

УДК 574.3 + 579.26

К ВОПРОСУ ИССЛЕДОВАНИЯ МЕТАГЕНОМНОЙ ДНК ИЗ ОБРАЗЦОВ ОКРУЖАЮЩЕЙ СРЕДЫ

¹Наумова Н.Б., ²Загребельный С.Н.

¹*Институт почвоведения и агрохимии СО РАН, Новосибирск, e-mail: nnaumova@mail.ru;*

²*Новосибирский государственный университет, Новосибирск*

В настоящее время основным методом оценки обилия разнообразных микроорганизмов в сообществах окружающей среды является определение разнообразия последовательностей определенных генов метагеномной ДНК путем секвенирования. Доступность секвенаторов и быстрое снижение стоимости анализа привели к экспоненциальному росту таких проектов и генерируемых ими данных. В связи с этим мы сочли необходимым привлечь внимание к нескольким моментам, принципиально важным для экологического осмысления метагеномных данных, а именно выбору образцов и стратегии их отбора, степени изученности местообитания и сбора соответствующих метаданных, установке соответствия между разнообразием последовательностей и разнообразием организмов.

Ключевые слова: метагеномная ДНК, сообщество микроорганизмов, биоразнообразие, образцы окружающей среды, стратегия отбора образцов

ABOUT SOME ASPECTS OF METAGENOMIC DNA STUDIES FROM ENVIRONMENTAL SAMPLES

¹Naumova N.B., ²Zagrebely S.N.

¹*Institute of Soil Science and Agrochemistry SBRAS, Novosibirsk, e-mail: nnaumova@mail.ru;*

²*Novosibirsk State University, Novosibirsk*

Currently the estimation of certain gene sequence abundance in metagenomic DNA from environmental samples is the mainstream tool to measure biodiversity of microbial communities. The increasing availability of DNA sequencers and decreasing analysis cost have resulted in exponential increase in the number of metagenomic research projects and the respective data. So the article focuses on some aspects that are critically significant for ecologically sensible interpretation of metagenomic data, namely the choice of samples and the strategy of their collection, availability of ecologically relevant information about the habitat and importance of metadata, the molecular sequence abundance and microbial community, i.e. organismal, abundance.

Keywords: metagenomic DNA, microbial community, biodiversity, environmental samples, sampling strategy

Применение методов быстрого и эффективного секвенирования и сборки последовательностей больших фрагментов ДНК вызвало революцию в геномике микроорганизмов, приведя к расшифровке десятков тысяч геномов. В большинстве таких проектов исходным материалом является ДНК, извлеченная из чистых лабораторных культур или микроорганизмов, живущих в ассоциации с животными или растительными клетками, с целью получения полной или почти полной последовательности генома исследуемого организма для ее последующего аннотирования и анализа.

Очевидно, что геномные исследования микроорганизмов отличаются от метагеномных исследований как минимум несколькими важными аспектами. По определению, исходным материалом в метагеномных проектах является смесь фрагментов ДНК, экстрагированной из сообщества микроорганизмов разных надцарств – бактерий, архей, эукариот и вирусов, живущих в определенном местообитании. Большинство микроорганизмов не поддаются культивированию в лабораторных условиях существующими методами. Кроме того, в исследуемом об-

разце может быть низкая численность микроорганизмов, а, следовательно, и низкая концентрация ДНК, и/или могут присутствовать вещества, затрудняющие экстракцию. Все это сильно осложняет получение геномной последовательности в метагеномных проектах. Однако целью исследований в таких проектах является оценка структуры и функционирования сообщества, и знание нуклеотидной последовательности всего генома для этого не нужно. Часто при исследовании сложных сообществ нужно ответить на вопрос, сколько последовательностей достаточно для сравнительной оценки биоразнообразия и видовой или функциональной структуры сообществ различных местообитаний, как правило, по градиенту какого-либо естественного или искусственного экологического фактора. Для этого анализируют разнообразие последовательностей фрагментов генов 16S рРНК и/или функциональных генов (например, *nif*), что дает огромное количество информации о видовом и функциональном составе микробных сообществ в окружающей среде.

