



Большинство бактериальных геномов имеют одну хромосому, которая может быть дополнена небольшими необязательными плазмидами. Однако около 10% бактерий с отсекуенными геномами содержат более одного стабильного репликона большого размера. Некоторые вторичные репликоны видоспецифичны, могут отвечать за патогенность или нести симбиотические гены. Другие репликоны являются общими для рода и содержат гены домашнего хозяйства. Мы рассматриваем два рода бактерий с нестандартной геномной архитектурой: *Azospirillum* и *Prevotella*.

Most bacterial genomes have a single chromosome, which can be supplemented with small optional plasmids. However, about 10% of bacteria with sequenced genomes contain more than one stable replicon of a large size. Some secondary replicons are species-specific, may be responsible for pathogenicity or carry symbiotic factors. Other replicons are common to the genus and contain household genes. We consider two genera of bacteria with non-standard genomic architecture, *Azospirillum* and *Prevotella*.

### Azospirillum

*Azospirillum* — род азотфиксирующих бактерий, чемпионов по количеству крупных репликонов в геноме (от 6 до 10). Мы получили от коллег полный геном *Azospirillum* sp., выделенного из почвы черневой тайги, остальные геномы были взяты из RefSeq.

*Azospirillum* is a genus of nitrogen-fixing bacteria, champion in the number of large replicons in the genome (from 6 to 10). We were provided with the complete genome of *Azospirillum* sp. isolated from the soil of the Cherevaya taiga, the remaining genomes were taken from RefSeq.

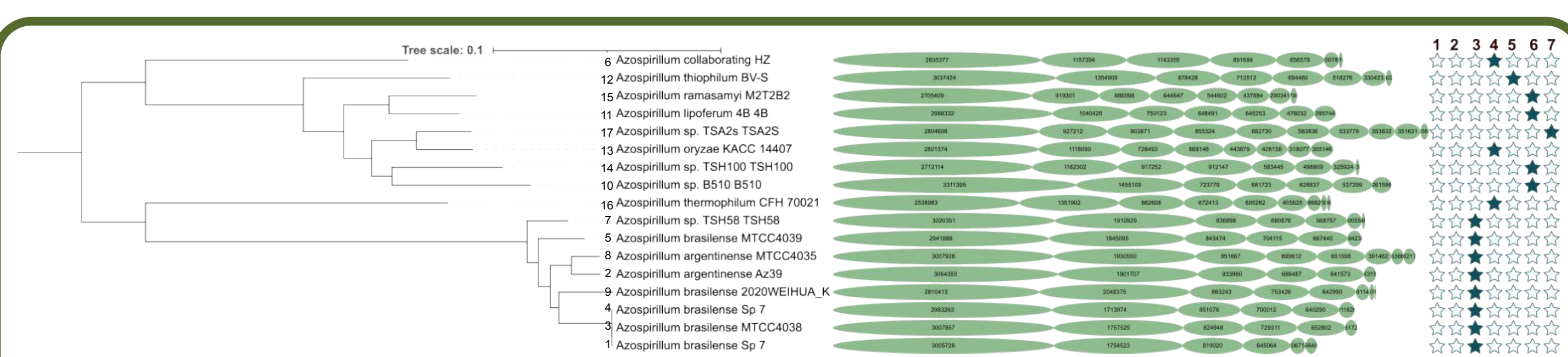
### Prevotella

*Prevotella* — часть здоровой микробиоты слизистых человека, но бактерии этого рода могут участвовать в воспалительных процессах. Некоторые виды содержат одну крупную хромосому, другие содержат два крупных репликона.

*Prevotella* is part of the healthy microbiota of human mucous membranes, but may be involved in inflammatory processes. Some species contain one large chromosome, others contain two large replicons.

