

ЭВОЛЮЦИЯ СИСТЕМ РЕСТРИКЦИИ-МОДИФИКАЦИИ, СОДЕРЖАЩИХ ОДНУ ЭНДОНУКЛЕАЗУ РЕСТРИКЦИИ И ДВЕ ДНК-МЕТИЛТРАНСФЕРАЗЫ

© 2023 А.С. Фокина¹, А.С. Карягина^{2,3,4}, И.С. Русинов³, Д.М. Мошенский^{1,3},
С.А. Спирин^{3,5,6*}, А.В. Алексеевский^{1,3,6}

¹ Московский государственный университет имени М.В. Ломоносова,
факультет биоинженерии и биоинформатики, 119234 Москва, Россия

² Национальный исследовательский центр эпидемиологии и микробиологии
им. Н.Ф. Гамалеи Минздрава России, 123098 Москва, Россия

³ НИИ физико-химической биологии им. А.Н. Белозерского,
Московский государственный университет имени М.В. Ломоносова,
119992 Москва, Россия; электронная почта: sas@belozersky.msu.ru

⁴ Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии,
127550 Москва, Россия

⁵ НИУ «Высшая школа экономики», 109028 Москва, Россия

⁶ ФГУ ФНЦ НИИСИ РАН, 117218 Москва, Россия

Поступила в редакцию 28.11.2022

После доработки 15.01.2023

Принята к публикации 15.01.2023

Некоторые системы рестрикции-модификации содержат две ДНК-метилтрансферазы. В настоящей работе проведена классификация таких систем по присутствующим в белках систем каталитическим доменам, характерным для эндонуклеаз рестрикции и ДНК-метилтрансфераз. Подробно исследована эволюция белков из систем рестрикции-модификации, содержащих эндонуклеазный домен семейства NOV_C и две ДНК-метилтрансферазы, обе с доменами семейства DNA_methylase. Выяснено, что ДНК-метилтрансферазы таких систем разделяются на филогенетическом древе на две клады так, что ферменты одной системы оказываются в разных кладах, что свидетельствует о независимой эволюции двух метилтрансфераз. Обнаружены свидетельства множественных межвидовых горизонтальных переносов систем в целом, а также случаи переноса генов между системами.

КЛЮЧЕВЫЕ СЛОВА: система рестрикции-модификации, эндонуклеаза рестрикции, ДНК-метилтрансфераза, эволюционные домены, эволюция белков, горизонтальный перенос генов.

DOI: 10.31857/S0320972523020082, EDN: QGVWOU

ВВЕДЕНИЕ

Системы рестрикции-модификации (системы Р-М) защищают бактерии и археи от внедрения чужеродной ДНК, прежде всего от ДНК-содержащих бактериофагов [1]. Системы Р-М подразделяются на несколько типов [2], из которых самым хорошо изученным является тип II. Как правило, такие системы содержат белки с двумя ферментативными активностями: эндонуклеазу рестрикции (ЭР), узнающую специфическую последовательность ДНК и гидролизующую её, и ДНК-метилтрансферазу

(МТазу), модифицирующую хозяйскую ДНК в пределах узнаваемой последовательности, что препятствует её гидролизу ЭР. Модификация ДНК представляет собой метилирование цитозина с образованием С5-метилцитозина либо N4-метилцитозина, или же метилирование аденина с образованием N6-метиладенина.

Среди систем Р-М типа II можно выделить содержащие одну ЭР и две различных МТазы. Большинство таких систем относятся к подтипу IIA [3], т.е. их узнаваемая последовательность асимметрична (не является палиндромом), многие из них одновременно относятся

Принятые сокращения: МТаза – ДНК-метилтрансфераза; системы Р-М – системы рестрикции-модификации; ЭР – эндонуклеаза рестрикции.

* Адресат для корреспонденции.

к подтипу IIS, то есть гидролиз ДНК происходит вне узнаваемой последовательности [4]. Две отдельные МТазы обеспечивают метилирование обеих цепей асимметричных последовательностей, что предотвращает образование немодифицированных сайтов после репликации [5, 6]. Две МТазы могут метилировать различные типы оснований в различных цепях [5], как это было показано для систем MboII [7] и MnlII [8] — одна МТаз метилирует аденин в одной цепи, другая — цитозин в другой. В системе BfiI обе МТазы модифицируют цитозин с образованием C5-метилцитозина, но метилируемые цитозины расположены на разных цепях ДНК [9]; подобную же активность проявляют МТазы системы HgaI [10]. Такого рода системы могут представлять интерес с точки зрения их эволюции, а именно изучения связи между эволюционными путями двух МТаз в различных системах и роли двух отдельных МТаз в жизни бактериальной клетки и в ходе эволюции.

Целью работы было исследование эволюции систем Р-М, содержащих две МТазы. Была проведена классификация всех таких систем на основе семейств эндонуклеазных и метилтрансферазных доменов согласно банку Pfam [11]. Для подробного анализа мы выбрали один сравнительно небольшой класс таких систем. Для него мы показали, что две МТазы систем этого класса эволюционируют независимо друг от друга, и исследовали возможные горизонтальные переносы генов (ГПГ) компонентов системы.

