

Симбиогенетика

УДК 631.461.52:575.1:577.257.065

doi: 10.15389/agrobiologia.2017.5.995rus

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ФИЛОГЕНИЙ СИМБИОТИЧЕСКИХ ГЕНОВ КЛУБЕНЬКОВЫХ БАКТЕРИЙ С ИСПОЛЬЗОВАНИЕМ МЕТАДЕРЕВЬЕВ***Е.С. КАРАСЕВ, Е.П. ЧИЖЕВСКАЯ, Б.В. СИМАРОВ, Н.А. ПРОВОРОВ, Е.Е. АНДРОНОВ**

Для изучения закономерностей эволюции различных групп «симбиотических» генов (*nod*-генов, контролирующих образование клубеньков, *nif/fix*-генов, контролирующих симбиотическую азотфиксацию) был применен модифицированный нами метод филогенетического анализа — построение метадеревьев. Суть метода заключается в попарном сравнении топологий дендрограмм и построении комбинированных дендрограмм (метадеревьев), на которых относительное положение двух деревьев является мерой конгруэнтности филогений соответствующих генов. Для реализации данного метода были выбраны 18 симбиотических генов (*nodABCDIJN*, *nifABDEHKN*, *fixABC*, *fixB*), гомологи которых присутствуют у каждого из исследуемых организмов (9 штаммов, относящихся к родам *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, *Sinorhizobium* и *Neorhizobium*), а также ген 16S рРНК — традиционный хромосомный таксономический маркер. Для каждого из этих генов были построены и сопоставлены филогенетические деревья, затем были рассчитаны коэффициенты попарного сходства их топологий. По полученным данным было построено «метадревье», в пределах которого были выявлены два статистически различающихся кластера генов. В кластер 1 вошли преимущественно *nif- u fix*-гены, а в кластер 2 — преимущественно *nod*-гены, что согласуется с данными о раздельной локализации этих групп генов в геномах ризобий. Исключением было расположение генов *nifB* и *fixC* в кластере 2 вместе с геном *nodA*, а также локализация гена *nodI* в кластере 1 вместе с геном *nifD*. При анализе структуры выявленных кластеров не было обнаружено строгой зависимости между относительным положением изучаемых генов и особенностями их локализации в геномах клубеньковых бактерий. Важно отметить, что различия между кластерами 1 и 2 выражены не менее четко, чем различия между группами *nod*- и *nif/fix*-генов. Очевидно, что кластеры 1 и 2 на построенном нами метадревье отражают в первую очередь различие механизмов эволюции процессов образования клубеньков и симбиотической азотфиксации, связанное с независимым происхождением соответствующих групп генов, а возможно, и с их раздельным горизонтальным переносом между разными группами ризобий. Дальнейшее изучение эволюции симбиотических генов клубеньковых бактерий требует усовершенствования использованной методики филогенетического анализа, включая раздельный анализ метадеревьев для ризобий, представляющих разные этапы эволюции симбиоза.

Ключевые слова: филогенетический анализ, метадревья, клубеньковые бактерии, симбиотические гены.

Почвенные бактерии порядка *Rhizobiales* характеризуются способностью вступать в симбиоз с бобовыми растениями, обеспечивая их продуктами фиксации азота атмосферы, что сопровождается формированием на корнях клубеньков, в которых создаются условия для эффективной азотфиксации и ассимиляции аммония, образуемого в результате нитрогеназной реакции. Однако эта группа организмов представляет интерес не только для микробиологии, но и для эволюционной генетики симбиоза. Анализируемая модель позволяет изучать механизмы эволюции бактериальных генов, контролирующих взаимодействия с разными группами эукариот (1).

Образование клубеньков и симбиотическая азотфиксация — сложные многостадийные процессы, которые контролируются генами обоих партнеров. Симбиотические гены ризобий разделяют на две основные группы: гены вирулентности (*nod*-гены), ответственные за сигнальный диалог с растением-хозяином, в результате которого на его корнях образуются клубеньки (2), и гены, необходимые для симбиотической фиксации атмос-

* Работа выполнена в рамках проекта РНФ № 14-26-00094П и проводилась с использованием оборудования ЦКП «Геномные технологии, протеомика и клеточная биология» ФГБНУ ВНИИСХМ.

сферного азота, — *nif*-гены (контролируют синтез и регуляцию активности фермента нитрогеназы) и *fix*-гены (разнородная группа генов, вовлеченных в работу нитрогеназного комплекса, в первую очередь в его снабжение электронами и энергией) (3).