**Цель:** сравнительно-геномный анализ родов и реконструкция эволюции

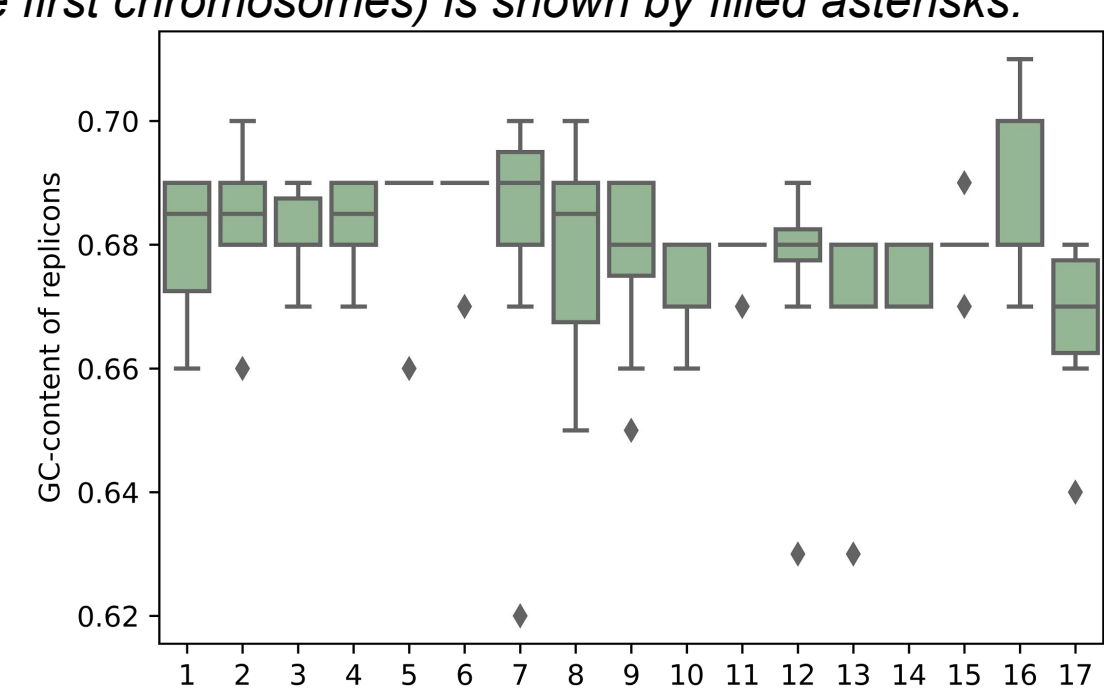
**Goal:** comparative genomic analysis of genera and reconstruction of their evolution



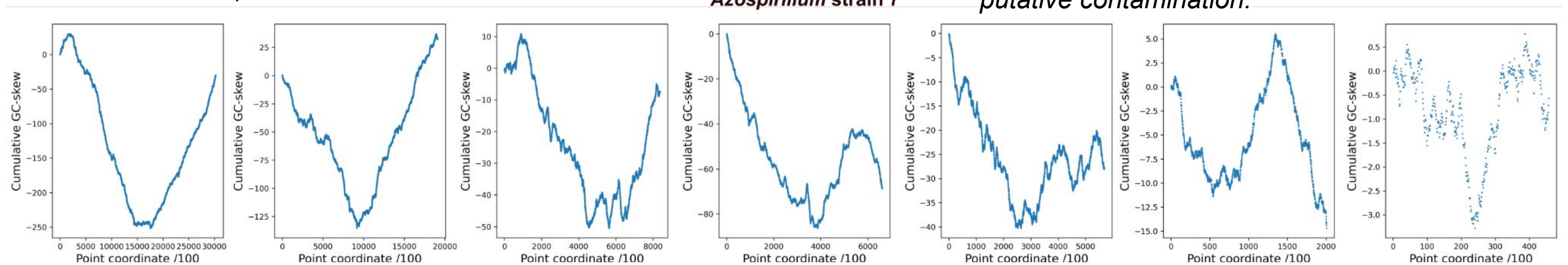
**Рис.1.** Филогенетическое дерево *Azospirillum*. Зеленые эллипсы показывают число репликонов и их размер. Звездочки отражают филогенетический паттерн 31 универсальной ортогруппы, все члены которой не лежат на первой хромосоме. Залюпка показывает, на каких репликонах расположены гены этих ортогрупп.

