Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Replicon-resolved pangenomics reveals functional specialisation and genome evolution in Vibrio coralliilyticus

Esther Rubio-Portillo

esther.portillo@ua.es

university of Alicante

Francisco Nadal-Molero

university of Alicante

Ana-Belen Martin-Cuadrado

university of Alicante

Article

Keywords: Vibrio coralliilyticus, Pangenome, Machine learning, Genome plasticity, Megaplasmids

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11507981/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Multipartite genomes define the genus Vibrio, yet replicon specialisation remains poorly understood, particularly in environmental species. We developed a machine-learning classifier to assign draft contigs to putative replicons with high accuracy, integrating complete and draft genomes in comparative analyses. Independent implementation in Vibrio cholerae yielded similar performance, supporting broader applicability. Analysis of 41 Vibrio coralliilyticus genomes revealed functional specialisation between chromosomes despite similar proportions of core and flexible gene families. Chromosome I was enriched in essential cellular processes, central metabolism, motility and regulation, whereas chromosome II was enriched in environmental sensing, signal transduction, transport and interaction-associated functions. Chromosome I also exhibited greater plasticity, with more extensive homologous recombination and higher densities of mobile genetic elements and defence systems. Two previously undescribed conjugative megaplasmid lineages (254–628 kb) displayed contrasting adaptive repertoires. One was broadly distributed across hosts and geographic regions, encoding a conserved accessory repertoire including a giant multidomain putative effector predicted to mediate host interactions. The other was restricted to Mediterranean coral isolates, with an expanded mobilome and a conserved Sofic-like anti-phage defence system. Together, these findings show that functional specialisation can coexist with similar gene conservation patterns, while chromosomes and megaplasmids make complementary contributions to adaptation and genome evolution.

INTRODUCTION

The genus *Vibrio* comprises more than 190 described species within the family *Vibrionaceae*, distributed into at least 51 clades (1). Members of this Gram-negative genus are predominantly heterotrophic bacteria inhabiting diverse aquatic environments, including estuaries, coastal waters, and marine sediments, where they occur as free-living organisms or in association with algae, invertebrates, and vertebrates, and can adopt mutualistic, commensal, or pathogenic lifestyles (2).

A distinctive, although not exclusive, feature of many *Vibrio* species is the presence of two chromosomes (Chr I and II) of unequal size (3). Chromosome II is thought to have originated from an ancestral megaplasmid because it retains a plasmid-type replication and partition system based on RctB (4), yet it also carries genes essential for growth and viability (5). This multipartite genome architecture is considered to enhance adaptation to diverse environmental conditions (3). The two chromosomes also differ in replication timing, gene dosage, and rates of evolution. Chromosome II initiates replication later than ChrI, although both chromosomes terminate replication synchronously (6–8). Consequently, genes on ChrI generally experience higher copy numbers and dosage-dependent expression during exponential growth (9), whereas genes on secondary chromosome have been proposed to evolve more rapidly (10).

Despite these well-established biological differences between chromosomes, most pangenome analyses treat *Vibrio* genomes as single entities, pooling genes from different replicons without

considering chromosome-specific evolutionary dynamics. A recent chromosome-resolved comparative analysis of 76 reference *Vibrio* genomes revealed pronounced functional asymmetry between chromosomes, with ChrI enriched in housekeeping functions and ChrII containing a larger proportion of accessory genes, including virulence-associated functions (11). These findings suggest that ChrII may play a major role in genome plasticity and ecological adaptation across the genus, potentially through the acquisition and reorganisation of genetic material mediated by horizontal gene transfer, recombination, and transposition. However, because that study included only a single representative genome per species, it did not address chromosome specialisation within species or determine whether ChrII consistently represents the principal reservoir of accessory genes and genome plasticity. Answering these questions requires complete genomes because fragmented assemblies hinder the accurate assignment of accessory genes and mobile genetic elements to individual replicons.

Niche adaptation and speciation within *Vibrio* are largely shaped by mobile genetic elements (MGEs), which can be shared between species through horizontal gene transfer (HGT) (12, 13). This genomic variability is mirrored in the pangenome structure, with many *Vibrio* species exhibiting open pangenomes (11), indicative of continuous gene acquisition and high genomic diversity across strains. However, most comparative genomic studies providing key insights into genome plasticity and the role of MGEs have largely focused on human pathogens (14–16). More recently, these analyses have expanded to environmentally and aquaculture-relevant species, including *V. anguillarum*, *V. splendidus*, *V. harveyi*, and *V. europaeus* (17–20).

Among environmental vibrios, *Vibrio coralliilyticus* is an important marine pathogen associated with disease outbreaks in reef-building corals worldwide, causing bleaching, tissue necrosis, and mortality (21–23). It is also a major cause of mortality in bivalve larvae, including Eastern and Pacific oysters, causing severe losses in hatchery production (24, 25). This broad host range highlights its ecological versatility and impact on both natural ecosystems and aquaculture. Importantly, virulence is strongly influenced by environmental conditions, particularly temperature, through the upregulation of virulence factors (26–28). Together, these characteristics suggest a genomic architecture that supports responses to environmental signals while facilitating the continuous acquisition of new functions, raising the question of how these two capacities are distributed across its replicons.

Here, we performed a replicon-resolved pangenome analysis of 41 *V. coralliilyticus* genomes to investigate how functional repertoires, genome plasticity, and MGEs are distributed across chromosomes and plasmids, and to determine how these replicons contribute to adaptation and genome evolution in this marine pathogen. Our findings reveal a multipartite genome shaped by both functional specialisation and differential genomic plasticity, challenging a simple core-flexible model of chromosome organisation and uncovering distinct contributions of chromosomes and megaplasmids to genome evolution and adaptive potential.

MATERIALS AND METHODS

Genomes used in this study

Seventeen *V. coralliilyticus* strains were newly sequenced (Supplementary Table 1). Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen), and hybrid genome assemblies were generated from Illumina short reads and Oxford Nanopore long reads using Unicycler v0.4.7 (29) with default parameters. Assembly quality was assessed with CheckM2 (30). An additional 10 complete *V. coralliilyticus* genomes were retrieved from NCBI and included in subsequent analyses (Supplementary Table 1).

Machine-learning classification of contigs from draft assemblies

Draft *V. coralliilyticus* genome assemblies were downloaded from NCBI and contigs were assigned to ChrI, ChrII or plasmids using supervised machine-learning models trained on the 27 complete genomes (17 newly sequenced and 10 retrieved from the NCBI). Replicons were randomly fragmented into sequences ranging from 200 bp to 3 Mb, and their 6-mer frequency profiles were used as classification features. The dataset was split into training and test subsets in a 67:33 ratio (random seed = 123) using a genome-wise strategy, so that all fragments from a given genome were assigned exclusively to one subset, preventing data leakage between training and test datasets. Eleven supervised classifiers implemented in scikit-learn v1.7.2 (31), including support vector machines, Naïve Bayes, k-nearest neighbours, decision trees and random forests, were evaluated. The Linear Support Vector Classifier (LinearSVC) achieved the highest accuracy and was selected for subsequent analyses. Only draft assemblies comprising less than 55 contigs were retained, yielding a final set of 14 draft genomes.

The complete dataset therefore comprised 41 genomes: 17 newly sequenced genomes and 24 genomes retrieved from NCBI (10 complete and 14 draft assemblies) (Supplementary Table 1).

To evaluate whether the k-mer-based replicon classification strategy was robust beyond *V. coralliilyticus*, an independent species-level validation was performed using 500 complete *V. cholerae* genomes with unambiguously resolved replicons. Replicons were fragmented and represented by 6-mer frequency profiles, and models were trained and evaluated *de novo* within the *V. cholerae* dataset.

