

# Поиск и описание разнообразия, эволюции систем рестрикции–модификации в гипертермофильных бактериях и археях

Попова А.<sup>1\*</sup>, Русинов И.<sup>1</sup>, Алексеевский А.<sup>1, 2</sup>

<sup>1</sup> Отдел математических методов в биологии НИИ ФХБ им. Белозерского МГУ, Москва, Россия

<sup>2</sup> НИИСИ РАН, Москва, Россия

\* diatrimma@mail.ru

**Ключевые слова:** Системы рестрикции-модификации; архен; бактерии; термозимы; эволюция

**Мотивация и цель:** Гипертермофильные бактерии и археи интересны для науки не только, как часть биоразнообразия живой природы. Ферменты, синтезируемые гипертермофилами, также называемые термозимами, термостабильны и максимально активны при высоких температурах 50–80 °С. Термозимы катализируют те же реакции, что и их мезофильные аналоги. Это позволяет применять их в различных экспериментах и лабораторных методиках и промышленных процессах, требующих поддержания высоких температур, которые не выдерживают мезофильные ферменты.

Эндонуклеазы рестрикции ферменты важные для молекулярной биологии, и термостабильные не являются исключением. Например, в работе Bonanno et al. 2007 года [1] предложен метод FLAG для выявления метилирования конкретных сайтов CpG в промоторах онкогенов. FLAG (генерация флуоресцентных ампликонов) представляет собой технологию генерации гомогенного сигнала, основанную на исключительно термостабильной эндонуклеазе PspGI. FLAG обеспечивает генерацию сигнала в реальном времени во время ПЦР путем PspGI-опосредованного расщепления погашенных флуорофоров на 5'-конце двухцепочечных продуктов ПЦР [1].

Помимо возможного практического применения термостабильные системы рестрикции-модификации (Р-М) интересны своей эволюцией, составом и распространенностью в контексте защитных систем прокариот.

Цель работы заключалась в изучении эволюции, родства, состава и распространенности систем Р-М в 7 родах гипертермофилов: *Thermocrinis*, *Sulfolobus*, *Pyrobaculum*, *Hyperthermus*, *Thermus*, *Thermotoga*, *Thermococcus*.

**Методы и алгоритмы:** Данные о системах рестрикции-модификации были получены из открытой базы данных REBASE (обновление от 6 июня 2023 г.). Белковые последовательности систем рестрикции-модификации выбранных нами семи родов гипертермофилов обрабатывались по отдельности в зависимости от типа белка: R/N – эндонуклеазы рестрикции и никазы, S – субъединицы S, узнающие последовательность, M – метилтрансферазы с помощью программы MEGA 11.

Последовательности анализировались с помощью построения множественных выравниваний и филогенетических деревьев. Были построены таблицы принадлежности систем штаммам и распространенности различных типов систем среди каждого вида принадлежащих 7 родам отобранных нами гипертермофилов.

**Результаты:** Среди систем были распространены все 4 типа.

Тип I. Эти ферменты представляют собой многосубъединичные белки, которые функционируют как единый белковый комплекс и обычно имеют типичный состав генов M|S|R (M – Mтаза, S – субъединица, узнающая сайт ДНК, R – ЭР). После узнавания мишени ферментативный комплекс перемещается по молекуле ДНК до момента встречи с каким-либо препятствием. Таким образом, разрыв вносится на большом произвольном расстоянии от сайта рестрикции. Для расщепления ДНК СР-М типа I необходим гидролиз АТФ. Это связано с тем, что для перемещения по молекуле ДНК и ее расплетания после распознавания сайта СР-М требуется энергия.

Тип II. Системы рестрикции типа II обычно имеют состав: R|M, R|M|M, RM (подтип IIG, один белок включает ЭР и MТаз каталитические домены). Эндонуклеазы разрезают ДНК в строго определенных позициях внутри или около сайта рестрикции. Метилтрансферазы обычно функционируют в форме мономера.

