

Супердерева пластид хлорофитной ветви водорослей и животных Ecdysozoa

А.С. Калинина¹, К.В. Михайлов², А.В. Селиверстов¹, В.А. Любецкий¹

¹ИППИ РАН, ²НИИФХБ МГУ

kalinina@iitp.ru, kv.mikhailov@gmail.com

Аннотация

Построены белковые семейства, выравнивания белков и супердерева видов. На этой основе изучена филогенетическая близость пластома паразита *Helicosporidium sp ex Simulium jonesii* к пластомам зеленых водорослей, включая требуксиевые водоросли, и пластомам простейших из родов *Euglena* и *Bigelowiella* (вторичных эндосимбионтов). А также изучена филогенетическая близость класса *Branchiopoda* (жаброногих) к надклассу *Hexapoda*, показана монофилетичность последней группы.

1. Введение

Филогенетическая группа *Helicosporidia* состоит из простейших паразитов беспозвоночных животных. Для них характерна стадия цисты. Нами рассмотрен вид, впервые выделенный из мухи *Simulium jonesi*, который может паразитировать и на других видах. Близким видом является паразитирующая водоросль *Prototheca wickerhamii*. В работах [1-3] изучается филогенетическая близость паразита *Helicosporidium sp ex Simulium jonesii* к зеленым водорослям. В [1] для ядерного рибосомного гена SSU-rDNA построено дерево, согласно которому *Helicosporidium sp* располагается внутри таксона требуксиевых водорослей. Однако в той же работе для конкатенированного выравнивания двух ядерных рибосомных генов SSU-rDNA и LSU-rDNA паразит *Helicosporidium sp* располагается внешней группой относительно этого таксона. По-видимому, вопрос о их точном взаимоотношении остается не ясным на уровне ядерных генов. В работе [4] изучается филогенетическая близость пластиды паразита *Helicosporidium sp ex Simulium jonesii* к пластидам зеленых водорослей и вторичных эндосимбионтов этих водорослей, включая пластиды требуксиевых

водорослей и простейших из рода *Euglena*. Мы также изучаем этот вопрос на уровне пластид.

В [5] по конкатенату из 37 пластоменных генов, кодирующих белки, показано, что пластиды водорослей (простейших) рода *Euglena* ближе к пластидам *Chlorophyceae* и *Trebouxiophyceae*, чем к пластидам *Prasinophyceae*. Наше исследование проведено на основе всех пластоменных белковых семейств и показало, что пластиды *Euglena spp.* близки к пластидам *Helicosporidium sp*. В нашей работе также изучается вопрос о доноре из числа зеленых водорослей пластид для водорослей рода *Euglena*. Нами изучаются этих двух вопросов.

Одновременно на одном конкретном примере (белков, кодируемых в ядерных геномах представителей группы *Ecdysozoa*) рассматривается вопрос о сравнении дерева, построенного по конкатенации белков, и супердерева, построенного по деревьям отдельных белковых выравниваний [6]. Предполагается продолжить исследование этого примера. Сам вопрос требует многих уточнений, и ответ, по-видимому, зависит от вида и числа выравниваний и их качества. Вопрос существенно зависит и от трудной проблемы укоренения дерева генов.

Известно исследование, которое говорит, что *Remipedia* является сестринской группой к группе *Hexapoda*, состоящей из насекомых (*Insecta*) и скрыточелюстных (*Entognatha*). Группа *Hexapoda* входит в группу *Pancrustacea* (насекомых и ракообразных), включенную в кладу членистоногих (*Arthropoda*), которая в свою очередь входит в *Ecdysozoa* (линяющие) – обширную кладу первичноротых животных [6, 7]. Совсем недавно результаты мультигенного анализа показали, что насекомые объединяются с «низшими» жаброногими ракообразными *Branchiopoda* [8], что соответствует и нашему результату.

2. Материалы и методы

Белковые последовательности взяты из пластов двух представителей рода эвглена *Euglena longa*, *Euglena gracilis*, из пласта простейшего *Bigelowiella natans* и из пластов 22-х зеленых водорослей, относящихся к классам Prasinophyceae, Mamiellophyceae, Chlorophyceae, Trebouxiophyceae, Ulvophyceae, включая паразита *Helicosporidium sp.* Это – все белковые последовательности, представленные в базе данных [9]. Из этих последовательностей образованы белковые семейства с помощью программы, описанной в [10]. Получено 131 семейство, состоящее из белков двух и более видов. Для каждого белкового семейства получено множественное выравнивание и построено укорененное дерево [11]. Затем по полученному набору белковых деревьев построено супердерево пластов (рис. 1) программой Super3GL, [12].