Genome characterisation and replicon comparative analyses

Plasmid sequences were identified using geNomad (32), retaining predictions with a classification score > 0.9. In draft genomes, plasmids were considered only for presence/absence analyses, whereas analyses requiring plasmid size or gene content were restricted to complete genomes with plasmids assembled as single contigs. Based on previous studies in *Vibrionaceae*, plasmids were classified as plasmids (< 250 kb) or megaplasmids (≥ 250 kb) (33).

Average nucleotide identity (ANI) was calculated using FastANI (34). Pairwise ANI values were first obtained from complete genomes to assess genomic relatedness and confirm taxonomic affiliation and

were subsequently calculated separately for ChrI and ChrII to compare sequence conservation between chromosomes.

GC skew analyses were restricted to complete genomes to avoid biases associated with fragmented assemblies (Supplementary Table 1). GC skew was calculated for each replicon using a custom Python 3 script with 10 kb sliding windows and 1 kb step as (G − C)/(G + C), and cumulative GC skew was obtained as the running sum of window values. To quantify strand asymmetry, we defined two metrics. *Skew strength* was calculated as the sum of the mean positive GC skew and the absolute mean negative GC skew, providing a direction-independent measure of strand asymmetry. *Skew symmetry* was calculated as the smaller of the two mean magnitudes divided by their sum (range 0-0.5), where 0.5 indicates a perfectly balanced bidirectional GC skew profile. Replicons lacking values of either sign were assigned a symmetry value of 0. Differences among ChrI, ChrII, plasmids (< 250 kb) and megaplasmids (≥ 250 kb) were assessed using the Kruskal-Wallis test followed by pairwise Wilcoxon rank-sum tests with Benjamini-Hochberg correction.

Codon usage was analysed separately for ChrI, ChrII and plasmids from complete genomes. Relative synonymous codon usage (RSCU) values were calculated from all predicted protein-coding sequences according to Sharp and Li (35). Principal component analysis (PCA) of centred and scaled RSCU values was performed in R v4.5.0 to compare codon usage biases among replicons.

Pangenome analysis

Gene families were partitioned using PPanGGOLiN (36) into persistent, shell and cloud categories. Hereafter, persistent families are referred to as the core genome, whereas shell and cloud families constitute the flexible genome. Protein sequences from the core and flexible genomes of each chromosome were functionally annotated using BlastKOALA (37). Annotated proteins were assigned Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) identifiers and classified into KEGG functional categories. The relative abundance of each category was calculated separately for the core and flexible fractions of ChrI and ChrII.

Proteins carrying N-terminal signal peptides were identified using SignalP 6.0 (38). Predicted proteins were classified into five categories according to their secretion or processing pathway: Sec-dependent (SP), lipoprotein (LIPO), Tat-dependent (TAT), Tat-dependent lipoprotein (TATLIPO) and type IV pilin-like (PILIN) signal peptides. For each genome, the abundance of each category was normalised to the total number of predicted proteins encoded by the corresponding replicon.

Defence systems

Anti-phage defence systems were identified using PADLOC (39). For each genome, the number of systems in each category was normalised by replicon length and expressed as systems per Mb.

Virulence-associated genes

Open reading frames (ORFs) were predicted using Prodigal (40). Protein sequences were searched against the Virulence Factor Database (VFDB; 41) using BLASTp with thresholds of ≥ 80% sequence identity and ≥ 60% query coverage. For each genome, the number of hits assigned to each virulence category was normalised to the total number of predicted proteins encoded by the corresponding chromosome.

Homologous recombination analysis

To investigate the relative contributions of homologous recombination and mutation to genome diversification, recombination analyses were performed independently for ChrI and ChrII using concatenated nucleotide alignments of the core genes shared by all isolates. Phylogenetic trees were inferred from each alignment using FastTree v2.1.11 (42).

Homologous recombination was inferred using ClonalFrameML (43). For each chromosome, the corresponding phylogenetic tree and core genome alignment were analysed using default parameters to estimate the ratio of recombination to mutation rates (R/θ), the mean length of recombination fragments, and the relative effect of recombination versus mutation (r/m). These parameters were used to assess the contribution of homologous recombination to genome evolution.

Mobile genetic elements (MGEs)

Insertion sequences (ISs) were identified using ISEScan (44). Integrated prophages and plasmids were predicted using geNomad (32). Regions of genomic plasticity (RGPs) were identified using the panRGP module of PPanGGOLiN (36), and integron-associated attC recombination sites were detected using IntegronFinder (45). Transposase proteins were clustered at 90% sequence identity and 70% alignment coverage to reduce redundancy.

To compare the distribution of MGEs between chromosomes, the abundance of each category was normalised by chromosome length and expressed as elements per Mb.

Plasmid analysis

Plasmids were classified into distinct groups based on sequence similarity and genomic architecture. Sequence similarity was evaluated using BLASTn (≥ 90% identity and ≥ 60% query coverage). Plasmids were annotated using PlasAnn (46), and conjugative systems were identified using CONJScan v2.0.1 (47). Structural comparisons were performed with GBDraw (48). These analyses were restricted to plasmids assembled as single contigs in complete genomes to avoid biases associated with fragmented assemblies.

To assess the phylogenetic distribution of plasmid lineages, a genome phylogeny was inferred using the Type (Strain) Genome Server (TYGS) (49) following the standard Genome BLAST Distance Phylogeny (GBDP) workflow. The analysis included the 27 complete *V. coralliilyticus* genomes together with *Vibrio neptunius* KCTC 12702 and was rooted with *Vibrio vulnificus* NBRC 15645.

An unusually large hypothetical protein encoded by one plasmid was further characterised using HHpred (50) to identify structural homologues and conserved domains. The central region of the protein was modelled with AlphaFold 3 (51).

Relative transcriptional activity of chromosome I and chromosome II

RNA-seq datasets generated in a previous study on the effects of interspecies interactions on the virulence of *Vibrio* coral pathogens (28) were reanalysed to evaluate chromosome-specific transcriptional activity. Transcriptomes from pure and mixed cultures grown at 20°C and 28°C were analysed as described previously (28). Sequencing reads from each treatment were mapped to the reference genome of *V. coralliilyticus* strain Vic-OC-068 using Bowtie (52), the strain analysed in the original study. The proportion of covered positions (depth > 0) was calculated for each chromosome as a proxy for the fraction of the replicon covered by RNA-seq reads. Fold changes in chromosome coverage (coculture/monoculture) were then calculated separately for each temperature.

Statistical analysis

Statistical analyses were performed in R v4.5.0. Unless otherwise indicated, paired comparisons between ChrI and ChrII were performed using two-sided Wilcoxon signed-rank tests. P values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure.

Data availability

Genome accession numbers for all publicly available *Vibrio coralliilyticus* genomes analysed in this study are listed in Supplementary Table 1. The newly generated genome sequences have been deposited in NCBI under BioProject accession number PRJNA1523816. The individual BioSample accession numbers for the 17 newly sequenced strains are publicly accessible and are provided in Supplementary Table 1.

RESULTS

Classification of draft genome contigs into chromosomal and plasmid replicons

The eleven classifiers achieved accuracies ranging from 80.5% to 96% on the held-out test set. Across all models, ChrI contigs were classified almost perfectly (F1: 0.99), indicating a highly distinctive k-mer signature. The main classification challenge was distinguishing ChrII from plasmids (F1: 0.94), for which performance was lower across most classifiers (Supplementary Table 3).

The Linear Support Vector Classifier (LinearSVC) outperformed all other classifiers, achieving 96% overall accuracy together with the highest precision, recall, and F1-score across the three classes. It was followed by the k-nearest neighbour classifiers (k = 1, 3 and 5), with accuracies of 95.9%, 95.6% and 95.3%, respectively. In contrast, the radial basis function support vector machine achieved 91% accuracy, whereas the AdaBoost decision-tree classifier yielded the lowest performance (80.5%; Supplementary Table 3). LinearSVC was therefore selected for the classification of draft genome contigs in subsequent analyses.

