

# Генное досье микробиома

© Е. В. Першина,  
Е. Е. Андронов,  
Г. Г. Самосоров,  
А. Н. Семенов,  
2013

Последние годы ознаменовались революционными преобразованиями в технологиях исследования огромного мира микроорганизмов – обитателей почвы. Широкое использование высокопроизводительного секвенирования бактериальных геномов вызвало качественный скачок в понимании устройства и функционирования почвенного микробного сообщества. Однако новые возможности породили и новые проблемы, главная из которых – исключительно высокая численность и беспрецедентный уровень генетического разнообразия почвенных микроорганизмов, подавляющее большинство которых нельзя культивировать обычными способами. Как следствие – огромные массивы экспериментальных данных, трудно поддающихся анализу, тогда как от исследователей ждут не только фундаментальных работ, раскрывающих основы устройства живых систем, но и их практического применения

*«Everything is everywhere but the environment selects»  
(«Везде есть все, но отбирает среда»)  
М. Бейеринк (1913)*

В 2013 г. исполняется сто лет знаменитому тезису голландского микробиолога М. Бейеринка, в котором постулируется ведущая роль окружающей среды в формировании микробных сообществ. И хотя современная наука далека от буквального понимания этого утверждения, его актуальность сегодня не вызывает сомнений. По сути, этот тезис задает обширную программу исследования природных микробиомов как универсального экологического индикатора, в котором, как в зеркале, отражаются все особенности среды и даже ее история.

Действительно, более чем за столетнюю историю микробной экологии это направление всегда было одним из наиболее популярных, однако реальная возможность детального исследования «частных» микробных сообществ появилась лишь в самое последнее время.

Дело в том, что до 99 % микроорганизмов, обитающих в окружающей среде (например, в почве) являются некультивируемыми, т.е. не могут расти на питательных средах, используемых в лабораторной практике. Именно по этой причине традиционная микробиология



ПЕРШИНА Елизавета Владимировна – младший научный сотрудник лаборатории микробиологического мониторинга и биоремедиации почв Всероссийского научно-исследовательского института сельскохозяйственной микробиологии (Санкт-Петербург). Автор и соавтор 10 научных публикаций

АНДРОНОВ Евгений Евгеньевич – кандидат биологических наук, заведующий лабораторией микробиологического мониторинга и биоремедиации почв и руководитель отделения геномных технологий ЦКП Всероссийского научно-исследовательского института сельскохозяйственной микробиологии (Санкт-Петербург). Автор и соавтор более 30 научных публикаций

САМОСОРОВ Георгий Германович – кандидат геолого-минералогических наук, руководитель подразделения ФГКУ «Войсковая часть 68240». Автор и соавтор более 40 научных работ

СЕМЕНОВ Андрей Николаевич – сотрудник ФГКУ «Войсковая часть 68240». Автор и соавтор трех научных работ

**Ключевые слова:** микробное сообщество, микробиом, 16S rPHK, пиросеквенирование  
**Key words:** microbial community, microbiome, 16S rRNA, pyrosequencing

всегда имела дело лишь с вершиной микробиологического «айсберга» окружающей среды. Например, оценка численности микроорганизмов в почве с помощью посева на питательные среды дает значения около  $10^7$  клеток/г, в то время как с помощью молекулярно-генетических методов эти значения возрастают на два-три порядка!

Такой «молекулярный» подход к анализу микробных сообществ почвы открывает возможность создания совершенно новых систем агроэкологического мониторинга. Действительно, каждое из изменений, касающееся состояния почв, несомненно, приводит к изменению в структуре почвенного микробиома. Если мы научимся их регистрировать и интерпретировать, то в наших руках окажется один из самых эффективных и универсальных экологических индикаторов, грамотное использование которого может способствовать прогрессу в самых разных областях: от земледелия до поиска полезных ископаемых и криминалистики.

