

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Bipartite genomes in Enterobacterales: Independent origins of chromids, elevated openness and donors of horizontally transferred genes

Cecilie Bækkelund Sonnenberg¹ and Peik Haugen^{1*}

¹Department of Chemistry and the Center for Bioinformatics (SfB), Faculty of Science and Technology, UiT The Arctic University of Norway, N-9037, Tromsø, Norway

*Authors for Correspondence:

Peik Haugen: E-mail: peik.haugen@uit.no

Abstract

Multipartite bacteria have one chromosome and one or more chromid. Chromids are believed to have properties that enhance genomic flexibility, making them a favored integration site for new genes. However, the mechanism by which chromosomes and chromids jointly contribute to this flexibility is not clear. To shed light on this, we analyzed the openness of chromosomes and chromids of the two bacteria, *Vibrio* and *Pseudoalteromonas*, both in the Enterobacterales order of gamma-proteobacteria, and compared it with monopartite genomes in the same order. We applied pan-genome analysis, codon usage analysis and the HGTeator software to detect horizontally transferred genes. Our findings suggest that the chromids of *Vibrio* and *Pseudoalteromonas* likely originated from two separate plasmid acquisition events. Bipartite genomes were found to be more open compared to monopartite. We found that the shell and cloud pan-gene categories drive the openness of bipartite genomes in *Vibrio* and *Pseudoalteromonas*. Based on this and our two recent studies, we propose a hypothesis that explains how chromids and the chromosome terminus region contribute to the genomic plasticity of bipartite genomes.

Keywords

Vibrionaceae, *Pseudoalteromonas*, multipartite, bipartite, pan-genome, horizontal gene transfer, codon usage bias, chromid

Introduction

Multipartite genomes refer to the presence of multiple replicons in a single bacterial cell and include one large chromosome, as well as one or more replicons (typically average size of 1,5 Mb), called chromids [1,2]. Bacteria with multipartite genomes are commonly found as pathogens or symbionts in animals, humans, and plants, as well as free-living bacteria [3,4]. Although multipartite genomes are found throughout Bacteria (GTDB “Domain”; NCBI “Superkingdom”), 92% of the currently known, are found in Proteobacteria (newly proposed renamed to “Pseudomonadota” by International Committee on Systematics of Prokaryotes [5]). They are distributed among alfa-, beta- and gamma-proteobacteria, with 25%, 46% and 28% of multipartite bacteria found in each group, respectively [4]. Out of all multipartite bacteria, the majority (88%) are bipartite, i.e., they consist of one chromosome and one chromid.

The prevailing theory for the origin of bipartite genomes is that chromids have their origin from plasmids or megaplasmids that have been captured and domesticated by the ancestral host (the plasmid hypothesis) [1]. However, alternative hypotheses exist, such as that chromids can arise from a split of the chromosome (the schism hypothesis) [6], that the entire chromid is acquired through conjugation from another bacterium [7], or that the chromid arises through recombination between a chromid and a plasmid (chromid “rebirth”) [1]. The majority of known chromids have originated from a plasmid or megaplasmid and have plasmid-like replication machineries. For example, in beta-proteobacteria the majority of chromids are found within the *Burkholderiaceae* family [8], and are thought to have originated from two ancestral plasmids. Similarly, in alpha-proteobacteria, most chromids are found within *Rhizobiaceae* and are believed to originate from a relatively small number of plasmids [1].

Exactly why 10% of the currently available bacterial genomes are multipartite, and which purpose the extra replicons may serve is still unclear. Several hypotheses have been suggested [1,2]. One hypothesis is that chromids acquire and loose genes more rapidly, thus providing bacteria with an increased genetic plasticity. This can be advantageous in terms of environmental specialization and niche-specificity [8–10]. E.g., studies have suggested that the gene content of chromids varies more than in chromosomes [7,11], and thus evolve more rapidly and acquire new genes at a faster rate [8], and finally, experience more relaxed selection pressure (i.e., greater evolutionary plasticity) [12]. This hypothesis is also known as the test bed hypothesis [11]. Other suggested hypotheses are that chromids can contribute with replicon-specific gene regulation and expression [13,14], and that extra replicons are responsible for larger genomes and increased genome content [15].

Several different calculations can be performed to provide new insights into the plasticity of multipartite genomes, and potentially differentiate between the alternative hypotheses of their existence. One commonly used approach is to estimate the rate of growth of the so-called pangenome of a species (or genus or a family), also known as the “openness” of a genome [16]. The open or closed state of a pangenome depends on the ability of the bacteria to acquire new genes, for example through horizontal gene transfer. In an open pangenome, new genes are added to the pangenome as more genomes are sequenced or added to the analysis. In contrast, a closed pangenome approaches a constant size as more genomes are added. Heap's law can be used to describe the pangenome size and number of new genes added for each new genome sequences [17] and is formulated as: $n = kN^\gamma$, where n is the pangenome size, N is the number of genomes used and k and γ are the fitting parameters. If $\gamma < 0$, the pangenome is closed, and if $\gamma > 0$, the pangenome is open.

Another frequently used method to study the flexibility of genomes and horizontal gene transfer, is through calculation of codon usage. Codon usage can differ between organisms, as well as between genes of the same genome [18,19]. The typical codon usage of an organism i. e, the preferential use of certain synonymous codons in typical genes, can be distinguished from the codon usage of highly expressed genes (optimal codon usage), and codon usage of horizontally transferred genes (HTGs) (atypical codon usage)[20,21]. Optimal codon usage corresponds to the use of the most abundant tRNAs in the organism, thus leading to faster translation (protein synthesis) [18]. HTGs on the other hand have a codon usage similar to its donor organism. To what extent the codon usage of a HTG deviates from the recipient genomes, depends on how distantly related the donor and recipient genomes are. Variations in relatedness between the donor and recipient, as well as amelioration (that codon usage evolves towards that of the typical genome over time) are limitations that can lead to underestimation of HTGs [22]

Within gamma-proteobacteria, bipartite genomes are exclusively found in *Vibrionaceae* and *Pseudoalteromonas*, both of which belong to the Enterobacterales order (according to the Genome Taxonomy database (GTDB) [23]. *Vibrionaceae* consists of eight genera, all of which have bipartite genomes, whereas *Pseudoalteromonas* is the only bipartite genus among the 44 genera within *Alteromonadaceae*. According to estimates of time since divergence, *Pseudoalteromonas* is much younger than *Vibrionaceae* [24,25]. Both the *Vibrionaceae* and the *Pseudoalteromonas* chromids are believed to have originated from plasmids from the same order [24,26–30]. The replication of chromosomes and chromids of *Vibrionaceae* have been heavily studied, with research showing that both replicons are

bidirectionally replicated, and the replication is highly coordinated with synchronized termination of the replicons [31–33]. Replication of most *Pseudoalteromonas* chromids occur in an unidirectionally manner, while some are replicated bidirectionally. Additionally, the replication termination has been proposed to be synchronized [25]. We recently studied the global gene distribution and gene expression in *Vibrionaceae* [34] and *Pseudoalteromonas* [30]. Briefly, we calculated the pangenomes of 124 *Vibrionaceae* and 25 *Pseudoalteromonas* genomes, mapped the pangene categories on the genomes and compared the gene distribution with gene expression under fast and slow growth conditions. In both cases, core and softcore genes were overrepresented around the origin of replication (*ori1*), whereas shell and unique genes densely populated the regions surrounding the replication terminus (*ter1*). Gene expression strongly correlated with the distance to *ori1*, with higher expression levels closer to *ori1*. The *Vibrionaceae* chromids did not display any distinct gene distribution pattern. In contrast, the core genes of *Pseudoalteromonas* chromids were found to have a strong correlation with *ter2*, regardless of the chromid was replicated bi- or unidirectionally. Gene expression in chromids did not correlate with distance to *ori* or *ter*. Based on the subcellular organization of chromosome and chromid in *Vibrio cholerae* [31,35–37] we found that core/softcore and shell/cloud was spatially separated into separated intracellular regions (the poles of *V. cholerae*). This led us to propose a hypothesis that the bipartite genome structure enables intracellular spatial separation of different pangene categories and that there is a connection between gene placement and gene function.

Research suggests that chromids possess features that promote increased genomic plasticity and that they are a preferred location for horizontally transferred genes. However, the extent to which chromosomes and chromids contribute to the overall plasticity and openness of bipartite genomes is not well understood. Our study aims to address this knowledge gap by calculating the openness of chromids and chromosomes of the bipartite bacteria *Vibrio* and *Pseudoalteromonas*, as well as monopartite genomes, and use codon usage and horizontal gene transfer analysis to determine which genes that contribute to the openness. Based on our data and two recent studies, we propose a hypothesis that describe how chromids and a specific region of the chromosomes appear to contribute to the genomic plasticity of bipartite genomes. Additionally, we establish the origin of *Vibrionaceae* and *Pseudoalteromonas* chromids.

Results

***Vibrio* and *Pseudoalteromonas* belong to the same bacterial order**

The only known cases of bacteria with bipartite genomes within the class of gamma-proteabacteria are *Pseudoalteromonas* and *Vibrionaceae*. According to the NCBI taxonomy classification, *Vibrionaceae* and *Pseudoalteromonas* belong to separate orders (i.e., Vibrionales and Pseudoalteromonadales). Interestingly, according to new phylogenomics-based data included in the Genome Taxonomy database (GTDB) [23] (Oren et al 2015), *Pseudoalteromonas* and *Vibrionaceae* both group within the order Enterobacterales. Figure 1 shows the overall phylogenetic relationship between bacterial families and their respective genera that form the order Enterobacterales, based on information derived from GTDB release 89. Lineages with bipartite genomes are highlighted.

