

Минимизация ошибок реконструкции филогении программой FastME

Научный руководитель – Спирин Сергей Александрович

Панова Вера Викторовна

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия

E-mail: nooroka17@gmail.com

Использование программ реконструкции филогении с различными опциями зачастую приводит к разным результатам. Разработка метода, позволяющего наиболее правильно реконструировать дерево, является актуальной задачей.

Одним из способов поиска наилучшего филогенетического дерева является выращивание: находится лучшее дерево для части последовательностей, затем к этому дереву добавляются листья по одному, каждый раз в наилучшее место [2]. При этом результат выращивания зависит от порядка добавления последовательностей. Для улучшения качества реконструкции в программе FastME [2] предусмотрена возможность оптимизировать выращенное филогенетическое дерево, используя локальные перестройки топологии — так называемые Nearest Neighbor Interchange (NNI) [1].

Мы рассмотрели другой способ улучшения качества реконструкции филогенетического дерева. Выращивание повторяется много раз с разным порядком добавления листьев, а далее берется неразрешенное дерево — консенсус результата выращивания. Был написан комплекс программ на языках Python и Bash, позволяющий реконструировать филогенетическое дерево по заданному выравниванию последовательностей белков с использованием этого подхода, а также протестировать качество реконструкции при каждом из двух подходов. Создан также веб-интерфейс к программе реконструкции филогении.

С целью оценки уровня ошибок дерева, реконструируемые по ортологическим рядам белков, сравнивались с эталонным деревом видов. В качестве критерия сравнения использовалось «симметрическое расстояние Робинсона – Фоулдса» — количество ветвей (разбиений множества таксонов на два), которые присутствуют в одном дереве, но отсутствуют в другом. Кроме того, сравнивались два критерия качества филогенетического дерева, реализованных в программе FastME, а именно: BME (balanced minimum evolution, метод сбалансированной минимальной эволюции) и OLS (ordinary least squares, метод наименьших квадратов). В качестве наборов для тестирования использовались наборы выравниваний ортологических рядов из 25 животных (Metazoa) — 970 выравниваний, 45 грибов (Fungi) — 829 выравниваний и 45 протеобактерий (Proteobacteria) — 780 выравниваний. Тестирование проводилось как на полных выравниваниях, так и на выборках из 15 и (кроме животных) 30 последовательностей.

Показано, что в большинстве случаев в консенсусном дереве меньше ошибок, чем в оптимизированном дереве. Результаты при применении критериев OLS и BME близки, хотя результаты OLS немного хуже.

Работа поддержана грантом РНФ 16-14-10319. Автор выражает благодарность научному руководителю С.А.Спирину.

Источники и литература

- 1) Chernomor O., Minh B.Q., von Haeseler A. Consequences of Common Topological Rearrangements for Partition Trees in Phylogenomic Inference // Journal of Computational Biology. 2015. 22(12). Pp. 1129–1142.

- 2) Desper R. and Gascuel O. Fast and Accurate Phylogeny Reconstruction Algorithms Based on the Minimum-Evolution Principle // Journal of Computational Biology. 2002. 9(5). Pp. 687–705.