Р-М Типа III включают две субъединицы (Res и Mod), которые объединяются в гетеротетрамер (Res<sub>2</sub>Mod<sub>2</sub>) с эндонуклеазной и метилтрансферазной активностями. Системы типа III имеют состав R|M, отличие от типа II в необходимости двух близко расположенных сайтов для расщепления ДНК ЭР и механизме действия ЭР.

Тип IV. Система Тип IV имеет состав R или R|R, состоящие из одного или двух генов, кодирующих белки, которые расщепляют только модифицированную ДНК, включая метилированные, гидроксиметилированные и глюкозил-гидроксиметилированные

Одиночные MТазы (в REBASE отнесены к системам типа II с составом M) также распространены во всех видах [2].

Кроме перечисленных систем, встречались системы R-М с нетипичными составами M|S|S|R – 6 систем, M|S|S|S|R – 1, M|V – 5 (V – никаза, расщепляющая одну цепочку ДНК), M|R|V – 3, M|S – 5, M|M – 3, S – 3, S|R – 1. В дальнейшей работе планируется разобраться, являются ли эти случаи ошибками предсказания или же это особые системы.

Была составлена таблица, содержащая 65 видов, 174 штаммов и 578 систем. Построены сводные таблицы представленности по видам и родам. В дальнейшем, планируется посмотреть представленность среди гипертермофилов систем Р-М согласно классификации Р-М по каталитическим доменам ЭР и Mтаз, которая разработана в текущих лабораторных исследованиях.

**Выводы:** В роде *Thermococcus* самыми распространенными являлись системы первого типа, в роде *Thermotoga* самыми распространенными были системы состава R|M, в роде *Thermus* самыми распространенными были системы состава R|M и RM, *Thermocrinis*, *Sulfolobus*, *Pyrobaculum* самыми распространенными были системы состава R|M.

Нетипичные системы встречались чаще всего у штаммов вида *Thermus thermophilus* и штаммов с неопределенным видом у рода *Thermococcus*.

**Финансирование:** Исследование поддержано грантом РНФ № 21-14-00135.

# Search and description of diversity, evolution of restriction–modification systems in hyperthermophilic bacteria and archaea

Popova A.<sup>1\*</sup>, Rusinov I.<sup>1</sup>, Alexeevski A.<sup>1, 2</sup>

<sup>1</sup> Department of Mathematical Methods in Biology A.N. Belozersky Research Institute of Physico-Chemical Biology MSU, Moscow, Russia

<sup>2</sup> FSI FSC SRISA RAS, Moscow, Russia

\* diatrimma@mail.ru

**Key words:** Restriction modification systems; archaea; bacteria; thermowinters; evolution

**Motivation and Aim:** Hyperthermophilic bacteria and archaea are of interest to science not only as part of the biodiversity of living nature. Enzymes synthesized by hyperthermophiles, also called thermozymes, are thermostable and are most active at high temperatures of 50–80 °C. Thermozymes catalyze the same reactions as their mesophilic counterparts. This allows them to be used in a variety of experiments and laboratory techniques and industrial processes that require maintaining high temperatures that mesophilic enzymes cannot withstand.

Restriction endonucleases are important enzymes in molecular biology, and heat-stable enzymes are no exception. For example, Bonanno et al. In 2007, the FLAG method was proposed to detect methylation of specific CpG sites in oncogene promoters. FLAG (Fluorescent Amplicon Generation) is a homogeneous signal generation technology based on the exclusively thermostable endonuclease PspGI. FLAG enables real-time signal generation during PCR by PspGI-mediated cleavage of quenched fluorophores at the 5' end of double-stranded PCR products [1].

In addition to their possible practical applications, thermostable restriction-modification (RM) systems are interesting for their evolution, composition, and prevalence in the context of prokaryotic defense systems.