Доступность секвенаторов и быстрое снижение стоимости секвенирования при-

вели к экспоненциальному росту проектов, основанных на секвенировании, и генерируемых ими данных. Однако, это привело и к тому, что в последнее время изучением микробных сообществ – собственно как сообществ! – занимаются изначально весьма далекие от экологической методологии специалисты. Именно в связи с этим мы сочли необходимым подчеркнуть несколько моментов, важных в плане интерпретации данных для эколого-микробиологических исследований. Строго говоря, эти моменты самоочевидны и/или хорошо известны, но почему-то многие исследователи упускают их из виду, занимаясь анализом разнообразия нуклеотидных последовательностей метагенома.

Первым важным моментом, к которому хотелось бы привлечь внимание, является выбор объектов окружающей среды для метагеномного анализа. Как известно, состав любого сообщества сильно зависит от условий его местообитания. Детальные знания об условиях местообитания являются весьма существенными для экологического осмысления полученных сведений о разнообразии нуклеотидных последовательностей метагенома.

Понятно, что выбор микробного сообщества (объекта) для изучения и, следовательно, образцов окружающей среды, которые нужно собрать, определяется тем научным вопросом, который исследователь ставит перед собой, т. е. целью и задачами исследования. Очевидно, однако, что чем больше самой разнообразной информации имеется об изучаемом местообитании – его физических, химических, экологических и других характеристиках, тем более содержательно можно будет интерпретировать полученные метагеномные данные [6]. Можно будет формулировать конкретные гипотезы и искать функциональные гены в геномных последовательностях сообщества хорошо изученного местообитания. Уже хрестоматийным примером является изучение микробного сообщества кислых дренажных вод металлодобывающих предприятий [9]: оценка баланса азота в изучаемом местообитании позволила выдвинуть гипотезу о его фиксации, а поиск соответствующих генов – найти оперон *nif* у одного из самых малочисленных микроорганизмов в сообществе. То есть исключительно благодаря наличию разнообразных сведений об изучаемом местообитании был выявлен очень интересный микроорганизм, чья значимость для устойчивой жизнедеятельности сообщества намного превышает его относительную численность. Поэтому для повышения эффективности метагеномных

исследований очень важно выбирать местообитания, которые еще будут либо уже детально изучены другими методами, т.е. планировать работу в составе междисциплинарных команд, а не как хобби для генетиков просто по причине незаполненности имеющихся секвенаторов и относительной дороговизны одного запуска этого прибора.

Детально изученные экосистемы часто сложно или даже невозможно найти. Тогда нужно собрать максимум метаданных [4], т.е. данных о данных, представляющих собой описание конкретного места отбора образцов и в целом той экосистемы, где они были отобраны. Важность метаданных для анализа и интерпретации метагеномных данных о последовательностях ДНК определяется двумя причинами. Во-первых, только полное описание образцов, для которых были получены метагеномные данные, обеспечивает возможность повторения исследования. Конечно, в строгом смысле образец окружающей среды или других биологических источников никогда нельзя повторить. Именно поэтому сложно переоценить важность детального описания образца. Во-вторых, без метаданных метагеномные данные не имеют экологического контекста, а выводы на основе анализа и интерпретации таких данных не имеют никакой экологической ценности.

Ясно, что метаданные зависят от типа метагеномного образца: так, метаданные для описания образца из кишечника человека будут значительно отличаться от метаданных, характеризующих образец донных отложений или почвы. Описание любых образцов окружающей среды должно включать, как минимум, такие характеристики¹, как детальное трехмерное географическое месторасположение участка отбора, в том числе глубину (для почв, донных отложений, водных экосистем) и высоту (для наземных или воздушных образцов), самое общее описание среды/экосистемы, откуда был отобран образец (озеро, почва, человек или насекомое, и т.п.); конкретные химические (рН, концентрация солей, и т.п.) и физические (температура, освещенность и т.п.) свойства среды, дату отбора, состояние организма-хозяина, его диету; подробное описание собственно метода отбора, размера отобранного образца, условий его обработки и хранения, длительность последнего до подготовки к анализу.

Некоторые из этих данных обычно получают или регистрируют во время отбора образца, другие доступны из геоинформа-

¹ Смотрите, например, <http://icomm.mbl.edu/microbis>, <http://darwin.nox.ac.uk/gsc/gcat>.

