



Палиндромные повторы в геномах бактерий филума Calditrichaeota

Palindromic repeats in the genomes of bacteria from the phylum Calditrichaeota

Alex Chistyakov, Olga Sigalova

Abstract

Palindromic repeats are widespread in bacterial genomes. Although their origin is typically associated with mobile DNA, they have been also shown to play an important role in bacterial genome evolution and regulation of gene expression. These repeats are often present in multiple copies within a genome but are not conserved between bacterial species, which complicates their analysis. Here, we identified two palindromic repeats in the genome of *Caldithrix abyssi* from the novel bacterial phylum Calditrichaeota and performed a detailed analysis of their properties.

Goals

- Идентифицировать и сравнить длинные палиндромы в трех доступных геномах бактерий филума Calditrichaeota,
- Проанализировать свойства наиболее часто встречающихся палиндромов, включая их локализацию, геновое окружение и распределение по цепям.
- Проверить, ассоциированы ли эти палиндромы с генами REP-транспозаз (механизм перемещения палиндромных повторов, описанный для энтеробактерий)

Methods

- Поиск и анализ мотивов (R Bioconductor packages)
- Выравнивание последовательностей (Muscle и JalView)
- Анализ геномного контекста (RAST server)
- Поиск по геномным базам данных (BLAST)

Results

Повтор 1: “одомашненный” мобильный элемент?

- Встречается в геноме 41 раз. Имеет палиндромное ядро и консервативный контекст около 100 нт, определяющий направленность
- Предпочитает располагаться на одной из цепей (лидирующая или запаздывающая)
- Располагается в межгенных интервалах стандартного размера (100-250 нт. до генов), часто около важных генов
- Нет предпочтений по типу межгенных интервалов (конвергоны, дивергоны, однонаправленные гены)
- Отсутствует в геномах *C. palaeochorensis* и *C. insularis*

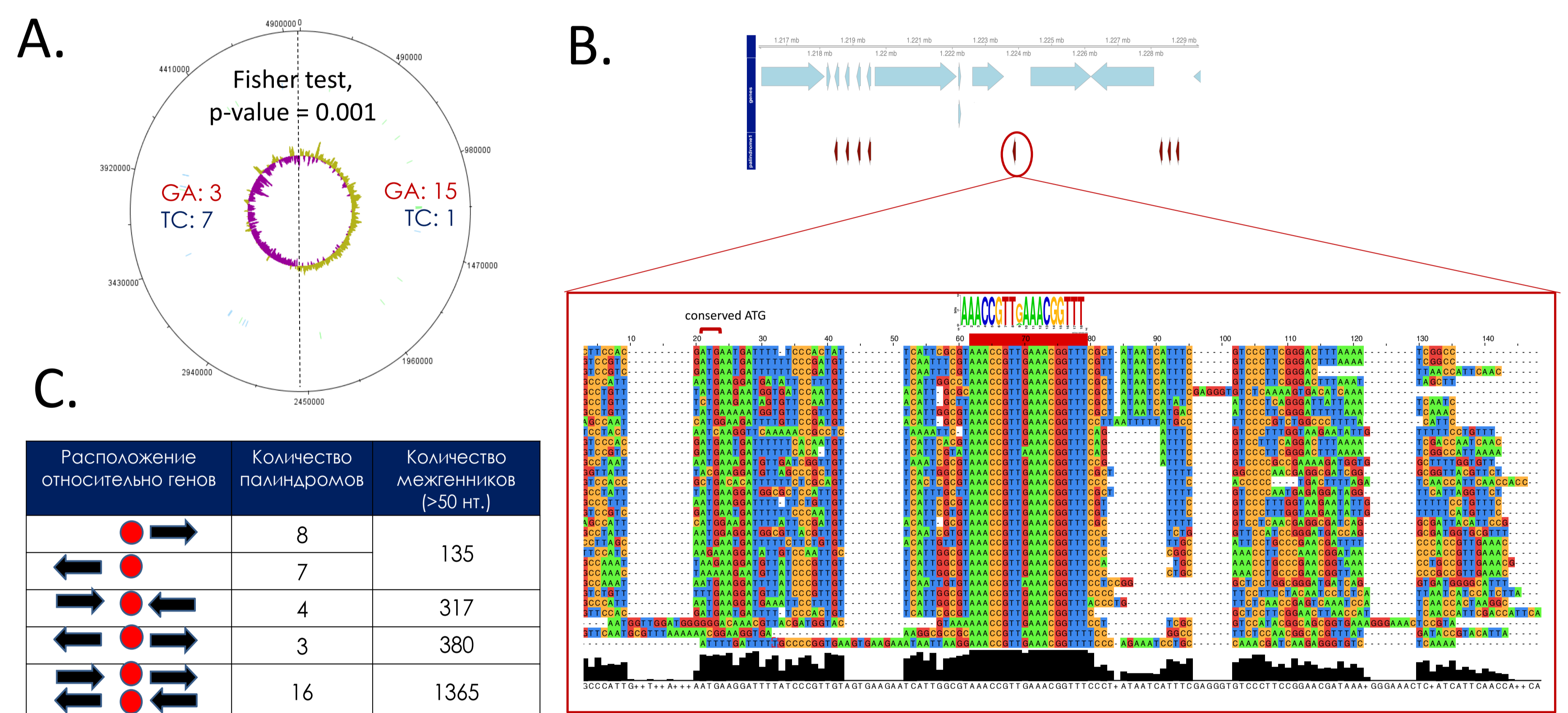
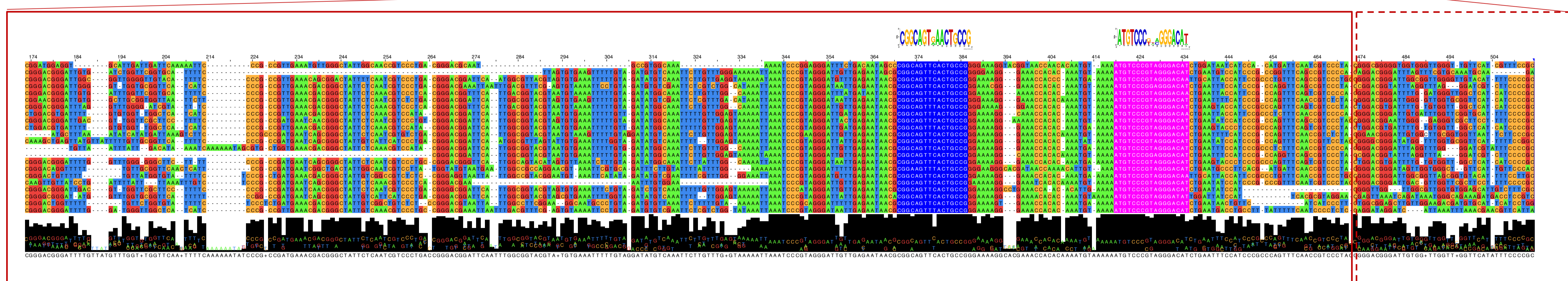


