



# Rearrangements in *Vibrio* genomes:

## ВАЖЕН ЛИ ДЛЯ ГЕНОВ РАЙОН ПРОЖИВАНИЯ

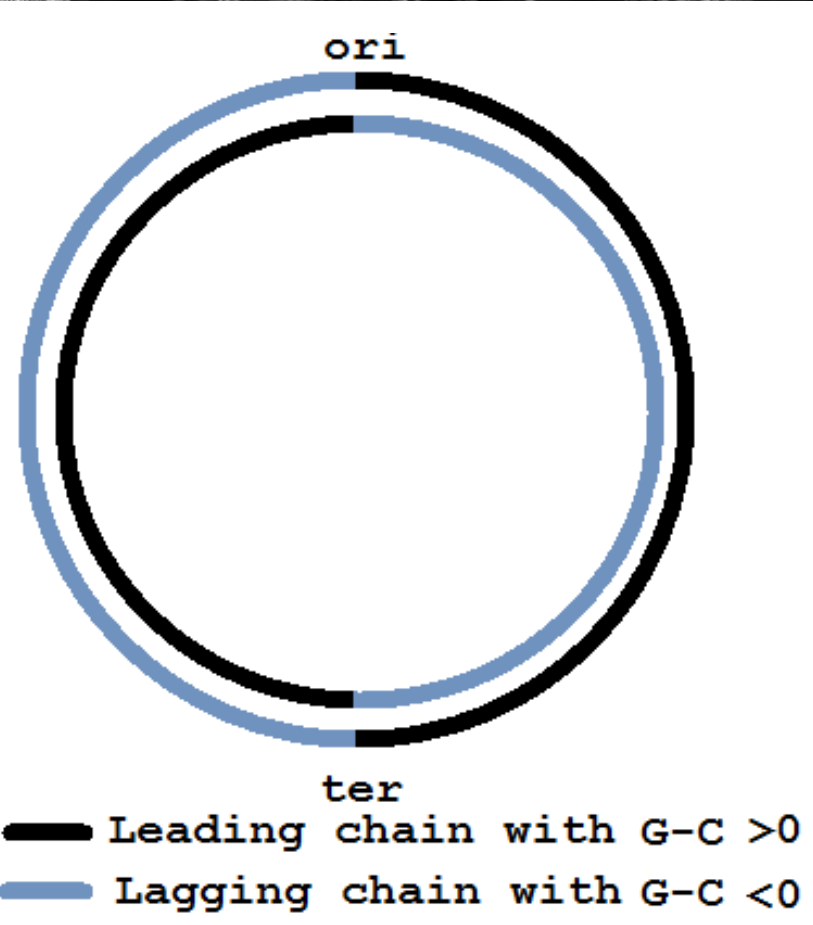
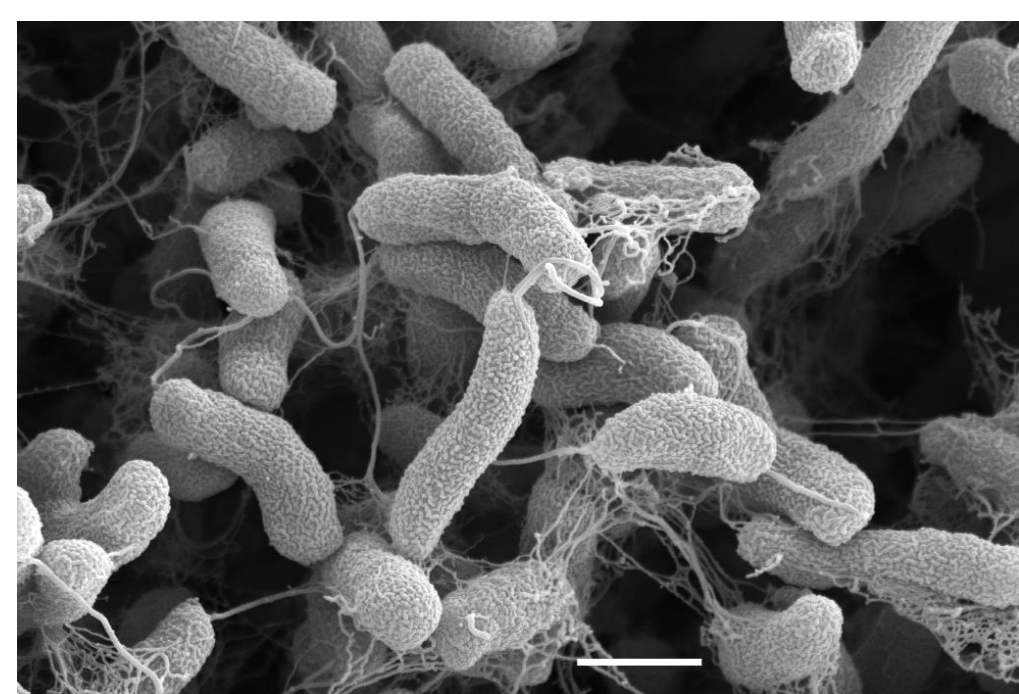
Regina Shaykhtudinova, Kristina Perevoshchikova, Olga Bochkareva

