

Эволюция бактериальных пангеномов

Е.Н.Гордиенко

Институт общей генетики им. Н.И.Вавилова РАН, Москва, Россия
egordienko@gmail.com

Аннотация

С увеличением количества секвенированных бактериальных геномов появляется возможность проводить сравнения групп штаммов друг с другом. При этом обнаруживается, что состав генома сильно варьирует даже внутри вида бактерий. Проведены попарное сравнение состава генома видов *Escherichia coli* (включая штаммы *Shigella*) и *Salmonella enterica* и функциональная категоризация наборов генов, характерных для каждой филогенетической группы. На основе полногеномного анализа деревьев генов были детектированы горизонтальные переносы, которые имели место после расхождения данных видов от общего предка, предложена гипотеза эволюции состава генома при разделении таксономических групп *Enterobacteriaceae*.

1. Введение

Анализ полностью секвенированных геномов бактерий указывает на большое разнообразие их размера и состава. Сравнение геномных последовательностей различных штаммов одного вида бактерий показывает, что варибельность состава генома обнаруживается даже внутри одного бактериального вида. Например, ранее было обнаружено, что только 39,2% генов являются общими для геномов трех штаммов кишечной палочки [5].

Для описания генома вида на основании геномных последовательностей его штаммов была предложена концепция пангенома [3, 4]. В этой концепции пангеном – это суммарный набор генов всех штаммов одного бактериального вида (рода). Пангеном разделяется на три части: универсальный геном (гены, которые встречаются во всех штаммах таксона), геном периферии (состоит из генов, которые отсутствуют в одном или более штаммах) и уникальные (штамм-специфичные) гены [1,3].

2. Результаты и обсуждение

В работе проведен анализ пангенома 48 штаммов семейства *Enterobacteriaceae* – использовано 25 геномов *Escherichia coli*, 7 геномов *Shigella* spp. и 16 геномов штаммов *Salmonella enterica*.

На первом этапе работы по аминокислотным последовательностям были построены таблицы ортологов для 48 штаммов.

Предложен метод аппроксимации распределения генов по встречаемости в штаммах как суммы двух степенных функций, его преимущества над ранее использовавшимися методами подтверждены статистически.

Сравнение наборов генов *E.coli* и *Shigella* spp. не позволяет обнаружить наборов генов, специфичных для той или иной группы. Также обнаруживается лишь небольшое количество периферических генов *E.coli*, отсутствующих в штаммах рода *Shigella*. Это подтверждает недавнее разделение *E.coli* и *Shigella* spp., обнаруженное в предыдущих исследованиях [2]. Можно говорить о существовании общего пангенома штаммов *E.coli* и *Shigella* spp. и провести его сравнение с пангеномом *S.enterica*.

Такое сравнение показывает, что каждая из этих групп имеет набор генов, специфичных только для данной группы. Для *E.coli+Shigella* spp. этот набор содержит 178 генов, большая часть которых связана с транспортом молекул. Среди 479 генов, специфичных только для *S. enterica*, перепредставлены следующие функциональные категории: патогенез и биосинтез кобаламина.

Среди генов периферии преобладают приближенные к пикам, например, присутствующие лишь в нескольких штаммах. Второй вариант — широко распространенные гены периферии, которые утрачены только в единичных случаях. Генов периферии, которые присутствуют в нескольких, но не в подавляющем большинстве или меньшинстве, штаммов в обеих таксономических группах, относительно мало. Функциональный анализ общей периферии практически не выявляет

статистически значимых перепредставленных категорий. При этом разделение общей периферии на подобласти редких и распространенных генов (аналогично предложенному методу аппроксимации) позволяет выявить в обеих подобластях функциональные категории. Иными словами, общая периферия естественным образом разделяется на две области – с редкими генами и с распространенными генами.

Анализ филогенетических деревьев подтверждает большое количество горизонтальных переносов среди редких генов общей периферии.

Суммируя полученные данные, можно представить общую периферию пангеномов *E.coli* и *S.enterica* как гетерогенную группу, которая состоит как из вертикально распространяемых генов, предположительно унаследованных от периферии общего предка, так и из генов, недавно перенесенных в группу и распространяемых горизонтально.

3. Благодарности

Автор благодарит М.С.Гельфанда за научное

руководство и М.Д.Казанова за содействие в построении ортологических рядов. Также автор благодарен Ю.И.Вульффу за плодотворные дискуссии.

4. Литература

- [1] Medini, D., et al., “Microbiology in the post-genomic era”, *Nat Rev Microbiol*, 2008. 6(6): p. 419-30.
- [2] Pupo, G.M., R. Lan, and P.R. Reeves, “Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics”, *Proc Natl Acad Sci U S A*, 2000. 97(19): p. 10567-72.
- [3] Tettelin, H., et al., “Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”, *Proc Natl Acad Sci U S A*, 2005. 102(39): p. 13950-5.
- [4] Tetz, V.V., “The pangenome concept: a unifying view of genetic information”, *Med Sci Monit*, 2005. 11(7): p. NY24-9.
- [5] Welch, R.A., et al., “Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*”, *Proc Natl Acad Sci U S A*, 2002. 99(26): p. 17020-4.