

Сравнительный анализ бактериальных пангеномов

Молдован Михаил Александрович

Студент (специалист)

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

E-mail: mika.moldovan@gmail.com

Пангеномом называется совокупность всех генов вида, рода или более крупной таксономической единицы. Условно пангеном можно разделить на четыре части: универсальные гены, входящие в геномы всех организмов, по которым строится пангеном, часто встречающиеся гены, которыми обладает более половины организмов из выборки, редко встречающиеся гены, входящие в состав геномов менее половины организмов и синглтоны, которые наблюдаются в составе генома лишь какого-либо одного организма. Целью работы является выяснение зависимости количественных и качественных характеристик пангеномов от длины генома.

Для построения пангеномов использовалась процедура BLAST всех-против-всех с последующим поиском связанных компонент в полученном графе гомологов. Поправка на таксономию осуществлялась с помощью метода независимых филогенетических контрастов.

Были построены пангеномы 115 видов, для которых имеются геномы как минимум 16 представителей. Для всех пангеномов были построены зависимости количества генов разных частей пангенома от длины генома. Оказалось, что количества генов из всех частей пангенома возрастает с увеличением длины генома как степенные функции, при этом доля редких генов возрастает, а доля частых — убывает. Далее были построены зависимости числа паралогов в разных частях пангенома от длины генома. Количество паралогов среди универсальных генов квадратично зависит от длины генома, в то время, как среди редких генов зависит от длины генома в степени 1.2, а количество паралогов среди частых генов зависит от длины генома в степени 0.8.

Количество редких генов отражает, с одной стороны, скорость приобретения генов, а с другой - скорость их удаления из генома вычищающим отбором, оба процесса физически от длины генома не зависят. В то же время, количество частых генов отражает скорее скорость нейтральной потери генов предкового штамма и является величиной, физически зависящей от длины генома. Полученные результаты позволяют говорить о том, что вычищающий отбор действует сильнее на малые геномы, а для больших геномов характерна меньшая скорость потери как недавно приобретенных генов, так и генов предкового штамма, что, вместе с фактом увеличения доли паралогов среди редких генов в больших геномах, говорит о тренде в сторону усложнения генома из-за ослабления влияния вычищающего отбора.