

Эволюция тандемных повторов на примере двух групп позвоночных животных

Остромышенский Д.И.^{1*}, Попова М.А.^{1,2}, Травина А.О.¹

¹ Институт цитологии РАН, Санкт-Петербург, Россия

² Центр молекулярной и клеточной биологии, Сколковский институт науки и технологий, Москва, Россия

* necroforus@gmail.com

Ключевые слова: тандемные повторы; гетерохроматин; хомяки; лягушки; эволюция

Мотивация и цель: Тандемные повторы (ТП) являются важной частью генома всех позвоночных животных. Конститутивный гетерохроматин, формирующий такие функционально значимые части хромосом, как центромерные, перицентромерные и субтеломерные районы, состоит в основном из ТП. ТП состоят из многократно повторяющихся мономеров длиной от нескольких п.н. до нескольких т.п.н. Эти мономеры формируют поля ТП длиной до нескольких десятков м.п.н. Из-за особенностей организации и сложности сборки ТП зачастую недопредставлены в геномных сборках. Эволюция ТП также является крайне малоизученным вопросом. Появление большого количества сборок геномов, сделанных с использованием технологий секвенирования с длинными ридами (PacBio и OxfordNanopore), позволяет более эффективно исследовать ТП и их эволюцию.

Методы и алгоритмы: Искали ТП в сборках геномов лягушки *Rana temporaria* (WGS project CAJIMO01), хомяков *Cricetulus griseus* (RAZU01), *Mesocricetus auratus* (JAFVMI01), *Phodopus sungorus* (JANBXA01) и *P. roborovskii* (CALSGD01) по ранее описанной методике [1]. Рассматривали ТП с длиной поля >10 т.п.н., кроме представителей *Phodopus* у которых рассматривали поля ТП с длиной >5 т.п.н. Найденные мономеры семейств ТП искали в доступных геномах всех видов семейства Ranidae для лягушки (*R. temporaria*, *R. kukuronis*, *R. muscosa*, *Lythobathes sylvaticus*, *L. septentrionalis*, *Aquarana catesbeiana*, *Staurois parvus* и *Glandirana rugosa*) и подсемейства Cricetinae для хомяков (указанные выше виды). Поиск осуществляли с помощью BLAST с параметрами “dust” “no”, “evaluate” “1-e4”. Из дальнейшего анализа удаляли хиты BLAST с bit-score < 50 и соотношением bit-score/длина мономера ТП < 1.

Результаты: В геноме *R. temporaria* найдено 77 семейств ТП. Среди них не нашли ярко выраженного мажорного по содержанию ТП. 6 семейств ТП примерно одинаковы по суммарной длине полей. Только 15 из найденных семейств ТП являются видоспецифичными. В том числе три из числа наиболее представленных в геноме Rtem1149A, Rtem138A и Rtem35A. Еще 14 семейств ТП найдены также у близкородственного вида *R. kukuronis*, в том числе один из мажорных для *R. temporaria* семейств Rtem494A. Одно из этих семейств – Rtem149A в геноме *R. kukuronis* присутствует в единичных копиях в отличие от генома *R. temporaria*. В геномах всех трех представителей *Rana*, но не других родов найдено 2 семейства ТП Rtem103A и Rtem965A, но у *R. kukuronis* и *R. muscosa* эти семейства представлены единичными копиями. Еще один из мажорных ТП *R. temporaria* Rtem27A представлен также большим числом копий (более 10 тысяч) и у других

Rana, но единичными копиями у видов других родов *Ranidae*. Около 10 семейств ТП *R. temporaria* имеют сходное содержание в исследованных геномах *Ranidae*, что может говорить о том, что эти ТП произошли от диспергированных повторов. У хомяков (*Cricetinae*) нашли от 45 (*P. roborovskii*) до 65 (*P. sungorus*) семейств ТП. У всех хомяков обнаружены ярко выраженные мажорные семейства ТП – Cgri33A (*C. griseus*), Maur49A (*M. auratus*) или несколько мажорных ТП – Prob863A, Prob18A и Prob19A (*P. roborovskii*); Psun20A, Psun1212A, Psun32A (*P. sungorus*). У всех исследованных видов около половины семейств ТП являются видоспецифичными и не найдены у других видов. В большинстве случаев к числу невидоспецифичных относятся семейства ТП с относительно небольшим содержанием в геноме. Исключением является один из мажорных ТП *P. roborovskii* Prob18A, который найден в геномах трех других видов, но в количестве не более 100 копий, что значительно меньше копийности в геноме *P. roborovskii*. Другие семейства ТП с большим содержанием в геномах либо не встречаются у других видов, либо представлены небольшим количеством копий, к числу таких семейств относятся, например, ТП *C. griseus* Cgri72A, Cgri23A; ТП *M. auratus* Maur24A, Maur20A и др. Для рода *Phodopus* нашли 5 родоспецифичных семейств ТП. Также у видов этого рода нашли значительное количество видоспецифичных ТП с мономером длиной около 1 т.п.н. и более, образующих поля длиной более 100 т.п.н. ТП с такими длинными мономерами часто являются производными диспергированных повторов [1, 2], но найденные нами ТП не показали гомологии с диспергированными повторами.

Выводы: Наиболее популярной моделью эволюции ТП является библиотечная гипотеза [3]. Появление сборок геномов близкородственных видов позволяет проверить эту гипотезу на полном наборе ТП видов разной степени филогенетической близости. Показано, что паттерн распределения ТП в геномах *Ranidae* в целом соответствует библиотечной гипотезе, в то время как у хомяков противоречит ей. Это может свидетельствовать о том, что эволюция ТП в разных группах позвоночных животных идет разными путями и обусловлена разными механизмами.