Figure 1. Structure and properties of the first repeat. A. Distribution of repeats with different directionality along the chromosome (as defined by the palindrome loop: GA or TC). The inner circle shows the GC-skew. The vertical dotted line shows the boundaries between the leading and lagging strands. Our results show that the repeat has a strand preference (Fisher test, p-value 0.001). B. An example of genomic context of several repeat clusters and the structure of a single repeat. The alignment shows that this repeat has conserved a palindromic core and less conserved surrounding region of around 100 nucleotides. A conserved ATG might explain multiple predicted short reading frames in the repeat region. C. Location of repeats relative to genes. From the table we see that this repeat shows no distinguishable preferences towards 3' or 5' ends of the genes.

Повтор 2: большие кластеры с регулярно расположенными шпильками

- Повтор длиной около 240 нт, включающий в себя несколько палиндромов
- В геноме лежит большими кластерами (3 кластера по 5-6 повторов, 5 кластеров по 2-3 повтора)
- Отсутствует в геномах *C. palaeochorensis* и *C. insularis*

Figure 2. Diagram of two clusters formed by the second repeat. Two conserved palindromes are shown with red and green arrows (directionality is defined by the palindrome loops). Zoom-in shows the structure of a single repeat within a cluster. Two most conserved palindromes within the repeat are highlighted with blue and purple. The dotted rectangle shows the beginning of the next repeat within the cluster.



Повторы в *C. abyssi* не ассоциированы с REP-транспозазами

- В геномах *C. abyssi* и *C. palaeochorensis* найдено несколько гомологов REP-транспозаз кластридий, которые потенциально могут участвовать в распространении палиндромов по геному.
- Палиндромы, фланкирующие гены транспозаз напоминают классические REP-элементы, но в геномах *Caldithrix* они больше нигде не встречаются

Figure 3. Diagrams of two candidate REP-transposases, one present in *C. abyssi* and *C. palaeochorensis* (A) and the other only in *C. abyssi* (B). Flanking palindrome sequences are shown with red arrows. Palindromes in A have no loop and therefore no directionality, while palindromes in B are directed. Both palindromes have properties reminiscent of classical REP elements but are not present in any other locations in the genome.