### Abstract

*Vibrio* is a genus of Gram-negative bacteria, belonging to  $\gamma$ -Proteobacteria. This genus contains species which are mainly found in marine environments, several species are human pathogens, notably particular serotypes of *V. cholerae* causing cholera, *V. parahaemolyticus* (gastroenteritis) and *V. vulnificus* (wound infections). *Vibrio* genomes consist of two chromosomes. The majority of genes necessary for the basic life processes are located on the first chromosomes; the second chromosomes contain few essential genes and are mainly composed of niche-specific genes.

The first goal of our research was to check the hypothesis that genes transferred between chromosomes keep their position on the lagging/leading strand. The second goal was to confirm that inversions with ends on different replichores are overrepresented and strongly symmetrical to balance the replichore size.

In the first chromosomes, about 60% genes are located on the leading strand. For the second chromosomes, the fraction of genes on the leading strand is not stable during evolution. Translocated genes do not keep their position on the leading or lagging strand. Breakpoints of large translocations, of 16 genes in *Vibrio alginolyticus* and of 79 genes in *Vibrio parahaemolyticus* are formed by repeats of 16S rRNA and 23S rRNA. All reconstructed inversions have breakpoints in different replichores. Moreover, all inversions are strongly symmetrical.



### Introduction/Введение

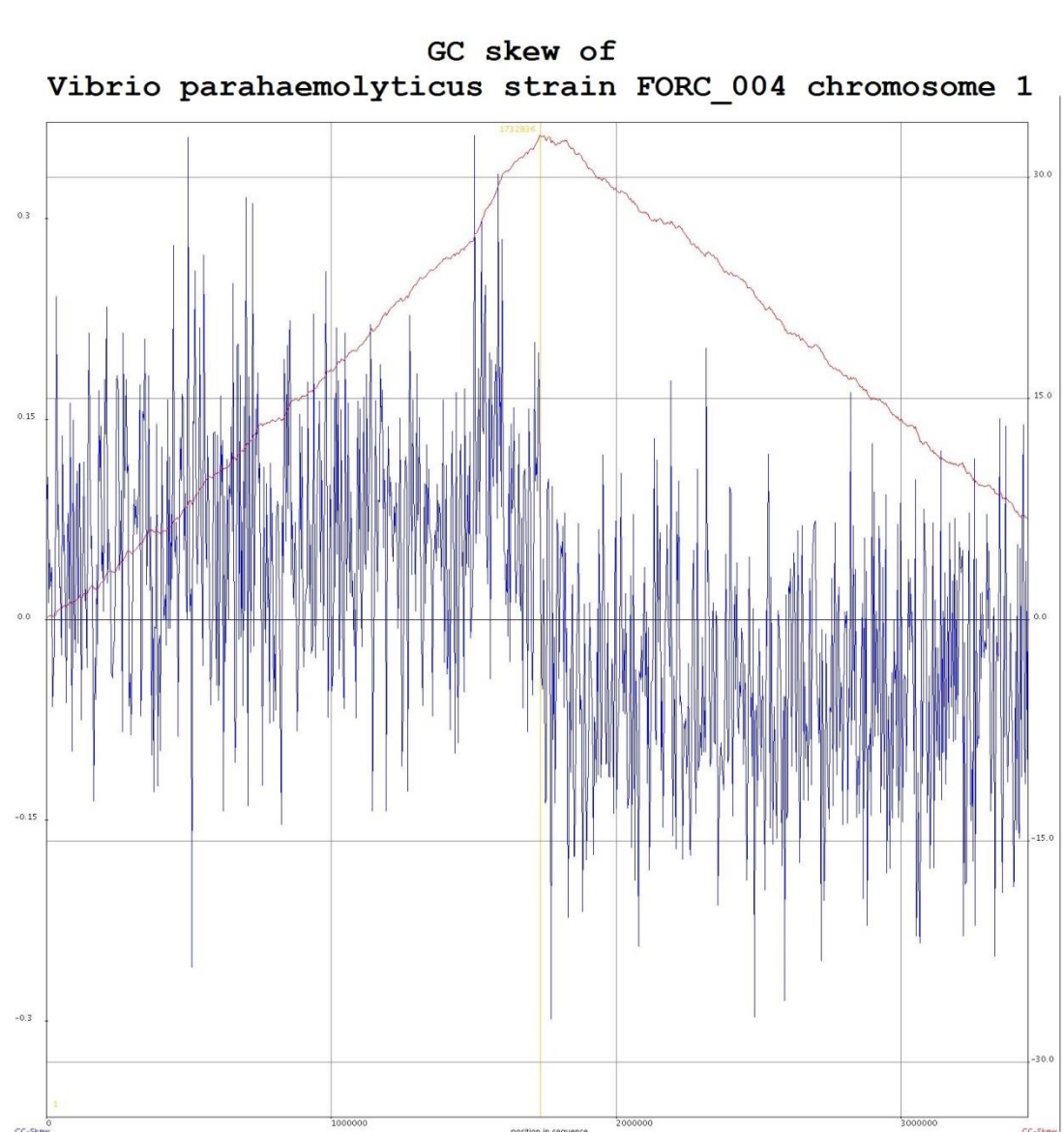
Вибрионы – род бактерий, относящийся к классу Gammaproteobacteria. Многие вибрионы являются патогенами, вызывающими заболевания у человека, рыб, моллюсков и других.

В отличие от большинства прокариот сохраняют на протяжении своей длительной эволюции две кольцевые хромосомы, реплицирующиеся независимо друг от друга. Большинство генов, отвечающих за базовые клеточные процессы, лежат на первой хромосоме, а вторая содержит гены, обеспечивающие адаптацию к условиям среды.

**Целью работы** было проверить гипотезу о том, что гены, транслоцирующиеся между первой и второй хромосомами, чаще сохраняют свое положение на лидирующей и запаздывающей цепи в силу действия отбора, а также показать, что отбор поддерживает инверсии, проходящие через сайт начала или конца репликации, и симметрично расположенные относительно этих точек.