The fact that *Vibrionaceae* and *Pseudoalteromonas* belong to the same order, raises the possibility, although unlikely, that their chromids originate from a single acquisition event in a common ancestor. Such a scenario would invoke a common origin followed by long-term retention of the chromid, and then massive losses in all representatives of Enterobacterales, except *Vibrionaceae* and *Pseudoalteromonas*. A more likely explanation is that the chromids originate from two separate acquisition events.

Separate origin of chromids in *Vibrionaceae* and *Pseudoalteromonas*

We used ParA and ParB as phylogenetic markers to discriminate between the two hypotheses i.e., a common or separate origin of the *Vibrionaceae* and *Pseudoalteromonas* chromids. ParA and ParB have fundamental roles in partitioning of replicons [38], and their conserved function and widespread distribution in Bacteria and Archaea make them suitable for establishing the origin of the chromids. A concatenated ParA-ParB alignment was created from sequences identified by BLASTp when using ParA and ParB sequences from *Pseudoalteromonas* and *Vibrionaceae* chromids as queries against the nr. protein database. The final dataset included a total of 376 residues from ParA and 313 residues from ParB (few residues were kept due to highly divergent regions that could not be reliably aligned).

Figure 2 shows the resulting maximum likelihood tree (WAG+G+I model). Chromosomal sequences were used as the outgroup. Here, chromidial ParA-ParB from *Vibrionaceae* branches together with plasmid sequences from *Alteromonas*, *Pseudoalteromonas* and *Paraglaciecola* (Plasmid group 2), whereas chromidial *Pseudoalteromonas* ParA-ParB form a sister group with another set of plasmids, i.e., from *Shewanella*, *Vibrio* and *Pseudoalteromonas* (Plasmid group 1). These relationships are

supported by bootstrap values of 90% and 75%, respectively. In summary, our result agrees with separate origins of the *Vibrionaceae* and *Pseudoalteromonas* chromids and suggests that both chromids were acquired from plasmids belonging to the Enterobacterales gene pool.

The chromids in *Pseudoalteromonas* and *Vibrio* play a significant role in the openness of the two genomes

It has been proposed that the main advantage of keeping multiple replicons is increased genetic flexibility, often termed “openness” (e.g., [8,11,12,30]). A commonly used method to estimate the openness of a pangenome, is to perform curve fitting of the pangenome size versus number of genomes using Heap’s law [16,17]. Heaps law is formulated $n = kN^\gamma$, where an exponent $\gamma > 0$ indicates an open pangenome, i.e., the pangenome will grow/gain genes as new genomes are sequenced and added to the analysis. An exponent $\gamma < 0$ indicates a closed pangenome that will not grow in size as new genomes are added. To estimate to what extent the chromosome and the chromid contribute to the pangenome openness we made two separate datasets consisting of 50 complete *Vibrio* and 26 complete *Pseudoalteromonas* genomes. The datasets are non-redundant, meaning that only one complete genome per available species was included (see Table S1 for complete list of genomes). We then calculated the pangenome size and Heap’s exponent for the chromosome, chromid and total genome (see Table S2). The pangenome of *Vibrio* consists of 822 core (encoded by all 50 genomes), 1505 softcore (encoded by 47 genomes), 8463 shell (encoded by 46 and 3 genomes), and 37,177 cloud (encoded by 2 genomes). The *Pseudoalteromonas* pangenome consists of 1386 core (encoded by all 26 genomes), 1787 softcore (encoded by 24 genomes), 5096 shell (encoded by 23 and 3 genomes), and finally 20,635 cloud (encoded by 2 genomes).

Figure 3 shows the calculated pangenome sizes relative to the number of added genomes (median of 100 randomly generated combinations of genome datasets). For both *Vibrio* and *Pseudoalteromonas* the size of the chromosomal, chromidial and total genomes increase as more genomes are added to the analysis, more in the beginning of the curve and less after 10 genomes are added. The Heap’s exponent associated with the *Vibrio* chromid (0.668 ± 0.001) and the chromosome (0.660 ± 0.003) are virtually identical. This means that the two replicons are equally “open”, but because of its bigger size, the chromosome hosts the majority of new genes. For *Pseudoalteromonas* the chromid exponent (0.685 ± 0.007) is considerably larger than that of the chromosome (0.594 ± 0.002) and total genome (0.601 ± 0.003). With the highest Heap’s exponent, the chromid contributes considerably to the openness of the

Pseudoalteromonas genome. In summary, we have used Heap's law to evaluate the openness of the chromosome and chromid of *Vibrio* and *Pseudoalteromonas* by calculating the pangenome sizes and Heap's exponents. The *Vibrio* chromosome and chromid are equally open, whereas the *Pseudoalteromonas* chromid is more open than the chromosome.

Bipartite genomes are more open compared to monopartite genomes

Next, we compared the openness of the *Pseudoalteromonas* and *Vibrio* genomes to that of monopartite genomes of closely related genera. Hypothetically, the structural organization of genomes into one or multiple replicons can have a major impact on the flexibility of the genomes. The four relatively closely related genera *Alteromonas*, *Idiomarina*, *Rodentibacter* and *Yersinia* (all from Enterobacteriales) with monopartite genomes were chosen for the analysis, for comparison to bipartite genomes. For each genera, the Heap's exponent was calculated from a random combination of an increasing number of genomes (using seven permutations) (see Table S1). This was done to test what effect the number of genomes and genome combinations have on the resulting Heap's exponent. A dataset consisting of 27 *Escherichia coli* (species level) genomes was added as a control.

Figure 4A shows plots with Heap's exponent for *Pseudoalteromonas* and *Vibrio* relative to the number of genomes. Here, the Heap's exponent is widely distributed when only a few numbers of genomes are included in the datasets. As the number of genomes increases, the exponents are less distributed (see Table S2 for complete list of Heap's exponents). Similarly, the calculations for *Pseudoalteromonas* chromids vary greatly for small datasets but become more stable as the number of included genomes increases. These results show, as expected, that larger dataset (>10 genomes) result in more stable Heap's values. Figure 4B shows the corresponding plots for genera with monopartite genomes. When the number of genomes is small, the distribution of Heap's exponent is wide for *Yersinia*, *Alteromonas* and *Rodentibacter*, whereas for *Idiomarina*, the distribution is smaller.

Figure 4C shows a summary of the results from Figure 4AB through curve fitting of the Heap's exponents. All bipartite replicons have larger Heap's exponents compared to the monopartite genomes. For example, at 10 genome datasets the lowest Heap's value for bipartite are 0.618, whereas the highest Heap's value for monopartite are 0.572. These results show that, with the currently available genomes, bipartite genomes have more open pangenomes, and thus appear more genetically flexible than monopartite counterparts. Chromids have the most open state of all replicons compared. Notably, how the exponent will change when more genomes become available is however unclear.

In summary, we plotted the Heap's exponent relative to the size of genome datasets to compare openness of monopartite versus bipartite genomes. With the currently available datasets, bipartite genomes appear more open than that of closely related monopartite bacteria.

Codon usage is specific for each pangene category rather than for each replicon type

Next, we used codon usage bias calculations to further explore the plasticity of bipartite genomes. Newly acquired genes are expected, in general, to have different codon usage profiles compared to those of most genes, especially genes with essential cellular roles (e.g., for cellular growth). Codon bias analyses are therefore used for exploring evolutionary aspects, including lateral transfer of genes.

Therefore, we first measured the relative synonymous codon usage (RSCU) for all individual genes in each of the 50 *Vibrio* and 26 *Pseudoalteromonas* genomes and performed a correspondence analysis of the RSCU values. Variations in codon usage among different pangene categories were explored by dividing the gene datasets into *core*, *softcore*, *shell* and *cloud* genes, and visualize the gene categories in different colors. Axis1 and Axis2 correlate with the two main influencing factors of codon usage bias. They represent 10.98% and 8.07% of the total variation for *Vibrio* and 10.97 % and 7.52% of the total variation for *Pseudoalteromonas*, respectively.

Figure 5AB shows a broad distribution of codon usage in both *Vibrio* and *Pseudoalteromonas* that are to a great extent specific for each pangene category. In *Vibrio*, core and softcore genes are densely clustered toward the upper and lower right quadrants, whereas the shell and especially cloud genes are distributed towards upper left quadrant. In *Pseudoalteromonas*, core and softcore genes are distributed densely in upper left quadrant, shell genes toward the lower quadrants and in upper left quadrant.

Figure 4CD shows PCA plots of the RSCU data described above (from Fig. 4AB). Codon usage clusters based on pangene categories and not on the type of replicon. This result is supported by correlation analysis of the RSCU values for each pangene category and analysis of median effective number of codons (ENC) for each pangene category (see Table S3 for global RSCU values and Table S4 for correlation plot and ENC values).

In summary, we performed COA and PCA on RSCU values to identify major trends of codon usage patterns in *Vibrio* and *Pseudoalteromonas*. Both type of plots show that codon usage is specific for each pangene category rather than type of replicon. This is valid for both *Pseudoalteromonas* and *Vibrio*. Similar codon usage for each pangene category indicates that they also have different evolutionary trajectories, which we explore further (see below).