У большинства клубеньковых бактерий *nod*- и *nif/fix*-гены собраны в группы, расположенные в геноме в виде отдельных кластеров, структура которых варьирует у разных видов. Например, у клубеньковых бактерий люцерны (*Sinorhizobium meliloti* и *S. medicae*), гороха (*Rhizobium leguminosarum* bv. *viciae*), клевера (*R. leguminosarum* bv. *trifolii*), козлятника (*Neorhizobium galegae*) и сои (*Bradyrhizobium japonicum*) основные *nod*-гены тесно сцеплены, при этом гены *nodABC* организованы в один оперон (рис. 1). У многих ризобий структурные гены нитрогеназы *nifHDK* представляют собой один оперон и тесно сцеплены с генами *nifEN*. Такая организация генов азотфиксации характерна для представителей *Sinorhizobium*, *R. leguminosarum*, *N. galegae* и *Mesorhizobium loti*. Однако у *B. japonicum* дополнительная копия гена *nifH* расположена вне *nifHDK*-оперона, а ген *nifA* — за пределами *nif/fix*-кластера совместно с *nod*-генами.

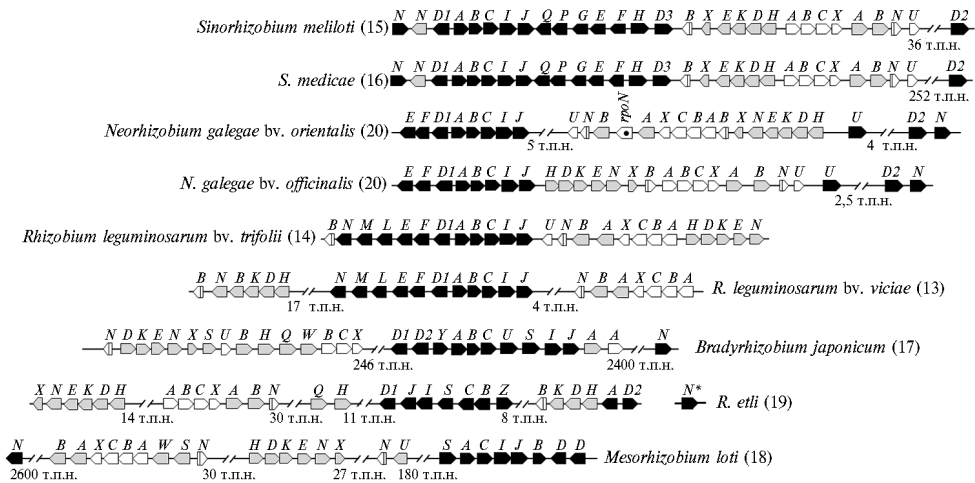


Рис. 1. Расположение симбиотических генов у различных видов клубеньковых бактерий: ■ — *nod*-гены, □ — *nif*-гены, ◐ — *fix*-гены, ◑ — *fdx*-гены, ◒ — несимбиотические гены; ген *nodN* в геноме *R. etli* (отмечен звездочкой) имеет хромосомную локализацию. Схемы построены по данным секвенирования геномов изучаемых ризобияльных штаммов (4–11).

У эволюционно продвинутых видов клубеньковых бактерий (например, *R. leguminosarum*) симбиотические гены расположены на высокомолекулярных плаزمидах (12). В то же время их хромосомная локализация характерна для анцестральной формы — *B. japonicum*, а также для *M. loti* (13, 14). Есть ризобии, у которых на хромосоме могут располагаться только некоторые симбиотические гены. Это свойственно, например, для *R. etli*: большинство симбиотических генов находятся на плазмиде, а ген *nodN* — на хромосоме (15) (см. рис. 1).

Долгое время молекулярные механизмы эволюции бобово-ризобияльного симбиоза, насчитывающей 60–70 млн лет (16), оставались неясными. Лишь благодаря методам высокопроизводительного секвенирования, когда началось быстрое накопление данных о составе и строении геномов различных видов бактерий, пангеномному и метагеномному анализу удалось реконструировать тот длинный путь, который прошли микро- и макросимбионты от рекрутирования в бактериальные клетки генов азотфиксации (*nif*-генов) до развития современных высокоспециализированных бобово-

ризобияльных симбиотических систем. Этот процесс, по-видимому, происходил в несколько этапов, на которых рассматриваемые группы генов были приобретены из разных источников: *nif*-гены и большинство *fix*-генов — получены анцестральными медленно растущими ризобиями (*Bradyrhizobium*) от своих свободноживущих предков (*Rhodopseudomonas*) посредством вертикального наследования, тогда как гены синтеза липо-хито-олигосахаридных Nod-факторов, отвечающих за образование клубеньков (*nod*-гены), — приобретены от грибов либо от грамположительных бактерий при горизонтальном переносе (17). Возникшие позднее быстро растущие ризобии (*Rhizobium*, *Sinorhizobium*, *Neorhizobium*) получили *nod*- и *nif/fix*-гены от анцестральных ризобий при горизонтальном переносе генов (ГПГ), что привело к образованию особых плазмид (*pSym*), содержащих обе группы *sym*-генов.