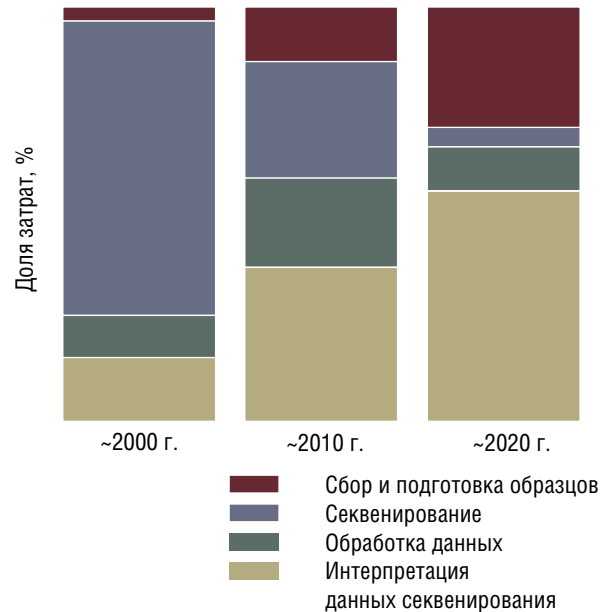
## Неопознанные Микробные Объекты

Сегодня для высокопроизводительного секвенирования нуклеотидных последовательностей почвенных микроорганизмов используется популярный таксономический маркер – ген *16S rPHK*. Этот ген кодирует одну из РНК, составляющих основу бактериальных рибосом – многочисленных клеточных органелл, отвечающих за синтез белков. Такой выбор обусловлен рядом причин. Например, этот ген имеется в геномах всех известных микроорганизмов, однако отсутствует у вирусов и высших организмов. Кроме того, он достаточно консервативен, однако имеет вариабельные видоспецифичные участки, что позволяет различить организмы разной систематической принадлежности.

Схема самого исследования проста: на первом этапе из почвы выделяют ДНК, затем получают так называемую *геномную библиотеку*, содержащую копии гена *16S rPHK*, принадлежащие различным почвенным бактериям. Библиотеку «читают» с использованием высокопроизводительных секвенаторов, обеспечивающих получение нескольких тысяч нуклеотидных последовательностей гена *16S rPHK* для каждого из образцов.

Следующий этап – анализ огромного массива полученных данных с помощью методов биоинформатики. Результаты представляют способом, наиболее подходящим в каждом конкретном случае, например, в виде традиционного филогенетического дерева.

На сегодняшний день в мире существует несколько баз данных, аккумулирующих информацию по нуклеотидным последовательностям бактериального гена *16S rPHK*. Прежде всего это три открытые «общающие»