МАТЕРИАЛЫ И МЕТОДЫ

Отбор из REBASE систем Р-М, включающих одну эндонуклеазу рестрикции и две МТазы. Из всех 499 311 белков REBASE [12] (версия 206 от 02.06.2022) были отобраны 2002 системы, включающие одну ЭР и две МТазы. Доменная архитектура всех белков была определена с использованием HMM-профилей из базы данных Pfam [11]. В дальнейшем «классом системы» мы называем список каталитических доменов ЭР и МТаз, найденных в последовательностях белков системы. Каталитическими мы считали 99 Pfam-доменов, встречающихся в ЭР, и 15 доменов, характерных для МТаз, отобранных на основании экспертной оценки, полный список см.: https://github.com/belozersky321/NOV_C-DNA_methylaseX2/blob/main/Domains.xlsx. Было обнаружено семь классов систем, удовлетворяющих следующим условиям: (1) класс содержит не менее

15 систем с подтвержденным составом из ЭР и двух МТаз (возможно наличие других белков в системе), (2) обе МТазы содержат каталитические домены из одного и того же семейства согласно Pfam. Эти семь классов представлены в таблице, название класса составлено из идентификаторов (согласно Pfam) семейств каталитических доменов ЭР и МТаз.

На рисунках для систем Р-М, ЭР и МТаз используются названия из REBASE. Многие белки в REBASE (в том числе все, которые исследовались в нашей работе) содержат в названии суффикс «Р», что означает, что они были предсказаны по последовательности генома и их функциональность не доказана экспериментально. Для части систем REBASE содержит информацию о сайтах узнавания, в том числе о предсказанных сайтах для некоторых из предсказанных белков. В ряде случаев указаны также места гидролиза и метилирования ДНК и тип модификации азотистого основания.

Поиск систем Р-М класса NOV_C/DNA_methylase×2, белки которых в REBASE не объединены в систему. В REBASE были найдены системы, состоящие из одной ЭР, содержащей домен NOV_C, и одной МТазы с доменом DNA_methylase. Для тех из таких систем, у которых ген ЭР непосредственно соседствовал в геноме с геном ещё одной МТазы, были найдены эволюционные домены этой МТазы. В случае, если среди этих доменов присутствовал домен семейства DNA_methylase, мы считали такую МТазу входящей в состав системы указанного класса и ошибочно не приписанной к системе в REBASE.

Филогенетические деревья строились в сервисе NGPhylogeny.fr [13] с использованием выравнивания программой Muscle и филогенетической реконструкции программой FastME со 100 бутстреп-репликами. Изображения деревьев редактировались и укоренялись в среднюю точку в iTOL [14]. Дерево систем строилось как консенсусное дерево ЭР и 3'-МТаз (см. в разделе «Результаты и обсуждение»), к которому были добавлены положения тех 5'-МТаз, чье положение на дереве отличалось от положения соответствующих 3'-МТаз.

Кластеризация систем. Последовательности белков систем класса NOV_C/DNA_methylase×2 кластеризовались программой CD-HIT [15] на уровне 98% идентичности при 100%-ном покрытии более короткой последовательности. Системы объединялись в один кластер, если кластеризовались все три белка, то есть в одном кластере оказывались ЭР этих систем, также в одном кластере — по одной МТазе из каждой системы, и ещё в одном кластере — по другой МТазе из каждой системы.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Классы систем Р-М с двумя метилтрансферазами. В таблице приведены семь классов систем, отобранных по указанным в разделе «Материалы и методы» параметрам. Все они содержат системы, включающие одну ЭР, две МТазы и, возможно, другие белки.

Системы Р-М этих классов содержат МТазы четырёх разных семейств (Pfam ID: N6_Mtase, MethyltransfD12, DNA_methylase, N6_N4_Mtase) и ЭР шести семейств (HSDR_N, RE_AlwI, NOV_C, HNH_2, RE_NgoFVII, RE_EcoO109I). Самый распространённый класс HSDR_N/N6_Mtase×2 содержит 1306 систем.

Для дальнейшего анализа был выбран класс NOV_C/DNA_methylase×2. Этот класс был выбран, так как количество систем в нём достаточно, чтобы проследить эволюционные закономерности, но не так велико, чтобы существенно затруднить визуальный анализ филогенетических деревьев и выравниваний. В дальнейшем планируется исследование остальных классов с использованием подходов, отработанных на классе NOV_C/DNA_methylase×2. Для исследования двух наиболее многочисленных классов потребуется автоматизация филогенетического анализа и выявления горизонтальных переносов, что также входит в планы авторов.

Белки систем класса NOV_C/DNA_methylase×2. Этот класс содержит 57 систем из REBASE. Из дальнейшего рассмотрения мы исключили систему, в которой ген ЭР оказался разделён транспозоном на две части (соответствующие белки в REBASE обозначены как Spn11901ORF903AP и Spn11901ORF903BP). К оставшимся 56 системам мы добавили ещё девять систем, для которых в REBASE была указана только одна МТаза, и включили в со-

став этих систем закодированную рядом «одиночную» МТазу (см. «Материалы и методы»), всего получилось 65 систем. Полученный список был сокращён за счёт оставления по одной системе из каждого 98%-ного кластера по всем белкам (см. «Материалы и методы»); таких кластеров оказалось 38, и вся дальнейшая работа проводилась с получившимся списком из 38 систем. Все кластеры включают системы из близких видов, за одним исключением, из одного рода, поэтому ограничение на 38 систем не должно отразиться на эволюционных выводах. Список систем см.: https://github.com/belozersky321/NOV_C-DNA_methylaseX2/blob/main/Systems.xlsx. В колонке «System in REBASE» на первом листе этой книги значением «ИСТИНА» отмечены системы, включающие, согласно REBASE, две МТазы, а значением «ЛОЖЬ» — девять систем, в чей состав мы включили МТазы, аннотированные в REBASE как одиночные.