Shewanella* represents the top donor of HTGs to *Vibrio* and *Pseudoalteromonas

To identify putatively horizontally transferred genes (HTGs) in *Vibrio* and *Pseudoalteromonas*, we used HGTeator [39], which is a software for genome-wide detection of horizontal gene transfer events based on homology searches. For *Pseudoalteromonas* we defined horizontally transferred genes as all genes that originate from a donor outside of *Alteromonadaceae*, whereas for *Vibrio* horizontally transferred genes come from outside *Vibrionaceae*.

Figure 6AB shows the number of HTGs detected for each pangene category on each replicon. HTGs comprise 11% and 23% of the total number of genes in the pangenomes in *Vibrio* [24,529 genes / 7308 gene clusters (12 core, 32 softcore, 1496 shell, 4765 cloud)] and *Pseudoalteromonas* [19,970 genes / 4310 gene clusters (309 core, 424 softcore, 2510 shell, 2389 cloud)], respectively. In *Vibrio*, the majority of HTGs (98%) are shell or cloud genes. These are distributed on the chromosome, where they make up 15% of shell and 13% of cloud genes, and on the chromid where they make up 20% (shell) and 16% (cloud). Notably, the *Vibrio* dataset contains 35 plasmids (from 19 genomes), of which 27% of shell genes and 13% of cloud genes are HTGs. For *Pseudoalteromonas*, about half of the HTGs are core and softcore genes. Of these, 15% and 18% of softcore genes are distributed on chromosomes and chromids, respectively. The other half of HTGs corresponds to chromosomal genes where they make up 24% of shell and 12% of cloud genes, respectively, and the corresponding numbers for chromidial genes are 30% (shell) and 13% (cloud). Six genomes contain one plasmid each. Here, 30% of HTGs represent shell and 14% represent cloud genes.

To summarize, in *Vibrio* the identified horizontally transferred genes are typically shell and cloud genes located on both the chromosomes and chromids. In *Pseudoalteromonas*, the HTGs are more evenly distributed among all pangene categories from both chromosomes and chromids.

Figure 6CD shows the phylogenetic distribution of the bacterial gene donors, i.e., the bacterial families from where the predicted HTGs in *Vibrio* and *Pseudoalteromonas* originated from. In both *Vibrio* and *Pseudoalteromonas* the main contributors are families within the gamma-proteobacteria orders Enterobacterales and Pseudomonadales (according to GTDB classification), accounting for 66% and 22% of the total HTGs in *Vibrio* and 61% and 21% in *Pseudoalteromonas*, respectively. For *Pseudoalteromonas*, the top three donor genera are *Shewanella* (17%; *Shewanellaceae*), followed by *Vibrio* (11%; *Vibrionaceae*) and *Photobacterium* (5%; *Vibrionaceae*). Similarly, for *Vibrio* the top three donors are *Shewanella*

(13%; *Shewanellaceae*), *Marimonas* (6%; *Marinomonadaceae*), and *Psychromonas* (6%; *Psychromonadaceae*).

In summary, we found that the majority of HTGs in *Vibrio* and *Pseudoalteromonas* originates from Enterobacterales and Pseudomonadales, with *Shewanella* representing the top donor of all genera.

Discussion

Here we continue our studies on the bipartite genomes of *Vibrionaceae* and *Pseudoalteromonas*. According to GTDB, *Vibrionaceae* and *Pseudoalteromonas* both belong to Enterobacterales [23]. Based on an inferred ParAB phylogeny, we first established that the *Vibrio* and *Pseudoalteromonas* chromids do not share the same last common ancestor. It is therefore more likely that their chromids originate from two separate plasmid acquisition events. The two plasmids are however likely from the same Enterobacterales gene pool. We then calculated the pangenome and openness of the *Vibrio* and *Pseudoalteromonas* genomes and found that the *Vibrio* chromosome and chromid are equally open (i.e., the chromosome and chromid pangenome size increase at a similar rate as more genomes are added to the analysis), whereas the *Pseudoalteromonas* chromid is more open than the chromosome. Compared with monopartite genomes, bipartite are more open, at least based on today's available genome datasets. We next used codon usage bias calculations to elucidate which type of genes are more likely to have been acquired horizontally, thus leading to open bipartite genomes in *Vibrio* and *Pseudoalteromonas*. The data support that codon usage is specific to each pangene category regardless of which replicon they reside in. The vast majority of HTGs in *Vibrio* are shell or cloud genes, whereas HTGs in *Pseudoalteromonas* are more evenly distributed among all pangene categories.

By comparing the bipartite genomes of *Vibrio* and *Pseudoalteromonas* with monopartite genomes of related bacterial families, we showed that bipartite genomes appear more open than monopartite. The increased openness suggests that bipartite genomes have a higher capacity to acquire genes [40]. Using codon usage bias calculations and the HGTeator tool we therefore set out to identify which type of genes are typically horizontally acquired by vibrios and pseudoalteromonases. We found that the codon usage in both *Vibrio* and *Pseudoalteromonas* group based on which pangene category genes belong to, and not based on which replicon genes reside on (chromidal or chromosomal placement). Notably, codon usage of cloud genes differs most from that of core genes (compared to shell genes), which are typically more highly expressed and therefore assumed to use codons better adapted to the

translation machinery (adaption) [18,21]. This supports that cloud genes include a higher portion of more recently acquired genes. A similar pattern was reported for the multipartite bacterium *Sinorhizobium meliloti*, where codon usage of core genes on the chromosome and chromid were more similar than when compared to unique genes on the same replicons [41]. To conclude, less optimal codon usage of shell and cloud genes agree with data from our HGTeator analysis, which suggests that as much as 98% of the detected HTGs in vibrios are either cloud or shell genes.

For *Pseudoalteromonas* the general picture is similar, but here the HGTeator result suggests that about half of the HTGs are core/softcore genes, whereas the other half corresponds to shell and cloud genes. The high proportion of HTGs among core/softcore is somewhat puzzling to us. To be detected as HTG, BLAST searches must identify the closest hit outside of *Alteromonadaceae*. We speculate that this result can be explained by the fact that *Pseudoalteromonas* is relatively young compared to *Vibrio* [502–378 vs 1100–900 million years ago [24,25], respectively], and more genes will thus potentially be identified as HTG among core/softcore. The rationale is that HTGs in the last common ancestor (LCA) of extant *Pseudoalteromonas* bacteria have had approx. 500 million fewer years to adapt to the translation machinery than the corresponding genes in *Vibrio*. Also, *Pseudoalteromonas* have had less time to diverge from the LCA into different species, which subsequently can occupy various biological niches (like *Vibrio* that comprises at least 140 species). Consequently, our pangenome analyses identified 1386/1787 and 822/1505 core/softcore genes in *Pseudoalteromonas* and *Vibrio*, respectively. To summarize, HTGs in *Vibrio* are almost exclusively from the shell and cloud categories, whereas about half of HTGs in *Pseudoalteromonas* are shell and cloud genes.

Based on the results presented above, a new question arises: If a significant portion (>98% and >50%) of HTGs belong to the shell and cloud categories, where in the genomes are they typically located, and could their location explain why bipartite genomes are more flexible than monopartite genomes? In the light of this and previous studies, we suggest that the chromid and the lower half of the chromosome are particularly available for integration of new genes, and thus contribute to the elevated flexibility/openness of bipartite genomes (Figure 7). We recently mapped the pangenome categories on the genomes of *Vibrionaceae* [34] and *Pseudoalteromonas* [30] and discovered distinct distribution patterns. On the chromosomes, core and softcore genes are overrepresented around the origin of replication (*oriI*), whereas shell and unique genes densely populate the regions surrounding the replication terminus (*terI*). The *Vibrionaceae* chromids showed no clear gene distribution

pattern, but for *Pseudoalteromonas* the distribution of core genes strongly correlates with *ter2*, regardless of its position [i.e., *Pseudoalteromonas* chromids are replicated bi- or unidirectional, hence the position of *ter2* varies [25]]. Other studies have also found a correlation between density of mobile genetic elements and proximity to the *ter* region. Kopetja et al., discovered that in *Rhodobacterales* core genes are located near *oriC*, whereas phages are located near the terminus [42]. A similar finding was reported by Oliveira et al. [43]. Using a diverse genome dataset, they found a higher frequency of "hot-spots" for horizontal gene transfer that contained prophages near *terC*. The evolutionary process responsible for this distribution pattern is discussed elsewhere [25,29], but from the current results we conclude that chromids and the lower halves of chromosomes appear to be favored "landing sites" for gene acquisition in bipartite genomes.

Material and methods

Enterobacterales reference tree

The phylogenetic tree of Enterobacterales was made using Annotree[44], which is based on phylogeny and taxonomic nomenclature from the Genome Taxonomy database (GTDB)[23]. Notably, in addition to multipartite genomes in *Vibrionaceae* and *Pseudoalteromonas*, there are reports of single strains with chromids in *Alteromonas mediterranea* [45] and in *Plesiomonas shigella* [46].

ParAB phylogenetic tree

BLASTp was used to compile ParA and ParB protein sequences from the databases using ParA and ParB from *Vibrionaceae* and *Pseudoalteromonas* as queries. The protein sequences were aligned using MUSCLE [47]. The alignment was manually adjusted using BioEdit [48], and only unambiguously aligned positions were kept for phylogenetic inference. A total of 689 aa positions were kept. MEGA11 was used to generate a Maximum Likelihood (ML) tree using the WAG model, Gamma distribution of evolutionary rates among sites, with invariant sites allowed (WAG+G+I) [49,50]. Bootstrap analysis with the same parameters as described above was performed with 1000 pseudoreplicates.