Although contigs from all 26 draft genomes were classified into ChrI, ChrII and plasmid replicons, only assemblies containing ≤ 55 contigs were retained for downstream analyses. This filtering step yielded 14 draft genomes and minimised potential biases associated with assembly fragmentation.

To determine whether this classification approach was applicable beyond *V. coralliilyticus*, the workflow was independently applied to *V. cholerae*. Classifiers accuracies ranged from 85.3% to 96.7%, with Random Forest (100 trees) showing the highest performance (accuracy = 0.967; F1-score = 0.97), including accurate classification of plasmid contigs (F1-score = 0.97) despite their under-representation in the test set. The LinearSVC, the best-performing classifier in *V. coralliilyticus*, also performed well in *V. cholerae* (accuracy = 0.953), with F1-scores of 0.95, 0.95 and 0.94 for ChrI, ChrII and plasmids, respectively (Supplementary Table 4). These results support the applicability of the k-mer-based replicon classification strategy beyond *V. coralliilyticus*.

Distinct genomic organisation and compositional features characterise the replicons of Vibrio coralliilyticus

A total of 41 *V. coralliilyticus* genomes were analysed (Supplementary Table 1). Genome sizes ranged from 5,295,939 to 6,542,744 bp, with GC contents between 45.3% and 46.0%. ChrI ranged from 2,498,117 to 3,873,393 bp, whereas ChrII ranged from 1,809,248 to 2,132,499 bp. Because plasmid sequences in draft assemblies may be fragmented across multiple contigs, analyses of plasmid size and structural organisation were restricted to the 27 complete genomes. Plasmids were detected in 24 of these genomes (88.8%) and ranged from 52,449 to 628,063 bp, with megaplasmids (≥ 250 kb) present in 79% of plasmid-bearing genomes (Supplementary Table 2).

More than 80% of pairwise genome comparisons fell within a narrow ANI interval (96.8–97.3%), whereas the shared genome fraction ranged from 74.4% to 100%, revealing accessory-genome variation among closely related strains (Supplementary Fig. 1). Both chromosomes displayed comparable ANI values despite differences in shared genome fraction, which was more variable in ChrI than in ChrII. In contrast, plasmids were markedly more heterogeneous than either chromosome in both ANI and shared genome fraction (Supplementary Fig. 1).

Replication-associated compositional asymmetry differed significantly among replicons, as reflected by GC skew strength and symmetry (Kruskal-Wallis test, $p < 0.001$). Both chromosomes displayed highly symmetric GC skew profiles, although ChrI exhibited significantly stronger and more symmetric signals

than ChrII. In contrast, plasmids showed a clear size-dependent pattern. Small plasmids (< 250 kb) displayed low and highly variable symmetry values, whereas megaplasmids exhibited more symmetric, chromosome-like GC skew profiles, although both remained below the symmetry observed in ChrI and ChrII (Fig. 1).

The distribution of ribosomal and transfer RNA genes was strongly biased toward ChrI. All eleven rRNA operons were located on this replicon and showed a highly conserved arrangement relative to *dnaA* across all genomes, clustering near the origin of replication. Likewise, ChrI contained at least one copy of every tRNA gene, whereas ChrII encoded only a reduced subset (5–6 tRNAs) corresponding to a limited number of amino acids (Ser, Gly, Leu, Cys and occasionally Arg), all of which were also present on ChrI (Supplementary Table 2). Consistent with this organisation, all single-copy marker genes used by CheckM2 were exclusively located on ChrI.

Consistent with the biased distribution of tRNA genes, principal component analysis of relative synonymous codon usage (RSCU) revealed subtle but consistent differences in codon preferences between replicons. The reduced tRNA repertoire of ChrII matched its preferred codons, suggesting a non-random distribution that may reflect adaptation to optimise translational efficiency. In contrast, plasmid-encoded genes exhibited substantially greater heterogeneity in codon usage, with no clear clustering by plasmid size (Supplementary Fig. 2).

Pangenome analyses reveal functional partitioning between chromosomes

The *V. coralliilyticus* pangenome comprised 12,881 gene families and exhibited an open structure (Heaps' law, $\alpha = 0.276$, $R^2 = 0.996$, $p < 0.001$). Of these, 4,244 belonged to the core genome and 8,637 to the flexible genome, including 1,291 shell and 7,346 cloud gene families. The ChrI pangenome comprised 128,319 genes grouped into 7,393 families, whereas ChrII contained 71,495 genes distributed across 3,860 families (Fig. 2A, B). Similar proportions of core and flexible gene families were observed in both chromosomes, with only 545 families (4.23% of the pangenome) detected on both replicons. Of these, only 14 were recurrently present in more than 20 genomes and were predominantly associated with central metabolism (Supplementary Table 5).

The flexible genome of both chromosomes was dominated by unclassified proteins (> 70%), whereas this proportion decreased markedly in the core genome (< 3%). ChrI was enriched in genes involved in genetic information processing (12.49% vs. 5.30% in ChrII), consistent with its role in housekeeping functions. Cell motility and glycan biosynthesis were detected exclusively in ChrI, whereas ChrII contained larger proportions of genes involved in cellular community-prokaryotes, signal transduction, and membrane transport (Fig. 2C, D).

These functional differences extended to cell-environment interactions. ChrII encoded a significantly higher proportion of proteins carrying predicted signal peptides than ChrI (16.88% vs. 13.61%, paired Wilcoxon test, $p < 0.0001$), primarily due to enrichment of Sec-dependent, lipoprotein, and Tat-dependent

signal peptides, whereas pilin-like proteins were more abundant in ChrI (Supplementary Fig. 3). Functional specialisation also extended to ABC transport systems and nitrogen metabolism. ChrI predominantly encoded transporters for essential nutrients, together with assimilatory nitrate reduction pathways. In contrast, ChrII contained more transporters for more variable substrates, including ribose, maltose, urea and nitrate/nitrite, and encoded dissimilatory nitrate/nitrite reduction pathways associated with facultative anaerobic respiration.

This partitioning was particularly evident in quorum-sensing and c-di-GMP signalling pathways. The ChrI core genome encoded the central signal integration and regulatory machinery, while ChrII encoded the major signal detection systems, including the AI-2 receptor complex LuxP/LuxQ, the CAI-1 synthase-receptor pair CqsA/CqsS, and the high-cell density regulator LuxR (Supplementary Fig. 4A). Notably, the AI-1 (AHL) synthase-receptor pair LuxM/LuxN was only detected in 25 strains and always as part of the flexible genome, occurring on either ChrI or ChrII depending on the strain. A similar partitioning was observed for c-di-GMP-associated pathways, with ChrII harbouring multiple sensory and regulatory proteins, whereas ChrI encoded additional downstream components together with elements of the MSHA pilus system (Supplementary Fig. 4B).

Similar chromosome-specific trends were observed for virulence-associated genes. ChrI was significantly enriched in genes related to motility, biofilm formation, adherence, effector delivery systems, and regulatory functions, whereas ChrII contained higher proportions of genes associated with stress survival, antimicrobial activity, and exoenzyme production (Fig. 3). These patterns indicate that ChrI preferentially encodes functions involved in host colonisation and virulence, whereas ChrII contributes more extensively to environmental adaptation and extracellular interactions.

To explore whether the distinct functional profiles of the two chromosomes were reflected at the transcriptional level, we reanalysed previously published RNA-seq datasets obtained from *V. coralliilyticus* grown in monoculture or in coculture with *Vibrio mediterranei* at 20°C and 28°C (28). We compared transcriptomic coverage between monoculture and coculture separately at each temperature to distinguish the effect of interspecific interaction from that of temperature. At both temperatures, coculture substantially increased the fraction of each chromosome covered by RNA-seq reads relative to the corresponding monoculture, but this increase was consistently greater for ChrII. At 20°C, the proportion of covered positions increased approximately 20-fold in ChrI and 123-fold in ChrII during coculture, whereas at 28°C the corresponding increases were approximately 9-fold and 34-fold, respectively (Supplementary Fig. 5).