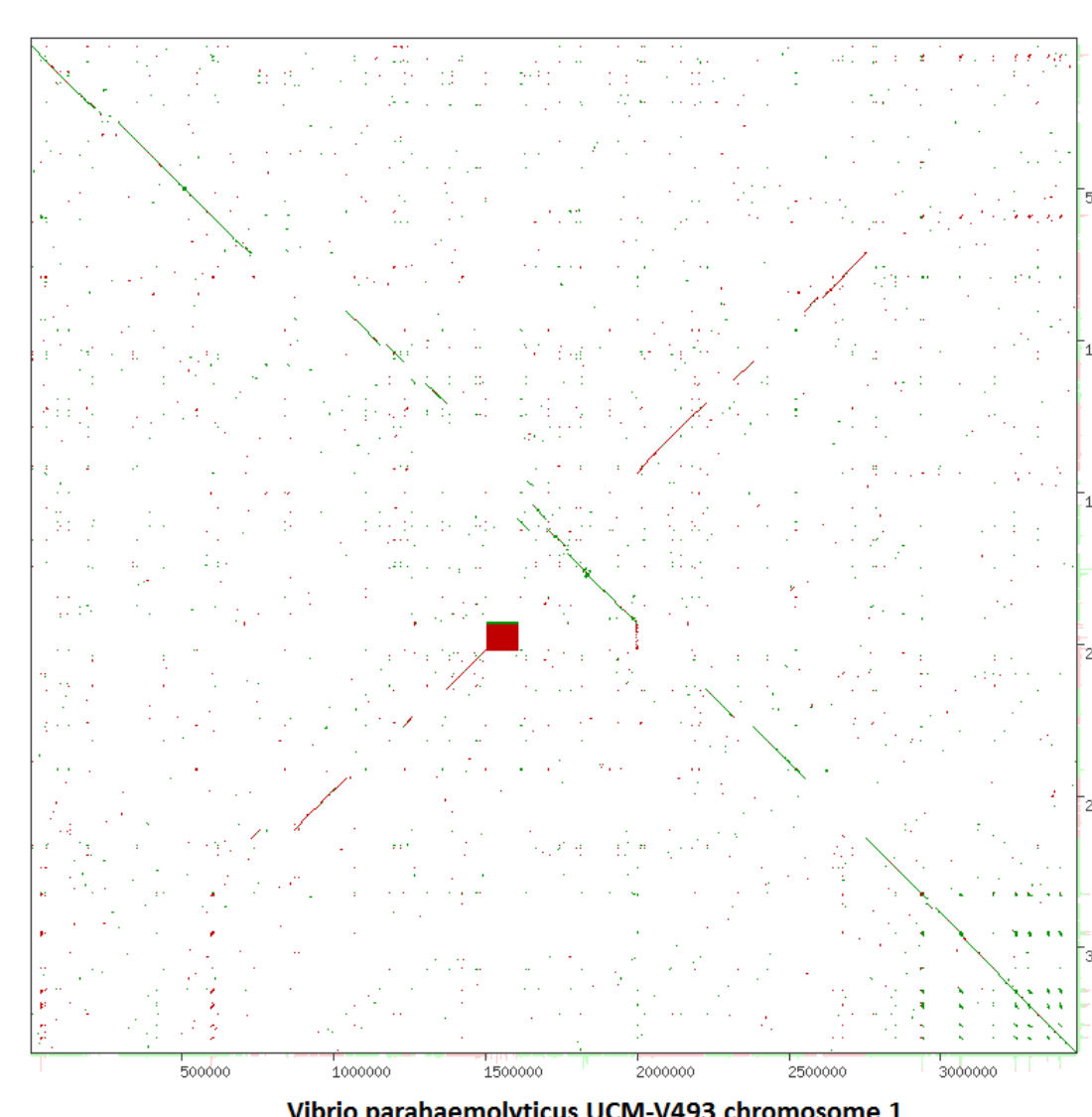
### Methods/Методы

**1. Позиции начала и конца репликации** на каждой хромосоме были определены на основе анализа пиков GC-skew при помощи программ на Python. Метод основывается на наблюдении, что на лидирующей цепи содержится больше гуанина и тимина, в то время запаздывающая цепь богата цитозином и аденином.



Identification of the origin and terminus of DNA replication using cumulative GC-skew plots. This method uses the sliding window strategy and calculates the GC-skew =  $(G - C)/(G + C)$  where G and C are the counts of guanine and cytosine in the sequence.

**2. Полногеномные выравнивания** для штаммов видов *V. parahaemolyticus*, *V. cholerae* и *V. alginolyticus* были получены с помощью алгоритма Sibelia.



Dot plot compares two sequences by organizing one sequence on the x-axis, and another on the y-axis. When fragments of both sequences match a dot is drawn at the respective position of the plot.