Genome retrieval and gene annotation

One dataset for each of the genera *Pseudoalteromonas*, *Vibrio*, *Alteromonas*, *Yersinia*, *Idiomarina* and *Rodentobacter* and *E. coli* was made based on taxonomy of Genome Taxonomy database [23], (see Table S1 for complete lists of genomes). The genomes were downloaded

from the RefSeq database at National Center for Biotechnology Information (NCBI) [51]. All *Vibrio* and *Pseudoalteromonas* genomes were complete. We allowed draft genomes with up to 200 contigs to be included for the other datasets. All genomes were re-annotated using RAST (Rapid Annotation using Subsystem Technology) version 2.0 [52]. To make the datasets non-redundant, FastANI [53] was used to calculate average nucleotide identity values for all genomes against all genomes to select one genome per species.

Pangenome calculation

To classify the annotated protein sequences of each of the seven datasets from *Pseudoalteromonas*, *Vibrio*, *Alteromonas*, *Yersinia*, *Idiomarina*, *Rodentobacter* and *E. coli* into four pangenome categories, we performed pangenome analysis using the clustering algorithm MCL in the software package GET_HOMOLOGUES (v3.1.0 (20180103)) [54]. The parameter “minimum percent sequence identity” was set to 50 and “minimum percent coverage in BLAST query/subj pairs” was set to 75 (default). To calculate the openness of pangenomes, pangenome analysis was performed using 100 permutations (for each datapoint). The median values of the combinations was used to perform curve fitting and calculate Heap’s exponent using power-law regression in the “aomisc package” in R v.4.0.3 [55] (see Table S2).

Calculation of codon usage

To investigate codon usage bias, codonW [56] was used to calculate relative synonymous codon usage (RSCU) and perform correspondence analysis of all genes in *Pseudoalteromonas* and *Vibrio*. Correspondence analysis (COA) was used to identify the major trends of codon usage among the four pangene categories. Each gene is described by a vector of 59 variables (codons) that correspond to the RSCU value of each synonymous codon. Codons without synonymous alternatives were excluded from the analysis (methionine, tryptophane and stop codons UAA, UAG, UGA). CodonW was also used to calculate global RSCU values of the pangenome categories separated based on their respective replicon (either chromosome, chromid or plasmid). The RSCU values were then plotted on a principal component analysis (PCA) (see Table S3 for global RSCU values). Effective number of codons was calculated using the R package “vhcub” [57] (see Table S4). ENC is used to estimate the overall codon bias for each gene in a dataset. ENC values range from 20 to 61, where all synonymous codons are used equally at 61 and only one codon used at 20 [58].

Prediction of horizontally transferred genes

HGTector v2.0b3 [39] was used to identify putatively horizontally transferred genes in *Vibrio* and *Pseudoalteromonas*. A database consisting of 25,859 bacterial RefSeq proteins was downloaded from NCBI [51] and compiled using DIAMOND [59]. DIAMOND BLASTP searches with *Vibrio* pangenomes and *Pseudoalteromonas* pangenomes as queries was performed with the parameters e-value < 1e-05, sequence identity > 30%, and sequence coverage > 50%. To search for horizontally transferred genes in *Pseudoalteromonas*, the parameter “self group” was set to *Pseudoalteromonas* (TaxID: 53246) and “close group” to *Alteromonadaceae* (TaxID: 226, 2848171, 135575, 28228, 1621534, 2071980, 336830, 2800384, 67575, 89404, 1249554, 111142, 2800384, 907197, 1518149, 366580, 1751872, 249523, 265980, 1407056, 2834759, 2125985, 296014, 1406885, 1172191, 137583, 2848177, 2661818, 2798470, 2851088). To search for horizontally transferred genes in *Vibrio*, the parameter “self group” was set to *Vibrio* (TaxID: 662) and “close group” was set to *Vibrionaceae* (TaxID: 641).

Statistical analysis

Statistical analysis was performed using R in RStudio [60]. Correlation analysis was performed using the cor() function with Pearsons correlation.

Supplementary Materials

Table S1: List of all genomes used in pangenome analysis and Heap’s law, Table S2: Results from calculations of Heap’s law, Table S3: Global RSCU values, Table S4: Correlation analysis of RSCU values and median effective number of codons.

Funding

This work was supported by the UiT The Arctic University of Norway. The publication charges for this article have been funded by UiT The Arctic University of Norway. The funder had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

Author Contributions

PH and CBS designed the study and wrote the manuscript. CBS performed all bioinformatics analysis. Both authors contributed to proofreading and approved on the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Figure legends:

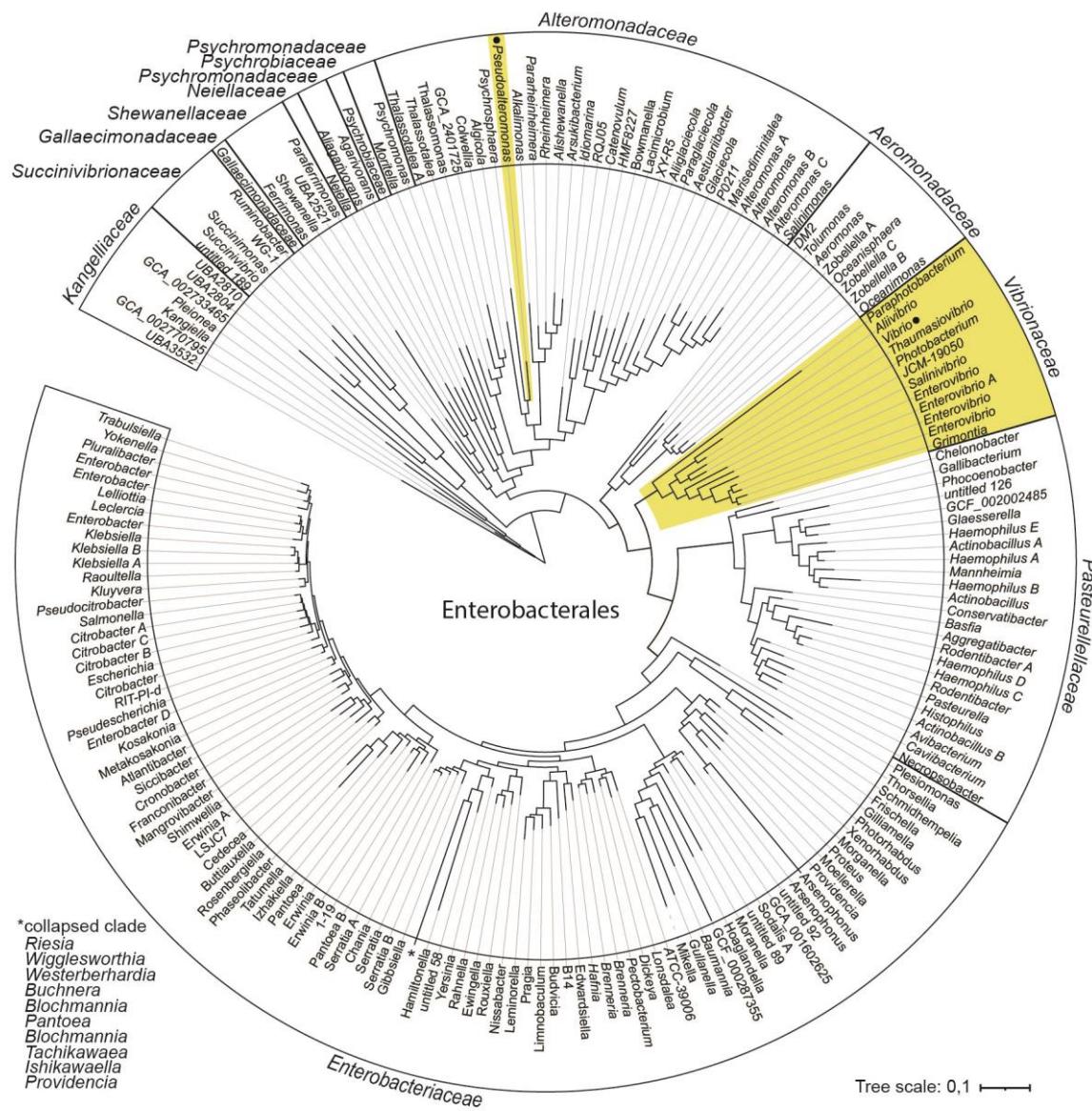


Figure 1: Phylogeny and distribution of bipartite genomes within Enterobacterales. Phylogenetic relationship between bacterial families and their respective genera are derived from the Genome Taxonomy database (GTDB). Lineages with bipartite genomes are highlighted in yellow, and genera investigated in this study are indicated with black dots.

