

## Поиск и описание хромосомных перестроек в хромосоме 1А у большой синицы (*Parus major*) с помощью метода Hi-C

Кораблёва С.Ю.<sup>1, 2\*</sup>, Гридина М.М.<sup>1</sup>, Торгашева А.А.<sup>1</sup>, Нурисламов А.Р.<sup>1</sup>,  
Задесенец К.С.<sup>1</sup>, Малиновская Л.П.<sup>1</sup>, Фишман В.С.<sup>1</sup>

<sup>1</sup> Институт цитологии и генетики СО РАН, Новосибирск, Россия

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

\* s.kozyreva@g.nsu.ru

**Ключевые слова:** Hi-C; инверсия; геномные перестройки; большая синица

**Мотивация и цель:** Инверсии играют важную роль в эволюции хромосом, ограничивая поток генов в популяции. Они являются наиболее распространенными хромосомными перестройками в достаточно консервативных кариотипах птиц. Исследование инверсий важно для понимания механизмов видообразования и адаптации.

Ранее в популяции на территории Нидерландов был выявлен полиморфизм по комплексной перестройке в хромосоме 1А большой синицы (*Parus major*) [1]. Согласно данным анализа гетерозиготности и распределения количества однонуклеотидных полиморфизмов, перестройка предположительно включает в себя большую (более 90 % длины хромосомы) инверсию. Один из ее концов, вероятно, находится вблизи комплекса, для которого характерна вариация числа копий (copy number variants, CNV) [1]. При этом у 96 % гетерозигот по комплексной перестройке в хромосоме 1А число копий этого комплекса было существенно выше, чем у гомозигот по нормальной хромосоме 1А. Общая длина последовательностей, дополнительно присутствующих в перестроенной хромосоме, предположительно могла составлять около 3.5 Мб [1], что затрудняет определение точных границ инверсии.

С помощью методов цитогенетического анализа наши коллеги из Института цитологии и генетики СО РАН обнаружили полиморфизм, предположительно, по той же хромосомной перестройке в хромосоме 1А, у синиц из новосибирской популяции. Однако длина перестроенной хромосомы была примерно в 1.5 раза больше, чем нормальной, т. е. увеличение длины гомолога, несущего инвертированный аллель, составляло до 30 Мб – существенно больше, чем увеличение на 3.5 Мб, описанное раньше. Таким образом, структура этой перестройки – точные границы инверсии, а также генетическое содержание инсерции, неизвестны.

**Методы и алгоритмы:** Для исследования структуры сложной перестройки использовались методы Hi-C-секвенирования [2] и полногеномного секвенирования парными короткими прочтениями.

Для того чтобы определить нуклеотидную последовательность неизвестной инсерции, был разработан подход для поиска дополнительных последовательностей, отсутствующих в референсном геноме, с использованием данных полногеномного и Hi-C-секвенирования. В его основе лежит анализ отношения количества пар сегментов выравнивания, в которых одно из прочтений не картируется на геном, к парам, в которых оба прочтения выравниваются на геном.

Этот метод подходит для поиска неизвестных высокоповторенных последовательностей с помощью доступных методов секвенирования второго поколения, не прибегая к дорогостоящим методам третьего поколения.

**Результаты:** Впервые были построены Hi-C-карты для большой синицы дикого типа и с комплексной хромосомной перестройкой.

Нами были разработаны две версии алгоритма поиска неизвестных высокоповторенных последовательностей. Первый основан на сравнительном анализе Hi-C-данных животных с инверсией и дикого типа, второй – на анализе исключительно Hi-C-данных животных с инверсией и может применяться, если данные для животных дикого типа недоступны. С помощью этих алгоритмов была обнаружена последовательность длиной ~700 п. о., специфически амплифицированная у животных с инверсией. По предварительной оценке, суммарная длина копий данной последовательности составляет ~15 Мб. Эксперимент FISH с найденной последовательностью показал, что это повтор, диспергированный по перестроенному плечу хромосомы 1A.

С помощью секвенирования Nanopore удалось подтвердить результаты, полученные с помощью Hi-C секвенирования, и установить последовательность границ инверсии.

**Выводы:** с помощью метода Hi-C была построена карта контактов хроматина для синиц, гетерозиготных по комплексной перестройке в хромосоме 1A, и синиц дикого типа.

Уточнена сборка хромосомы 1A большой синицы, с нуклеотидной точностью определены границы инверсии и установлено, что границы инверсии фланкированы простыми повторами.

С помощью разработанного подхода и анализа копийности CNV обнаружены отсутствующие в референсном геноме синицы высокоповторенные последовательности ДНК, суммарная длина которых составляет около 3 Мб в геноме животных дикого типа и около 14 Мб в геноме животных с комплексной хромосомной перестройкой. Таким образом, объяснено увеличение гомолога хромосомы 1A с комплексной хромосомной перестройкой на 11 Мб по сравнению с гомологом дикого типа.