Chromosome I concentrates defence-system diversity

Restriction-modification (RM) and dXTPase were the most widespread defence systems, whereas the remaining mechanisms showed a more heterogeneous distribution across isolates. Overall, defence systems were significantly enriched in ChrI relative to ChrII (paired Wilcoxon test, adjusted $p = 0.0065$). ChrI contained higher densities of RM, Mokosh type II, dXTPase, and abortive infection systems, whereas

ChrII harboured a more limited repertoire, with DISARM/DMS being the only defence system enriched on this replicon (Supplementary Fig. 6).

Hierarchical clustering based on the presence and absence of defence-system types revealed two major isolated groups. These groups showed no obvious association with geographic origin or host but largely coincided with the two megaplasmid groups described below. Cluster separation was primarily driven by differences in accessory defence-system repertoires, with the Sofic system largely defining one of the two clusters (Fig. 4).

Chromosome I concentrates genome plasticity and horizontal genetic exchange

Recombination contributed substantially to sequence evolution in both chromosomes, with estimated r/m values of 3.25 for ChrI and 2.57 for ChrII, indicating that nucleotide substitutions introduced through homologous recombination exceeded those generated by mutation alone. Moreover, inferred recombinant intervals were significantly longer in ChrI than in ChrII (median length, 382 vs. 298 bp; Wilcoxon rank-sum test, $P < 0.001$; Supplementary Fig. 7). The mean divergence between imported and recipient sequences was similar in both chromosomes (4.5%), which exceeded the average divergence among the genomes analysed here (2.7–3.2%), consistent with recombination involving DNA from donors more divergent than those represented in the sampled population.

To assess whether this trend extended to other signatures of genome plasticity, we compared the density of MGEs and diversification hotspots between chromosomes. ChrI exhibited significantly higher densities of insertion sequences (ISs), transposases, regions of genomic plasticity (RGPs), and integron-associated *attC* sites than ChrII (paired Wilcoxon tests, adjusted $p < 0.001$; Fig. 5). This enrichment was most pronounced for attC sites, which were almost exclusively detected in ChrI.

Two megaplasmid lineages differ in ecological distribution, conjugation systems, and accessory functions

Comparative analyses identified two major megaplasmid lineages, hereafter referred to as Type1 and Type2 plasmids. Type1 plasmids were detected in eleven isolates from all sampled regions and from diverse sources, including seawater, corals, oysters, and other hosts (Fig. 6; Supplementary Table 2). In contrast, Type2 plasmids were detected exclusively in ten Mediterranean coral isolates. Despite these contrasting distributions, both plasmid lineages encoded conjugation-associated Type IV secretion systems (T4SS), including relaxases, type IV coupling proteins (T4CPs), and multiple *tra/trb* components, supporting their capacity for autonomous conjugative transfer, although no canonical oriT was identified. The conjugation modules belonged to distinct T4SS classes, with Type1 plasmids encoding an F-type system and Type2 plasmids a T-type system, indicating distinct conjugation machinery.

Only about 25% of Type1 plasmid proteins could be assigned to KEGG functional categories, mainly related to genetic and environmental information processing, and no complete metabolic pathways were identified. Despite the limited functional annotation, several proteins suggested roles in host interaction, interbacterial competition, signalling, and stress adaptation, including hemolysin-related proteins, Rhs-family proteins, the RstAB two-component system, and multiple oxidative stress proteins. A particularly notable feature was a highly conserved multidomain protein of approximately 5,000 amino acids (Fig. 6). HHpred identified N-terminal sensor-like domains, a central Lymphostatin/LifA-like region, and C-terminal RTX-associated domains. Adjacent RTX secretion-associated ATPases were encoded within the same plasmid region, suggesting a secreted host-interaction effector.

In contrast, Type2 plasmids displayed a markedly different functional profile. Approximately 32% of their proteins were annotated, including a complete homoprotocatechuate degradation pathway, a nearly complete ascorbate degradation module, and multiple ABC transporters involved in the uptake of osmoprotectants, complex carbohydrates, antimicrobial peptides, and signalling-related molecules. These plasmids also carried an exceptionally high density of MGEs, including numerous TnpA-family transposases and IS-associated elements, with some plasmids encoding more than 20 copies. This was accompanied by a highly conserved transposase repertoire across Type2 plasmids, indicating the persistence of a shared mobilome within this lineage. Consistent with this pattern, isolates carrying Type2 plasmids also exhibited higher densities of insertion sequences and transposases on ChrI than isolates carrying Type1 plasmids or lacking plasmids. Remarkably, approximately one-third of all chromosomal transposases belonged to families also encoded by the resident Type2 plasmid (Supplementary Fig. 8), indicating extensive overlap between plasmid and chromosomal mobilomes. In addition, all Type2 plasmids encoded a Sofic defence system, and every isolate carrying a Type2 plasmid also possessed a chromosomal Sofic homolog, whereas this system was absent from all strains lacking Type2 plasmids (Fig. 4). Because Type2 plasmids were detected in isolates from multiple phylogenetic lineages (Supplementary Fig. 9), this association is unlikely to be explained solely by shared ancestry.

DISCUSSION

While multipartite genome organisation is a hallmark of many members of the family *Vibrionaceae*, the extent to which individual replicons are functionally specialised remains poorly understood in most non-human pathogenic species. Replicon-resolved comparative analyses are limited by the relatively low availability of complete genomes for many environmental species and by the predominance of draft assemblies, in which the chromosomal or plasmid origin of individual contigs is not directly preserved. To overcome this limitation, we developed a k-mer-based machine-learning approach that assigned *V. coralliilyticus* draft contigs to ChrI, ChrII or plasmids with high accuracy, allowing complete and draft genomes to be integrated within the same replicon-resolved framework. Independent implementation of the workflow in *V. cholerae* achieved similarly high classification performance, indicating that the ability of k-mer composition to discriminate replicons is not restricted to *V. coralliilyticus* and supporting its potential applicability in other members of the genus.

Preserving replicon identity in pangenome analysis was biologically informative, revealing a pronounced functional compartmentalisation of the multipartite genome of *V. coralliilyticus*, in which chromosomes and megaplasmids contribute differently to cellular function, adaptation, and genome evolution. Consistent with the genomic architecture originally described for *Vibrio cholerae* (5), ChrI concentrates the core genomic features of the species, including the strongest and most symmetric GC-skew patterns, all rRNA operons, and a complete set of tRNA genes. Although a small subset of tRNAs was also detected on ChrII, these corresponded to the most frequently used codons. Notably, rRNA operons were consistently located near the origin of replication, an arrangement commonly observed in fast-growing bacteria as an optimisation strategy to satisfy the high demand for translational machinery during fast growth (53, 54).

A widely accepted model for multipartite genomes in *Vibrionaceae* proposes a strong asymmetry between chromosomes, with most core genes residing on ChrI, whereas ChrII is enriched in species- and strain-specific genes and represents the main source of adaptive innovation (55, 56). However, our results only partially support this model. Although the two chromosomes differed markedly in their functional repertoires, they contained remarkably similar proportions of core and flexible gene families, indicating that chromosome specialisation in *V. coralliilyticus* is determined primarily by functional partitioning rather than differences in gene conservation. Accordingly, ChrI is specialised in genetic information processing, central metabolism, and cell motility, whereas ChrII encodes a conserved repertoire of genes involved in environmental sensing, signal transduction, membrane transport, and other adaptive processes. Consistent with this functional specialisation, the RNA-seq coverage analyses suggested that ChrII responded more strongly during interspecific interactions, consistent with its role in sensing and responding to the biotic environment.