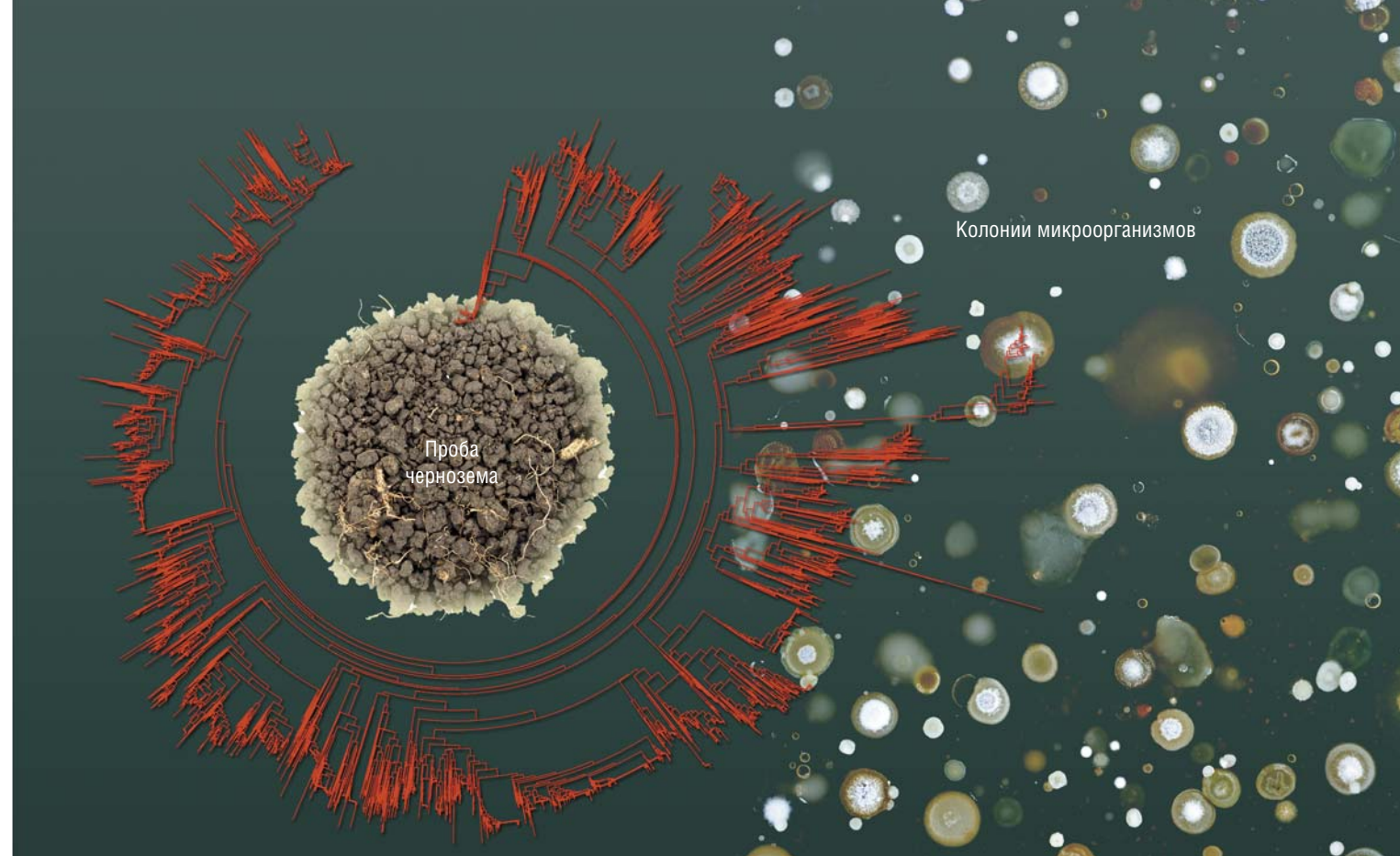
За последнее десятилетие процедура секвенирования (расшифровки) ДНК стала рутинным делом, при том что затраты на хранение и анализ полученных данных резко возросли. И в ближайшее десятилетие эта тенденция только усилится.  
По: (Sbonner et al., 2011)

### В ИНФОРМАЦИОННОМ ПОЛОВОДЬЕ

**Характерный для последних лет обвалный рост количества добываемых научных данных уже привел к возникновению глобальной проблемы, которую авторитетный журнал «The Economist» (2010) назвал «The data deluge» (информационный потоп).**

**Это в полной мере касается высокопроизводительного секвенирования генетического материала: хранение данных, полученных этим методом, скоро может стать проблемой, так как цена секвенирования снижается гораздо быстрее, чем цена самого хранения (Stein, 2010). Более того, к 2020 г. доля расходов в научном исследовании, связанная с собственно секвенированием, составит не более одной десятой общих затрат, тогда как более половины из них придется на анализ данных (Sbonner et al., 2011).**

**В этой ситуации с каждым годом будет возрастать роль биоинформатики, задача которой – разработать методы анализа, позволяющие эффективно выявлять биологически значимую компоненту в громадных массивах информации, получаемых в ходе экспериментальных исследований**



Это филогенетическое дерево демонстрирует богатство и структуру микробного сообщества одной пробы чернозема, оцененные по данным анализа 4,3 тыс. нуклеотидных последовательностей гена *16S rPHK* с использованием программы *UGENE*.  
Справа – результат посева почвенной суспензии на питательную среду

» базы *GenBank/EMBL/DBJ*, где собран весь массив имеющихся экспериментальных данных, а также ряд специализированных баз данных, в которые попадают только тщательно отобранные «опознанные» генные последовательности.

Из последних баз следует отметить *RDP*, в которой на сегодняшний день имеется свыше 2,5 млн последовательностей гена *16S rPHK*. С использованием этого ресурса можно наиболее точно определить систематическую принадлежность больших массивов нуклеотидных последовательностей этого гена, получаемых в результате высокопроизводительного секвенирования генных библиотек.