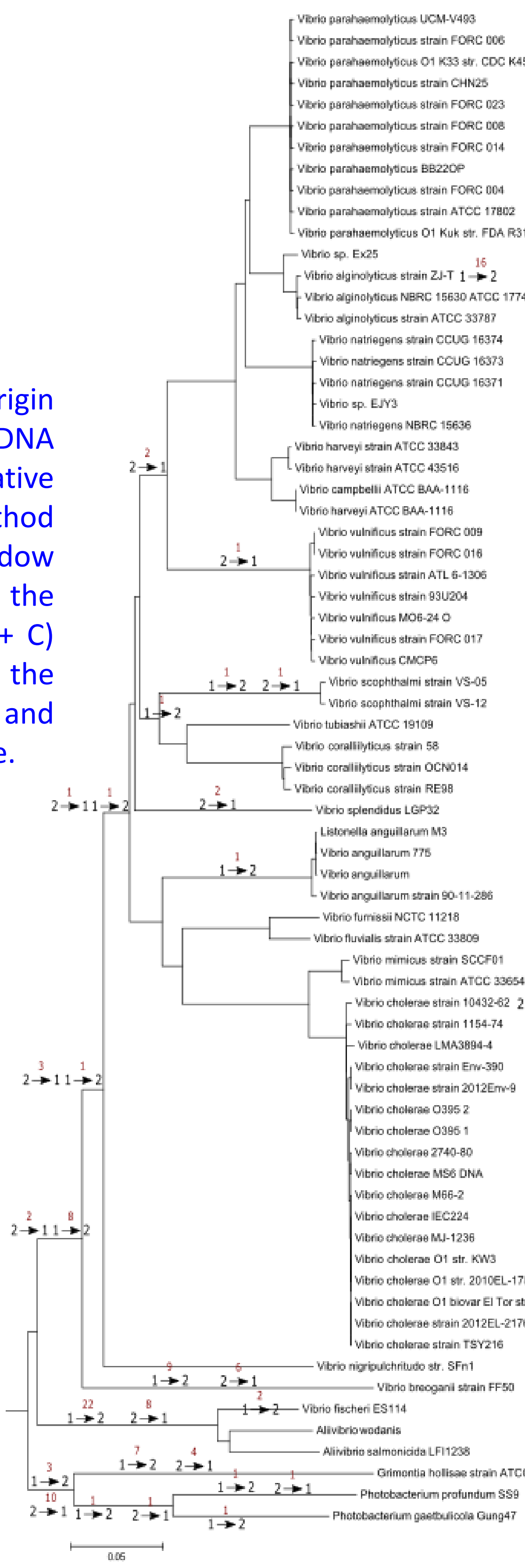


Fig.6 phylogenetic tree of *vibrio* spp.

### Results/Результаты

Fig. 1. Distribution of chromosome lengths

The first chromosomes of all strains have the length of about 3 Mb, whereas the second chromosomes differ dramatically.

Первые хромосомы разных штаммов рода *Vibrio* имеют длину  $3 \times 10^6$  п.о., в то время как длины вторых хромосом существенно различаются у штаммов разных видов.