Our results suggest that the functional partitioning between chromosomes extends beyond gene content to encompass the spatial organisation of major regulatory networks. In *V. coralliilyticus*, environmental signal detection and signal integration appear to be distributed across different replicons, a genomic organisation that may enhance regulatory flexibility by facilitating coordinated responses through interactions between chromosomes, as previously proposed for *V. cholerae* (5). This organisation was consistently observed for both quorum sensing and c-di-GMP signalling pathways. Although the modular architecture of c-di-GMP signalling is well established in vibrios and other bacteria (57), to our knowledge, the chromosomal segregation of c-di-GMP signalling components has not been previously reported and may represent an additional level of regulatory organisation in vibrios.

Other comparative studies of multipartite bacterial genomes have suggested that secondary chromosomes experience higher rates of horizontal gene acquisition and incorporation of adaptive functions (58). However, our results challenge this paradigm. Several indicators of genome plasticity, including homologous recombination, insertion sequences, transposases, integron-associated *attC* sites, and regions of genomic plasticity, were consistently enriched in ChrI. The near-exclusive localisation of *attC* sites on ChrI contrasts with the canonical genome organisation of *Vibrionaceae*, in which the chromosomal superintegron is typically located on ChrII (59). Our findings indicate that this association

is not universal within the family. Together, these findings indicate that, despite concentrating the core cellular functions of the species, ChrI also exhibits the strongest signatures of genome plasticity, including increased recombination and the accumulation of mobile and horizontally acquired genetic elements. A similar chromosomal bias has been reported in *Vibrio splendidus*, where genomic islands are more abundant on ChrI than on ChrII (18), suggesting that the preferential accumulation of mobile and recombinant DNA within the primary chromosome may occur more broadly among environmental vibrios than previously appreciated.

An additional major finding of this study is the identification of two major megaplasmid lineages not previously described in *V. coralliilyticus*. To our knowledge, neither lineage showed close similarity to megaplasmids previously reported in other vibrios. The apparent absence of closely related plasmids in public databases may not necessarily reflect their true rarity, but rather the historical challenges associated with reconstructing complete megaplasmids from short-read sequencing data, which frequently yield fragmented plasmid assemblies due to repetitive sequences (60). The increasing availability of complete genomes generated using long-read sequencing technologies is therefore likely to reveal a much larger and previously hidden diversity of megaplasmids within *Vibrionaceae* and other marine bacterial groups.

The broad geographic and ecological distribution of Type1 megaplasmids is consistent with the long-term persistence of this lineage across diverse environmental and host-associated contexts. Consistent with this, Type1 plasmids encode an F-type conjugation system, typically associated with large conjugative plasmids carrying extensive accessory gene repertoires and exhibiting long-term persistence within bacterial populations (61). One hallmark of the Type1 lineage is a highly conserved giant multidomain protein containing a lymphostatin/LifA-like region. LifA/lymphostatin proteins are among the largest bacterial effectors known and have been implicated in host colonisation and modulation of host cellular responses (62, 63). Recent structural studies revealed a multidomain architecture specialised for interactions with eukaryotic cells (64), while the adjacent RTX secretion-associated ATPases suggest that this protein is secreted as a multifunctional effector. Although its biological role remains unknown, its conservation across geographically distant isolates suggests that it confers a strong selective advantage to the Type1 lineage.

In contrast to the broadly distributed Type1 lineage, Type2 megaplasmids were detected only among the sampled Mediterranean coral isolates and exhibited a markedly different functional profile. In addition to encoding an expanded repertoire of transposases and insertion sequence-associated elements, all Type2 plasmids carried a Sofic defence system. Remarkably, a chromosomal Sofic homolog was also present in every isolate harbouring a Type2 plasmid, whereas this system was absent from strains lacking this plasmid lineage. Sofic was recently characterised as an anti-phage defence system that triggers abortive infection through a toxin-antitoxin-like AMPylation mechanism (65). Beyond phage defence, toxin-antitoxin systems are also known to stabilise MGEs, including plasmids (66, 67). Although the biological significance of the strict co-occurrence of plasmid- and chromosome-encoded Sofic

systems remains unknown, this association raises the possibility that Sofic contributes to both phage defence and the persistence of Type2 plasmids.

Conclusions

Our comparative genomic analysis shows that the multipartite genome of *V. coralliilyticus* is organised into functionally specialised yet complementary genomic compartments. Unlike the prevailing model for *Vibrionaceae*, chromosome specialisation reflects functional partitioning rather than differences in core and flexible gene content, with ChrI combining essential cellular functions and genome plasticity, whereas ChrII is specialised in environmental sensing and adaptive responses. Genome evolution therefore operates at multiple hierarchical levels, with chromosome I acting as the principal chromosomal reservoir of genome plasticity and conjugative megaplasmids providing additional adaptive potential.

Declarations

Conflicts of interest

The authors declare that there are no conflicts of interest.

Funding information

This research was supported by the projects METACIRCLE PID2021-126114NB-C41 and CONNECTIVITY PID2024-158829NB-C41 (Spanish Ministry of Science and Innovation).

Author contributions

FNM: Methodology, Software, Validation, Formal analysis, Writing-review and editing. ABMC: Methodology, Writing-original draft, Writing-review and editing. ERP: Conceptualization, Investigation, Formal analysis, Data curation, Visualization, Writing-original draft, Writing-review and editing, Supervision.

Acknowledgements.

We thank Tobias Busche and Jörn Kalinowski with the NGS team of the Bielefeld University Core Facility Omics NGS Unit and CeBiTec as well as technical staff of the CeBiTec Technology Platform Genomics, particularly Eva Schulte-Bernd, Yvonne Kutter, and Katharina Hanuschka for sequencing. We also thank Pepa Antón for her critical reading of the manuscript and valuable comments.

References

1. Jiang, C., Tanaka, M., Nishikawa, S., Mino, S., Romalde, J.L., Thompson, F.L., Gomez-Gil, B., Sawabe, T.: *Vibrio* clade 3.0: new Vibrionaceae evolutionary units using a genome-based approach. Curr Microbiol 79:10. (2022)
2. Thompson, F.L., Iida, T., Swings, J.: Biodiversity of vibrios. Microbiol. Mol. Biol. Rev. **68**, 403–431 (2004)
3. Okada, K., Iida, T., Kita-Tsukamoto, K., Honda, T.: Vibrios commonly possess two chromosomes. J. Bacteriol. **187**, 752–757 (2005)
4. Egan, E.S., Waldor, M.K.: Distinct replication requirements for the two *Vibrio cholerae* chromosomes. Cell. **114**, 521–530 (2003)
5. Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., et al.: DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature. **406**, 477–483 (2000)
6. 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. **26**, 3124–3131 (2007)
7. Val, M.E., Marbouty, M., de Lemos Martins, F., Kennedy, S.P., Kemble, H., Bland, M.J., et al.: A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. Sci. Adv. **2**, e1501914 (2016)
8. Kemter, F.S., Messerschmidt, S.J., Schallopp, N., Sobetzko, P., Lang, E., Bunk, B., et al.: Synchronous termination of replication of the two chromosomes is an evolutionarily selected feature in *Vibrionaceae*. PLoS Genet. **14**, e1007251 (2018)
9. Dryselius, R., Izutsu, K., Honda, T., Iida, T.: Differential replication dynamics for large and small *Vibrio* chromosomes affect gene dosage, expression and location. BMC Genom. **9**, 559 (2008)
10. Cooper, V.S., Vohr, S.H., Wrocklage, S.C., Hatcher, P.J.: Why genes evolve faster on secondary chromosomes in bacteria. PLoS Comput. Biol. **6**, e1000732 (2010)
11. Volodarskiy, M., Rayko, M.: The evolution of the *Vibrio* genome: the discovery of the pan-genome and the functional specialization of chromosomes. Mol. Genet. Genomics. **301**, 52 (2026)
12. Hazen, T.H., Pan, L., Gu, J.D., Sobecky, P.A.: The contribution of mobile genetic elements to the evolution and ecology of vibrios. FEMS Microbiol. Ecol. **74**, 485–499 (2010)
13. Le Roux, F., Blokesch, M.: Eco-evolutionary dynamics linked to horizontal gene transferin vibrios. Annu. Rev. Microbiol. **72**, 89–110 (2018)
14. Boucher, Y., Cordero, O.X., Takemura, A.F., Hunt, D.E., Schliep, K., Bapteste, E., et al.: Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. mBio. **2**, e00335–e00310 (2011)
15. Dutilh, B.E., Thompson, C.C., Vicente, A.C.P., Marin, M.A., Lee, C., Silva, G.G.Z., et al.: Comparative genomics of 274 *Vibrio cholerae* genomes reveals mobile functions structuring three niche dimensions. BMC Genom. **15**, 654 (2014)