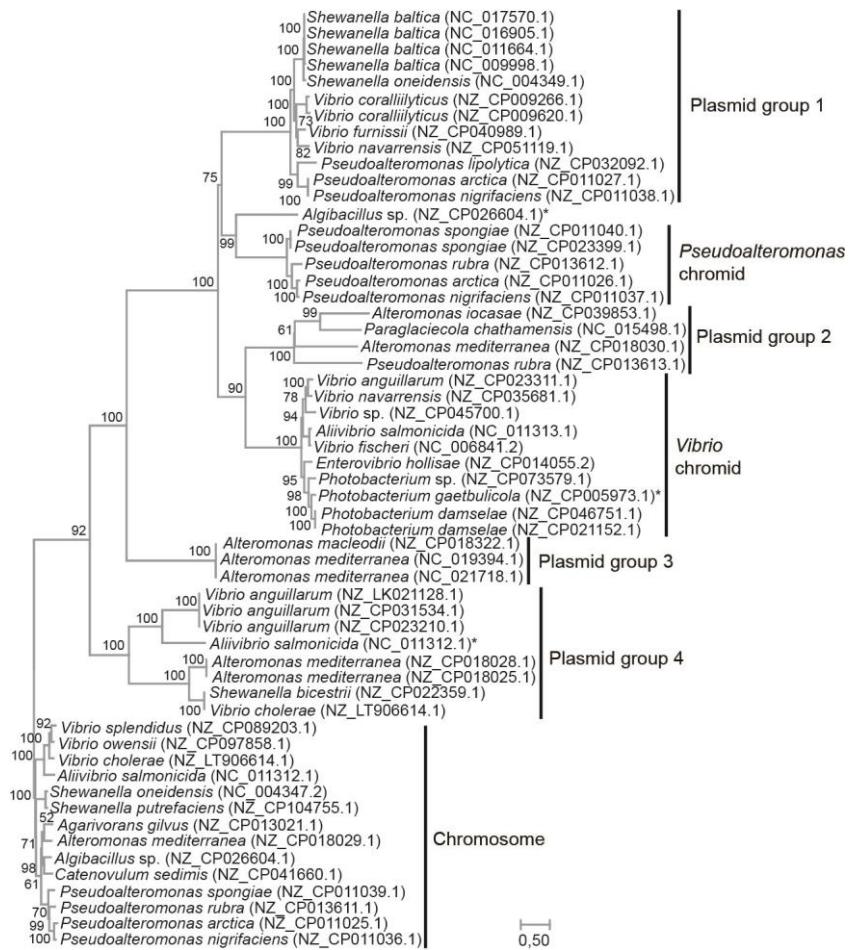


Figure 2: ML-tree based on the concatenated protein sequences of ParA and ParB and the WAG+G+I model. The tree shows the evolutionary relationships between chromosomal sequences from *Vibrio* and *Pseudoalteromonas*, and sequences from plasmids carrying related ParA and ParB pairs. Chromosomal sequences were used as the outgroup. Clades containing plasmid sequences were designated Plasmid group 1-4 for clarity. Asterix denotes chromosomal sequences with an auxiliary pair of ParA and ParB. Bootstrap values (ML method, WAG+G+I model, 1,000 pseudoreplicates) are associated with the nodes. Branch lengths are proportional to the number of substitutions per site (see scale).

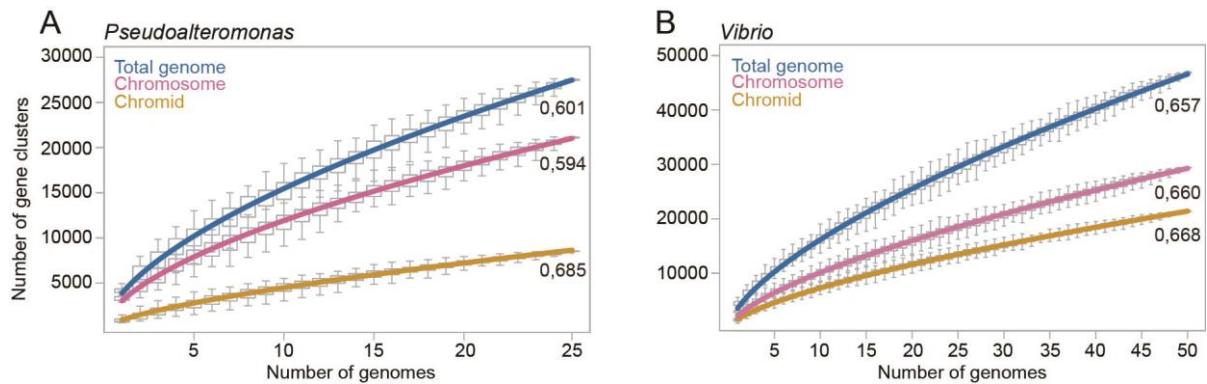


Figure 3: Graphs showing the calculated pangenome sizes of *Pseudoalteromonas* and *Vibrio* relative to the number of added genomes. For *Pseudoalteromonas* (A) and *Vibrio* (B) the number of gene clusters continues to grow as more genomes are added to the analysis, which shows that the chromids, chromosomes and total genomes are open. Each data point in the graph is based on the median of pangenome size of 100 randomly generated datasets (strain orders). The Heaps' exponents are shown associated with each graph and are used to evaluate the openness of the genomes.

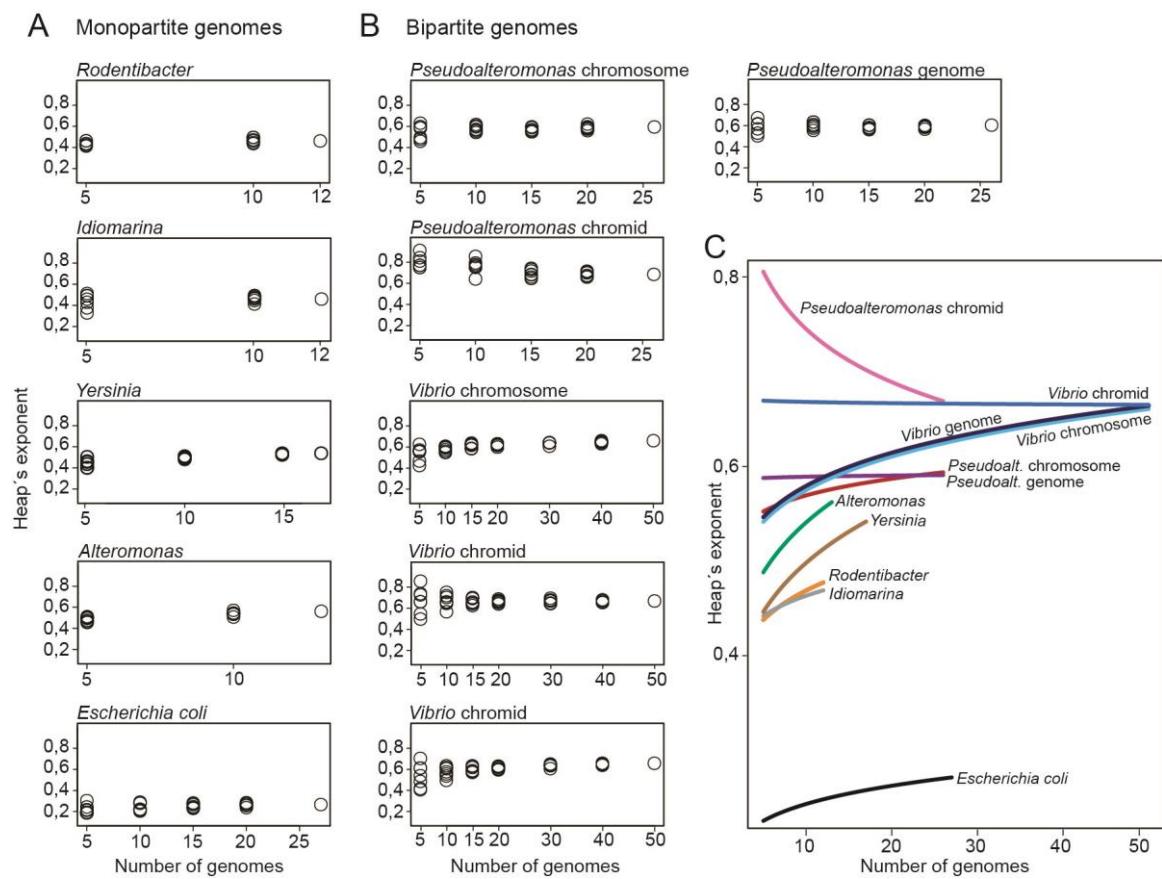


Figure 4: Plots of Heap's exponents against the number of genomes. The analysis was done for datasets with monopartite (A) or (B) bipartite genomes. Each of the Heap's exponents are made from the median number of pangenome sizes from 100 randomly generated strain orders. (C) Rarefaction curves of Heap's exponents plotted against number of genomes. The curves can be regarded as a summary and of the results from A and B through curve fitting of the Heap's exponents.