Связь филогении *sym*-генов с механизмами их происхождения и эволюции, а также с локализацией в геномах ризобий исследована недостаточно. Для ее изучения мы использовали новый метод филогенетического анализа с построением «метадеревьев» — комбинированных дендрограмм, на которых относительное положение двух деревьев служит мерой конгруэнтности филогений индивидуальных генов. Ранее этот метод был разработан и использован для анализа альтернативных филогений, полученных при изучении функционально разнородных генов в различных группах эукариот (дрожжи, рыбы), в которых ГПГ ограничен (18).

В нашем исследовании названный метод был применен для филогенетического анализа двух функционально различающихся групп генов (*nod* и *nif/fix*), которые имеют неодинаковое происхождение и распространялись среди таксономически неродственных групп ризобий посредством как вертикального наследования, так и ГПГ. Выявленные нами особенности топологий филогений, построенных для генов образования клубеньков и симбиотической азотфиксации, позволяют существенно дополнить возникшие ранее представления о направлениях и механизмах эволюции генных систем симбиоза у клубеньковых бактерий.

Целью работы было использование метадеревьев для анализа участия горизонтального переноса в процессе эволюционной сборки кластера симбиотических генов у ризобий.

Методика. Для анализа использовали нуклеотидные последовательности симбиотических генов из полногеномных сиквенсов 9 представителей клубеньковых бактерий: *R. leguminosarum* bv. *viciae* 3841 (GenBank GI:115259115) (4), *R. leguminosarum* bv. *trifolii* WSM2304 (GI:209537694) (5), *S. meliloti* 1021 (GI:25168258) (6), *S. medicae* WSM419 (GI:150031715) (7), *B. japonicum* USDA110 (GI:47118316) (8), *M. loti* MAFF303099 (GI:47118328) (9), *R. etli* CFN42 (GI:89213252) (10), а также *N. galegae* bv. *officinalis* HA-MBI1141 (GI:659665307) и *N. galegae* bv. *orientalis* HAMBI540 (GI:659657635) (11).

Кластерный анализ выполняли в компьютерной программе MEGA 5.1 (<http://www.megasoftware.net/>). Выравнивание нуклеотидных последовательностей проводили с помощью алгоритма ClustalW; математическая модель при составлении древ для каждого исследуемого гена — *p*-distance, механизм кластеризации — Neighbor-Joining. Численное выражение сходств топологий филогенетических деревьев рассчитывали с применением статистических подходов (реализованы в программе, представленной на сайте <http://www.mas.ncl.ac.uk/~ntmwn/compare2trees/index.html>) (19). Для построения метадрева и анализа достоверности кластеризации использовали коэффициенты различия, которые вычисляли как разность единицы (100 %, уменьшаемое) и соответствующего коэффициента сходства (вычитаемое).

Результаты. На начальном этапе работы были построены филоге-

нетические деревья для каждого из следующих 18 генов: генов вирулентности *nodA*, *nodB*, *nodC* (отвечают за синтез коровой части Nod-фактора) (20–22), *nodD* (при наличии нескольких копий *nodD*, это конститутивно экспрессируемый ген флавоноид-чувствительного активатора транскрипции *nod*-генов) (23), *nodI*, *nodJ* (гены мембранных транспортеров Nod-фактора) (24), *nodN* (кодирует фермент дегидратазу) (25); генов симбиотической азотфиксации *fixA*, *fixB*, *fixC* (компоненты электронно-транспортной цепи нитрогеназы) (26), *nifA* (ген транскрипционного регулятора *nif*-генов), *nifB*, *nifN*, *nifE* (отвечают за синтез Fe-Мо кофактора нитрогеназы), *nifD*, *nifH*, *nifK* (структурные компоненты нитрогеназы) (27), *fdxN* (ферредоксин). Выбор был обусловлен тем, что гомологи указанных генов присутствуют у всех представителей клубеньковых бактерий, отобранных для анализа. Кроме того, в исследование включили традиционный хромосомный таксономический маркер — ген 16S-рРНК в связи с консервативностью его нуклеотидной последовательности и низкой частотой горизонтального переноса.

Методология сравнения топологий основана на представлении о том, что горизонтальный перенос генов от одного вида ризобий к другому в ходе эволюции вносит существенные различия в топологии филогений этих генов. В том случае, когда топология деревьев двух генов не различается, предполагают, что они либо не участвуют в горизонтальном переносе совсем, либо переносятся совместно (рис. 2, А).