Финансирование: Исследование поддержано грантом Российского научного фонда (проект № 24-24-00480).

The evolution of tandem repeats using the example of two groups of vertebrates

Ostromyshenskii D.I.^{1*}, Popova M.A.^{1,2}, Travina A.O.¹

¹ Institute of Cytology, RAS, St. Petersburg, Russia

² Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology, Moscow, Russia

* necroforus@gmail.com

Key words: tandem repeat; heterochromatin; hamster; frog; evaluation

Motivation and Aim: Tandem repeats (TR) are an important part of the genome of all vertebrates. Constitutive heterochromatin, which forms such functionally significant parts of chromosomes as centromeric, pericentromeric and subtelomeric regions, consists mainly of TR. TR consist of multiply repeating monomers from several bp up

to several kbp. These monomers form fields up to several tens of mbp long. Specific features of the TR structural organization complicate the assembly, annotation, and mapping of heterochromatic chromosome regions that contain TR as the main component. The evolution of TR is also an extremely poorly studied issue. The emergence of a large number of genome assemblies made using sequencing technologies with long reads (PacBio and OxfordNanopore) makes it possible to more effectively study TR and their evolution.

Methods and Algorithms: We looked for TRs in the genome assemblies of the frog *Rana temporaria* (WGS project CAJIMO01) and hamsters *Cricetulus griseus* (RAZU01), *Mesocricetus auratus* (JAFVMI01), *Phodopus sungorus* (JANBXA01) and *P. roborovskii* (CALSGD01) according to a previously described method [1]. TRs with a field length > 10 kb were considered, except for species of genus *Phodopus* in which TR fields with a length > 5 kb were considered. The founded monomers of TR families were searched in the available genomes of all species of the family Ranidae for the frog (*R. temporaria*, *R. kukuronis*, *R. muscosa*, *Lythobathes sylvaticus*, *L. septentrionalis*, *Aquarana catesbeiana*, *Staurois parvus* and *Glandirana rugosa*) and the subfamily Cricetinae for hamsters (indicated above species). The search was carried out using BLAST with the parameters “dust” “no”, “eval” “1-e4”. BLAST hits with a bit-score < 50 and a bit-score/TR monomer length ratio < 1 were removed from further analysis.

Results: 77 TR families were found in the *R. temporaria* genome. We did not find a clearly expressed major TR. 6 TR families are approximately the same in total fields length. Only 15 of the found TR families are species specific. Including three of the most represented in the *R. temporaria* genome Rtem1149A, Rtem138A and Rtem35A. Another 14 TR families were also found in the closely related species *R. kukuronis*, including one of the major families for *R. temporaria* Rtem494A. One of these families, Rtem149A, is present in single copies in the *R. kukuronis* genome, in contrast to the *R. temporaria* genome. In the genomes of all three representatives of *Rana*, but not of other genera, two TR families Rtem103A and Rtem965A were found, but in *R. kukuronis* and *R. muscosa* these families are represented by single copies. Another major TR of *R. temporaria* Rtem27A is represented by a large number of copies (more than ten thousand) in other *Rana* and single copies in species of other genera. About 10 families of TRs have similar contents in the studied genomes of Ranidae, which may indicate that these TRs originated from interspersed repeats.

In hamsters (Cricetinae), from 45 (*P. roborovskii*) to 65 (*P. sungorus*) TR families were found. In all hamsters, major TR families Cgri33A (*C. griseus*), Maur49A (*M. auratus*) or several major TR families were found – Prob863A, Prob18A and Prob19A (*P. roborovskii*); Psun20A, Psunh1212A, Psun32A (*P. sungorus*). In all studied species, about half of the TR families are species specific and are not found in other species. In most cases, TR families with a relatively small content in the genome are considered non-species-specific. The exception is one of the major TRs of *P. roborovskii* Prob18A, which is found in the genomes of three other species, but in an amount of no more than 100 copies, which is significantly less than the content in the genome of *P. roborovskii*. Other families of TRs with a high content in genomes are either not found in other species or are represented by a small number of copies; such families include, for example, TRs *C. griseus* Cgri72A, Cgri23A; *M. auratus* TR families Maur24A, Maur20A and others. For the genus *Phodopus*, 5 genus-specific TR families were found. Also, in species of this genus, a significant number of species-specific TRs with a monomer about 1 kb and more in length were found. These TRs forming fields more than

100 kbp in length. TRs with such long monomers are often derivatives of interspersed repeats [1, 2], but the TRs we found did not show homology with interspersed repeats.

Conclusion: The most popular model of TR evolution is the library hypothesis [3]. The emergence of genome assemblies of closely related species makes it possible to test this hypothesis on the full set of TRs of different species. It was shown that the distribution pattern of TRs in the genomes of Ranidae generally corresponds to the library hypothesis, while in hamsters it contradicts it. This may indicate that the evolution of TR in different groups of vertebrates follows different paths and is caused by different mechanisms.

Funding: The study is supported Russian Science Foundation (project No. 24-24-00480).

Список литературы/References

1. Komissarov A.S., Gavrilova E.V., Demin S.J., Ishov A.M., Podgornaya O.I. Tandemly repeated DNA families in the mouse genome. *BMC Genomics*. 2011;12:531
2. Ivanova N.G., Stefanova V.N., Ostromyshenskii D.I., Podgornaya O.I. Tandem repeats in the genome of *Sus scrofa*, their localization on chromosomes and in the spermatogenic cell nuclei. *Russ J Genet*. 2019;55:835-846
3. Fry K., Salser W. Nucleotide sequences of HS- α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell*. 1977;12(4):1069-1084