

Эволюционное пространство гена 16S рРНК: *a simplex problem*

Евгений Андронов

Лаборатория микробиологического мониторинга ГНУ ВНИИСХМ

Современный этап развития биологической науки характеризуется своеобразным «глобализмом» – переходом от характеристики отдельных объектов и явлений к описанию общей совокупности генов, белков и других соединений у всех существ, живущих на Земле. Одним из таких стремительно развивающихся направлений является метагеномика, оперирующая массивами нуклеотидных последовательностей, принадлежащих сообществам организмов в данной среде обитания. Размеры метагеномов не просто велики, а громадны и, несмотря на бурное развитие технологий высокопроизводительного секвенирования, все еще представляют собой объекты, недоступные для анализа во всей полноте. Так, например, размер метагенома 1 г почвы (преимущественно бактериального происхождения) может достигать 10^{16} пн, что соответствует примерно 10^7 человеческих геномов. Очевидно, что для исчерпывающего секвенирования (в разумные сроки) этого мельчайшего фрагмента глобального метагенома может не хватить всего доступного на сегодняшний день мирового парка секвенаторов. Однако гораздо более серьезную проблему, чем технологии секвенирования метагеномов, представляет собой анализ и интерпретация полученных данных, создание биологических, эволюционно осмысленных построений.

В настоящем докладе будут рассмотрены непростые, отчасти космологические проблемы, которые встречают исследователя в одном из разделов метагеномики – анализе таксономической структуры микробиомов по данным глубокого секвенирования библиотек гена 16S рРНК. Особый интерес, который вызывает этот популярный таксономический маркер, объясняется следующими обстоятельствами: его универсальностью (наличие в любом геноме, кроме вирусного), наличие колоссальной мировой базы данных по разнообразию этого гена (не исключено, что это ген-чемпион по представленности в базах данных) и, наконец, тем, что это – один из возможных кандидатов в родственники первых прагенов «мира РНК».

Основной темой доклада будут проблемы и первые результаты построения так называемого «эволюционного пространства» гена 16S рРНК. Определение эволюционного пространства гена весьма лапидарно: «Эволюционное пространство гена – это метрическое пространство, в котором все возможные варианты гена представлены

точками, а расстояния между точками соответствуют эволюционным дистанциям между вариантами». Данная концепция не нова, идеологически она восходит к проблеме «естественной системы организмов», о которой писал еще Карл Линней в 1783 году. Одно из наиболее ясных определений естественной системы цитирует Чарльз Дарвин в «Происхождении видов»: естественная система – это «схема, по которой соединяются более сходные организмы и разделяются наиболее несходные». Очевидно, что «эволюционное пространство гена» имеет ясное сходство именно с этой трактовкой естественной системы и представляет собой в некотором смысле естественную систему гена, которая отличается от всех остальных классификаций именно системностью, т.е. тем, что в ней есть место для любого варианта гена 16S рРНК – существовавшего, существующего или потенциально возможного, но (еще) не реализованного в ходе эволюции.

Как нам представляется, особый интерес представляет отображение в эволюционном пространстве эволюционных процессов, а именно, эволюции монофилетических бактериальных (архейных) таксонов. В ходе таких эволюционных процессов должна реализовываться схема радиального расширения от общего предкового аллеля (рис. 1).