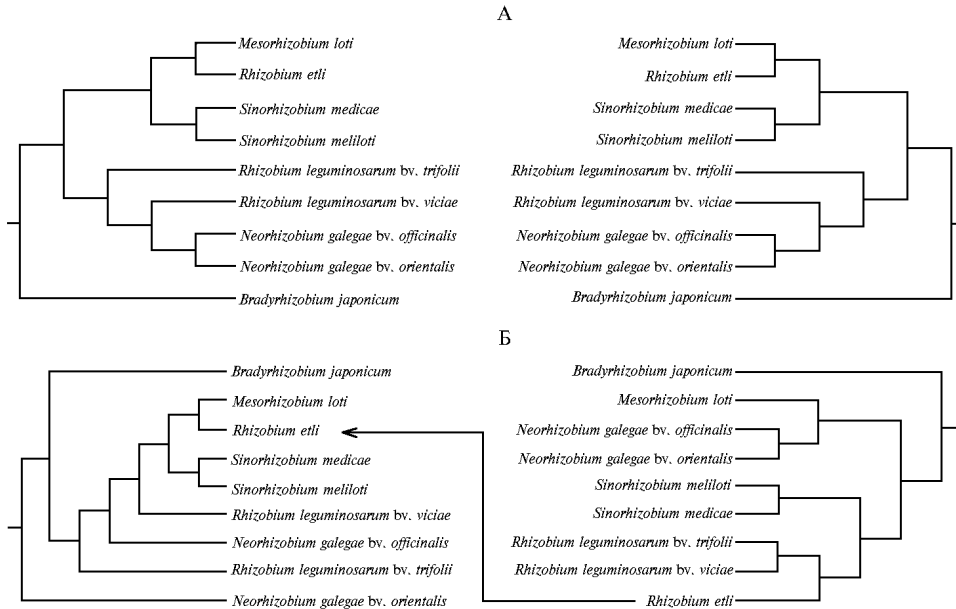


Рис. 2. Сравнения топологии филогенетических деревьев для генов *nifH* (слева) и *nifN* (справа) со степенью сходства 100 % (А) и для генов *fdxN* (слева) и 16S-рРНК (справа) со степенью сходства 50,5 % (Б). Стрелкой отмечено наиболее заметное различие в топологии: перемещение *Rhizobium etli* из кластера *R. leguminosarum* на древе гена 16S-рРНК в кластер *M. loti* на древе *fdxN*.

В то же время несоответствие в топологии деревьев, построенных на основании анализа сиквенсов одного и того же набора штаммов, представляет собой следствие горизонтального переноса симбиотических генов (см. рис. 2, Б). По этим различиям в топологии мы можем предполагать, как происходил ГПГ в ходе эволюции. Например, видно, что в древе по гену 16S-рРНК в одном кластере находятся два биовара *R. leguminosarum* и *R. etli*. Так как ген 16S-рРНК — хромосомный маркер, то допустимо утвер-

ждать, что кластеризующиеся вместе представители эволюционно близки. В то же время на древе плазмидного маркера *fdxN* ризобии *R. etli* группировались в одном кластере с *M. loti*, хотя *R. etli* и *M. loti* относятся к разным семействам — *Rhizobiaceae* и *Phyllobacteriaceae*. Этот факт позволяет предположить, что имел место горизонтальный перенос гена *fdxN* между клубеньковыми бактериями *M. loti* и *R. etli* либо при горизонтальном переносе указанные виды получили этот ген из одного источника.

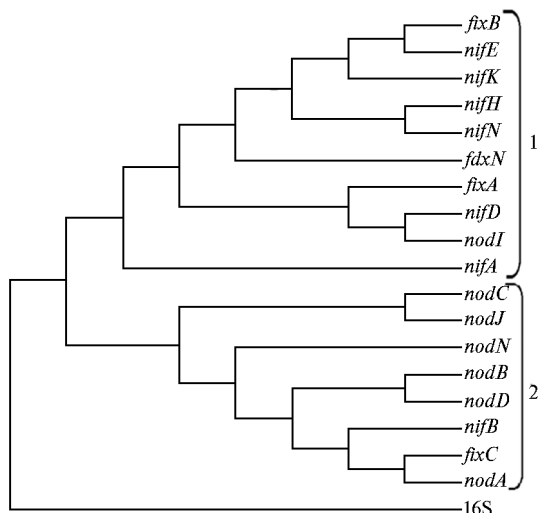


Рис. 3. Основные кластеры метадрева по матрице сходств топологий в филогенетических деревьях, построенных по данным о нуклеотидном полиморфизме симбиотических генов: 1 — *nif/fix*-кластер, 2 — *nod*-кластер.

Видно, что в метадреве достоверно выделились два кластера: группа *nif/fix*-генов оказалась преимущественно сосредоточена в кластере 1, *nod*-генов — в кластере 2, тогда как хромосомный маркер (ген 16S-рРНК) не вошел ни в один из кластеров (табл.).