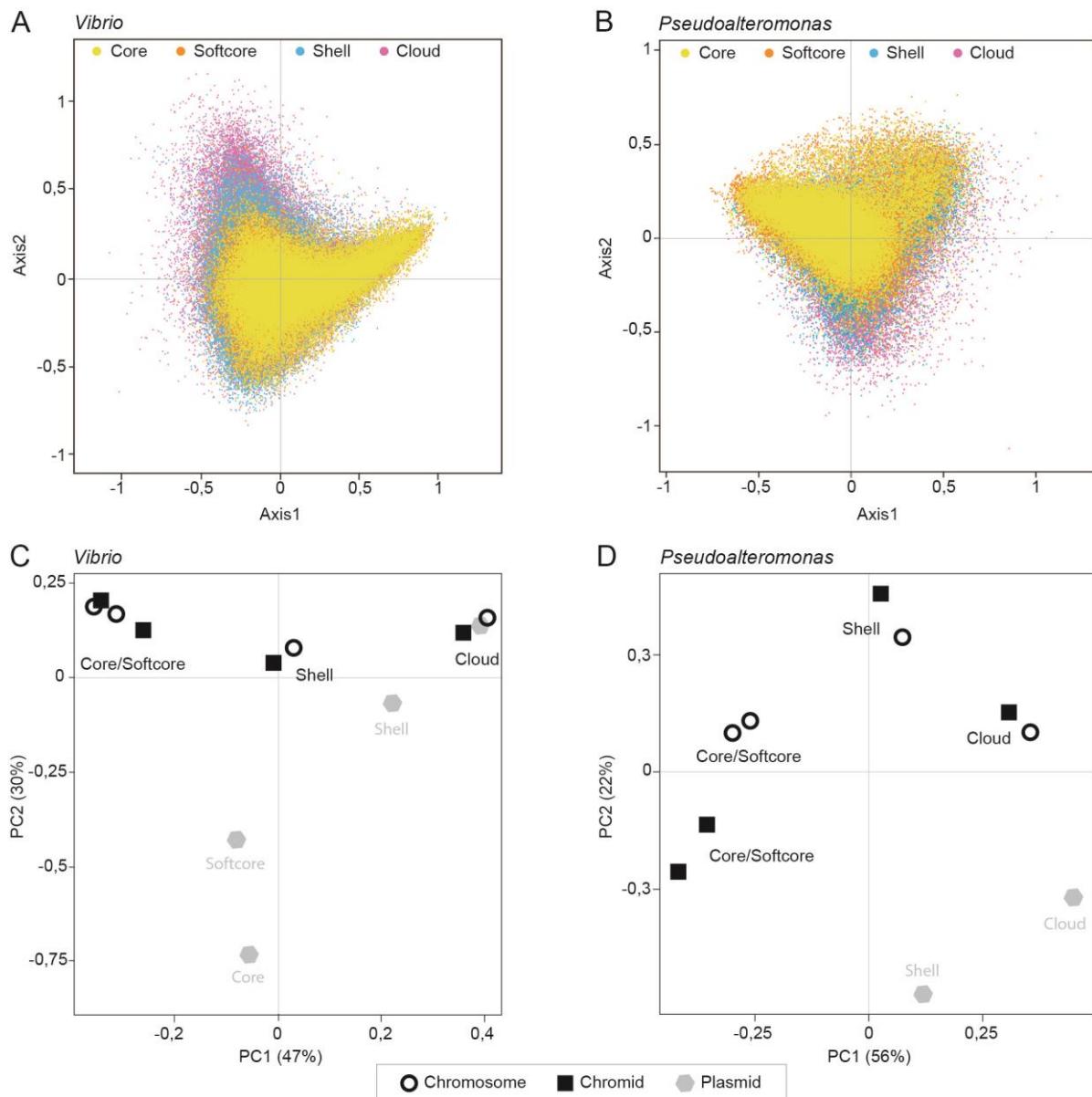


Figure 5: Correspondence analysis of relative synonymous codon usage (RSCU). The analyses are based on 50 *Vibrio* (A) and 26 *Pseudoalteromonas* (B) genomes. Core, softcore, shell and cloud genes are indicated with yellow, orange, blue and pink colors, respectively. The genes are distributed on primary and secondary axes which account for 10,98% and 8,07% in *Vibrio* and 10,97 % and 7,52% *Pseudoalteromonas* of the total variation. Principal component analysis (PCA) plots of the RSCU data from *Vibrio* (C) and *Pseudoalteromonas* (D) are shown. Both type of plots show that codon usage is specific for each pangene category rather than type of replicon.

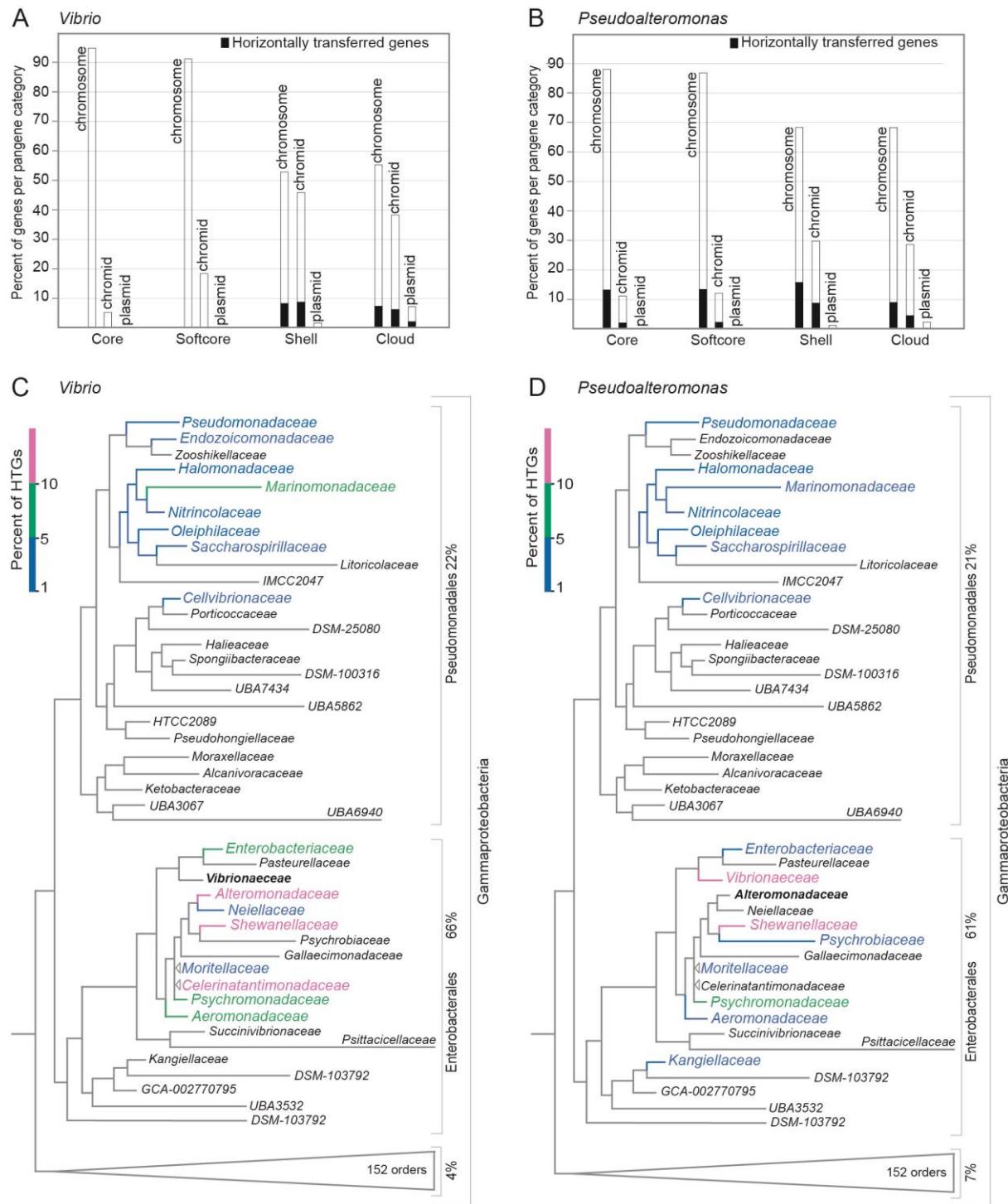


Figure 6: Horizontally transferred genes in *Vibrio* and *Pseudoalteromonas*, and the phylogenetic distribution of their donors. The number of HTGs in *Vibrio* (A) and *Pseudoalteromonas* (B) were predicted using the HGTeator software. The data is shown as percentage of HTGs in each pangene category (core, softcore, shell and cloud), and also they are distributed among the three types of replicons (chromosomes, chromids and plasmids).

HTGs were defined as genes with closest BLASTp hits outside of its family (i.e., *Vibrionaceae* and *Alteromonadaceae*, respectively). Next, the predicted bacterial donors of HTGs that reside in *Vibrio* (C) and *Pseudoalteromonas* (D) are shown mapped onto a phylogeny of Gamma-proteobacteria. The top donors are shown in colorblindness-friendly color codes, from 1-5% (blue), 5-10% (green) and 10-15% (reddish purple). The majority of HTGs originates from other families within Bacteriales, with *Shewanella* (at genus level) as the top donor to both *Vibrio* and *Pseudoalteromonas*.