ЭР систем этого класса характеризуются наличием домена из семейства, имеющего в банке Pfam идентификатор NOV_C, accession PF13020, название «Protein NO VEIN C-terminal». Ранее этот домен имел в Pfam идентификатор DUF3883 и с этим идентификатором упоминался в литературе [16]; многие ЭР систем нашего класса описаны в REBASE и GenBank как «DUF3883 domain-containing proteins». В последовательностях всех ЭР исследуемого класса диагностируется также домен BpuJI_N, который, согласно статье Sukackaite et al. [17], отвечает за распознавание нуклеотидной последовательности. Домены других семейств Pfam в последовательностях этих ЭР не находятся: все ЭР систем этого класса содержат два эволюционных домена, BpuJI_N в N-концевой части и NOV_C — в C-концевой части белка.

Классы систем Р-М из REBASE, включающие одну ЭР и две МТазы

Класс	Число систем Р-М	Тип метилирования
HSDR_N/N6_Mtase×2	1306	N6-аденин
RE_AlwI/MethyltransfD12×2	494	N6-аденин
NOV_C/DNA_methylase×2	57	C5-цитозин
HNH_2/DNA_methylase×2	42	C5-цитозин
RE_AlwI/N6_N4_Mtase×2	30	N4-цитозин, N6-аденин
RE_NgoFVII/DNA_methylase×2	28	C5-цитозин
RE_EcoO109I/DNA_methylase×2	17	C5-цитозин

Несмотря на то что в описании домена NOV_C в Pfam указано, что его функция неизвестна, в случае ЭР из систем Р-М домен NOV_C, несомненно, является каталитическим доменом с эндонуклеазной активностью. Это следует из сравнения последовательностей ЭР с этим доменом с последовательностью ЭР BpuJI. Для BpuJI в работе Sukackaite et al. [18] экспериментально показана эндонуклеазная активность и определены остатки активного сайта: D348, E367, K369, входящие в домен NOV_C. В последовательности ЭР BpuJI диагностируются домены BpuJI_N и NOV_C, однако в REBASE для неё не указаны МТазы, входящие с ней в одну систему, поэтому нельзя достоверно сказать, относится ли соответствующая система Р-М к нашему классу. Нами было построено совместное выравнивание последовательностей 38 ЭР из систем Р-М класса NOV_C/DNA_methylase×2 и ЭР BpuJI. Три остатка активного центра ЭР BpuJI оказались в абсолютно консервативных колонках выравнивания, что свидетельствует в пользу того, что у всех этих ЭР сохранена эндонуклеазная активность.

МТазы данного класса содержат домен DNA_methylase (PF00145, C-5 cytosine-specific DNA_methylase). Этот домен характерен для всех МТаз, метилирующих цитозин с образованием C5-метилцитозина. Других достоверных находок доменов Pfam в МТазы систем этого класса нет, а домен DNA_methylase, за несколькими исключениями, занимает практически весь белок. Таким образом, подавляющее большинство МТаз данного класса, по всей видимости, содержат только один домен.

Семь систем этого класса содержат С-белки — факторы транскрипции. В этих системах встречаются С-белки с находками профилей семейств Pfam HTH_19, HTH_26 и HTH_3. В большинстве найденных С-белков с достаточно высоким весом находятся все три профиля. В N-концевой части последовательности одной из МТаз одной из систем этого класса, а именно M2.Gva7778BORF1255P, найден домен HTH_17, также характерный для факторов транскрипции, однако вес этой находки ниже порога достоверности. Возможно, эта МТаз имеет дополнительную функцию регуляции транскрипции белков системы. Три системы содержат никазы с каталитическим доменом Vsr.

Для восьми систем класса NOV_C/DNA_methylase×2 в REBASE хотя бы для одного белка системы указана последовательность (сайт) узнавания CCCGT, для 13 других — CCCGC, для остальных сайт узнавания в REBASE не

указан. При этом для ЭР Bal380ORF12195P сайт CCCGT, указанный в REBASE, предсказывается с высокой достоверностью благодаря почти 100%-ной идентичности её последовательности с ЭР BpuJI, для которой такой сайт был подтверждён экспериментально [18]. ЭР BpuJI гидролизует ДНК вне сайта узнавания — непосредственно после, через один или через два нуклеотида после узнаваемой последовательности. В систему FauTH167ORF820 входит МТазы M2.FauTH167ORF820P — близкий гомолог МТазы M.FauI, сайт которой, CCCGC, тоже был определён экспериментально в работе Degtyarev et al. [19]. Тем не менее в REBASE для белков системы FauTH167ORF820 не указан сайт узнавания.

По данным REBASE, среди МТаз с сайтом узнавания CCCGT позиции метилируемых цитозinov в сайтах узнавания экспериментально определены только для метилтрансфераз системы BscGI. Эта система, подобно исследованному нами, содержит две МТазы: M1.BscGI, которая метилирует второй с 5'-конца цитозин в последовательности CCCGT, и M2.BscGI, метилирующую единственный цитозин в комплементарной последовательности ACGGG. К сожалению, данные о последовательностях генов и белков системы BscGI в REBASE и других банках данных отсутствуют, и это не позволяет проверить принадлежность этой системы исследуемому классу. Относительно специфичности метилтрансфераз с сайтом CCCGC экспериментальных данных в REBASE не представлено. Имеются лишь данные об ЭР Bme585I, BstFZ438I, FauI и SmaI, имеющих сайт узнавания CCCGC, при этом гидролиз одной цепи ДНК происходит через четыре нуклеотида, а другой цепи — через шесть нуклеотидов от узнаваемой последовательности («CCCGC (4/6)» в стандартных для систем Р-М обозначениях). Аминокислотные последовательности этих ЭР в REBASE также не представлены. Для ЭР FauI и SmaI представлены данные о чувствительности к метилированию, в частности, о защите от гидролиза при метилировании CG-специфичной метилтрансферазой SssI.