Оценка достоверности кластеризации метадрева в соответствии со средним расстоянием внутри и между кластерами / группами генов с использованием *t*-критерия Стьюдента

Группа сравнения	TD	SD	SE	Достоверность различий по <i>t</i> -критерию Стьюдента при уровне значимости <i>p</i>
Внутри кластера 1	0,197	0,094	0,013	< 0,001
Внутри кластера 2	0,216	0,104	0,020	< 0,01
Между кластерами	0,269	0,107	0,012	
Внутри группы <i>nif-fix</i>	0,194	0,084	0,011	< 0,001
Внутри группы <i>nod</i>	0,201	0,082	0,018	< 0,01
Между группами симбиотических генов	0,270	0,099	0,011	

Примечание. TD — среднее различие топологий, SD — стандартное отклонение, SE — стандартная ошибка. Проводили сравнение среднего расстояния между топологиями деревьев внутри одного кластера/группы, а также между деревьями одного кластера/группы и каждым деревом другого кластера/группы. Структура кластеров приведена на рисунке 3.

Такая кластеризация может указывать на то, что в эволюционной истории группы генов *nod* и *nif/fix* не связаны между собой и с хромосомными генами. Она определяется их самостоятельным происхождением и отдельным горизонтальным переносом. Действительно, различие групп *nod*- и *nif/fix*-генов выражено не менее четко, чем несходство между кластерами 1 и 2. Однако были выявлены и исключения: ген *nodI*, который находится в *nif/fix* кластере 1 и переносится вместе с геном *nifD*, а также

Затем мы сравнили все полученные 19 филогенетических деревьев для выявления различий в их топологиях. Биоинформатический анализ полученных данных позволил рассчитать коэффициенты попарного сходства топологий для всех деревьев, варьирующие от 1 (полная конгруэнтность двух филогений) до 0 (полное отсутствие конгруэнтности). Затем построенная матрица сходств была преобразована в матрицу различий вычитанием соответствующих коэффициентов сходства из единицы. По данным этой матрицы было построено метадрево (рис. 3).

гены *nifB* и *fixC*, попадающие в *nod*-кластер 2 вместе с *nodA*.

Как отмечалось выше, в геномах многих клубеньковых бактерий симбиотические гены объединены в кластеры. В связи с этим совместный горизонтальный перенос групп *nod*- и *nif/fix* генов может быть обусловлен особенностями их кластерного расположения в геноме (см. рис. 1). Тем не менее, объяснить полученную картину только спецификой расположения генов на генетических картах не представляется возможным. Так, в кластере 1 были обнаружены все структурные гены нитрогеназы, причем гены *nifH*, *nifK*, *nifE* и *nifN* группировались друг с другом. Подобная кластеризация генов соответствует их расположению в геномах, поскольку у всех изучаемых организмов гены *nifHDK* составляют один оперон, к которому прилегают гены *nifE* и *nifN*. Единственное исключение — локализация гена *nifN* за пределами *nif/fix*-региона у клубеньковых бактерий люцерны (*Sinorhizobium meliloti*). Вместе с теми же *nod*-генами в кластере 2 был обнаружен *nodN*, однако внутри кластера он располагался отдельно. Следует отметить, что локализация *nodN* в геноме может быть различной. У некоторых представителей он находится в одном кластере с генами *nodABC* (*S. meliloti*, *S. medicae*, *R. leguminosarum*), у других — расположен отдельно от общих *nod*-генов. Например, у *R. galegae* *nodN* находится в другом *nod*-кластере, а у *B. japonicum*, *M. loti* и *R. etli* — вообще вне кластера симбиотических генов. Кроме того, у *R. etli* ген *nodN* имеет хромосомную локализацию.

Интересным результатом анализа стало попадание генов *nifB* и *fixC* в кластер 2 и их группировка вместе с геном *nodA*. В настоящее время известно, что гены, отвечающие за синтез корового Nod-фактора (*nodABC*), имеют неризобияльное происхождение. Так, А. Hirsch с соавт. (28) проанализировали последовательности бактериальных и грибных генов, кодирующих ферменты с каталитическими центрами, схожими с таковыми у продуктов генов *nodABC*. Было сделано предположение, что гены *nodB* и *nodC* могли быть получены ризобиями в процессе горизонтального переноса от грамположительных бактерий, у которых они участвуют в биогенезе муреиновой клеточной стенки. Кроме этого указанные гены могли быть получены клубеньковыми бактериями от грибов, активно синтезирующих хитиноподобные вещества. Известно, что многие грибы вступают с бактериями в тесные симбиотические отношения, при которых вполне возможен горизонтальный перенос генов. Однако для гена *nodA* гомологичных последовательностей в других организмах не обнаружили. Таким образом, ген *nodA* имеет иное происхождение, нежели *nodB* и *nodC* (28). В связи с этим его кластеризация вместе с генами *nifB* и *fixC* может свидетельствовать об их изначально совместном горизонтальном переносе из не идентифицированных пока организмов.