The purpose of the work was to study the evolution, relationship, composition and prevalence of R-M systems in 7 genera of hyperthermophiles: *Thermocrinis*, *Sulfolobus*, *Pyrobaculum*, *Hyperthermus*, *Thermus*, *Thermotoga*, *Thermococcus*.

**Methods and Algorithms:** Data on restriction modification systems were obtained from the open database REBASE (updated June 6, 2023). Protein sequences of restriction-modification systems of the seven genera of hyperthermophiles selected by us were processed separately depending on the type of protein: R\N – restriction endonucleases and nickases, S – subunits S that recognize the sequence, M – methyltransferases using the MEGA 11 program.

Sequences were analyzed using multiple alignments and phylogenetic trees. Tables were constructed of the affiliation of systems to strains and the prevalence of various types of systems among each species belonging to the 7 genera of hyperthermophiles we selected.

**Results:** All 4 types were common among systems.

Type I. These enzymes are multi-subunit proteins that function as a single protein complex and usually have a typical M|S|R gene composition (M – Mtase, S – DNA site recognition subunit, R – ER). After recognizing the target, the enzymatic complex moves along the DNA molecule until it encounters some obstacle. Thus, the break is introduced at a large random distance from the restriction site. DNA cleavage by CP-M type I

requires ATP hydrolysis. This is due to the fact that energy is required to move along the DNA molecule and unwind it after recognition of the CP-M site.

Type II. Type II restriction systems typically have the following composition: R|M, R|M|M, RM (sub-type IIG, one protein includes ER and MTase catalytic domains). Endonucleases cut DNA at strictly defined positions within or near the restriction site. Methyltransferases typically function in monomer form.

Type III PMs include two subunits (Res and Mod), which combine into a heterotetramer (Res2Mod2) with endonuclease and methyltransferase activities. Type III systems have an R|M composition, differing from type II in the need for two closely located sites for ER DNA cleavage and the mechanism of action of ER.

Type IV. The Type IV system has an R or R|R composition, consisting of one or two genes encoding proteins that cleave only modified DNA, including methylated, hydroxymethylated and glucosyl-hydroxymethylated

Single MTases (referred to as type II systems with M composition in REBASE) are also common in all species [2].

In addition to the listed systems, there were systems of the R-M system with atypical compositions: M|S|S|R – 6 systems, M|S|S|S|R – 1, M|V – 5 (V – nickase, cleaving one DNA chain), M|R|V – 3, M|S – 5, M|M – 3, S – 3, S|R – 1. Future work is planned to determine whether these cases are prediction errors or whether they are special systems. A table was compiled containing 65 species, 174 strains and 578 systems. Summary tables of representation by species and genera were constructed. In the future, it is planned to look at the representation of R-M systems among hyperthermophiles according to the R-M classification according to the catalytic domains ER and Mt3, which was developed in ongoing laboratory studies.

*Conclusion:* In the genus *Thermococcus* the most common systems were the first type, in the genus *Thermotoga* the most common systems were the composition R|M, in the genus *Thermus* the most common systems were the composition R|M and RM, *Thermocrinis*, *Sulfolobus*, *Pyrobaculum* the most common systems were the composition R|M.

Atypical systems were most often found in strains of the species *Thermus thermophilus* and strains of uncertain species in the genus *Thermococcus*.

*Funding:* The study is supported by RSF grant No. 21-14-00135.

#### *Список литературы/References*

1. Bonanno C., Shehi E., Adlerstein D., Makrigiorgos G.M. MS-FLAG, a novel real-time signal generation method for methylation-specific PCR. *Clin Chem.* 2007;53(12):2119-2127. doi 10.1373/clinchem.2007.094011
2. Официальный сайт базы данных REBASE [Электронный ресурс]. URL: <http://rebase.neb.com/rebase/rebhelp.html> (дата обращения: 20.02.2024)
3. Официальный сайт MEGA 11 [Электронный ресурс]. URL: [https://www.megasoftware.net/downloads/dload\\_win\\_gui](https://www.megasoftware.net/downloads/dload_win_gui)