Для первых шагов в исследовании вопроса, сформулированного в конце Введения, нами рассмотрен пример, который сам по себе представляет существенный интерес. А именно, нами получены 102 выравнивания аминокислотных последовательностей. Из них взяты 52 более длинных выравнивания, с длинами от 163 до 807 аминокислот, которые относятся к белкам, кодируемым в ядерных геномах беспозвоночных животных. А именно, – к белкам в геномах 90 видов из группы Ecdysozoa и четырех видов, составляющих внешнюю группу (представители Lophotrochozoa): *Capitella*, *Helobdella robusta*, *Aplysia californica* и *Crassostrea gigas*. Эти выравнивания доступны по ссылке <http://lab6.iitp.ru/dl/ask/ecdysozoa52.zip>.

Выравнивания составлены нами из отобранных для целей филогенетического анализа ортологичных белковых последовательностей и последовательностей EST, собранных в контиги и транслированных, которые найдены в доступных базах данных при помощи программы BLAST. Выравнивания подготовлены программой MUSCLE и скорректированы вручную нами. Для каждого выравнивания был проведен предварительный филогенетический анализ с целью выявить и элиминировать как случаи подмены ортологичного белка паралогом, так и случаи контаминации EST библиотек. Конкатенацию индивидуальных выравниваний в единую матрицу проводили с помощью программы SCaFoS.

Для определения матрицы аминокислотных замещений (LG, WAG или JTT) нами использова-

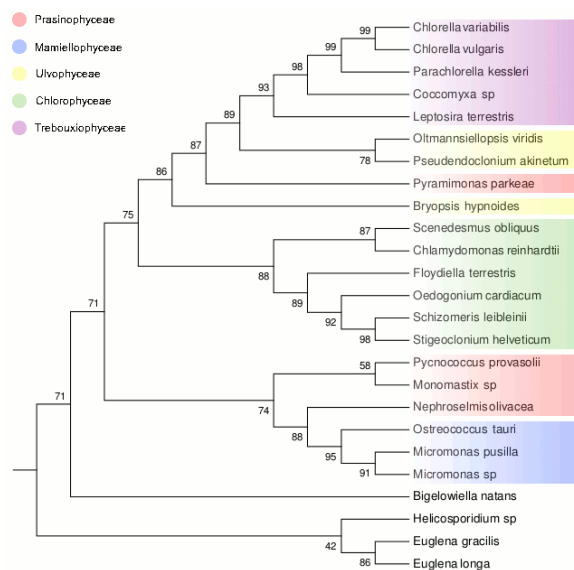


Рис. 1. Дерево пластид хлорофитной ветви, построенное по всем белковым семействам.

лась программа ProtTest, [13]. Для построения дерева белкового семейства использовался пакет PhyloBayes, [14]. Для этого по каждому выравниванию запускалось по 4 цепи с опциями -cat, -dgam 6, -matrix, где matrix – матрица аминокислотных замещений, определенная как сказано выше. Для проверки состояния сходимости цепей использовалась программа burnp из этого пакета со значением параметра burn-in порядка 2000. Счет останавливался при maxdiff < 0.3. Программа burnp выдает дерево белкового семейства, построенное методом major-rule consensus tree и включает в него клады с порогом встречаемости больше 0.5, которое в нашем случае содержало сильно политомические вершины. Если модифицировать программу burnp, опустив условие на порог встречаемости, то получаются в основном бинарные деревья. Полученные деревья укоренялись программой retree из пакета PHYLIP, [15]. По всем полученным деревьям построено супердерево (рис. 2) программой Super3GL, [12].

3. Результаты

На полученном супердереве пластид (рис. 1) выделяются три клады, хорошо соответствующие классам Prasinophyceae вместе с Mamiellophyceae, Chlorophyceae и Trebouxiophyceae вместе с Ulvophyceae, что соответствует филогенетическому дереву из [16]. Виды из класса Chlorophyceae собраны в одну кладу. Виды из Prasinophyceae, Mamiellophyceae и Ulvophyceae

применена также к деревьям генов, полученных на основе выравниваний перечисленных в предыдущем разделе (и доступных по ссылке <http://lab6.iitp.ru/dl/ask/eccdysozoa52.zip>).

В супердереве на рис. 2 правильно определилась внешняя группа, состоящая из четырех видов *Capitella*, *Helobdella robusta*, *Aplysia californica* и *Crassostrea gigas*. Класс Branchiopoda оказался сестринской группой к надклассу Hexapoda, который выделился в качестве монофилетичной группы. Класс Malacostraca не восходит к корню дерева и т.д.

В тоже время дерево на рис. 2, в его высветленной части, содержит существенные артефакты, связанных с трудностью укоренения некоторых из полученных деревьев белковых семейств.

4. Выводы

1. Программа Super3GL успешно применена для построения двух супердеревьев: пластид хлорофитной ветви и беспозвоночных животных из группы Ecdysozoa.

2. Показано, что использование большого числа индивидуальных деревьев, каждое из которых значительно отличается от дерева видов, позволяет строить супердерево, соответствующее дереву видов.