**Финансирование:** Работа выполнена при поддержке гранта РНФ 23-14-00182.

## **Dissecting the structure of the chromosomal rearrangements in chromosome 1A in great tits (*Parus major*) using Hi-C technique**

Korableva S.Y.<sup>1, 2\*</sup>, Gridina M.M.<sup>1</sup>, Torgasheva A.A.<sup>1</sup>, Nurislamov A.R.<sup>1</sup>, Zadesenets K.S.<sup>1</sup>, Malinovskaya L.P.<sup>1</sup>, Fishman V.S.<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

\* s.kozyreva@g.nsu.ru

**Key words:** Hi-C; inversion; genome rearrangements; great tit

**Motivation and Aim:** Inversions play an important role in chromosome evolution, limiting gene flow within populations. They are the most common chromosomal

rearrangements in relatively conservative avian karyotypes. Studying inversions is crucial for understanding the mechanisms of speciation and adaptation. Previously, polymorphism due to a complex rearrangement on chromosome 1A of the great tit (*Parus major*) population in the Netherlands [1] was identified. According to the analysis of heterozygosity and the distribution of single nucleotide polymorphisms, the rearrangement presumably involves a large inversion (>90 % of the chromosome length). One of the inversion breakpoint is likely to be near a complex characterized by copy number variation (CNV) [1]. Notably, in 96 % of individuals heterozygous for the complex rearrangement on chromosome 1A, the copy number of this complex was significantly higher than in individuals homozygous for the normal chromosome 1A. The total length of additional sequences present in the rearranged chromosome was estimated to be about 3.5 Mb [1], making it difficult to determine the exact boundaries of the inversion.

Using cytogenetic analysis methods, our colleagues from the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences discovered polymorphism, presumably due to the same chromosomal rearrangement on chromosome 1A, in a population of tits from Novosibirsk. However, the length of the rearranged chromosome was approximately 1.5 times longer than normal, indicating an increase in the length of the homolog carrying the inverted allele, up to 30 Mb – significantly more than the previously described increase of 3.5 Mb. Thus, the structure of this rearrangement – the exact inversion breakpoints as well as the genetic content of the insertion – remains unknown.

**Methods and Algorithms:** Hi-C sequencing methods [2] and whole-genome sequencing with paired short reads were employed to investigate the structure of the complex rearrangement. To determine the nucleotide sequence of the unknown insertion, an approach was developed for searching additional sequences absent in the reference genome, utilizing data from whole-genome and Hi-C sequencing. It is based on analyzing the ratio of read pairs, where one read is unmapped to the genome, to pairs where both reads are mapped. This method is suitable for identifying unknown highly repetitive sequences using available second-generation sequencing methods, without resorting to expensive third-generation methods.

**Results:** Hi-C maps were constructed for both wild-type great tits and those with complex chromosomal rearrangements for the first time. Two versions of an algorithm for identifying unknown highly repetitive sequences were developed. The first is based on comparative analysis of Hi-C data from animals with the inversion and wild type, while the second solely analyzes Hi-C data from animals with the inversion and can be applied when data for wild-type animals are unavailable. Using these algorithms, a sequence of approximately 700 base pairs long, specifically amplified in animals with the inversion, was discovered. The preliminary estimation of the copy number of this sequence is approximately 15 Mb. FISH experiment with the discovered sequence indicated it to be a repeat dispersed throughout the rearranged arm of chromosome 1A.

Sequencing with Nanopore technology confirmed the results obtained with Hi-C sequencing and established the sequence of inversion boundaries.

**Conclusion:** Using the Hi-C method, a chromatin contact map was constructed for tits heterozygous for the complex rearrangement on chromosome 1A and wild-type tits.

The assembly of chromosome 1A of the great tit has been refined, with nucleotide-level precision defining the boundaries of the inversion. It has been established that the inversion boundaries are flanked by simple repeats.

Using the developed approach and CNV copy number analysis, high-repeat DNA sequences absent in the reference genome of tits were discovered, with a total length of about 3 Mb in the genome of wild-type animals and about 14 Mb in the genome of animals with the complex chromosomal rearrangement on chromosome 1A. Thus, the increase in the homolog of chromosome 1A with a complex chromosomal rearrangement was explained by 11 Mb compared to the wild-type homolog.

*Funding:* The study is supported was supported by RSF grant No. 23-14-00182.

#### *Список литературы/References*

1. Da Silva V.H. The genomic complexity of a large inversion in great tits. *Genome Biol Evol.* 2019;11(7):1870-1881
2. Lieberman-Aiden E. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009;32(5950):289-293. doi 10.1126/science.1181369