ционных, погодно-климатических, океанических и других баз данных. Инфраструктура метаданных должна существовать в каждой лаборатории; и при этом желательно опираться на соответствующие национальные и международные стандарты. При неимении таковых лаборатории должны поддерживать и вести собственные максимально детальные реестры метагеномных образцов.

Релевантность, воспроизводимость, повторяемость и значимость результатов любых экологических исследований практически полностью зависит от стратегии отбора образцов, который почти всегда сопряжен с серьезными проблемами. Каждое решение о типе, размахе, объеме выборки, числе индивидуальных повторностей, сроках отбора влияет на выводы и заключения, которые можно будет сделать по окончании работы при интерпретации данных. Трудоемкость получения и анализа метагеномных данных значительно обостряет связанные с отбором образцов проблемы, присущие всем экологическим исследованиям. Если по окончании работы нужно будет сделать выводы о микробном сообществе местообитания, то образцы должны адекватно представлять это местообитание, т.е. быть репрезентативными. Для получения репрезентативных образцов очень важно знать вариабельность условий среды данного местообитания. Сообщества почвенных микроорганизмов, например, изменяются на расстояниях порядка микрометра, в соответствии с физической и химической гетерогенностью минеральной и биологической матрицы почвы. Почвенный агрегат объемом 1 см³ может содержать аэробные и анаэробные зоны; частицы разного размера; растительный материал на различных стадиях разложения, а также множество беспозвоночных организмов, каждый из которых с большой вероятностью обладает собственной ассоциированной микрофлорой. Какого размера образец будет репрезентативным для такого местообитания? Можно ли учесть, да и нужно ли учитывать все микрзоны?

Изменение микробного сообщества с течением времени является одним из самых интересных аспектов изучения сообществ, так как реакция на изменение экологических условий является очень важным моментом для понимания структуры, функционирования и устойчивости сообщества. Периодический отбор образцов в одном и том же местообитании нужен для изучения динамики микробных сообществ в развитии и обеспечении экосистемных функций почв, здоровья организма-хозяина и т.п.

Также нужно принимать решение о том, каким образом следует учитывать различные типы изменений – например, естественные циклы и катастрофические события. Долгосрочные изменения типа изменения климата представляют еще больше проблем с точки зрения выработки стратегии отбора образцов для метагеномного анализа: такого рода изменения не только сами влияют на микробные сообщества, но и микробные сообщества, в свою очередь, влияют на них. Каков объем выборки и, следовательно, объем работы, необходимый для того, чтобы выявить такого рода изменения на фоне обычной вариабельности?

Непростой задачей является и определение пространственной вариабельности сообществ. Необходимо оценить, сколько образцов нужно отобрать и проанализировать, чтобы полученная выборка представляла все разнообразие условий обитания сообщества. Однако многие исследователи делают выводы об особенностях видового разнообразия бактериальных сообществ на основе однократного анализа одного отдельного образца, без отбора индивидуальных повторностей образцов конкретного объекта окружающей среды [1, 2, 7].

Ответы на эти вопросы зависят от многих факторов, в том числе сложности микробного сообщества, изменчивости местообитания в пространстве и во времени и необходимой детальности различий. Стремительное повышение эффективности биологических, биоинформационных и компьютерных методов анализа увеличивает надежность выводов о более сложных сообществах в более вариабельных местообитаниях, однако ни в коем случае не отменяет – а даже наоборот, повышает! – значение тщательного обдумывания стратегии отбора образцов и связанные с ним проблемы и ограничения, выработывая в итоге такую схему, которая бы существенно облегчала интерпретацию данных планируемого метагеномного анализа.