Однако ни одно из этих хранилищ информации не может претендовать на полный охват генетического материала почвенной микробиоты, что и неудивительно. Ведь 1 г почвы может содержать до  $10^{16}$  н\* генетической информации, что превосходит объем не только *GenBank* ( $1,5 \times 10^{11}$  н на конец 2012 г.), но даже архива *SRA* ( $1,1 \times 10^{15}$  н на конец 2012 г.), в котором аккумулируются все данные, полученные на секвенаторах нового поколения.

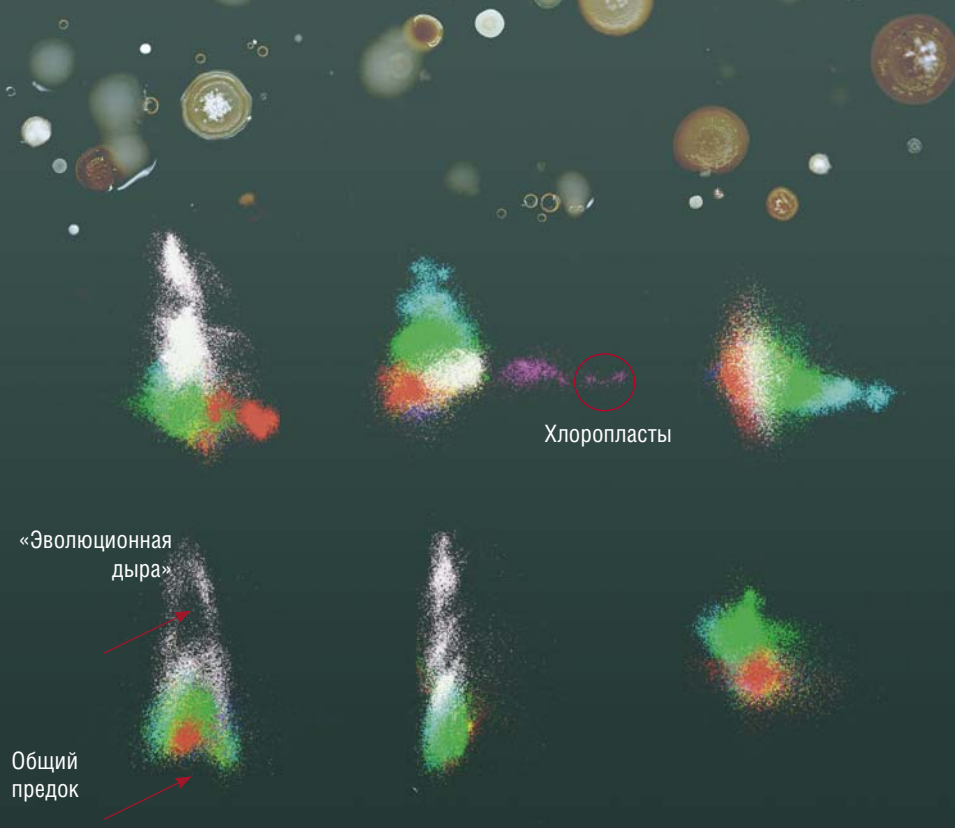
\* Единица измерения – 1 нуклеотид

Не будет преувеличением сказать, что исчерпывающее секвенирование генетического материала, содержащегося в 1 г богатой почвы, сегодня невозможно в принципе, даже с использованием всего мирового парка секвенаторов. Приблизительная оценка видового разнообразия микроорганизмов в таком крошечном образце почвы дает цифру в тысячи и даже десятки тысяч видов, свидетельствующую о том, что именно почва является крупнейшим в биосфере депозитарием генетического материала.

Поэтому неудивительно, что для большого числа нуклеотидных последовательностей из почвенных библиотек гена *16S rPHK* не удастся подобрать известных «родственников» в базах данных. Из-за этого до 30–60 % секвенированных последовательностей не могут быть определены до уровня рода, а часть из них – даже до уровня крупнейших таксономических рангов. И такие «неопознанные» последовательности не являются результатом ошибки секвенирования: в базах данных в большинстве случаев удается обнаружить схожие последовательности, которым в свое время также не был поставлен «таксономический диагноз».