Филогения эндонуклеаз рестрикции. На рис. 1 представлено филогенетическое дерево 38 ЭР из систем класса NOV_C/DNA_methylase×2, на котором обозначены бутстреп-поддержка ветвей, узнаваемые последовательности ЭР, согласно REBASE, и принадлежность ЭР органнизмам из разных отделов, а также отмечены ЭР, входящие в состав систем с С-белком и/или никазой.

На рис. 1 обращают на себя внимание несколько фактов.

2. Четыре ЭР из систем с С-белками образуют единую кладу, включающую ещё ЭР FbaHQM9ORFGP, входящую в систему без аннотированного в REBASE С-белка. Однако

непосредственно перед генами этой системы находится ген С-белка C.FbaHQM9ORFFP, приписанный в REBASE к другой системе, в которой, кроме этого белка, нет никаких других. Можно предположить, что для этих пяти систем (все они — из бактерий отдела Bacteroidetes) важную роль играет регуляция С-белками. Три других системы с С-белками не родственны этим пяти и друг другу. Три системы, относящиеся к двум разным прикорневым кладам, содержат никазы семейства Vpr. Роль этих белков в обсуждаемых системах Р-М неясна.

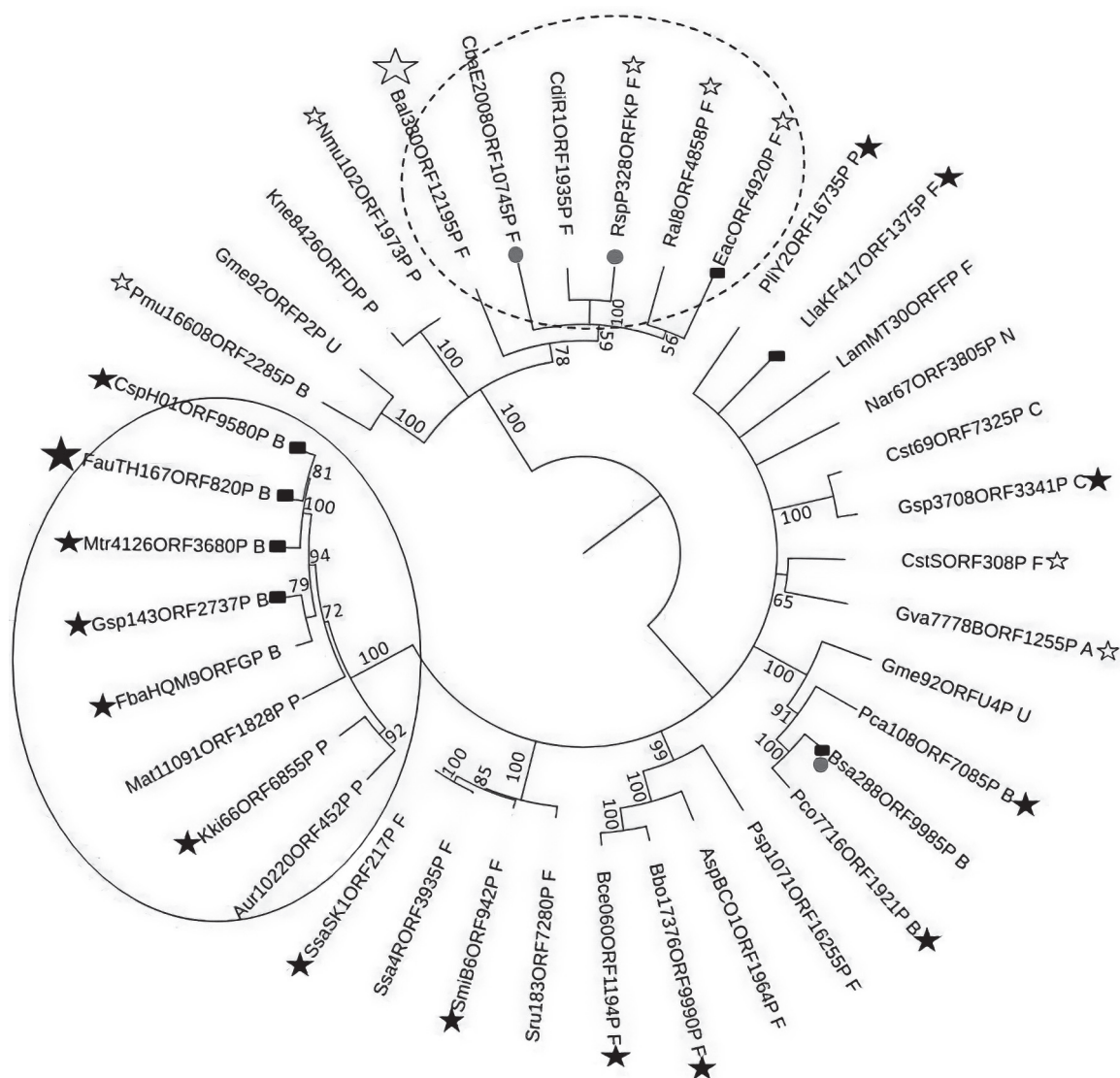


Рис. 1. Филогенетическое дерево 38 ЭР класса NOV_C/DNA_Methylase×2. Названия ферментов из систем с сайтами, указанными в REBASE, отмечены: чёрной звездой – CCCGC и контурной звездой – CCCGT. Звёзды большего размера отмечают ЭР из систем, чьи белки имеют близкие гомологи с экспериментально определёнными сайтами. Овалами обведены клады из ЭР с относительно достоверной аннотацией сайтов по гомологии, прерывистый овал соответствует сайту CCCGT, сплошной – сайту CCCGC. Белки из систем с С-белком отмечены чёрными прямоугольниками, белки из систем с никазой – серыми кружками. Буквы после названия белка соответствуют отделу, к которому принадлежит хозяин системы: F – Firmicutes, C – Cyanobacteria, P – Proteobacteria, B – Bacteroidetes, A – Actinobacteria, N – Nanoarchaeota (Archaea) и U – кишечный метабеном. Ветви, поддерживаемые менее чем половиной бутстреп-реплик, сжаты (т.е. удалены путём сжатия в точку). Числа на ветвях обозначают проценты бутстреп-реплик, поддерживающих данную ветвь.