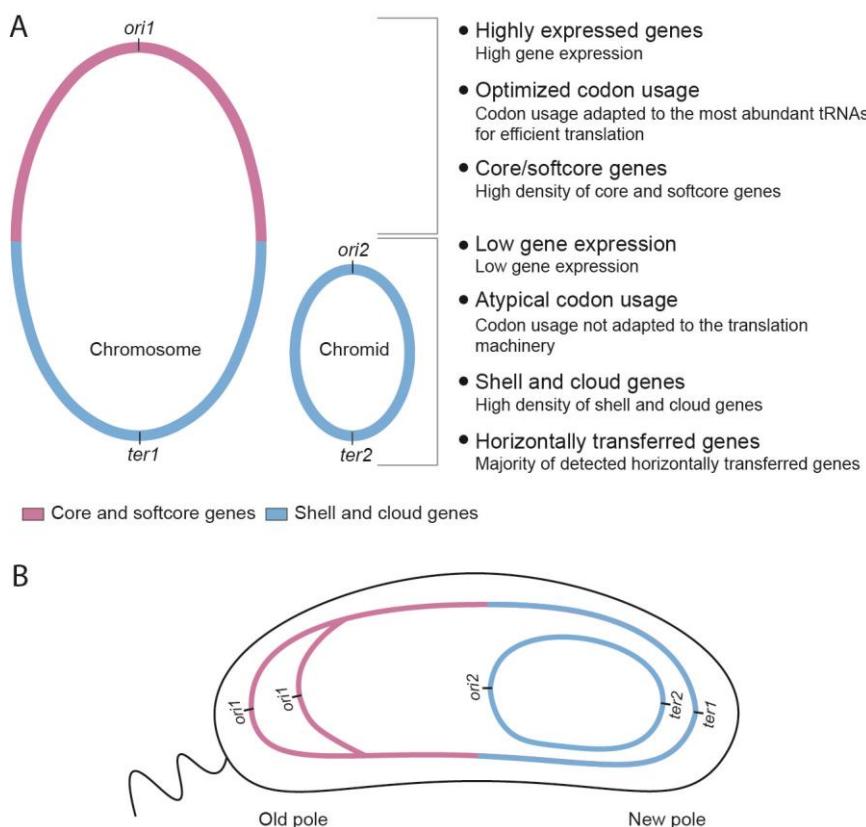


Figure 7: Summary of key characteristics of bipartite genomes in *Vibrio* and *Pseudoalteromonas*, and a putative model for accepted landing sites of HTGs. (A) Genes on the upper half of the chromosome are statistically more highly expressed, more likely to be core or softcore genes, and the codon usage is well adapted to the translational machinery. Genes located on the lower half of the chromosome, or the chromid, are statistically lower expressed, more likely to be shell or cloud genes, and have atypical codon usage less adapted to the translational machinery (compared to core/softcore). (B) Sketch of a hypothetical cell with a bipartite genome, and depicting the subcellular location of a chromosome and a chromid. The model is based on our pangenome calculations and genomic mapping of pangene types [30,34], and data from *V. cholerae* where the subcellular position of replicons have been determined [31,35–37]. Based on the genomic characteristics described in A, we hypothesize that chromids

and the lower halves of the chromosomes are favored “landing sites” for gene acquisition in bipartite genomes.

References

1. Harrison, P.W.; Lower, R.P.J.; Kim, N.K.D.; Young, J.P.W. Introducing the Bacterial “Chromid”: Not a Chromosome, Not a Plasmid. *Trends Microbiol.* **2010**, *18*, 141–148, doi:10.1016/j.tim.2009.12.010.
2. DiCenzo, G.C.; Finan, T.M. The Divided Bacterial Genome. *Microbiol. Mol. Biol. Rev.* **2017**, *81*, e00019-17, doi: 10.1128/MMBR.00019-17.
3. Misra, H.S.; Maurya, G.K.; Kota, S.; Charaka, V.K. Maintenance of Multipartite Genome System and Its Functional Significance in Bacteria. *J. Genet.* **2018**, *97*, 1013–1038, doi:10.1007/s12041-018-0969-x.
4. Almalki, F.; Choudhary, M.; Azad, R.K. Analysis of Multipartite Bacterial Genomes Using Alignment Free and Alignment-Based Pipelines. *Arch. Microbiol.* **2023**, *205*, 25, doi:10.1007/s00203-022-03354-2.
5. Oren, A.; Arahal, D.R.; Rosselló-Móra, R.; Sutcliffe, I.C.; Moore, E.R.B. Emendation of Rules 5b, 8, 15 and 22 of the International Code of Nomenclature of Prokaryotes to Include the Rank of Phylum. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 004851 R, doi:10.1099/ijsem.0.004851.
6. Egan, E.S.; Fogel, M.A.; Waldor, M.K. MicroReview: Divided Genomes: Negotiating the Cell Cycle in Prokaryotes with Multiple Chromosomes. *Mol. Microbiol.* **2005**, *56*, 1129–1138, doi:10.1111/j.1365-2958.2005.04622.x.
7. Choudhary, M., Cho, H., Bavishi, A., Trahan, C., Myagmarjav, B. Evolution of Multipartite Genomes in Prokaryotes. In *Evolutionary Biology: Mechanisms and Trends*. Pontarotti, P., Springer, Berlin, Heidelberg.; 2012: pp. 301–323.
8. Dicenzo, G.C.; Mengoni, A.; Perrin, E. Chromids Aid Genome Expansion and Functional Diversification in the Family *Burkholderiaceae*. *Mol. Biol. Evol.* **2019**, *36*, 562–574, doi:10.1093/molbev/msy248.
9. Galardini, M.; Pini, F.; Bazzicalupo, M.; Biondi, E.G.; Mengoni, A. Replicon-Dependent Bacterial Genome Evolution: The Case of *Sinorhizobium meliloti*. *Genome Biol. Evol.* **2013**, *5*, 542–558, doi:10.1093/gbe/evt027.
10. diCenzo, G.C.; MacLean, A.M.; Milunovic, B.; Golding, G.B.; Finan, T.M. Examination of Prokaryotic Multipartite Genome Evolution through Experimental

Genome Reduction. *PLoS Genet.* **2014**, *10*, e1004742, doi:10.1371/journal.pgen.1004742.

11. Cooper, V.S.; Vohr, S.H.; Wrocklage, S.C.; Hatcher, P.J. Why Genes Evolve Faster on Secondary Chromosomes in Bacteria. *PLoS Comput. Biol.* **2010**, *6*, e1000732, doi:10.1371/journal.pcbi.1000732.

12. Feng, Z.; Zhang, Z.; Liu, Y.; Gu, J.; Cheng, Y.; Hu, W.; Li, Y.; Han, W. The Second Chromosome Promotes the Adaptation of the Genus *Flammeovirga* to Complex Environments. *Microbiol. Spectr.* **2021**, *9*, e00980-21, doi:10.1128/SPECTRUM.00980-21.

13. Dryselius, R.; Izutsu, K.; Honda, T.; Iida, T. Differential Replication Dynamics for Large and Small *Vibrio* Chromosomes Affect Gene Dosage, Expression and Location. *BMC Genomics* **2008**, *9*, 559, doi:10.1186/1471-2164-9-559.

14. Couturier, E.; Rocha, E.P.C. Replication-Associated Gene Dosage Effects Shape the Genomes of Fast-Growing Bacteria but Only for Transcription and Translation Genes. *Mol. Microbiol.* **2006**, *59*, 1506–1518, doi:10.1111/j.1365-2958.2006.05046.x.

15. Slater, S.C.; Goldman, B.S.; Goodner, B.; Setubal, J.C.; Farrand, S.K.; Nester, E.W.; Burr, T.J.; Banta, L.; Dickerman, A.W.; Paulsen, I.; et al. Genome Sequences of Three *Agrobacterium* Biovars Help Elucidate the Evolution of Multichromosome Genomes in Bacteria. *J. Bacteriol.* **2009**, *191*, 2501–2511, doi:10.1128/JB.01779-08.

16. Tettelin, H.; Masianni, V.; Cieslewicz, M.J.; Donati, C.; Medini, D.; Ward, N.L.; Angiuoli, S. V.; Crabtree, J.; Jones, A.L.; Durkin, A.S.; et al. Genome Analysis of Multiple Pathogenic Isolates of *Streptococcus Agalactiae*: Implications for the Microbial “Pan-Genome.” *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13950–13955, doi:10.1073/pnas.0508532102.

17. Tettelin, H.; Riley, D.; Cattuto, C.; Medini, D. Comparative Genomics: The Bacterial Pan-Genome. *Curr. Opin. Microbiol.* **2008**, *11*, 472–477, doi:10.1016/j.mib.2008.09.006.

18. Ikemura, T. Codon Usage and tRNA Content in Unicellular and Multicellular Organisms. *Mol. Biol. Evol.* **1985**, *2*, 13–34, doi:10.1093/oxfordjournals.molbev.a040335.

19. Plotkin, J.B.; Kudla, G. Synonymous but Not the Same. *Natl. Rev. Genet.* **2011**, *12*, 32–42, doi:10.1038/nrg2899.Synonymous.