16. López-Pérez, M., Jayakumar, J.M., Haro-Moreno, J.M., Zaragoza-Solas, A., Reddi, G., Rodriguez-Valera, F., et al.: Evolutionary model of cluster divergence of the emergent marine pathogen *Vibrio vulnificus*: from genotype to ecotype. mBio. **10**, e02852–e02818 (2019)
17. Hansen, M.J., Kudirkiene, E., Dalsgaard, I.: Analysis of 44 *Vibrio anguillarum* genomes reveals high genetic diversity. PeerJ. **8**, e10451 (2020)
18. Jiang, C., Kasai, H., Mino, S., Romalde, J.L., Sawabe, T.: The pan-genome of Splendidus clade species in the family *Vibrionaceae*: insights into evolution, adaptation, and pathogenicity. Environ. Microbiol. **24**, 4587–4606 (2022)
19. Triga, A., Issa, Z.A., Smyrli, M., Fenske, L., Katharios, P.: Virulence and pangenome analysis of *Vibrio harveyi* strains from Greek and Red Sea marine aquaculture. Aquaculture. **587**, 740839 (2024)
20. Rodriguez, S., Rey-Varela, D., Blanco-Hortas, A., Martinez, C., Martínez, P., Travers, M.A., Barja, J.L., Dubert, J.: Deciphering the pangenome of the shellfish pathogen *Vibrio europaeus*: evolutionary history and functional impact of core and accessory genes in aquaculture. Microb. Genom. **12**, 001682 (2026)
21. Ben-Haim, Y., Thompson, F.L., Thompson, C.C., Cnockaert, M.C., Hoste, B., Swings, J., Rosenberg, E.: *Vibrio coralliilyticus* sp. nov., a temperature-dependent pathogen of the coral *Pocillopora damicornis*. Int. J. Syst. Evol. Microbiol. **53**, 309–315 (2003)
22. Sussman, M., Willis, B.L., Victor, S., Bourne, D.G.: Coral pathogens identified for white syndrome epizootics in the Indo-Pacific. PLoS One. **3**, e2393 (2008)
23. Ushijima, B., Videau, P., Poscablo, D., Vine, V., Salcedo, M., Aeby, G., Callahan, S.M.: Complete genome sequence of *Vibrio coralliilyticus* strain OCN014, isolated from a diseased coral at Palmyra Atoll. Genome Announc. **2**, e01317–e01314 (2014)
24. Richards, G.P., Watson, M.A., Needleman, D.S., Church, K.M., Häse, C.C.: Mortalities of eastern and Pacific oyster larvae caused by the pathogens *Vibrio coralliilyticus* and *Vibrio tubiashii*. Appl. Environ. Microbiol. **81**, 292–297 (2015)
25. Kim, H.J., Jun, J.W., Giri, S.S., Chi, C., Yun, S., Kim, S.G., et al.: Identification and genome analysis of *Vibrio coralliilyticus* causing mortality of Pacific oyster (Crassostrea gigas) larvae. Pathogens. **9**, 206 (2020)
26. Kimes, N.E., Grim, C.J., Johnson, W.R., Hasan, N.A., Tall, B.D., Kothary, M.H., et al.: Temperature regulation of virulence factors in the pathogen *Vibrio coralliilyticus*. ISME J. **6**, 835–846 (2012)
27. Ushijima, B., Richards, G.P., Watson, M.A., Schubiger, C.B., Häse, C.C.: Factors affecting infection of corals and larval oysters by *Vibrio coralliilyticus*. PLoS One. **13**, e0199475 (2018)
28. Rubio-Portillo, E., Martin-Cuadrado, A.B., Caraballo-Rodríguez, A.M., Rohwer, F., Dorrestein, P.C., Antón, J.: Virulence as a side effect of interspecies interaction in *Vibrio* coral pathogens. mBio. **11**, e00260–e00220 (2020)
29. Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E.: Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. **13**, e1005595 (2017)

30. Chklovski, A., Parks, D.H., Woodcroft, B.J., Tyson, G.W.: CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat. Methods. **20**, 1203–1212 (2023)
31. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
32. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., et al.: Identification of mobile genetic elements with geNomad. Nat. Biotechnol. **42**, 1303–1312 (2024)
33. Hall, J.P.J., Botelho, J., Cazares, A., Baltrus, D.A.: What makes a megaplasmid? Philos. Trans. R Soc. B Biol. Sci. **377**, 20200472 (2022)
34. 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. **9**, 5114 (2018)
35. Sharp, P.M., Li, W.H.: An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. **24**, 28–38 (1986)
36. Gautreau, G., Bazin, A., Gachet, M., Planel, R., Burlot, L., Dubois, M., et al.: PPanGGOLiN: depicting microbial diversity via a partitioned pangenome graph. PLoS Comput. Biol. **16**, e1007732 (2020)
37. Kanehisa, M., Sato, Y., Morishima, K.: BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. **428**, 726–731 (2016)
38. Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., et al.: SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. **40**, 1023–1025 (2022)
39. Payne, L.J., Meaden, S., Mestre, M.R., Palmer, C., Toro, N., Fineran, P.C., Jackson, S.A.: PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. Nucleic Acids Res. **50**, W541–W550 (2022)
40. Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J.: Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. **11**, 119 (2010)
41. Zhou, S., Liu, B., Zheng, D., Chen, L., Yang, J.: VFDB 2025: an integrated resource for exploring anti-virulence compounds. Nucleic Acids Res. **53**, D871–D877 (2025)
42. Price, M.N., Dehal, P.S., Arkin, A.P.: FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One. **5**, e9490 (2010)
43. Didelot, X., Wilson, D.J.: ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput. Biol. **11**, e1004041 (2015)
44. Xie, Z., Tang, H.: ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics. **33**, 3340–3347 (2017)
45. Cury, J., Jové, T., Touchon, M., Néron, B., Rocha, E.P.C.: Identification and analysis of integrons and cassette arrays in bacterial genomes. Nucleic Acids Res. **44**, 4539–4550 (2016)