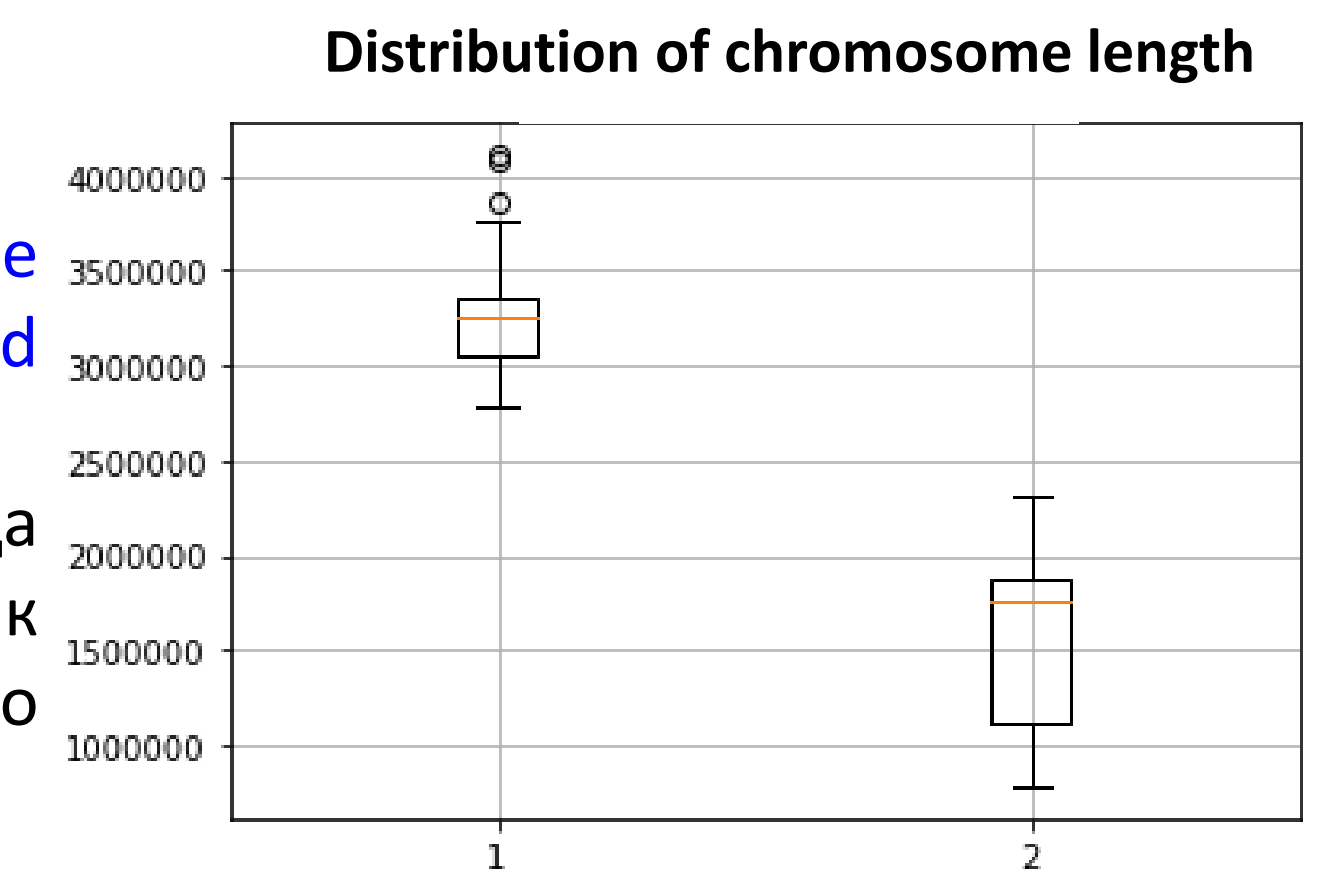


Fig. 2. Distribution of replichore balance

The replichore balance is more stable for the first chromosomes than for the second ones.



Fig. 2. Distribution of replichore balance

Длины реплихор на первых хромосомах более сбалансированы, чем длины реплихор на вторых хромосомах.

Fig.3 Distribution of gene fraction on the leading strand.

In the first chromosomes, about 60% genes are located on the leading strand. For the second chromosomes the fraction of genes on the leading strand is not stable during evolution. Доля генов, лежащих на лидирующей цепи, для первых хромосом всех видов вибрионов составляет 58-62%, в то время как для вторых хромосом перепредставленность генов на лидирующей цепи менее выражена.