- Proacteria
- Bacteroidetes
- Chloroflexi
- Cyanobacteria
- Acidobacteria
- Actinobacteria
- Firmicutes
- Другие

На избранных ортогональных срезах многомерного (13D) эволюционного пространства гена 16S рРНК бактерий, проходящих через геометрический центр, различным цветом показаны основные бактериальные филы. Отмечены предполагаемая «эволюционная дыра» в филе протеобактерий и возможное место локализации общего предка бактерий. По: (Дольник и др. 2012)



## В эволюционном пространстве гена

Учет «безродных» микроорганизмов является важной задачей как по причине их высокой численности в почве, так и по тому, что именно среди них нередко обнаруживаются важные экологические индикаторы.

Традиционно для этой цели проводится *выравнивание* всего экспериментального массива последовательностей (поиск одинаковых участков) и *кластерный анализ*. В результате выявляются группы, объединяющие последовательности с высоким уровнем сходства, близким к внутривидовой изменчивости.

Этот подход, реализованный в ряде популярных программных пакетов и сетевых ресурсов (*MOTHUR*, *QIIME*, *VAMPS*), позволяет дать «неопознанным» последовательностям формальные идентификаторы. Однако у него имеется ряд существенных недостатков: при добавлении новых массивов данных приходится заново выравнивать совокупный массив, что часто требует больших вычислительных ресурсов. Кроме

## ЭВОЛЮЦИЯ В МНОГОМЕРНОМ ПРОСТРАНСТВЕ ГЕНА

**Самая неожиданная и интересная особенность нового подхода – его эволюционный характер. Глобальный эволюционный процесс, представленный в пространстве гена 16S рРНК бактерий, должен, вероятно, представлять собой необратимое радиальное расширение от исходной точки (с учетом предположения о наличии общего предка для всех бактерий).**

**Ряд эволюционных феноменов очевиден уже в первом (13D) варианте такого пространства, построенном с учетом всех имеющихся данных. Так, фила цианобактерий имеет ярко выраженную вытянутость, при этом на ее «хвосте» группируются последовательности хлоропластов (клеточных органелл фотосинтеза). Эта картина напоминает «эволюционную пушку», нацеленную на грядущую эукариотическую клетку, хлоропласты которой, как принято считать, являются результатом интеграции цианобактерий и предков эукариотической клетки.**

**Другой феномен – «эволюционные дыры», области эволюционного пространства, где ранее находились предковые формы гена 16S рРНК микроорганизмов, которые были утрачены в ходе эволюции. При этом в силу необратимости и радиальности эволюционного расширения образовавшие «дыры» уже никогда не могут быть заполнены**

того, в одном эксперименте невозможно объединять нуклеотидные последовательности, представляющие различные фрагменты гена 16S рРНК, так как в этом случае выравнивание в принципе невозможно.

Но что если создать систему, в которой любая возможная нуклеотидная последовательность гена 16S рРНК получит фиксированный идентификатор вне зависимости от того, имеются ли родственные ей последовательности в базах данных или в современной биосфере, и были ли они вообще реализованы в ходе эволюции? По сути дела, речь идет об извечной проблеме создания «естественной системы организмов», которая в нашем частном случае решается упорядочиванием нуклеотидного разнообразия в пределах одного бактериального гена.

Такая концепция «эволюционного пространства» гена 16S рРНК была предложена коллективом специалистов из Всероссийского научно-исследовательского института сельскохозяйственной микробиологии РАСХН (Санкт-Петербург), Санкт-Петербургского государственного университета и Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (Дольник и др. 2012). В этом многомерном пространстве любая нуклеотидная последовательность гена отображается в виде точки с фиксированными координатами, а расстояния между точками соответствуют эволюционным расстояниям между последовательностями.

В качестве первой версии разработчики представили 13-мерное метрическое эволюционное пространство, в котором поместили представительную выборку последовательностей гена 16S рРНК, принадлежащих всем известным таксонам. Эти построения были визуализированы с использованием «томографических» сечений в ортогональных или параллельных плоскостях.