3. ЭР из отделов Bacteroidetes, Firmicutes и Proteobacteria не группируются на дереве в одну кладу. Те клады дерева, которые содержат ЭР из близких бактерий, включают намного меньше ЭР, чем имеется бактерий той же степени родства с секвенированными геномами. Эти два обстоятельства указывают, на наш взгляд, на особенности эволюции и экологии систем данного класса, а именно: такие системы активно участвуют в ГПП, но в то же время легко утрачиваются их хозяевами.

Филогения МТаз. На рис. 2 изображено филогенетическое дерево МТаз класса NOV_C/DNA_Methylase×2. МТазы из 38 систем формируют две клады со 100%-ной бутстреп-поддержкой их разделения, причём МТазы одной системы всегда принадлежат разным клadam. Это означает, что две МТазы систем данного класса эволюционировали независимо. Непалиндромность сайтов узнавания систем требует различной функциональной роли двух МТаз (вероятно, узнавая один и тот же сайт, две МТазы модифицируют основания разных цепей ДНК), что объясняет такую особенность эволюции. Гены белков всех систем класса NOV_C/DNA_Methylase×2 всегда расположены на одной цепи ДНК в порядке: 5'-ген

МТазы – ген ЭР – ген МТазы-3'. При этом все МТазы, чьи гены расположены до гена соответствующей ЭР, относятся к одной клade дерева, а все МТазы, чьи гены расположены после гена ЭР – к другой. На рис. 2 эти клады обозначены, соответственно, как 5'М и 3'М. Далее мы называем МТазы первой клады 5'-МТазами, а МТазы второй – 3'-МТазами. Префиксы «М1» и «М2» в названиях МТаз не связаны с разделением на клады и отражают порядок расположения генов МТаз в записи банка последовательностей, который может отличаться от порядка в направлении транскрипции, что приводит к тому, что белки с разными префиксами оказываются в одной клade.

Топология клад 3'-МТаз практически (с точностью до небольших различий, которые, вероятнее всего, вызваны ошибками филогенетической реконструкции) совпадает с топологией дерева ЭР (см. рис. 1). В то же время в клade 5'-МТаз четыре белка, обведённые рамками на рис. 2, сближены с белками других систем, чем на дереве ЭР и в клade 3'-МТаз. Из этого можно сделать вывод, что в эволюционной истории систем нашего класса были события переноса белков в другую систему.

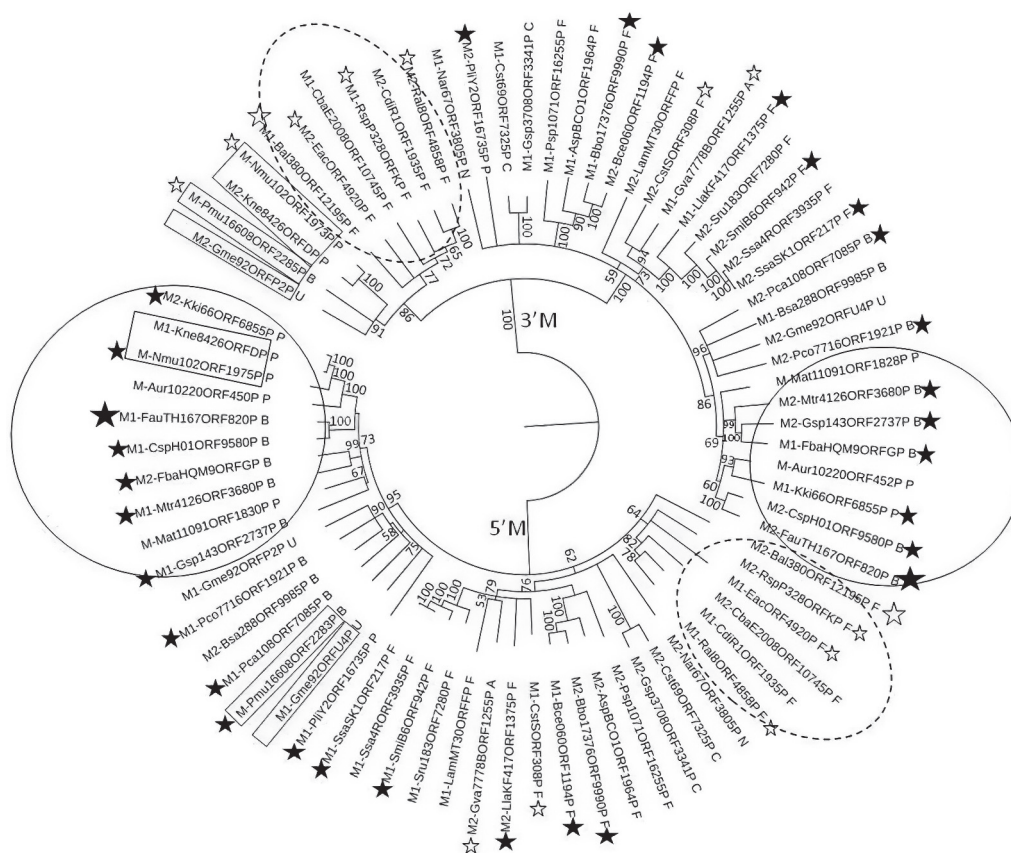


Рис. 2. Филогенетическое дерево МТаз класса NOV_C/DNA_Methylase×2. Три группы МТаз, чьё положение различается в клдах 5'М и 3'М, обведены рамками. Другие обозначения как на рис. 1

Реконструкция филогении систем Р-М.

