

Анализ представленности семейств транспозонов в геномах домашней и дикой формы северного оленя *Rangifer tarandus*

Коноров Е.^{1, 2*}, Семина М.¹, Онохов А.¹, Лайшев К.^{1, 3}

¹ Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия

² Федеральный научный центр пищевых систем им. В.М. Горбатова РАН, Москва, Россия

³ Санкт-Петербургский федеральный исследовательский центр РАН, Санкт-Петербург, Россия

* konorov@vigg.ru

Ключевые слова: *Rangifer tarandus*; LINE ретротранспозоны; доместикация

Мотивация и цель: На сегодняшний день доступно несколько сборок геномов северного оленя в открытых базах данных, в том числе одна на уровне хромосом [1–3]. В данных работах рассматривались специфичные для *Rangifer tarandus* гены, в том числе быстро меняющиеся и относящиеся к перепредставленным геномным онтологиям. Около 38 % генома составляли повторяющиеся последовательности, но различия между особями из разных популяций не были предметом геномных исследований. Фракция повторяющихся последовательностей в геноме может разительно отличаться между подвидами одного вида копытных животных [4], при этом данные различия могут приводить к адаптации [5] и коррелировать с хозяйственно значимыми признаками. В данной работе мы сравнили два генома диких и два генома домашних северных оленей по распространенности повторяющихся элементов и, в частности, по числу копий LINE ретротранспозонов.

Методы и алгоритмы: С помощью Illumina NovaSeq секвенированы полные геномы двух домашних и двух диких особей северного оленя для последующей сборки и анализа повторяющихся последовательностей для характеристики различий между дикой и домашней морфами. Глубина секвенирования была выше 30х для каждого из геномов. Проведена очистка геномных прочтений для каждого образца с помощью trimmomatic [6], при этом были удалены геномные прочтения с низким качеством, адаптеры и другие артефакты секвенирования, а также произведен поиск контаминации. Произведен процесс сборки генома с помощью SPAdes [7]. Поиск повторяющихся последовательностей осуществлялся с помощью RepeatMasker и hmmer.

Результаты: N50 полученных сборок составляли 10.3 и 9.7 kВ для домашних северных оленей и 11.7 и 12.2 kВ для диких оленей, GC% состав не отличался и составлял 41.55 %.

Отдельно было изучено число копий BovB LINE элементов и L1 LINE. У домашних особей было обнаружено 101 и 97 копий BovB (DF000280309.1 RanTar-1.51) в геноме, в то время как в геномах диких оленей было найдено 120 и 127 полных BovB LINE. L1 LINE, специфичный для *R. tarandus* (DF000280306.1 RanTar-1.226) был найден в примерно одинаковом числе копий для всех четырех особей (121 и 125 для домашних, 125 и 128 для диких). По другим видоспецифичным семействам L1 (DF000280305, DF000280315 и DF000280316) также не обнаружено значительных различий.

Вместе с этим можно отметить, что различия внутри семейства ретротранспозонов у диких оленей (по р-дистанциям) в среднем выше, чем различия между таковыми копиями у домашних оленей. Это согласуется с неопубликованными данными по тем же выборкам, отсеквенированным с меньшим покрытием, но в большем объеме, что как ядерный, так и митохондриальный геном диких оленей обладает большей изменчивостью по сравнению с домашней морфой.

Выводы: Нами были аннотированы геномы повторяющихся последовательностей двух диких оленей из Магаданской области и двух домашних оленей из Мурманской области. Дикие олени обладают большим количеством копий отдельных семейств BovB LINE ретротранспозонов, а также большей вариативностью между копиями. Поскольку, как показано в других работах по домостикации и локальным адаптациям у млекопитающих, LINE могут менять паттерн изменчивости для последующего отбора, в дальнейшем планируется расширить материал на нескольких популяциях с более низкой глубиной секвенирования

Финансирование: Работа выполнена при финансовой поддержке гранта РНФ № 22-16-00062.