3. Полученные данные позволяют предположить, что донором пластид эвгленовых водорослей является вид, близкий к общему предку *Helicosporidia*. При этом наличие фотосистемы у *Eglena gracilis* свидетельствует о приобретении пластид до потери способности к фотосинтезу как паразитом *Helicosporidium sp.*, так и *Euglena longa*.

4. Получены 102 множественных выравнивания и их деревья. Проведено сравнение деревьев: построенного программой PhyloBayes на основе конкатенации этих выравниваний (данные не приводятся) и построенного программой Super3GL на основе 52 выравниваний и соответствующих деревьев выравниваний. Дерево, построенное на основе конкатената, в лучшей степени согласуется с результатами предыдущих работ.

Авторы глубоко благодарны В.В. Алешину за глубокое обсуждение результатов и методов. Работа выполнена при частичной финансовой поддержке госконтрактами 14.740.11.0624, 14.740.11.1053, 14.740.12.0830 Министерства образования и науки РФ.

Список литературы

- [1] Tartar A, Boucias DG, Adams BJ, Becnel JJ, *Phylogenetic analysis identifies the invertebrate pathogen Helicosporidium sp as a green alga (Chlorophyta)*, Int J Syst Evol Micr 2002, 52:273-279.
- [2] Pombert JF, Keeling PJ, *The Mitochondrial Genome of the Entomoparasitic Green Alga Helicosporidium*, PloS ONE, January 2010, Volume 5, Issue 1.
- [3] Tartar A, Boucias DG, *The non-photosynthetic, pathogenic green alga Helicosporidium sp. has retained a modified, functional plastid genome*, FEMS Microbiology Letters 233 (2004) 153–157.
- [4] de Koning AP, Keeling PJ, *The complete plastid genome sequence of the parasitic green alga Helicosporidium sp. is highly reduced and structured*, BMC Biology, 4:12, 2006.
- [5] Turmel M, Otis C, Lemieux C, *The complete chloroplast DNA sequence of the green alga Nephroselmis olivacea: Insights into the architecture of ancestral chloroplast genomes*, Proc. Natl. Acad. Sci. USA. 96, 10248–10253, 1999.
- [6] von Reumont BM et al., *Pancrustacean Phylogeny in the Light of New Phylogenomic Data: Support for Remipedia as the Possible Sister Group of Hexapoda*, Mol. Biol. Evol., Nov 1, 2011.
- [7] Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA, *Evidence for a clade of nematodes, arthropods and other moulting animals*, Nature, 1997, vol. 387, pp. 489-493.
- [8] В.В. Алешин и др. *О положении насекомых в классе Pancrustacea*, Молекулярная биология, 2009, 43(5), 866-881.
- [9] Зверков О.А., Селиверстов А.В., Любецкий В.А., *Семейства белков, кодируемых в пластомах Chlorophyta, Euglenozoa и Rhizaria*, Информационные технологии и системы – 2012.
- [10] Зверков О.А., Селиверстов А.В., Любецкий В.А., *Белковые семейства, специфичные для пластомав небольших таксономических групп водорослей и простейших*, Молекулярная биология, 2012 (в печати).
- [11] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG, *Clustal W and Clustal X version 2.0.*, Bioinformatics, 2007, 23, 2947-2948.
- [12] Горбунов К.Ю., Любецкий В.А., *Быстрый алгоритм построения супердеревья видов по набору белковых деревьев*, Молекулярная биология, 2012, том 46, № 1, 176–183.
- [13] Abascal F, Zardoya R, Posada D, *ProtTest: Selection of best-fit models of protein evolution*, Bioinformatics:21(9):2104-2105, 2005.
- [14] Lartillot N, Philippe H, *A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process*, Molecular Biology and Evolution. 2004 21(6):1095-1109.
- [15] Felsenstein J, *PHYLP - Phylogeny Inference Package (Version 3.2)*, Cladistics 5:164-166, 1989.

- [16] Fang Lu, Wei Xu, Chao Tian, Guangce Wang, Jiangfeng Niu, Guanghai Pan, Songnian Hu, *The Bryopsis hypnoides Plastid Genome: Multimeric Forms and Complete Nucleotide Sequence*, PLoS ONE, Feb 2011, Volume 6, Issue 2.
- [17] Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI, *Complete nucleotide sequence of the chlorarachniophyte nucleomorph: Nature's smallest nucleus*, Proc. Natl. Acad. Sci. U.S.A. 103(25), 9566–9571, 2006.
- [18] Turmel M, Gagnon M, O'Kelly CJ, Otis C, Lemieux C, *The Chloroplast Genomes of the Green Algae Pyramimonas, Monomastix, and Pycnococcus Shed New light on the Evolutionary History of Prasinophytes and the Origin of the Secondary Chloroplasts of Euglenids*, Mol Biol Evol (2009) 26 (3): 631-648, 2009.
- [19] Rubanov LI, Gorbunov KYu, Lyubetsky VA, *Super3GL: a cubic time mathematically rigorous reconciliation of large tree sets*, In Silico Biology, 2012 (представлена).