Филогенетический анализ показал две крупные клады. Крупнейший репликон стабилен по размеру и по составу генов, остальные репликоны более разнообразны, встречаются небольшие уникальные репликоны. У штаммов азоспириллума GC состав репликонов ~70%, но встречаются выбросы с 62-64%. Это может свидетельствовать о контаминации сборок, так как у других почвенных бактерий GC состав ~60%.

Phylogenetic analysis revealed two large clades. The largest replicon has a conserved size and gene composition, other replicons are more diverse, some small replicons are unique. GC-content of *Azospirillum* replicons is 70% with several outliers of 62-64%, indicative of a possible contamination (GC-content of other soil bacteria is ~60%).



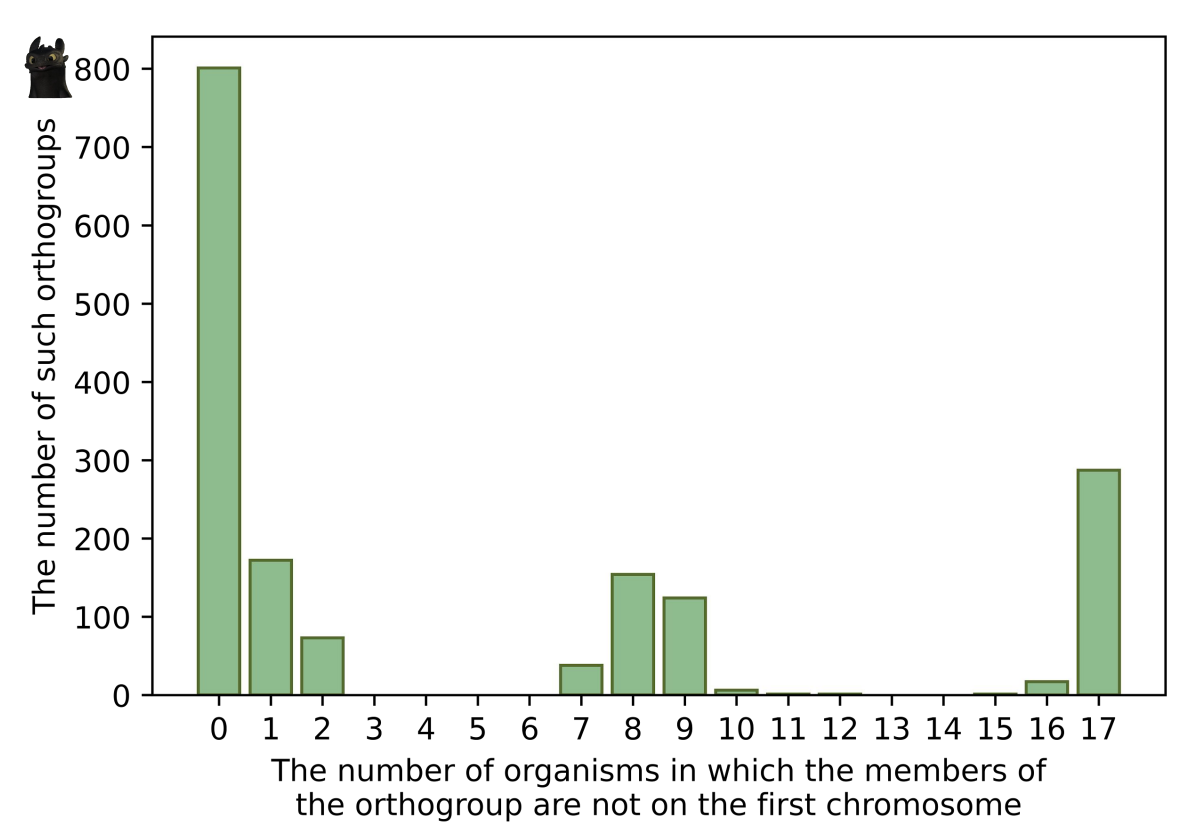
**Рис.2.** GC-состав репликонов. Выбросы могут означать контаминацию.



**Рис.3.** Пример накопительных GC-skew [(G-C)/(G+C)] для *Azospirillum* с семью репликонами, отсортированными по длине.

При сравнении репликонов внутри генома виден поэтапный переход формы накопительного GC-skew от типичного для хромосомы к шумному (типичному для плазмиды). Из 1675 однокопийных универсальных ортогрупп ~800 всегда лежат на хромосоме, 287 всегда лежат на вторичных репликонах. Мы выявили предпочтительные направления рекомбинации: между вторым и третьим репликонами и между репликонами с четвертого и меньше.

