

УДК 575.852

## ЭВОЛЮЦИЯ ТРАНСКРИПЦИОННОЙ РЕГУЛЯЦИИ СИНТЕЗА ПРОЛИНА У ГАММА-ПРОТЕОБАКТЕРИЙ

К.В. Лопатовская, К.Ю. Горбунов, Л.Ю. Русин, А.В. Селиверстов, В.А. Любецкий

(Институт проблем передачи информации имени А.А. Харкевича УРАН;  
лаборатория математических методов и моделей в биоинформатике;  
e-mail: kristina@iitp.ru)

Найден консервативный мотив перед генами *proA* и *proB* у многих  $\gamma$ -протеобактерий, наиболее широко представленный у *Pseudomonas*, *Marinobacter* и *Shewanella*. Перед многими генами имеются пары сайтов кооперативного связывания белка обычно из 8 п.н. На основе разработанного алгоритма предположено, что соответствующим транскрипционным фактором является белок из семейства TetR, ортологичный белку NP\_249058 из *P. aeruginosa* PAO1. Построено дерево видов, имеющих найденную нами регуляцию генов *pro* и три вложения в него: дерева самих генов *pro*, дерева доменов транскрипционных факторов и дерева сайтов связывания этих факторов. Выведен сценарий совместной эволюции взаимодействующих элементов: генов, белковых факторов регуляции и сайтов связывания.

**Ключевые слова:** синтез пролина, белковое семейство TetR, регуляция транскрипции,  $\gamma$ -протеобактерии, эволюция, горизонтальный перенос генов.

У *Pseudomonas aeruginosa* экспериментально была показана зависимость транскрипции гена *proA* синтеза пролина от его концентрации [1, 2]. Гены *proA* и *proB* кодируют два фермента ( $\gamma$ -глутамилкиназу и  $\gamma$ -глутамилфосфатредуктазу). Предварительный анализ небольшого числа геномов у небольшого числа представителей родов *Pseudomonas* и *Shewanella* позволил выявить потенциальный консервативный мотив перед этими генами [3]. Нами проведен поиск такого мотива перед генами *proA* и *proB* у всех протеобактерий.

### Методы и данные

Данные взяты из базы GenBank (www.ncbi.nlm.nih.gov). Для поиска фактора связывания, соответствующего найденному мотиву, использована разработанная авторами программа, которая основана на сравнении филогенетического профиля сайтов с филогенетическими профилями всех бактериальных белков [4, 5]. В кратком изложении программа состоит в следующем. Для такого сравнения профиль мотива и профиль каждого белка представляются в виде векторов. Филогенетический профиль мотива в  $i$ -й позиции содержит +1, если соответствующий сайт присутствует в  $i$ -м геноме; или -1, если регуляторный сайт отсутствует в  $i$ -м геноме. Вместо такой дискретной характеристики присутствия сайта в  $i$ -й позиции можно указывать в ней качество соответствующего сайта. Профиль производного белка в  $i$ -й позиции содержит сумму характеристик сходства соответствующих аминокислот (по матрице BLOSUM62) в выравнивании этого

белка с его ближайшим гомологом из  $i$ -го генома. Характеристика *cos близости векторов* вычисляется как косинус угла между этими векторами. Качество сайта длины 8 п.н. вычислялось как *расстояние* между консенсусом мотива и сайтом. Для этого консенсус и сайт отождествлялись с векторами в 32-мерном пространстве и затем определялось расстояние между векторами как сумма модулей разностей координат. Вектор, соответствующий консенсусу, состоит последовательно из 4 частот встречаемости нуклеотидов в каждой из восьми позиций сайта. Вектор, соответствующий сайту, определяется аналогично: значения частот равны одной единице и трем нулям.

### Результаты

Найден консервативный мотив перед генами *proA* и *proB* у многих  $\gamma$ -протеобактерий, наиболее широко мотив представлен у родов *Pseudomonas*, *Marinobacter* и *Shewanella*. (Иллюстративный материал данной работы, занимающий большой объем, представлен на сайте Proline лаборатории математических методов и моделей в биоинформатике Института проблем передачи информации имени А.А. Харкевича УРАН [4].) На рис. 1 [4] показано дерево видов, у которых предсказано белок-ДНК взаимодействие, обеспечивающее регуляцию. Гены *pro* обычно образуют оперон *proBA*, но у *Pseudomonas* они транскрибируются независимо, нами не найден убедительный случай регуляции транскрипции гена *proB*, не входящего в состав такого оперона. Соответствующее множественное выравнивание приведено на рис. 2 [4]. Консервативные сайты связы-

вания белка с ДНК имеют длину 8 п.н. и обычно повторяются дважды на расстоянии от 10 до 16 п.н. К 3'-концу консервативного сайта на смысловой цепи примыкает слабо консервативный участок, одинаковый по длине у всех изученных видов (рис. 3 [4]). На основе разработанного нами метода предположено, что соответствующим транскрипционным фактором является белок из семейства TetR, ортологичный белку NP\_249058 из *P. aeruginosa* PAO1 (см. табл. 2 [4]). Этот белок на N-конце содержит гипотетический ДНК-связывающий домен, характерный для семейства TetR (Pfam номер PF00440, для участка от 11-й до 57-й аминокислоты с уровнем *expect value* отличия от консенсуса  $1,5 \cdot 10^{-13}$ ) [6]. Ортологи этого белка с достаточно высоким значением *expect value* присутствуют у всех видов, в которых предсказана регуляция; небольшое число исключений относится к низко консервативным сайтам. Однако найденный белок имеет паралог NP\_253746 в том же виде *P. aeruginosa*; близкие гомологи этих белка и паралога обнаружены у *Vibrio* и у других родов; в том числе у видов, не содержащих предсказанного нами мотива, табл. 2 [4]. Можно думать, что в последнем случае эти гомологи имеют другую специфичность. Алгоритмом, описанным в работе [7], построены три вложения в дерево *S* видов: дерева *G* самих генов *pro* (рис. 4 [4]); дерева *F* доменов транскрипционных факторов,