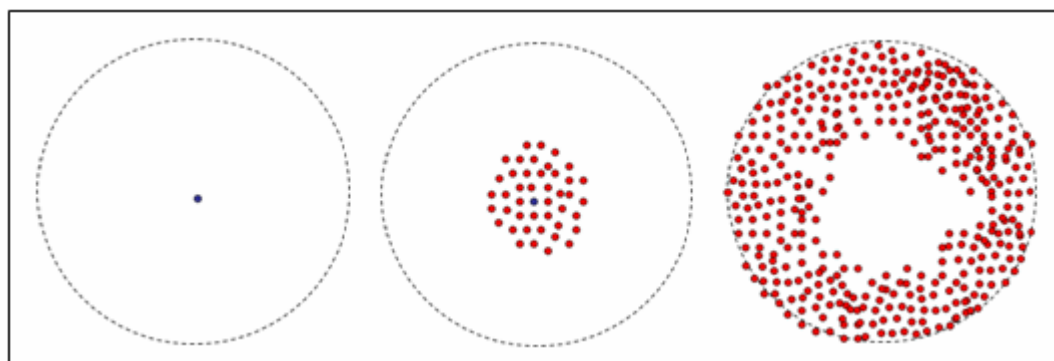


Рис. 1. Три стадии эволюции бактериального монофилетического таксона, представленные в эволюционном пространстве.

Важными особенностями такого процесса должны являться 1) радиальность и необратимость расширения; 2) наличие внешнего функционального предела расширения (не бесспорно) и «сплющивание» таксона в ходе эволюции при приближении к этим границам с изменением типа отбора от нейтрального к отсекающему (также не бесспорно); 3) образование «эволюционных полостей» за счет вымывания предковых

аллелей. Конечно, представленная схема весьма груба и далека от реальности, так как не учитывает ни разной скорости эволюции в различных группах организмов, ни действия факторов селекции. Несмотря на это, такие базовые характеристики, как радиальность, необратимость и наличие полостей должны выявляться экспериментально. Почему же, при достаточной по нашему мнению очевидности таких построений, ни один из этих паттернов до сих пор не был идентифицирован? Ведь детекция эволюционных полостей – фактически «мертвых дыр» эволюционного космоса, т.е. участков эволюционного пространства, которые уже никогда и ничем не будут заполнены, могли бы представлять громадный интерес для науки, так как имеют прямое отношение к вопросу происхождения жизни на Земле, так как открывают путь к идентификации предковых вариантов гена, позволяют определить возраст «ископаемых» ДНК и т.д. Ответ, как нам представляется, довольно прост: размерность пространства, в котором развиваются все эти события, чрезвычайно велика. Это обстоятельство делает *принципиально невозможным* наблюдение указанных паттернов при использовании традиционных эволюционных и филогенетических подходов. По нашим оценкам, размерность «истинного эволюционного пространства» гена 16S рРНК, возможно, составляет более 100 000 измерений, и, по всей видимости, оно пока недоступно для анализа. В то же время, возможно, что основные топологические особенности могут раскрыться в пространствах с меньшей размерностью. С этой целью нами были предприняты вычисления, в ходе которых был проанализирован один из последних релизов базы данных по гену 16S рРНК. В этих вычислениях мы опирались на оказавшееся очень эффективным как для поиска путей разрешения данной проблемы, так и просто для ее понимания, понятие о «симплексах» – многомерных обобщениях треугольника. В ходе вычислений было показано, что минимальная размерность пространства, в которое возможно «вложение» матрицы попарных эволюционных дистанций последнего релиза базы данных по гену 16S рРНК, равно 13. Нами была предпринята первая попытка создания 13-мерной модели эволюционного пространства гена 16S рРНК, результаты которой будут продемонстрированы в ходе доклада.

В предлагаемом докладе будут обсуждены полученные результаты и основные трудности, связанные с эволюционными, геометрическими и вычислительными проблемами таких построений. Будет продемонстрировано, что в данной области исследований весьма заметны феномены, сходные с классическими соотношениями неопределенностей. Также будут обрисованы чрезвычайно интересные перспективы данного подхода, как в вопросах происхождения жизни, общей эволюционной биологии, так и частных ее областях, таких как экологическая метагеномика.

Мы приглашаем принять участие в семинаре (обсуждении и дискуссии) биологов, математиков, программистов – всех интересующихся вопросами эволюции микробиомов, экологии, метагеномики. Кроме того, мы надеемся найти среди слушателей единомышленников, которые захотят присоединиться к практическим исследованиям в данной области на базе отлично оборудованного отделения геномных технологий ЦКП ГНУ ВНИИСХМ (г. Пушкин).