Чтобы лучше понять эволюцию систем данного класса, мы реконструировали общее филогенетическое дерево систем (рис. 3), основываясь на деревьях ЭР, 5'- и 3'-МТаз. На этом дереве ветви, различающиеся в трёх деревьях предположительно вследствие ошибок реконструкции, сжаты, а случаи предположительно обмена генами между системами отмечены различным положением разных ферментов и стрелками.

Все переносы генов между системами затрагивают системы, ЭР и 3'-МТазы которых на соответствующих деревьях находятся в одной кладе (верхние четыре системы на рис. 3). Эта клада является сестринской по отношению к кладе с относительно достоверно предсказанным сайтом CCCGT (обведена пунктирным овалом на рис. 1, 2 и 3, ЭР Bal380ORF12195P из этой клады имеет почти 100%-ную идентичность с экспериментально охарактеризованной ЭР BpuJI). При этом 5'-МТазы этих систем

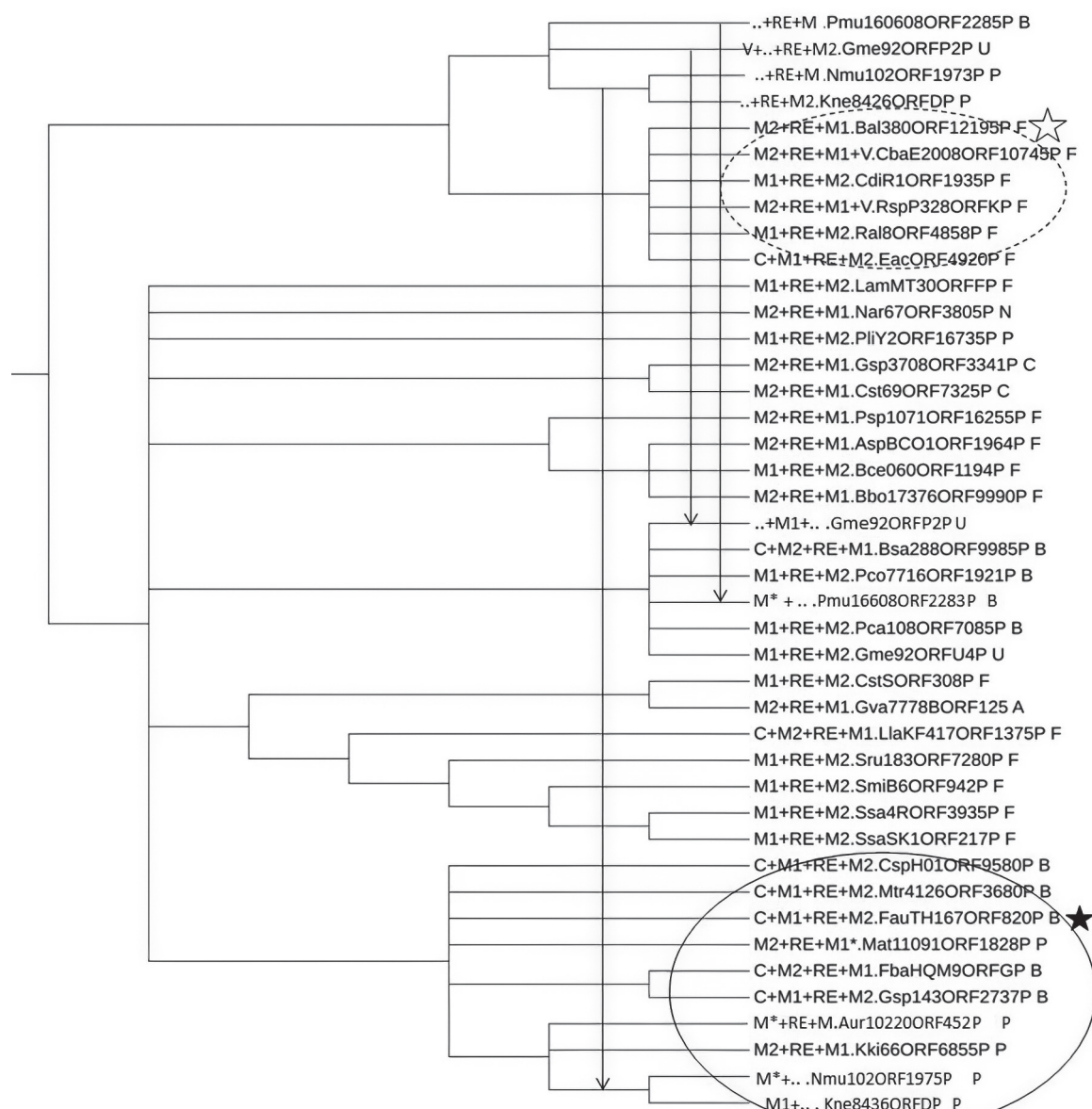


Рис. 3. Филогенетическое дерево систем Р-М класса NOV_C/DNA_Methylase×2. Перед названием каждой системы приведены обозначения белков системы: М, М1 и М2 – для МТаз, чьи названия в REBASE имеют соответствующие префиксы; RE – для ЭР; С – для С-белков, V – для никак. Порядок обозначений белков каждой системы соответствует порядку генов систем на кодирующей цепи ДНК, то есть 5'-МТазы – слева, 3'-МТазы – справа. Звёздочка (*) после обозначения белка указывает, что в REBASE этот белок не приписан к системе. Стрелки обозначают предполагаемые события ГПГ. Остальные обозначения как на рис. 1 и 2