46. Islam, H., Sharma, A., Blair, J., Lopatkin, A.J.: PlasAnn: a curated plasmid-specific database and annotation pipeline for standardized gene and function analysis. Nucleic Acids Res. **54**, gkaf1507 (2026)
47. Cury, J., Abby, S.S., Doppelt-Azeroual, O., Néron, B., Rocha, E.P.C.: Identifying conjugative plasmids and integrative conjugative elements with CONJscan. In de la Cruz F (ed), *Horizontal Gene Transfer: Methods and Protocols*, Methods in Molecular Biology, vol 2075, p 265–283. Humana, New York, NY. (2020)
48. Kawato, S.: gbdraw: a genome diagram generator for microbes and organelles. bioRxiv. (2026). 10.64898/2026.04.07.716863
49. Meier-Kolthoff, J.P., Göker, M.: TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat. Commun. **10**, 2182 (2019)
50. Söding, J., Biegert, A., Lupas, A.N.: The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. **33**, W244–W248 (2005)
51. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al.: Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. **630**, 493–500 (2024)
52. Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L.: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10**, R25 (2009)
53. Slager, J., Veening, J.W.: Hard-wired control of bacterial processes by chromosomal gene location. Trends Microbiol. **24**, 788–800 (2016)
54. Sonnenberg, C.B., Kahlke, T., Haugen, P.: *Vibrionaceae* core, shell and cloud genes are non-randomly distributed on chromosome 1: a hypothesis that links the genomic location of genes with their intracellular placement. BMC Genom. **21**, 695 (2020)
55. Reen, F.J., Almagro-Moreno, S., Ussery, D., Boyd, E.F.: The genomic code: inferring *Vibrionaceae* niche specialization. Nat. Rev. Microbiol. **4**, 697–704 (2006)
56. DiCenzo, G.C., Finan, T.M.: The divided bacterial genome: structure, function, and evolution. Microbiol. Mol. Biol. Rev. **81**, e00019–e00017 (2017)
57. Conner, J.G., Zamorano-Sánchez, D., Park, J.H., Sondermann, H., Yildiz, F.H.: The ins and outs of cyclic di-GMP signaling in *Vibrio cholerae*. Curr. Opin. Microbiol. **36**, 20–29 (2017)
58. DiCenzo, G.C., Mengoni, A., Perrin, E.: Chromids aid genome expansion and functional diversification in the family Burkholderiaceae. Mol. Biol. Evol. **36**, 562–574 (2019)
59. Rowe-Magnus, D.A., Guerout, A.M., Biskri, L., Bouige, P., Mazel, D.: Comparative analysis of superintegrons: engineering extensive genetic diversity in the *Vibrionaceae*. Genome Res. **13**, 428–442 (2003)
60. Orlek, A., Stoesser, N., Anjum, M.F., Doumith, M., Ellington, M.J., Peto, T., et al.: Plasmid classification in an era of whole-genome sequencing: application in studies of antibiotic resistance epidemiology. Front. Microbiol. **8**, 182 (2017)

61. Cazares, A., Moore, M.P., Hall, J.P.J., Wright, L.L., Grimes, M., Emond-Rhéault, J.G., Pongchaikul, P., Santanirand, P., Levesque, R.C., Fothergill, J.L., Winstanley, C.: A megaplasmid family driving dissemination of multidrug resistance in Pseudomonas. Nat. Commun. **11**, 1370 (2020)
62. Klapproth, J.M.A.: The role of lymphostatin/EHEC factor for adherence-1 in the pathogenesis of Gram-negative infection. Toxins (Basel). **2**, 954–962 (2010)
63. Ruamsap, N., Riyapa, D., Janesomboon, S., Stevens, J.M., Pichyangkul, S., Pattanapanyasat, K., et al.: Lymphostatin, a virulence factor of attaching and effacing *Escherichia coli*, inhibits proliferation and cytokine responses of human T cells in a manner associated with cell cycle arrest but not apoptosis or necrosis. Front. Cell. Infect. Microbiol. **12**, 941939 (2022)
64. Griessmann, M., Rasmussen, T., Flegler, V.J., Kraft, C., Schneider, R., Hateley, M., et al.: Structure of lymphostatin, a large multifunctional virulence factor of pathogenic *Escherichia coli*. Nat. Commun. **16**, 5389 (2025)
65. Wu, Z., Chen, G., He, L., Guo, H., Yuan, R., Su, H., et al.: Structural insights into the regulatory mechanisms of the toxic activity of Sofic in anti-phage defence systems. Int. J. Mol. Sci. **26**, 6074 (2025)
66. Wozniak, R.A.F., Waldor, M.K.: A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. PLoS Genet. **5**, e1000439 (2009)
67. Van Melderen, L., Saavedra De Bast, M.: Bacterial toxin-antitoxin systems: more than selfish entities? PLoS Genet. **5**, e1000437 (2009)
68. Haro-Moreno, J.M., Díaz-Arinero, E., Aldeguer-Riquelme, B., Rubio-Portillo, E.: Effects of marine heatwaves on the dynamics of marine coastal microbial communities. Environ. Microbiome. **21**, 51 (2026)
69. Rubio-Portillo, E., Gago, J.F., Martínez-García, M., Vezzulli, L., Rosselló-Móra, R., Antón, J., Ramos-Esplá, A.A.: *Vibrio* communities in scleractinian corals differ according to health status and geographic location in the Mediterranean Sea. Syst. Appl. Microbiol. **41**, 131–138 (2018)
70. Rubio-Portillo, E., Izquierdo-Muñoz, A., Gago, J.F., Rosselló-Mora, R., Antón, J., Ramos-Esplá, A.A.: Effects of the 2015 heat wave on benthic invertebrates in the Tabarca Marine Protected Area (southeast Spain). Mar. Environ. Res. **122**, 135–142 (2016)
71. Sugumar, G., Nakai, T., Hirata, Y., Matsubara, D., Muroga, K.: *Vibrio splendidus* biovar II as the causative agent of bacillary necrosis of Japanese oyster *Crassostrea gigas* larvae. Dis. Aquat. Organ. **33**, 111–118 (1998)
72. Ushijima, B., Videau, P., Burger, A.H., Shore-Maggio, A., Runyon, C.M., Sudek, M., Aeby, G.S., Callahan, S.M.: *Vibrio coralliilyticus* strain OCN008 is an etiological agent of acute *Montipora* white syndrome. Appl. Environ. Microbiol. **80**, 2102–2109 (2014)
73. Li, R., Huang, Y., Wu, J., Dang, H., Wang, L., Zhang, W., Ding, J.: Whole genome analysis and specific PCR primer development for *Vibrio coralliilyticus*, combined with transcription and metabolome analysis of red spotting disease in the sea urchin, Strongylocentrotus intermedius. Aquac Rep. **22**, 100957 (2022)

74. Spinard, E., Kessner, L., Gomez-Chiarri, M., Rowley, D.C., Nelson, D.R.: Draft genome sequence of the marine pathogen *Vibrio coralliilyticus* RE22. Genome Announcements. **3**(6), 10–1128 (2015)
75. Richards, G.P., Watson, M.A., Madison, D., Soffer, N., Needleman, D.S., Soroka, D.S., Uknalis, J., Baranzoni, G.M., Church, K.M., Polson, S.W., Elston, R., Langdon, C., Sulakvelidze, A.: Bacteriophages against *Vibrio coralliilyticus* and *Vibrio tubiashii*: Isolation, Characterization, and Remediation of Larval Oyster Mortalities. Appl. Environ. Microbiol. **87**, e00008–21 (2021)
76. Machado, H., Sonnenschein, E.C., Melchiorsen, J., Gram, L.: Genome mining reveals unlocked bioactive potential of marine Gram-negative bacteria. BMC Genom. **16**, 158 (2015)
77. Kehlet-Delgado, H., Häse, C.C., Mueller, R.S.: Comparative genomic analysis of Vibrios yields insights into genes associated with virulence towards C. gigas larvae. BMC Genom. **21**, 599 (2020)
78. Ushijima, B., Meyer, J.L., Thompson, S., Pitts, K., Marusich, M.F., Tittl, J., et al.: Disease diagnostics and potential coinfections by *Vibrio coralliilyticus* during an ongoing coral disease outbreak in Florida. Front. Microbiol. **11**, 569354 (2020)
79. Wang, Y., Wang, Q., Chen, L., Ding, R., Peng, Z., Li, B.: Pathogenicity and immune response of the sea urchin Mesocentrotus nudus to *Vibrio coralliilyticus* infection. Aquaculture. **572**, 739570 (2023)
80. Ushijima, B., Saw, J.H., Videau, P., Häse, C.C.: Comparison of *Vibrio coralliilyticus* virulence in Pacific oyster larvae and corals. Microbiol. (Reading). **168**, 001169 (2022)
81. Liu, G.F., Wang, J.Y., Xu, L.W., Ding, X., Zhou, S.N.: Sensitive and rapid detection of *Vibrio coralliilyticus* by loop-mediated isothermal amplification targeted to the alpha subunit gene of RNA polymerase. Lett. Appl. Microbiol. **51**, 301–307 (2010)
82. Hanna, K.D., Richards, G.P., Claudia, H., Mueller, R.S.: Three Draft Genome Sequences of *Vibrio coralliilyticus* Strains Isolated from Bivalve Hatcheries. Microbiol. Resour. Announc **5**(41). (2017)
83. Rubio-Portillo, E., Yarza, P., Peñalver, C., Ramos-Esplá, A.A., Antón, J.: New insights into *Oculina patagonica* coral diseases and their associated *Vibrio* spp. communities. ISME J. **8**, 1794–1807 (2014)