Важно отметить, что на специфику кластеризации генов на метаде-реве, несомненно, должно влиять не только их расположение у современных форм ризобий, но и происходившие ранее эволюционные события, включая независимые перестройки геномов (см. рис. 1), причем свою локализацию могли изменять как отдельные гены, так и содержащие их участки генома. В качестве примера таких перестроек можно привести инверсию *nif/fix* региона у *N. galegae* bv. *orientalis* (18) и транслокацию генов *nifHDKEN* у *R. leguminosarum* bv. *viciae* (11). Несомненно, что геномные перестройки оказывают существенное воздействие на порядок кластеризации генов, однако степень их влияния в рамках примененной модели недостаточно ясна. Существенную роль в возникновении различий в топологиях двух генов, вероятно, играет и неодинаковая скорость их эволюции.

Полученные в нашем исследовании результаты не согласуются с данными, полученными методами молекулярной филогении. Ранее было

показано, что полиморфизм *nod*-генов сформировался независимо от дивергенции коровых элементов бактериального генома, однако коррелирует с таксономией хозяев, тогда как полиморфизм *nif/fix*-генов коррелирует с дивергенцией коровой части генома бактерий, но не связан с дивергенцией хозяев (29). В нашем исследовании метод метадеревьев показал независимость эволюционной истории как *nod*-, так и *nif/fix*-генов от хромосомного маркера (ген 16S-rPHK). Возможно, использованный нами метод чувствительнее к влиянию эволюционных факторов, не связанных с ГПГ и не учитываемых в традиционных филогенетических подходах.

Таким образом, результаты филогенетического анализа *nod*- и *nif/fix*-генов, который мы провели с помощью метода построения и анализа метадеревьев, подтверждают выявленные ранее различия в эволюционных историях этих генов. Обнаружение двух кластеров, один из которых состоит преимущественно из *nif/fix*-генов (кластер 1), другой — из *nod*-генов (кластер 2), может отражать тот факт, что сопоставляемые группы генов возникли на разных этапах эволюции ризобий (в связи с чем они занимают различные участки на хромосомах или плаزمидях ризобий), а также экспрессию этих генов на разных стадиях развития симбиоза (инфицирование корневых волосков и освоение в растительную цитоплазму для *nod*-генов, симбиотическая фиксация азота — для *nif/fix*-генов). Присутствие некоторых *nod*-генов в кластере 1 и некоторых *nif/fix*-генов — в кластере 2, возможно, обусловлено тем, что на поздних этапах эволюции ризобий весь комплекс *sym*-генов переносился как единое целое в составе *Sym*-плазмид или геномных островов. Нельзя также исключить, что отсутствие корреляции между структурой кластеров, выявленных при анализе метадрева, и расположением изучаемых генов на генетических картах — следствие интенсивных внутригеномных перестроек, характерных для многих ризобий, и разной скорости эволюции первичной структуры у изучаемых генов.

Дальнейший анализ роли указанных эволюционных механизмов требует модификации использованного нами подхода, в частности раздельного анализа структур метадеревьев, построенных с использованием выборок штаммов, возникших на ранних стадиях эволюции ризобий (когда ГПГ был ограничен в связи с хромосомной локализацией *sym*-генов у *Bradyrhizobium*) либо на поздних стадиях (когда ГПГ был наиболее интенсивен и скорее всего захватывал весь комплекс *sym*-генов, локализованных на плазмидях *Rhizobium* и *Sinorhizobium*). Важная задача усовершенствования использованного метода состоит также в разработке статистических и биоинформационных критериев поддержки кластеров, выявляемых на метадревьях, поскольку на первом этапе работы мы применили для решения этой задачи лишь стандартные биометрические подходы.

Итак, филогенетический анализ *nod*- и *nif/fix*-генов, проведенный нами с помощью построения и сопоставления метадеревьев, подтвердил выявленные ранее различия в эволюционных историях этих генов. Наличие двух кластеров в метадреве, один из которых объединяет преимущественно *nif/fix*-, другой — *nod*-гены, может быть связано с тем, что этапы эволюции ризобий, на которых возникли сопоставляемые группы, не совпадали, а также с экспрессией этих генов на разных стадиях развития симбиоза. Отсутствие корреляции между структурой кластеров в метадреве и результатами генетического картирования, возможно, обусловлено внутригеномными перестройками и неодинаковой скоростью эволюции нуклеотидных последовательностей генов. Для усовершенствования примененного метода необходимо разработать статистические и биоинформационные критерии поддержки кластеров, выявляемых на метадревьях.