Analysis of the transposon families abundance in the genomes of domestic and wild reindeer *Rangifer tarandus*

Konorov E.A.^{1, 2*}, Semina M.T.¹, Onokhov A.A.¹, Layshev K.A.^{1, 3}

¹ Vavilov Institute of General Genetics, RAS, Moscow, Russia

² Gorbатов Federal Research Center for Food Systems, RAS, Moscow, Russia

³ St. Petersburg Federal Research Center, RAS, St. Petersburg, Russia

* konorov@vigg.ru

Key words: *Rangifer tarandus*; LINE retrotransposons; domestication

Motivation and Aim: To date, several genome assemblies of reindeer *R. tarandus* are available in open databases, including one at the chromosome level [1–3]. These studies examined species-specific genes, including those which are rapidly changing and with overrepresented gene ontologies. About 38 % of the reindeer genome was repetitive sequences, but differences between individuals from different populations have not been the subject of genomic studies. The fraction of repeated sequences in the genome can differ drastically between subspecies of the same species of ungulates [4], and these differences can lead to adaptation [5] and correlate with economically significant traits of domesticated animals. In present work, we compared two genomes of wild and two genomes of domestic reindeer on prevalence of repetitive elements and, in particular, the copy number of LINE retrotransposons.

Methods and Algorithms: Using Illumina NovaSeq, the complete genomes of two domestic and two wild reindeer were sequenced and assembled. Then we made repeat sequence analysis to characterize differences between wild and domestic morphs. The sequencing depth was above 30x for each of the genomes. The genomic reads for each sample were cleaned using trimmomatic [6], low-quality genomic reads, adapters and other sequencing artifacts were removed. The genomes were de novo assembled using SPAdes [7]. The search for repeated sequences was carried out using RepeatMasker and hmmer.

Results: The N50 of the resulting assemblies were 10.3 and 9.7 kB for domestic reindeer and 11.7 and 12.2 kB for wild deer, the GC% composition did not differ and was 41.55 %.

The copy numbers of BovB LINE elements and L1 LINE were studied separately. Domestic deer had 101 and 97 copies of BovB (DF000280309.1 RanTar-1.51) in the genome, while 120 and 127 complete BovB LINEs were found in wild deer genomes. L1 LINE specific for *R. tarandus* (DF000280306.1 RanTar-1.226) was found in approximately the same number of copies for all 4 individuals (121 and 125 for domestic, 125 and 128 for wild). No significant differences were also found for other species-specific L1 families (DF000280305, DF000280315 and DF000280316).

It should be noted that the differences within the family of retrotransposons in wild deer (according to p-distances) are on average higher than the differences between such copies in domestic deer. This is consistent with unpublished data from the same samples sequenced with less coverage, but in a larger volume, that both the nuclear and mitochondrial genomes of wild deer have greater variability compared to the domestic morph.

Conclusion: We have annotated repeat sequences in genomes of two wild deer from the Magadan region and two domestic deer from the Murmansk region. Wild deer have higher copy numbers of BovB LINE retrotransposon from several families and also have greater intercopy variability. As has been shown in other work on domestication and local adaptation in mammals LINEs can change the pattern of variation for subsequent selection, we are planning to expand the material on several populations with lower sequencing depth in the future.

Funding: The study is supported by Russian Science Foundation grant (No. 22-16-00062).

Список литературы/References

1. Taylor R.S., Horn R.L., Zhang X., Golding G.B., Manseau M., Wilson P.J. The caribou (*Rangifer tarandus*) genome. *Genes*. 2019;10(7):540
2. Poisson W., Prunier J., Carrier A., Gilbert I., Mastromonaco G., Albert V., Robert C. Chromosome-level assembly of the *Rangifer tarandus* genome and validation of cervid and bovid evolution insights. *BMC Genomics*. 2023;24(1):142
3. Weldenegodguad M., Pokharel K., Ming Y., Honkatukia M., Peippo J., Reilas T., Kantanen J. Genome sequence and comparative analysis of reindeer (*Rangifer tarandus*) in northern Eurasia. *Sci Rep*. 2020;10(1):8980
4. Petersen M., Winter S., Coimbra R., de Jong M.J., Kapitonov V.V., Nilsson M.A. Population analysis of retrotransposons in giraffe genomes supports RTE decline and widespread LINE1 activity in Giraffidae. *Mobile DNA*. 2021;12,1-13
5. Pan Z., Li S., Liu Q., Wang Z., Zhou Z., Di R., Li Y. Rapid evolution of a retro-transposable hotspot of ovine genome underlies the alteration of BMP2 expression and development of fat tails. *BMC Genomics*. 2019;20:261
6. Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120
7. Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Computat Biol*. 2012;19(5):455-477