Конечно, эта версия эволюционного пространства гена 16S рРНК является своего рода «эскизом», однако подобное построение может стать принципиально новым «операционным полем» для молекулярной экологии микроорганизмов, позволяя проводить анализ сложных микробных сообществ с использованием целого ряда ранее недоступных «интегральных» параметров, т. е. исследовать сообщества как единое целое.

## Пробный камень

Концепция эволюционного пространства была впервые применена на простой модельной системе, связанной с действием одного из самых мощных экологических факторов – засоленности. Для этого использовались образцы почв, взятые из природного солончака по градиенту засоленности, и темно-каштановой почвы, взятые в 200 км от солончака. Кроме того, был проведен опыт по искусственному засолению темно-каштановой почвы раствором солей такого же состава, как на наиболее засоленном участке солончака (Першина и др., 2013, *в печати*).

Традиционный микробиологический анализ показал, что образцы засоленной почвы различного происхождения различаются и по составу микроорганизмов, однако обнаруживают сходство в динамике структуры сообщества с ростом засоленности: бактерии, типичные для незасоленных участков, сменяются галофильными микроорганизмами. Однако это сходство выявилось лишь при подробном рассмотрении таксономической структуры сообществ, высокий же уровень биоразнообразия почвенных микроорганизмов не позволяет каждый раз проводить такой трудоемкий сравнительный анализ.

Концепция эволюционного пространства позволяет описать крупные сдвиги в структуре микробного сообщества с помощью интегральных параметров. Микробное сообщество в этом случае представляет собой «облако точек», изменение формы и положения которого будет отражать структурную динамику сообщества. На первом этапе можно использовать наиболее простые параметры: *центральную точку сообщества* и *вектор смещения*, соединяющий две центральные точки. Расстояние между центральными точками служит мерой сходства (или различия) сообществ, а вектор показывает направление его развития.