Distribution of gene fraction located on the leading strand

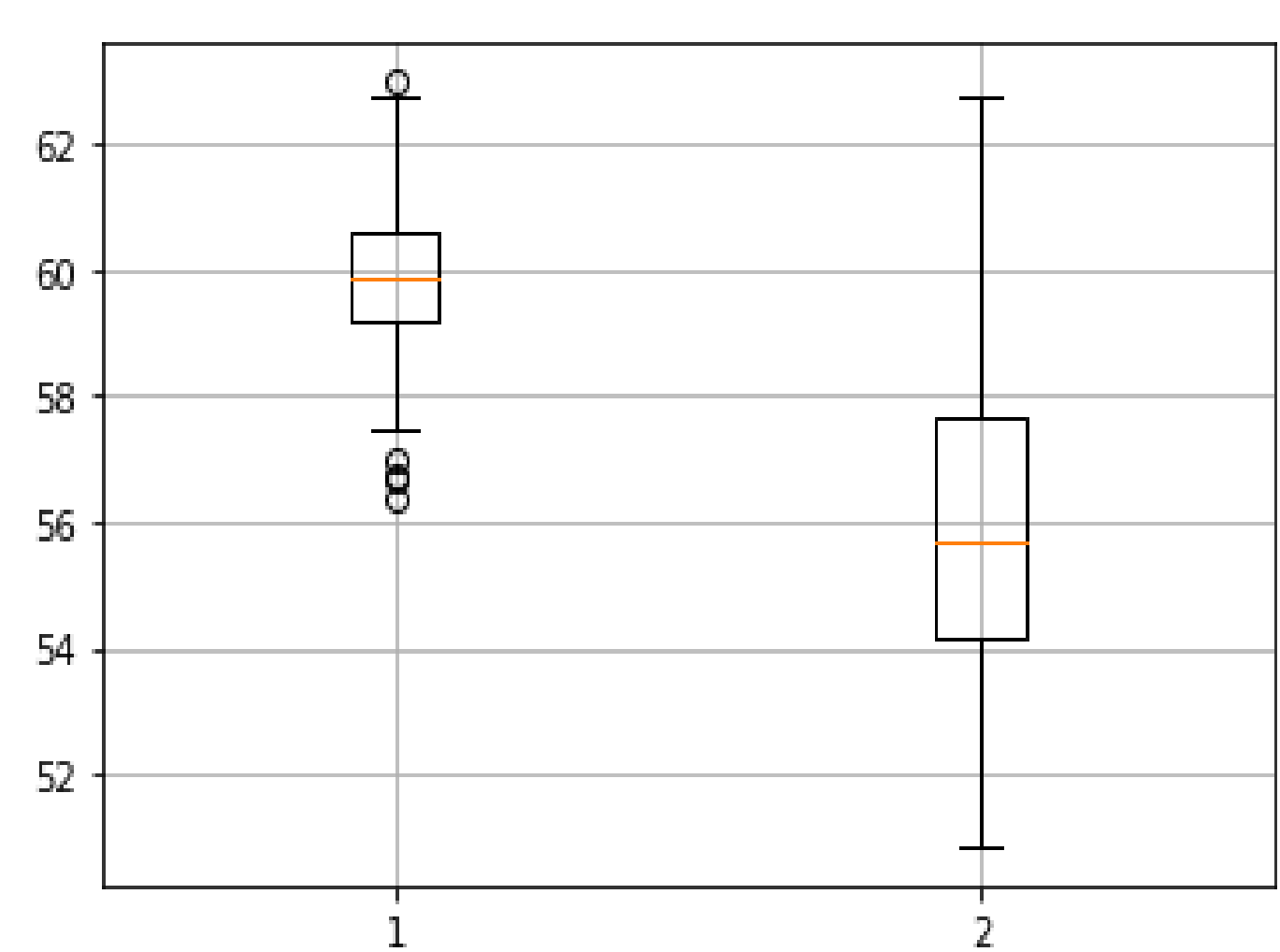


Fig. 4. Histogram of translocated blocks lengths

We identified 196 genes translocated between chromosomes. 29 of them were translocated several times during evolution. We did not observe over-representation of any GO terms in the set of translocated genes.

Translocated genes do not tend to keep their position on the leading (resp., lagging) strand. The breakpoints of large translocations, of 16 genes in *Vibrio alginolyticus* and of 79 genes in *Vibrio parahaemolyticus* are formed by genes of 16S rRNA and 23S rRNA.

Было выявлено 196 генов, в истории которых происходили транслокации между хромосомами. 29 из них транслоцировались несколько раз. Среди параллельных транслокаций не выявлено блоков, содержащих несколько генов.

Мы не выявили перепредставленности случаев сохранения цепи среди всех реконструированных транслокаций, однако транслоцированные участки, содержащие несколько генов, чаще сохраняют цепь, чем транслокации одного гена. Не было выявлено перепредставленности каких-либо GO категорий для генов, в истории которых происходили транслокации.