Comparison of cumulative GC-skews between replicons within a genome revealed transition from typical for a chromosome to more noisy/non-structured one (typical for plasmids). Of 1675 single-copy universal orthogroups, ~800 always placed in the chromosome, 287 always placed on secondary replicons. We detected the preferred directions of recombination: between the 2nd and 3rd replicons and between replicons from 4th and smaller.



**Рис.4.** Количество ортогрупп, у которых ровно в заданном числе штаммов представители расположены не на первом репликоне.

Гены 16S рРНК заметно различаются между видами, но внутри одного генома практически идентичны. С помощью выравнивания, мы обнаружили свидетельства горизонтального переноса генов 16S рРНК между штаммами.

The 16S rRNA genes differ between species, but within the same genome they are almost identical. We performed gene alignment and found the evidence of horizontal gene transfer of 16S pPHK genes between strains.

**Рис.5.** Фрагменты филогенетического дерева генов 16S рРНК. (А) На ветке кластеризовались 16S рнк гены одного штамма. (В) Клада содержит 16S рнк гены из разных штаммов, что позволяет предполагать горизонтальный перенос.

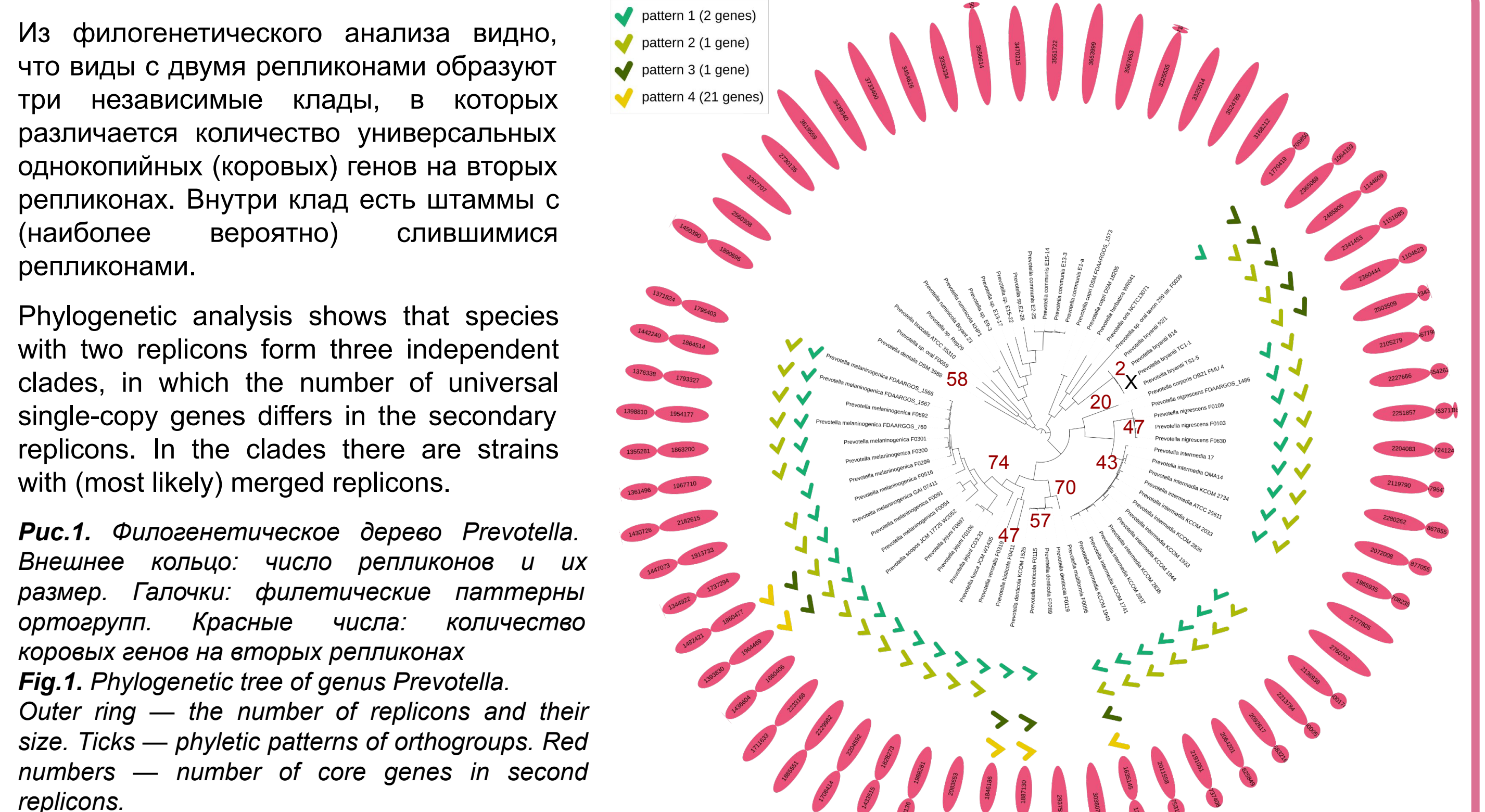
**Fig.5.** Clades of 16S rRNA genes tree. (A) The clade consists of the genes from the same strain. (B) The clade comprises genes from different strains indicating putative HGT.