COMPARATIVE PHYLOGENETIC ANALYSIS OF SYMBIOTIC GENES OF DIFFERENT NODULE BACTERIA GROUPS USING THE METATREES METHOD

E.S. Karasev, E.P. Chizhevskaya, B.V. Simarov, N.A. Provorov, E.E. Andronov

All-Russian Research Institute for Agricultural Microbiology, Federal Agency of Scientific Organizations, 3, sh. Podbel'skogo, St. Petersburg, 196608 Russia, e-mail evgenii1991.karasev@gmail.com (corresponding author)
ORCID:

Karasev E.S. orcid.org/0000-0003-4240-9587

Provorov N.A. orcid.org/0000-0001-9091-9384

Chizhevskaya E.P. orcid.org/0000-0002-7715-8696

Andronov E.E. orcid.org/0000-0002-5204-262X

The authors declare no conflict of interests

Acknowledgements:

The experiments were carried out using equipment of the ARRIAM Center of Genomic Technologies, Proteomics and Cell Biology.

Supported by Russian Science Foundation, project № 14-26-00094П

Received December 12, 2016

doi: 10.15389/agrobiol.2017.5.995eng

Abstract

We applied modified phylogenetic analysis method of building meta-trees to study the evolutionary patterns of various groups of symbiotic genes (*nod*-genes control the formation of nodules and *nif/fix*-genes control symbiotic nitrogen fixation). The method consists in the pairwise comparison of topologies and construction combined dendrograms («meta-trees»), where the relative position of the two trees is a measure of corresponding gene phylogenies congruence. Homologues of 18 symbiotic genes (*nodABCDIJN*, *nifABDEHKN*, *fixABC*, *fdxB*) are present in each test organism (9 strains belong to the genera *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, *Sinorhizobium* and *Neorhizobium*). These genes were selected for the implementation of this method, as well as the gene 16S rRNA as a traditional taxonomic chromosome marker. We constructed and compared phylogenetic trees for all these genes and then calculated the pairwise similarity coefficients for their topologies. According to the obtained data we built a meta-tree, and there were two statistically distinct gene clusters identified within. Cluster 1 includes mainly *nif*- and *fix*-genes and cluster 2 — mostly *nod*-genes, that is related to the data of separate localization of these gene groups in the rhizobial genomes. The exception is the arrangement of *nifB* and *fixC* genes with *nodA* in cluster 2, as well as co-localization of *nodI* and *nifD* in cluster 1. During the identified clusters structure analysis we found strong relationship between the relative gene position and the characteristics of their genome localization in nodule bacteria. Importantly, the differences between clusters 1 and 2 are not expressed less clearly than the differences between *nod*- and *nif/fix*-gene groups. It is obvious that clusters 1 and 2 of our meta-tree reflect primarily different mechanisms of nodulation evolution and symbiotic nitrogen fixation associated with the independent origin of the relevant gene groups, and possibly, with their separate horizontal transfer between different groups of rhizobia. Further study of the symbiotic gene evolution in nodule bacteria requires improvements used phylogenetic analysis techniques, including separate analysis meta-tree for rhizobia, representing different stages of symbiosis evolution.

Keywords: phylogenetic analysis, meta-trees, nodule bacteria, symbiotic genes.

REFERENCES

1. Provorov N., Tikhonovich I., Andronov E., Belimov A., Borisov A., Vorob'ev N., Dolgikh E., Zhernakov A., Zhukov V., Kimeklis A., Kopat' V., Kurchak O., Onishchuk O., Safronova V., Sulima A., Chizhevskaya E., Chirak E., Shtark O. *Geneticheskie osnovy evolyutsii bakterii — simbiotov rastenii* [Evolution of bacterial symbionts of plants — genetic aspects]. St. Petersburg, 2016 (in Russ.).
2. Douni Dzh. V knige: *Rhizobiaceae* [In: *Rhizobiaceae*]. St. Petersburg, 2002: 417-434 (in Russ.).
3. Kaminski P., Batut Zh., Boistard P. V knige: *Rhizobiaceae* [In: *Rhizobiaceae*]. St. Petersburg, 2002: 465-492 (in Russ.).
4. Young J., Crossman L., Johnston A. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.*, 2006, 7(4): R34 (doi: 10.1186/gb-