оказываются ближе к 5'-МТазам систем с предсказанным сайтом CCCGC. Это можно объяснить тем, что ЭР и 3'-МТазы этих четырёх систем объединились в одну систему с 5'-МТазой относительно далёких систем, например, благодаря повышенной способности участка ДНК, кодирующего ЭР и 3'-МТазу, к горизонтальным переносам. Число событий таких переносов невозможно точно установить по имеющимся данным, но нам представляются наиболее вероятными три события (обозначены стрелками на рис. 3), а именно: переносы белков системы Pmu16608ORF2285P, независимо — системы Gme92ORFP2P и независимо — общего предка систем Nmu102ORF1973P и Kne8426ORFDP. Возможен также сценарий с двумя событиями, включающий перенос белков общего предка систем Pmu16608ORF2285P и Gme92ORFP2P, или же с четырьмя независимыми событиями. Неясным остаётся различие сайтов узнавания между ближайшими родственниками ЭР и 3'-МТаз, с одной стороны, и 5'-МТаз — с другой, из систем, участвующих в переносе. Можно предложить два объяснения этому. Во-первых, не исключено, что все системы нашего класса, кроме обведённых на рис. 3 пунктирным овалом, имеют специфичность CCCGC. Во-вторых, можно предположить более широкую специфичность (например, CCCGY) для всех или большинства 5'-МТаз класса. Это предположение не противоречит экспериментально установленной специфичности CCCGC для МТазы FauI, родственной 3'-МТазе M2.FauTH167ORF820P.

Роль горизонтальных переносов генов в эволюции систем данного класса. Анализ филогенетических деревьев белков, составляющих системы, вместе с рассмотрением таксономии бактерий — их хозяев, показывает, что в эволюции систем данного класса происходили события двух родов.

Во-первых, это многочисленные горизонтальные переносы системы в целом. Например, системы из бактерий отдела Firmicutes представлены на дереве тремя неродственными кладами (включающими 4, 5 и 6 систем) и ещё двумя изолированными системами. Такое возможно только при условии активного горизонтального переноса систем в целом, сопряжённого с быстрой их утратой. Это подтверждается и другими наблюдениями. Так, в группе шести систем из отдела Firmicutes, обведённой пунктирной линией на рис. 3, хозяева этих систем относятся к трём разным семействам: Bacillaceae, Peptostreptococcaceae, Oscillospiraceae (см. https://github.com/belozersky321/NOV_C-DNA_methylaseX2/blob/main/Systems.xlsx).

Представляется маловероятным, что эти системы унаследованы от общего предка, учитывая отсутствие гомологичных систем в подавляющем большинстве бактерий. Следовательно, системы, скорее всего, получены в результате по крайней мере трёх событий горизонтального переноса.

Во-вторых, мы наблюдаем несколько (от двух до четырёх, см. предыдущий подраздел) событий приобретения системой белков из другой системы (вертикальные стрелки на рис. 3). ЭР и 3'-МТазы четырёх систем, демонстрирующих признаки таких событий, образуют на дереве отдельную кладу (верхние четыре системы на рис. 3). Невозможно установить, представляли ли собой эти события горизонтальный перенос гена 5'-МТазы или же совместный горизонтальный перенос пары генов ЭР и 3'-МТазы. Можно отметить, что хозяева всех этих систем относятся к микробиому человека.

ЗАКЛЮЧЕНИЕ

В настоящей работе выявлены классы систем Р-М, содержащие одну ЭР и две МТазы; классы выделяются в соответствии с каталитическими доменами, присутствующими в ЭР и МТазе. Подробно исследованы системы класса NOV_C/DNA_Methylase×2. МТазы этого класса формируют на филогенетическом дереве две прикорневые клады, при этом ферменты одной системы попадают в разные клады, что говорит об их независимом происхождении. Скорее всего, это отражает разную функциональную роль двух МТаз в системах, а именно метилирование разных цепей ДНК в асимметричном сайте.

Системы этого класса, общим числом 65, присутствуют в геномах бактерий пяти отделов и одной археи, что говорит о большой роли ГПГ и потери генов в эволюции этих систем. Обнаружены свидетельства переноса генов между системами с предположительно различной специфичностью узнавания.

Финансирование. Работа выполнена при финансовой поддержке Российского научного фонда (грант № 21-14-00135).

Вклад авторов. А.В. Алексеевский — концепция и руководство работой; А.С. Фокина — компьютерный анализ данных; И.С. Русинов — подготовка данных по доменам в белках REBASE; Д.М. Мошенский — помощь в поиске генов белков; А.С. Фокина, С.А. Спирин, А.С. Карягина — обсуждение

результатов, написание и редактирование текста статьи.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Соблюдение этических норм. Настоящая статья не содержит описания каких-либо исследований с участием людей или животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

