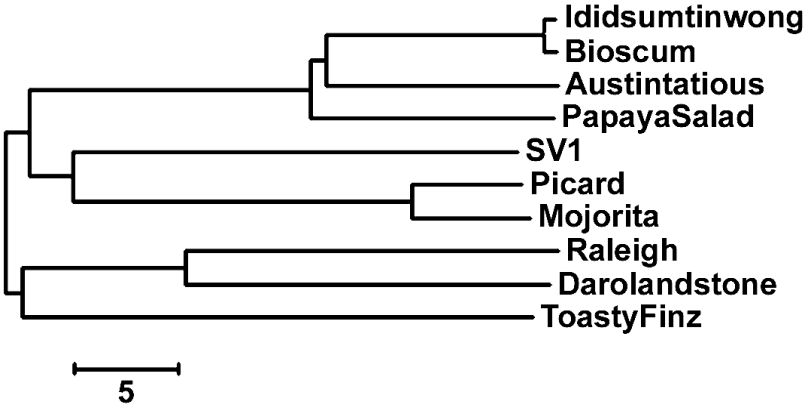

Рисунок 2 – Филогенетическое дерево, построенное с помощью метода присоединения ближайших соседей на основе попарного сравнения геномов актинофагов по используемым 9-мерам. На масштабной линейке отражено различие между геномами в процентах.

Деревья, приведённые на рис. 1 и рис. 2, незначительно отличаются по топологии. Геном фага SV1 попал в группу с 4 другими геномами (Ididsumtinwong, Bioscum, Austintatious, PapayaSalad) на дереве, полученном на основе полногеномного выравнивания с использованием метода максимального правдоподобия (рис. 1), причём уровень бутстреп-поддержки для данного кластера составил 100 %. Необходимо отметить, что группа из 4 указанных геномов удалена по числу накопившихся нуклеотидных замен от SV1 и для неё уровень бутстреп-поддержки тоже равен 100%. Однако фаг SV1 оказался внутри отдельной клады с фагами Picard и Mojorita на дереве, построенном на основе матрицы расстояний по используемым в геномах 9-мерам с помощью метода присоединения ближайших соседей (рис. 2). Оставшаяся клада из 3 геномов (Raleigh, Dorolandstone и ToastyFinz) выделилась на обоих деревьях. В свою очередь, если рассматривать кластеризацию на более крупном уровне, то группа из 7 геномов (Ididsumtinwong, Bioscum, Austintatious, PapayaSalad, SV1, Picard и Mojorita) тоже идентична по своему составу, но отличается по степени близости фага SV1 к той или иной подгруппе. Вполне вероятно, что этот эффект обусловлен небольшим количеством геномов анализируемых фагов, найденных в GenBank.

Выводы. Методы анализа нуклеотидных последовательностей, свободные от выравниваний, в частности, на основе попарных сравнений геномов по используемым *k*-мерам, можно применять для реконструкции филогенетических отношений между организмами.

Работа частично поддержана средствами гранта РФФИ (проект № 18-07-00899). Исследование для Зимина А.А. частично поддержано бюджетом ИБФМ РАН и частично поддержано для него же грантом РФФИ (проект № 20-54-53018 ГФЕН_а) и выполнено им в рамках этого проекта.

Список литературы

1. Vats S. Transductional analysis of chloramphenicol biosynthesis genes in *Streptomyces venezuelae* / S. Vats, C. Stuttard, L.C. Vining // J. Bacteriol. 1987. V. 169. P. 3809-3813.

2. Nguyen L.-T. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies / L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh // Mol. Biol. Evol. 2015. V. 32. P. 268-274.

3. Hoang D.T. UFBoot2: Improving the ultrafast bootstrap approximation / D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh // Mol. Biol. Evol. 2018. V. 35. P. 518-522.

4. Sorensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content / T. Sorensen // Kongelige Danske Videnskabernes Selskab. Biol. krifter. 1948. V. 4. P. 1-34.

5. Saitou N. The neighbor-joining method: A new method for reconstructing phylogenetic trees / N. Saitou, M. Nei // Mol. Biol. Evol. 1987. V. 4. P. 406-425.

6. Tamura K. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 / K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar // Mol. Biol. Evol. 2013. V. 30. P. 2725-2729.