- 2006-7-4-r34).
5. Reeve W., O'Hara G., Chain P. Complete genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM2304, an effective microsymbiont of the South American clover *Trifolium polymorphum*. *Stand Genomic Sci.*, 2010, 2(1): 66-76 (doi: 10.4056/signs.44642).
 6. Barnett M., Fisher R., Jones T. Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *PNAS USA*, 2001, 98(17): 9883-9888 (doi: 10.1073/pnas.161294798).
 7. Reeve W., Chain P., O'Hara G. Complete genome sequence of the *Medicago* microsymbiont *Ensifer (Sinorhizobium) medicae* strain WSM419. *Stand Genomic Sci.*, 2010, 2(1): 77-86 (doi: 10.4056/signs.43526).
 8. Kaneko T., Nakamura Y., Sato S. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.*, 2002, 9(6): 189-197 (doi: 10.1093/dnares./9.6.189).
 9. Kaneko T., Nakamura Y., Sato S. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.*, 2000, 7(6): 331-338 (doi: 10.1093/dnares./7.6.331).
 10. Girard M., Flores M., Brom S. Structural complexity of the symbiotic plasmid of *Rhizobium leguminosarum* bv. *phaseoli*. *J. Bacteriol.*, 1991, 173(8): 2411-2419.
 11. Usterman J., March J., Laine P.K. Genome sequencing of two *Neorhizobium galegae* strains reveals a *noeT* gene responsible for the unusual acetylation of the nodulation factors. *BMC Genomics*, 2014, 15: 500 (doi: 10.1186/1471-2164-15-500).
 12. Khaines M., Finan T. V knige: *Rhizobiaceae* [In: *Rhizobiaceae*]. St. Petersburg, 2002: 41-62 (in Russ.).
 13. Göttfert M., Röthlisberger S., Kündig C. Potential symbiosis-specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome. *J. Bacteriol.*, 2001, 183(4): 1405-1412 (doi: 10.1128/JB.183.4.1405-1412.2001).
 14. Sullivan J., Ronson C. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *PNAS USA*, 1998, 95(9): 5145-5149.
 15. González V., Santamaría R., Bustos P. The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons. *PNAS USA*, 2006, 103(10): 3834-3839 (doi: 10.1073/pnas.0508502103).
 16. Vorob'ev N., Provorov N. *Ekologicheskaya genetika*, 2013, 11: 73-85 (in Russ.).
 17. Provorov N., Andronov E. Evolution of root nodule bacteria: reconstruction of the speciation processes resulting from genomic rearrangements in a symbiotic system. *Microbiology*, 2016, 83(2): 131-139 (doi: 10.1134/S0026261716020156).
 18. Nye T. Trees of trees: an approach to comparing multiple alternative phylogenies. *Syst. Biol.*, 2008, 57(5): 785-794 (doi: 10.1080/10635150802424072).
 19. Nye T., Lio P., Gilks W. A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics*, 2005, 22(1): 117-119.
 20. Atkinson E., Palcic M., Hindsgaul O., Long S. Biosynthesis of *Rhizobium meliloti* lipooligosaccharide Nod-factors: NodA is required for an N-acyltransferase activity. *PNAS USA*, 1994, 91(18): 8418-8422.
 21. John M., Rohrig H., Schmidt J. *Rhizobium* NodB protein involved in nodulation signal synthesis is chitoooligosaccharide deacetylase. *PNAS USA*, 1993, 90: 625-629.
 22. Debelle F., Rosenberg C., Denarie J. The *Rhizobium*, *Bradyrhizobium* and *Azorhizobium* NodC proteins are homologous to yeast chitin synthases. *Mol. Plant-Microbe Interact.*, 1992, 5: 443-446.
 23. Peck M., Fisher R., Long S. Diverse flavonoids stimulate NodD₁ binding to *nod* gene promoters in *Sinorhizobium meliloti*. *J. Bacteriol.*, 2006, 188(15): 5417-5427 (doi: 10.1128/JB.00376-06).
 24. Evans I., Downie J. The NodI product of *Rhizobium leguminosarum* is closely related to ATP-binding bacterial transport proteins: nucleotide sequence of *nodI* and *nodJ* genes. *Gene*, 1986, 43: 95-101.
 25. Baev N., Shultze M., Barlier I. *Rhizobium nodM* and *nodN* genes are common nod genes: *nodM* encodes functions for efficiency of *nod* signal production and bacteroid maturation. *J. Bacteriol.*, 1992, 174(23): 7555-7565.
 26. Putnoky P., Grosskopf E., Cam D. *Rhizobium fix*-genes mediate at least two communication steps in symbiotic nodule development. *J. Cell Biol.*, 1988, 106: 597-607.
 27. Arnold W., Rump A., Klipp W. Nucleotide sequence of 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumonia*. *J. Mol. Biol.*, 1988, 203: 715-738.
 28. Hirsch A., Lum M., Downie J. What makes the rhizobia-legume symbiosis so special? *Plant Physiol.*, 2001, 127: 1484-1492 (doi: 10.1104/pp.010866).
 29. Provorov N., Vorob'ev N. *Geneticheskie osnovy evolyutsii rastitel'no-mikrobnogo simbioza* [Genetic aspects of plant-microbe symbiosis evolution]. St. Petersburg, 2012 (in Russ.).