### Вывод:

50% общих однокопийных генов лежат на основной хромосоме, 20% у всех штаммов на вторичных репликонах. Хромосома более стабильна по размерам в рамках рода, у вторичных репликонов много следов рекомбинации. Есть небольшие плазмиды с сильно отличающимся GC составом, возможно это контаминация.

### Conclusion:

50% of common single-copy genes are in the main chromosome, 20% of all genes are on secondary replicons. The chromosome has stable size within the genus, secondary replicons have many traces of recombination. There are small plasmids with low GC content (putative contamination).

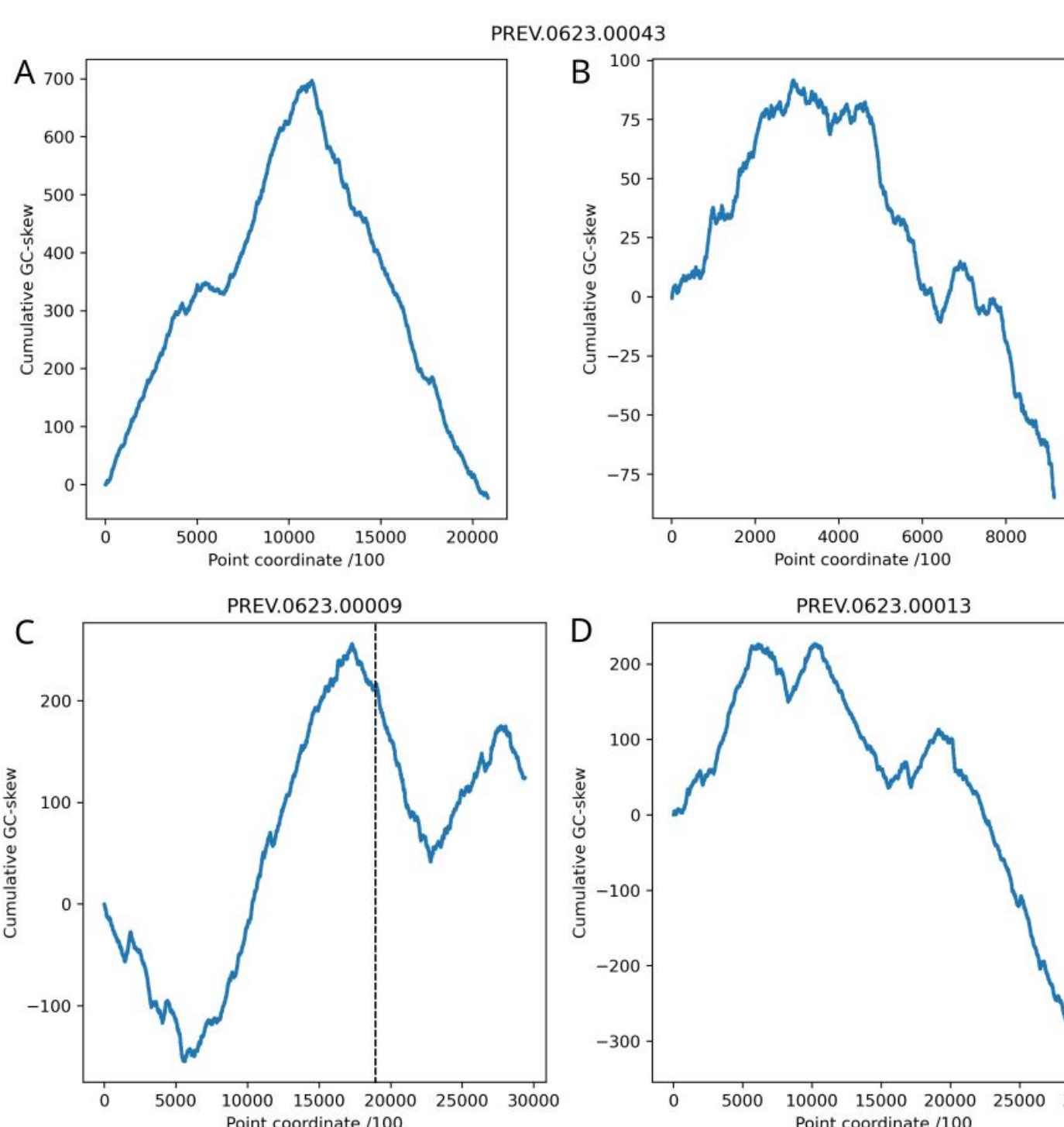


Из филогенетического анализа видно, что виды с двумя репликонами образуют три независимые клады, в которых различается количество универсальных однокопийных (коровых) генов на вторых репликонах. Внутри клады есть штаммы с (наиболее вероятно) слившимися репликонами.

Phylogenetic analysis shows that species with two replicons form three independent clades, in which the number of universal single-copy genes differs in the secondary replicons. In the clades there are strains with (most likely) merged replicons.

**Рис.1.** Филогенетическое дерево *Prevotella*. Внешнее кольцо: число репликонов и их размер. Галочки: филогенетические паттерны ортогрупп. Красные числа: количество коровых генов на вторых репликонах.

**Fig.1.** Phylogenetic tree of genus *Prevotella*. Outer ring — the number of replicons and their size. Ticks — phyletic patterns of orthogroups. Red numbers — number of core genes in second replicons.



На дереве показаны четыре филогенетических паттерна: (1) наиболее представленная на вторых хромосомах ортогруппа, (2-3) коровые гены, локализованные на второй хромосоме в кладе X и (4) немонафилический паттерн, общий для 21 ортогруппы.