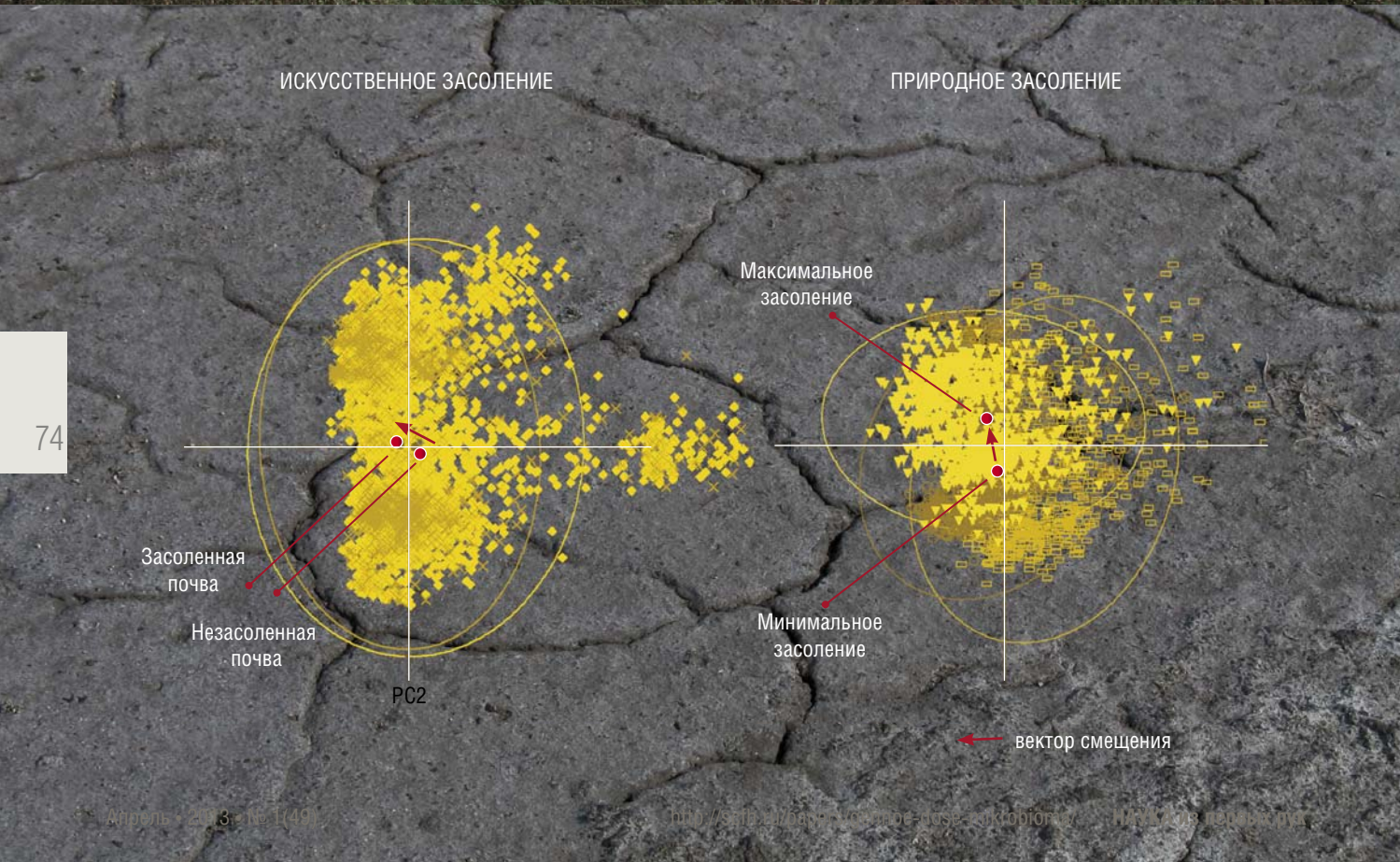
Для того чтобы ответить на вопрос о сходстве в протекании двух процессов (в нашем случае – природного и искусственного засоления) в эволюционном пространстве достаточно ввести третий интегральный параметр – угол между векторами смещения. Его значения, близкие к нулю, свидетельствуют об однонаправленности в развитии сообществ; при значениях угла, близких к 180°, сообщества развиваются в противоположных направлениях. Полученное значение угла, равное 73°, хорошо совпадает с описанием, данным в рамках традиционного подхода: динамика сообществ при природном и искусственном засолении имеет ряд общих особенностей (угол лежит в пределах 90°-го сектора), но при этом существенно различается (значения угла далеки от 0°).

Задача дальнейших исследований в этой области – накопление информации по действию разнообразных экологических факторов на микробное сообщество и построение шкалы для оценки углов между векторами смещения, чтобы избежать субъективной трактовки результатов. По завершению этой работы интегральные параметры эволюционного пространства могут стать в один ряд с традиционными экологическими показателями биоразнообразия и даже превзойти их.

## В качестве улики

Задача, о которой пойдет речь ниже, пришла из криминалистики. Дело в том, что объекты криминалистического исследования, в том числе и почвы, принципиально отличаются от научных образцов, отображенных в соответствии с принятыми стандартами. Они могут храниться и транспортироваться в неправильных условиях, быть пересушенными, загрязненными и т. д. Можно ли по такому образцу почвы судить о его географическом происхождении или экологической приуроченности?

Для решения таких практических вопросов экологии микроорганизмов в содружестве с новосибирской фирмой «Унипро» были разработаны программные продукты для анализа библиотек гена 16S рРНК, работающие



Засоленные образцы почв из природного солончака и эксперимента по искусственному засолению существенно различались по общему составу микроорганизмов (слева вверху). Но более подробный анализ выявил сходство в изменении структуры микробиомов с ростом засоленности (например, в них возрастала доля бактерий семейств Bacillaceae и Balneolaceae). Концепция эволюционного пространства гена 16S рРНК позволяет описать подобные сдвиги в структуре с помощью интегральных показателей, например, угла между векторами смещения «облаков точек» (слева внизу). В данном случае значение угла ( $73^\circ$ ), хорошо совпадает с результатами, полученными традиционным путем: динамика сообществ имеет ряд общих особенностей (угол  $< 90^\circ$ ), но при этом существенно различается (угол  $> 0^\circ$ )

в программной среде UGENE\*. На сегодняшний день эти программы являются простыми и эффективными инструментами исследования разнообразия почвенной (и не только почвенной) микробиоты.