- Williams, R. J. (2003) Restriction endonucleases: Classification, properties, and applications, *Mol. Biotechnol.*, **23**, 225-244, doi: 10.1385/mb:23:3:225.
- Roberts, R. J. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes, *Nucleic Acids Res.*, **31**, 1805-1812, doi: 10.1093/nar/gkg274.
- Pingoud, A., Wilson, G. G., and Wende, W. (2014) Type II restriction endonucleases – a historical perspective and more, *Nucleic Acids Res.*, **42**, 7489-7527, doi: 10.1093/nar/gku447.
- Szybalski, W., Kim, S. C., Hasan, N., and Podhajski, A. J. (1991) Class-IIS restriction enzymes – a review, *Gene*, **100**, 13-26, doi: 10.1016/0378-1119(91)90345-c.
- Madhusoodanan, U. K., and Rao, D. N. (2010) Diversity of DNA methyltransferases that recognize asymmetric target sequences, *Crit. Rev. Biochem. Mol. Biol.*, **45**, 125-145, doi: 10.3109/10409231003628007.
- Vasu, K., and Nagaraja, V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense, *Microbiol. Mol. Biol. Rev.*, **77**, 53-72, doi: 10.1128/mmbr.00044-12.
- Furmanek-Blaszk, B., Boratynski, R., Zolcinska, N., and Sektas, M. (2009) M1.MboII and M2.MboII type IIS methyltransferases: Different specificities, the same target, *Microbiology*, **155**, 1111-1121, doi: 10.1099/mic.0.025023-0.
- Kriukiene, E., Lubiene, J., Lagunavicius, A., and Lubys, A. (2005) MnlI – The member of H-N-H subtype of Type IIS restriction endonucleases, *Biochim. Biophys. Acta*, **1751**, 194-204, doi: 10.1016/j.bbapap.2005.06.006.
- Sapranaukas, R., Sasnauskas, G., Lagunavicius, A., Vilkaitis, G., Lubys, A., and Siksnys, V. (2000) Novel subtype of type IIS restriction enzymes, *J. Biol. Chem.*, **275**, 30878-30885, doi: 10.1074/jbc.m003350200.
- Sugisaki, H., Kita, K., and Takanami, M. (1989) The FokI restriction-modification system, *J. Biol. Chem.*, **264**, 5757-5761, doi: 10.1016/s0021-9258(18)83614-6.
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., and Bateman, A. (2020) Pfam: The protein families database in 2021, *Nucleic Acids Res.*, **49**, D412-D419, doi: 10.1093/nar/gkaa913.
- Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2014) REBASE – a database for DNA restriction and modification: Enzymes, genes and genomes, *Nucleic Acids Res.*, **43**, D298-D299, doi: 10.1093/nar/gku1046.
- Lemoine, F., Correia, D., Lefort, V., Doppl-Azeroual, O., Mareuil, F., Coen-Boulakia, S., and Gascuel, O. (2019) NGPhylogeny.fr: new generation phylogenetic services for non-specialists, *Nucleic Acids Research*, **47**, W260-W265, doi: 10.1093/nar/gkz303.
- Letunic, I., and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation, *Nucleic Acids Res.*, **49**, W293-W296, doi: 10.1093/nar/gkab301.
- Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinformatics*, **22**, 1658-1659, doi: 10.1093/bioinformatics/btl158.
- Steczkiwicz, K., Muszewska, A., Knizewski, L., Rychlewski, L., and Ginalski, K. (2012) Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily, *Nucleic Acids Res.*, **40**, 7016-7045, doi: 10.1093/nar/gks382.
- Sukackaite, R., Grazulis, S., Bochtler, M., and Siksnys, V. (2008) The recognition domain of the BpuJI restriction endonuclease in complex with cognate DNA at 1.3-Å resolution, *J. Mol. Biol.*, **378**, 1084-1093, doi: 10.1016/j.jmb.2008.03.041.
- Sukackaite, R., Lagunavicius, A., Stankevicius, K., Urbanke, C., Venclovas, Č., and Siksnys, V. (2007) Restriction endonuclease BpuJI specific for the 5'-CCCGT sequence is related to the archaeal Holliday junction resolvase family, *Nucleic Acids Res.*, **35**, 2377-2389, doi: 10.1093/nar/gkm164.
- Degtyarev, S. K., Netesova, N. A., Chizhikov, V. E., and Abdurashitov, M. (1998) Cloning and characterization of the gene encoding M.FauI DNA methyltransferase, *Biol. Chem.*, **379**, 567-568.

EVOLUTION OF RESTRICTION-MODIFICATION SYSTEMS WITH ONE RESTRICTION ENDONUCLEASE AND TWO DNA METHYLTRANSFERASES

A. S. Fokina¹, A. S. Karyagina^{2,3,4}, I. S. Rusinov³, D. M. Moshensky^{1,3},
S. A. Spirin^{3,5,6*}, and A. V. Alexeevski^{1,3,6}

¹ Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University,
119234 Moscow, Russia

² N.F. Gamaleya National Research Center for Epidemiology and Microbiology,
Ministry of Healthcare of the Russian Federation, 123098 Moscow, Russia

³ Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University,
Moscow, 119992 Moscow, Russia; e-mail: sas@belozersky.msu.ru

⁴ All-Russia Research Institute of Agricultural Biotechnology, 127550 Moscow, Russia

⁵ National Research University Higher School of Economics, 109028 Moscow, Russia

⁶ Federal Science Center System Research Institute of the Russian Academy of Sciences, 117218 Moscow, Russia

Some restriction-modification systems contain two DNA methyltransferases. In the present work, we have classified such systems according to the families of catalytic domains present in restriction endonucleases and both DNA methyltransferases. The evolution of restriction-modification systems of one class was studied in detail. Systems in this class include an endonuclease with a NOV_C family domain and two DNA methyltransferases, both with DNA_methylase family domains. The phylogenetic tree of DNA methyltransferases from systems of this class consists of two clades of the same size. Two DNA methyltransferases of each restriction-modification system of the class belong to different clades. This indicates independent evolution of the two methyltransferases. We detected multiple cross-species horizontal transfers of systems as a whole, as well as cases of gene transfer between systems.

Keywords: restriction modification system, DNA methyltransferase, restriction endonuclease, evolutionary domains, protein evolution, horizontal gene transfer