Второй, не менее важный, на наш взгляд, момент, касается числа копий генов 16S рРНК в водном геноме: часто исследователи ставят знак равенства между разнообразием филогенетически значимых фрагментов генов 16S рРНК прокариотических одноклеточных организмов и собственно биоразнообразием сообщества. Последнее обычно характеризуют списком видов (или разнообразием таксонов другого уровня) и обилием организмов разных таксонов. И если о списке таксонов корректно говорить на основании списка различных на соответствующем уровне нуклеотидных последовательностей, то о доминантной

структуре сообщества судить на основании относительного обилия разных последовательностей ДНК нельзя, поскольку на один геном (=один организм) может приходиться несколько копий генов 16S рРНК [5, 8]. В случае микроскопических грибов, т.е. мицелиальных многоклеточных организмов, разнообразии последовательностей 18S рРНК точно так же нельзя рассматривать как эквивалент разнообразия особей. Поэтому без применения методов учета числа копий генов на геном [5, 3] следует говорить об относительном обилии/доминировании/структуре последовательностей ДНК сообщества, а не об относительном обилии и доминировании особей разных таксонов в сообществе. Несмотря на это, многие исследователи часто ставят знак равенства между разнообразием генов 16S рРНК и биоразнообразием сообщества бактерий [1, 2, 10].

Заключение

Изучение любого свойства объектов окружающей нас среды по определению в первую очередь подразумевает отбор образцов и последующий анализ индивидуальных повторностей для характеристики пространственной вариабельности изучаемого свойства данной среды, или, в крайнем случае, анализ усредненного из нескольких индивидуальных повторностей образца. Только соблюдение этого правила, справедливого и для изучения биоразнообразия микробных сообществ молекулярными методами, совместно с наличием других сведений об изучаемом объекте окружающей среды и корректной трансляцией молекулярного разнообразия в видовое получаемые результаты будут иметь экологический смысл.

Список литературы

1. Ломакина А.В., Погодаева Т.В., Морозов И.В., Земская Т.И. Микробные сообщества зоны разгрузки газонетесодержащих флюидов ультрапресного озера Байкал // Микробиология. – 2014. – Т. 83. – С. 355–365.
2. Турова Т.П., Гречникова М.А., Кузнецов Б.Б., Сорокин Д.Ю. Исследование филогенетического разнообразия бактерий в стратифицированных осадках содовых озер // Микробиология. – 2014. – Т. 83. – С. 730–742.
3. DeAngelis K.M., Pold G., Topsuoglu B.D., van Diepen L.A., Varney R.M., Blanchard J.L., Melillo J., Frey S.D. Long-term forest soil warming alters microbial communities in temperate forest soils // *Frontiers in Microbiology*. 2015. 6. 104. <http://dx.doi.org/10.3389/fmicb.2015.00104>.
4. Gray J., Liu D. T., Nieto-Santisteban M., Szalay A.S., DeWitt D., Heber G. *Scientific Data Management in the Coming Decade*. Redmond, WA: Microsoft Corporation, 2005.
5. Kempbell S.W., Wu M., Eisen J.A., Green J.L. Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance // *PLOS Computational Biology*. 2012. V. 8. e1002743.
6. Reddy T.B.K., Thomas A., Stamatis D., Bertsch J., Isbandi M., Jansson J., Mallajosyula J., Pagani L., Lobos E., Kyripides N. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification // *Nucl. Acids Res.* 2014. doi: 10.1093/nar/gku950.
7. Roesch L.F.W., Fulthorpe R.R., Riva A., Casella G., Hadwin A.K.M., Kent A.D., Daroub S.H., Camargo F.A.O., Farmerie W.G., Triplett E.W. Pyrosequencing enumerates and contrasts soil microbial diversity // *The ISME Journal*. – 2007. – V. 1. – P. 283–290.
8. Stoddard S.F., Smith B.J., Hein R., Roller B.R.K., Schmidt T.M. *rmDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development* // *Nucl. Acids Res.* – 2015. – V. 43. – P. D593–D598.
9. Tyson G.W., Lo I., Baker B.J., Allen E.E., Hugenholtz P., Banfield J.F. Genome-Directed Isolation of the Key Nitrogenfixer *Leptospirillum ferrodiazotrophum* sp. nov. from an Acidophilic Microbial Community // *Appl. Environ. Microbiol.* – 2005. – V. 71. – P. 6319–6324.
10. Yun J., Zhang H., Deng Y., Wang Y. Aerobic Methanotroph Diversity in SanJiang Wetland, Northeast China // *Microbial Ecol.* – 2015. – V. 69. – P. 567–576.