Основной проблемой в таком анализе является выше упомянутое крайне высокое микробиологическое разнообразие и, как следствие, высокий уровень «шумов», затрудняющий выделение экологически значимого сигнала. Разумная стратегия – объединять образцы в пулы по экологическим признакам, а затем искать в них статистически значимые различия.

Для этой цели были созданы несколько небольших программ (DEREPLICATE для сокращения крупных массивов данных, RANDOMFILTER для нормализации массивов и др.), которые на сегодняшний день являются не столько средствами решения задач, сколько средствами научного поиска.

Так, программа REDUCE, позволяющая «вычитать» один массив из другого в соответствии с заданным уровнем сходства, используется для поиска последовательностей, которые встречаются лишь в одном исследуемом образце. И если из массива экспериментальных данных вычесть, к примеру, большой пул совокупной библиотеки почв различного происхождения (такая библиотека ВНИИСХМ уже содержит около 500 тыс. нуклеотидных последовательностей), то можно выявить в образце набор уникальных последовательностей. А если «вычесть» экологически приуроченные пулы, то можно выявить микроорганизмы, являющиеся в данном случае средовыми маркерами.

В качестве примера приведем результаты анализа почв сельскохозяйственного объекта, на котором практикуется органическое (без применения минеральных удобрений, гербицидов и пестицидов) и обычное земледелие, а также соседнего, частично сведенного леса.

\* Подробнее на сс. 60–67

На основе собранных проб почвы было получено девять библиотек гена 16S рРНК (около 27 тыс. нуклеотидных последовательностей). Библиотеки были объединены в пулы, соответствующие разным объектам, после чего было выполнено взаимное вычитание с пороговым значением сходства 97%. Результаты таксономического анализа показали, что для каждого участка имеются специфические последовательности, которые в сумме составили до 10% от суммарной библиотеки всех образцов почв. При этом почти половину этих «остатков» составили уникальные «неопознанные» последовательности, что в пять раз больше, чем в исходных библиотеках – этот факт требует дальнейшего изучения.

Перспективы использования такого подхода в криминалистике вполне ясны, хотя и требуют проведения дальнейших исследований. Не исключено, что уже в недалеком будущем результаты секвенирования почвенных проб смогут использовать современные «холмсы» в качестве достоверной и бесспорной улики.

Справедливости ради нужно отметить, что почти сто лет назад блестящий российский микробиолог и ботаник С. Н. Виноградский путем прямого подсчета микробных клеток в почве получил данные, по численности близкие к современным. Однако лишь молекулярно-генетические исследования дали реальный ключ к анализу разнообразия почвенной микробиоты во всей ее полноте.

Выделение и расшифровка ДНК из объектов окружающей среды сами по себе сегодня являются рутинными задачами, однако уже сейчас очевидно, что для поиска биологического смысла в громадных массивах данных требуются принципиально новые подходы, разрабатываемые в рамках биоинформатики.

#### Литература

Дольник А. С., Тамазян Г. С., Першина Е. В. и др. Концепция универсальной таксономической системы бактерий: эволюционное пространство гена 16S-РНК v. 1.0. // Сельскохозяйственная биология. 2012. № 5. С. 111–120.

Першина Е. В., Дольник А. С., Пинаев А. Г., Андронов Е. Е. Использование концепции эволюционного пространства для интегральной оценки сдвигов в структуре микробных сообществ по данным анализа библиотек гена 16S рРНК // Сельскохозяйственная биология. 2013. №4 (в печати).

К. Okonechnikov, O. Golosova, M. Fursov and the UGENE team. Unipro UGENE: a unified bioinformatics toolkit // Bioinformatics, 2012, 28: 1166–1167.

Авторы благодарят за сотрудничество в создании программного обеспечения фирму Унипро (Новосибирск). Работа поддержана ГК 2012/389, ГК 16.552.11.7085 и РФФИ 12-04-01371-а