20. Tuller, T.; Girshovich, Y.; Sella, Y.; Kreimer, A.; Freilich, S.; Kupiec, M.; Gophna, U.; Ruppin, E. Association between Translation Efficiency and Horizontal Gene Transfer within Microbial Communities. *Nucleic Acids Res.* **2011**, *39*, 4743–4755, doi:10.1093/nar/gkr054.
21. Komar, A.A. The Yin and Yang of Codon Usage. *Hum. Mol. Genet.* **2016**, *25*, R77–R85, doi.org/10.1093/hmg/ddw207.
22. Tuller, T. Codon Bias, tRNA Pools, and Horizontal Gene Transfer. *Mob. Genet. Elements* **2011**, *1*, 75–77, doi:10.4161/mge.1.1.15400.
23. Oren, A.; Da Costa, M.S.; Garrity, G.M.; Rainey, F.A.; Rosselló-Móra, R.; Schink, B.; Sutcliffe, I.; Trujillo, M.E.; Whitman, W.B. Proposal to Include the Rank of Phylum in the International Code of Nomenclature of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 4284–4287, doi:10.1099/ijsem.0.000664.
24. Liao, L.; Liu, C.; Zeng, Y.; Zhao, B.; Zhang, J.; Chen, B. Multipartite Genomes and the SRNome in Response to Temperature Stress of an Arctic *Pseudoalteromonas* Fuliginea BSW20308. *Environ. Microbiol.* **2019**, *21*, 272–285, doi:10.1111/1462-2920.14455.
25. Xie, B. Bin; Rong, J.C.; Tang, B.L.; Wang, S.; Liu, G.; Qin, Q.L.; Zhang, X.Y.; Zhang, W.; She, Q.; Chen, Y.; et al. Evolutionary Trajectory of the Replication Mode of Bacterial Replicons. *MBio* **2021**, *12*, e02745-20, doi:10.1128/mBio.02745-20.
26. Fournes, F.; Val, M.E.; Skovgaard, O.; Mazel, D. Replicate Once per Cell Cycle: Replication Control of Secondary Chromosomes. *Front. Microbiol.* **2018**, *9*, 1833, doi:10.3389/fmicb.2018.01833.
27. Heidelberg, J.F.; Elsen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.; et al. DNA Sequence of Both Chromosomes of the Cholera Pathogen *Vibrio Cholerae*. *Nature* **2000**, *406*, 477–483, doi:10.1038/35020000.
28. Médigue, C.; Krin, E.; Pascal, G.; Barbe, V.; Bernsel, A.; Bertin, P.N.; Cheung, F.; Cruveiller, S.; D'Amico, S.; Duilio, A.; et al. Coping with Cold: The Genome of the Versatile Marine Antarctica Bacterium *Pseudoalteromonas Haloplanktis TAC125*. *Genome Res.* **2005**, *15*, 1325–1335, doi:10.1101/gr.4126905.
29. Rong, J.C.; Liu, M.; Li, Y.; Sun, T.Y.; Pang, X.H.; Qin, Q.L.; Chen, X.L.; Xie, B. Bin Complete Genome Sequence of a Marine Bacterium With Two Chromosomes, *Pseudoalteromonas Translucida KMM 520T*. *Mar. Genomics* **2016**, *26*, 17–20, doi:10.1016/j.margen.2016.01.007.

30. Sonnenberg, C.B.; Haugen, P. The *Pseudoalteromonas* Multipartite Genome: Distribution and Expression of Pangene Categories, and a Hypothesis for the Origin and Evolution of the Chromid. *G3 Genes/Genomes/Genetics* **2021**, *11*, doi:10.1093/G3JOURNAL/JKAB256.
31. Rasmussen, T.; Jensen, R.B.; Skovgaard, O. The Two Chromosomes of *Vibrio Cholerae* Are Initiated at Different Time Points in the Cell Cycle. *EMBO J.* **2007**, *26*, 3124–3131, doi:10.1038/sj.emboj.7601747.
32. Kemter, F.S.; Messerschmidt, S.J.; Schalopp, N.; Sobetzko, P.; Lang, E.; Bunk, B.; Spröer, C.; Teschler, J.K.; Yildiz, F.H.; Overmann, J.; et al. Synchronous Termination of Replication of the Two Chromosomes Is an Evolutionary Selected Feature in *Vibrionaceae*. *PLoS Genet.* **2018**, *14*, e1007251, doi:10.1371/journal.pgen.1007251.
33. Val, M.-E.; Marbouty, M.; de Lemos Martins, F.; Kennedy, S.P.; Kemble, H.; Bland, M.J.; Possoz, C.; Koszul, R.; Skovgaard, O.; Mazel, D. A Checkpoint Control Orchestrates the Replication of the Two Chromosomes of *Vibrio Cholerae*. *Sci. Adv.* **2016**, *2*, e1501914, doi:10.1126/sciadv.1501914.
34. Sonnenberg, C.B.; Kahlke, T.; Haugen, P. Vibrionaceae Core, Shell and Cloud Genes Are Non-Randomly Distributed on Chr 1: An Hypothesis That Links the Genomic Location of Genes with Their Intracellular Placement. *BMC Genomics* **2020**, *21*, 695, doi:10.1186/s12864-020-07117-5.
35. Srivastava, P.; Chattoraj, D.K. Selective Chromosome Amplification in *Vibrio Cholerae*. *Mol. Microbiol.* **2007**, *66*, 1016–1028, doi:10.1111/j.1365-2958.2007.05973.x.
36. David, A.; Demarre, G.; Muresan, L.; Paly, E.; Barre, F.X.; Possoz, C. The Two Cis-Acting Sites, ParS1 and OriC1, Contribute to the Longitudinal Organisation of *Vibrio Cholerae* Chromosome I. *PLoS Genet.* **2014**, *10*, e1004448, doi:10.1371/journal.pgen.1004448.
37. Fogel, M.A.; Waldor, M.K. Distinct Segregation Dynamics of the Two *Vibrio Cholerae* Chromosomes. *Mol. Microbiol.* **2005**, *55*, 125–136, doi:10.1111/j.1365-2958.2004.04379.x.
38. Jalal, A.S.; Tran, N.T.; Le, T.B. ParB Spreading on DNA Requires Cytidine Triphosphate in Vitro. *eLife* **2020**, *20*, e53515, doi:10.7554/eLife.53515.

39. Zhu, Q.; Kosoy, M.; Dittmar, K. HGTeator: An Automated Method Facilitating Genome-Wide Discovery of Putative Horizontal Gene Transfers. *BMC Genomics* **2014**, *15*, 717, doi:10.1186/1471-2164-15-717.

40. Medini, D.; Donati, C.; Tettelin, H.; Massignani, V.; Rappuoli, R. The Microbial Pan-Genome. *Curr. Opin. Genet. Dev.* **2005**, *15*, 589–594, doi:10.1016/j.gde.2005.09.006.

41. López, J.L.; Lozano, M.J.; Lagares, J.A.; Fabre, M.L.; Draghi, W.O.; Del Papa, M.F.; Pistorio, M.; Becker, A.; Wibberg, D.; Schlüter, A.; et al. Codon Usage Heterogeneity in the Multipartite Prokaryote Genome: Selection-Based Coding Bias Associated with Gene Location, Expression Level, and Ancestry. *MBio* **2019**, *10*, e00505-19, doi:10.1128/mBio.00505-19.

42. Kopejtka, K.; Lin, Y.; Jakubovičová, M.; Koblížek, M.; Tomasch, J.; Moran, N. Clustered Core- And Pan-Genome Content on *Rhodobacteraceae* Chromosomes. *Genome Biol. Evol.* **2019**, *11*, 2208–2217, doi:10.1093/gbe/evz138.

43. Oliveira, P.H.; Touchon, M.; Cury, J.; Rocha, E.P.C. The Chromosomal Organization of Horizontal Gene Transfer in Bacteria. *Nat. Commun.* **2017**, *8*, 841, doi:10.1038/s41467-017-00808-w.

44. Mendlér, K.; Chen, H.; Parks, D.H.; Lobb, B.; Hug, L.A.; Doxey, A.C. AnnoTree: Visualization and Exploration of a Functionally Annotated Microbial Tree of Life. *Nucleic Acids Res.* **2019**, *47*, 4442–4448, doi:10.1093/NAR/GKZ246.

45. López-Pérez, M.; Ramon-Marco, N.; Rodriguez-Valera, F. Networking in Microbes: Conjugative Elements and Plasmids in the Genus *Alteromonas*. *BMC Genomics* **2017**, *18*, 36, doi:10.1186/S12864-016-3461-0.

46. Adam, Y.; Brezellec, P.; Espinosa, E.; Besombes, A.; Naquin, D.; Paly, E.; Possoz, C.; Dijk, E. van; Barre, F.X.; Ferat, J.L. *Plesiomonas Shigelloides*, an Atypical Enterobacterales with a *Vibrio*-Related Secondary Chromosome. *Genome Biol. Evol.* **2022**, *14*, 1–13, doi:10.1093/gbe/evac011.

47. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.

48. Hall, T.A. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. – ScienceOpen. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.

49. Stecher, G.; Tamura, K.; Kumar, S. Molecular Evolutionary Genetics Analysis (MEGA) for MacOS. *Mol. Biol. Evol.* **2020**, *37*, 1237–1239, doi:10.1093/molbev/msz312.

50. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549, doi:10.1093/molbev/msy096.

51. O'leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; Mcveigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2015**, *44*, D733-745, doi:10.1093/nar/gkv1189.

52. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsm, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST Server: Rapid Annotations Using Subsystems Technology. *BMC Genomics* **2008**, *9*, 75, doi:10.1186/1471-2164-9-75.

53. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *Nat. Commun.* **2018**, *9*, 5114, doi:10.1038/s41467-018-07641-9.

54. Contreras-Moreira, B.; Vinuesa, P. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. *Appl. Environ. Microbiol.* **2013**, *79*, 7696–7701, doi:10.1128/AEM.02411-13.

55. Onofri, A. The Broken Bridge between Biologists and Statisticians: A Blog and R Package. 2020. Available online: <https://www.statforbiology.com>

56. Peden, J.F. Analysis of Codon Usage. Doctoral dissertation, University of Nottingham, UK, 2000.

57. Anwar, A.M.; Soudy, M. vhcu: Virus-Host Codon Usage Co-Adaptation Analysis. 2019. Available online: <https://CRAN.r-project.org/package=vhcu>

58. Wright, F. The “effective Number of Codons” Used in a Gene. *Gene* **1990**, *87*, 23–29, doi: 10.1016/0378-1119(90)90491-9.

59. Buchfink, B.; Xie, C.; Huson, D.H. Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods* **2014**, *12*, 59–60, doi:10.1038/nmeth.3176.

60. RStudio Team. RStudio: Integrated Development for R. 2021. Available online: <http://www.rstudio.com/>