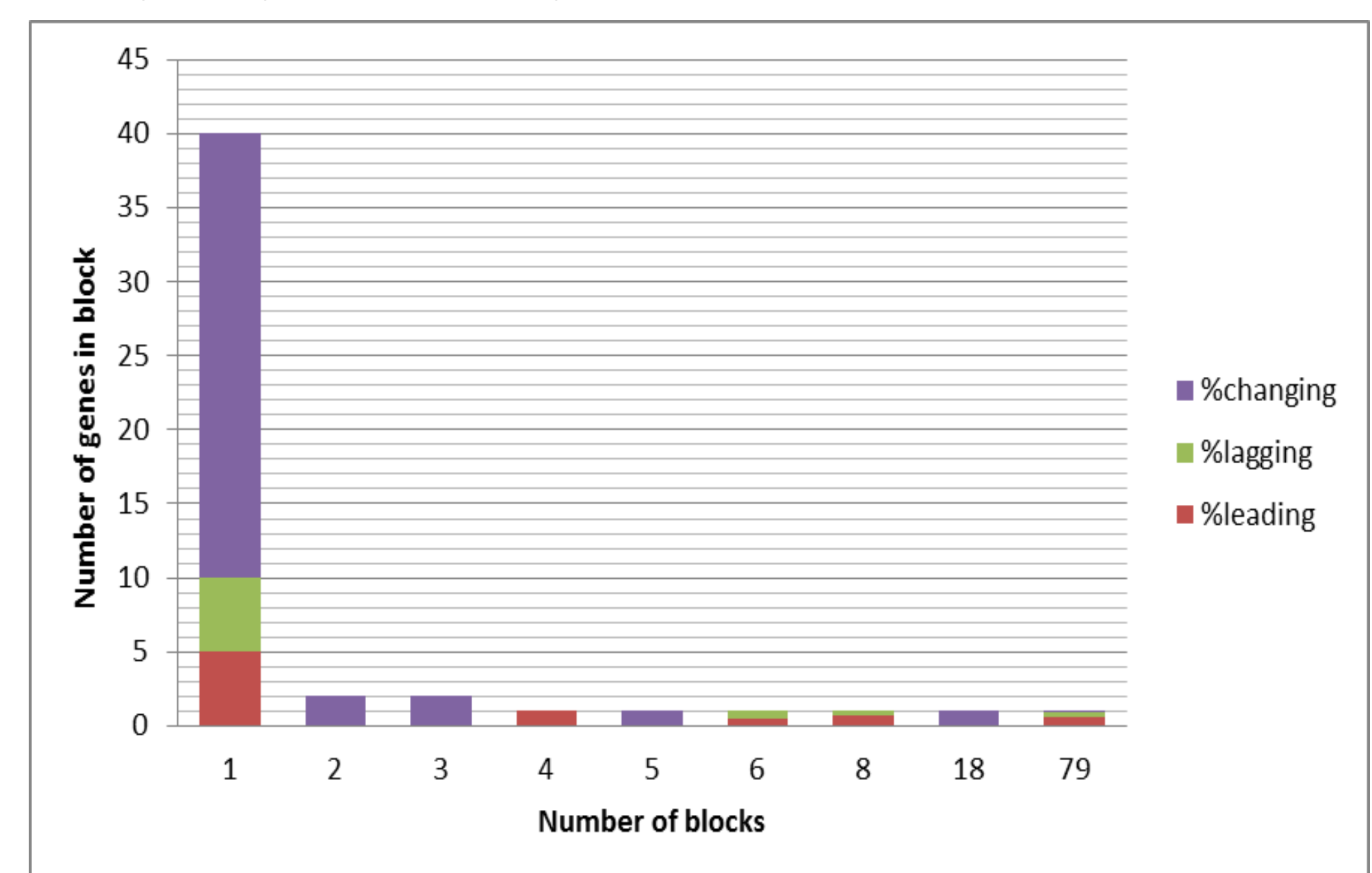


Fig. 5 Balance of inversion part placed on different replichores.

For twelve *V. parahaemolyticus* strains we reconstructed ten inversions in the first chromosomes and two, in the second ones. All reconstructed inversions have breakpoints in different replichores. Moreover, all inversions are strongly symmetrical.

Для 12 штаммов *V. parahaemolyticus* было реконструировано 10 инверсий на первых хромосомах и 2 – на вторых.

Все реконструированные инверсии включают начало или конец репликации, что подтверждает сильный отрицательный отбор на инверсии, лежащие внутри одной реплихоры  $p\text{-value} = 10^{-14}$ . Все инверсии расположены симметрично вокруг точек начала и конца репликации, что подтверждает отрицательный отбор на инверсии, нарушающие баланс реплихор.

Balance of inversion parts located in different replichores