рис. 5 [4]; дерева *R* сайтов связывания этих факторов, на рис. 6 [4] показан соответствующий лес [7]. Сайты вместе с областью промотора имеют небольшую длину (около 52 п.н.), поэтому их потери и возникновения могут происходить значительно чаще, чем соответствующих факторов и регулируемых генов. В нашей модели возникновение рассматривалось как перенос из внешней группы. Полученный сценарий совместной эволюции генов, факторов регуляции и сайтов связывания описан ниже и показан на рис. 1, 4–6 [4]. В этом сценарии наиболее важными являются события горизонтального переноса и связанные с ними. Понятие “труба” введено в [8] как ребро дерева видов вместе с эволюционными событиями, происшедшими на нем. Имеются пять событий переноса генов, шесть событий переноса факторов и два переноса сайтов с сохранением или возникновения сайтов вместе с их окрестностью (из-за короткой длины трудно отличить перенос сайта от его возникновения). Таким образом, с помощью оригинальных алгоритмов описана совместная эволюция найденных нами потенциальных сайтов связывания и регуляторных факторов генов *pro* и самих генов вдоль дерева видов. Насколько авторам известно, такая совместная эволюция всех компонент системы регуляции на основе белок-ДНК взаимодействий ранее не была представлена для какого-либо гена.

## СПИСОК ЛИТЕРАТУРЫ

1. Krishna R.V., Beilstein P., Leisinger T. Biosynthesis of proline in *Pseudomonas aeruginosa*. Partial purification and characterization of  $\gamma$ -glutamyl kinase // *Biochem. J.* 1979. Vol. 181. P. 215–222.
2. Krishna R.V., Beilstein P., Leisinger T. Biosynthesis of proline in *Pseudomonas aeruginosa*. Properties of gamma-glutamyl phosphate reductase and 1-pyrroline-5-carboxylate reductase // *Biochem. J.* 1979. Vol. 181. P. 223–230.
3. Селиверстов А.В., Любецкий В.А. Регуляция биосинтеза пролина у протеобактерий // *Молекулярная биология*. 2007. Т. 41(3). С. 572–574.
4. Официальный сайт Proline (URL: <http://lab6.iitp.ru/ru/proline/> 10.2.10).
5. Леонтьев Л.А., Любецкий В.А. Алгоритм определения белка, согласованного с заданным филогенетическим профилем // *Информационные процессы* ([www.jip.ru](http://www.jip.ru)). 2006. Т. 6(1). С. 24–32.
6. Kisker C., Hinrichs W., Tovar K., Hillen W., Saenger W. The complex formed between Tet repressor and tetracycline-Mg<sup>2+</sup> reveals mechanism of antibiotic resistance // *J. of Mol. Biol.* 1995. Vol. 247(2). P. 260–280.
7. Горбунов К.Ю., Любецкий В.А. Реконструкция эволюции генов вдоль дерева видов // *Молекулярная биология*. 2009. Т. 43(5). С. 946–958.

Поступила в редакцию  
20.04.10

## THE EVOLUTION OF PROLINE SYNTHESIS TRANSCRIPTIONAL REGULATION IN GAMMA PROTEOBACTERIA

K.V. Lopatovskaya, K.Yu. Gorbunov, L.Yu. Rusin, A.V. Seliverstov, V.A. Lyubetsky

We report a conserved motif upstream genes *proA* and *proB* widely represented among  $\gamma$ -proteobacteria, particularly in the genera *Pseudomonas*, *Marinobacter* and *Shewanella*. The conserved protein-DNA binding sites are 8 bp-long. Some genes have a duplex of sites for cooperative factor binding. The phylogenetic profiles of binding sites and all protein factors were compared. We identify a tetR family protein, an ortholog of the NP\_249058 protein from *P. aeruginosa* PAO1, as a transcription factor. Our algorithm was applied to construct the tree of species with predicted regulation of the *pro* genes. The phylogeny of *pro* genes, transcription factors phylogeny and their

binding sites phylogeny were mapped onto the tree of species with our algorithm. The obtained scenario displays important HGT and the related events.

**Key words:** *proline synthesis, TetR protein family, transcription regulation,  $\gamma$ -proteobacteria, evolutionary scenario, horizontal gene transfer.*

#### Сведения об авторах

*Лопатовская Кристина Викторовна* — мл. науч. сотр. Института проблем передачи информации имени А.А. Харкевича РАН. Тел. (485)699-83-54 (доб. 149); e-mail: kristina@iitp.ru

*Горбунов Константин Юрьевич* — канд. физ.-мат. наук, ст. науч. сотр. Института проблем передачи информации имени А.А. Харкевича УРАН. Тел. (485)699-83-54 (доб. 149); e-mail: gorbunov@iitp.ru

*Русин Леонид Юрьевич* — канд. биол. наук, науч. сотр. Института проблем передачи информации имени А.А. Харкевича УРАН. Тел. (485)699-83-54 (доб. 149); e-mail: roussine@uran-dex.ru

*Селиверстов Александр Владиславович* — канд. физ.-мат. наук, ст. науч. сотр. Института проблем передачи информации имени А.А. Харкевича УРАН. Тел. (485)699-83-54 (доб. 148); e-mail: slvstv@iitp.ru

*Любецкий Василий Александрович* — докт. физ.-мат. наук, зав. лабораторией Института проблем передачи информации имени А.А. Харкевича УРАН. Тел. (485)699-83-54 (доб. 148); e-mail: lyubetsk@iitp.ru