The tree shows four phyletic patterns: (1) the most represented orthogroup on the second chromosomes, (2-3) cortical genes localized on the second chromosome on clade X, and (4) a non-monophyletic pattern common to 21 orthogroups.

Из 523 универсальных однокопийных ортогрупп, гены 123 ортогрупп всегда лежат на первых хромосомах. Нет ортогрупп, у которых гены располагаются на второй хромосоме во всех кладах с двумя репликонами. Ситуации, когда сразу большая группа генов располагается на второй хромосоме у большого числа штаммов, очень редки.

Out of 523 universal single-copy orthogroups, genes of 123 orthogroups are always localized in the first chromosomes. There are no orthogroups whose genes are located on the second chromosomes in all clades with two replicons. Situations when a large group of genes is located on the second chromosome in the same large set of strains are very rare.

### Вывод:

Превотеллы с двумя репликонами образуют три клады на филогенетическом дереве. Размер вторых репликонов и количество общих однокопийных генов на них значительно варьируется между группами. Более того, нет ортогрупп, общих для всех вторых репликонов, что подтверждает их независимое образование в разных кладах.

### Conclusion:

*Prevotella* with two replicons form three clades on the phylogenetic tree. The size of the second replicons and the number of common single-copy genes on them varies significantly between groups. Moreover, there are no orthogroups common to all second replicons, which confirms their independent formation in different clades.