Figures

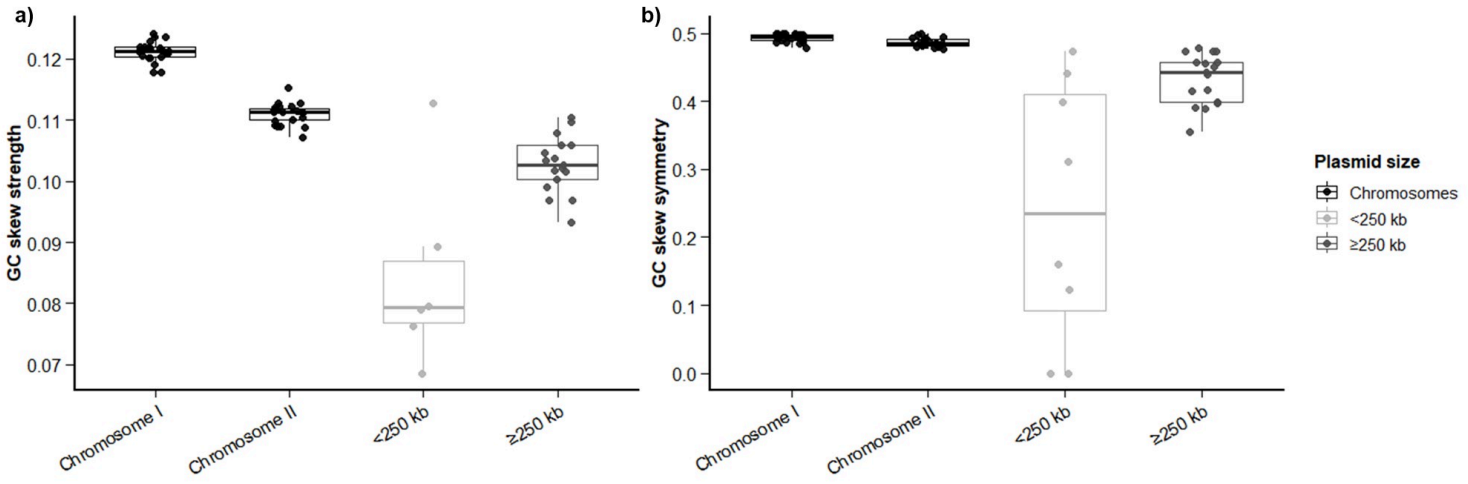


Figure 1

GC skew strength and symmetry across chromosomes and plasmid size classes. (A) GC skew strength, calculated as the sum of the mean positive GC skew and the absolute value of the mean negative GC skew for each replicon. (B) GC skew symmetry. Points represent individual replicons from the 27 complete genomes, and boxplots summarize their distributions.

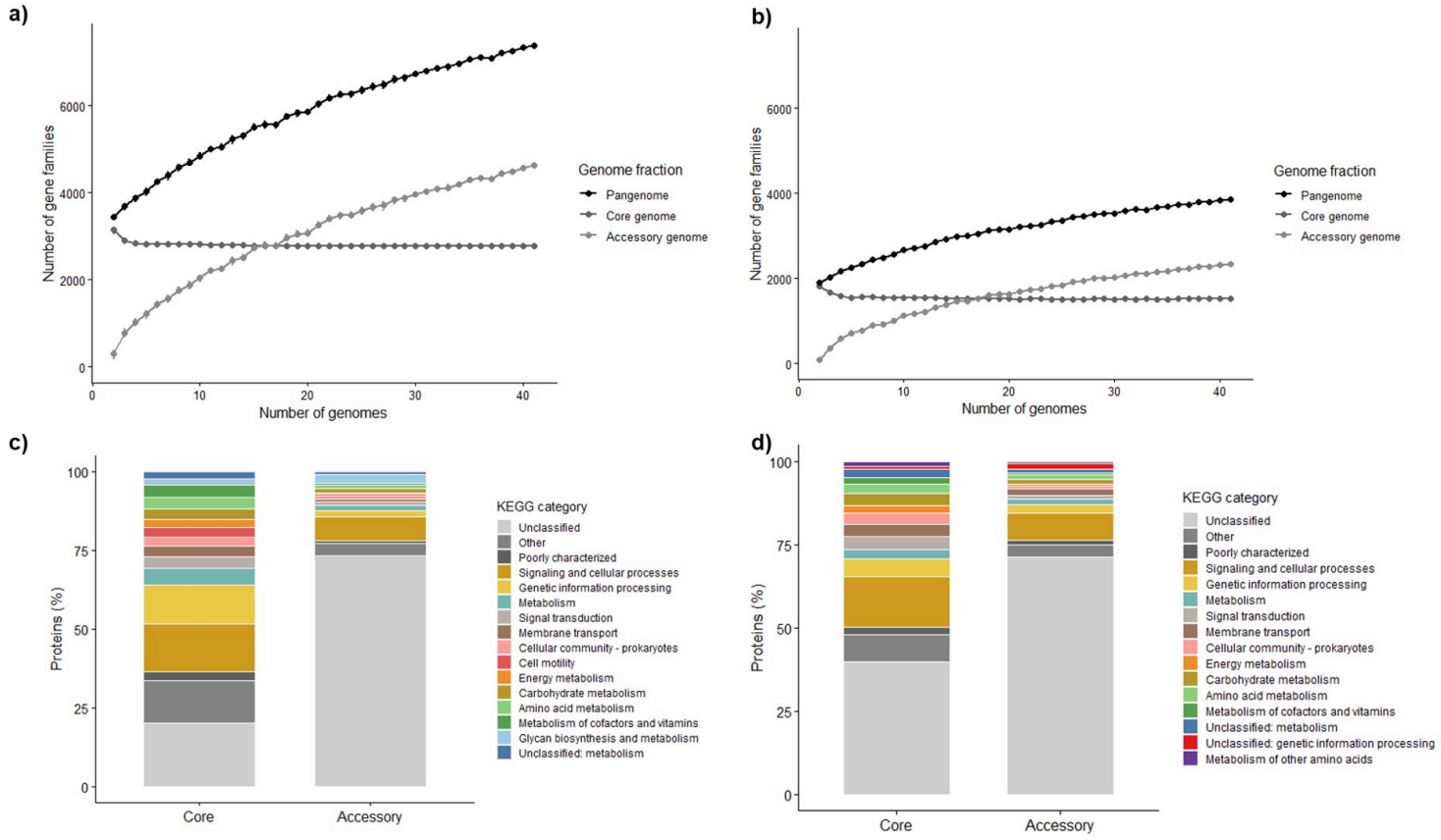


Figure 2

Pangenome structure and functional composition of *V. coralliilyticus* across chromosomes. (A, B) Pangenome accumulation curves for chromosome I (A) and chromosome II (B), showing the number of gene families assigned to the pangenome, core genome, and flexible (accessory) genome as a function of the number of genomes sampled (n = 41). (C, D) Relative distribution (%) of KEGG functional categories within the core and flexible fractions of chromosome I (C) and chromosome II (D). Gene families were partitioned with PPanGGOLiN and functionally annotated with BlastKOALA; proteins without a KEGG assignment are shown as "Unclassified".

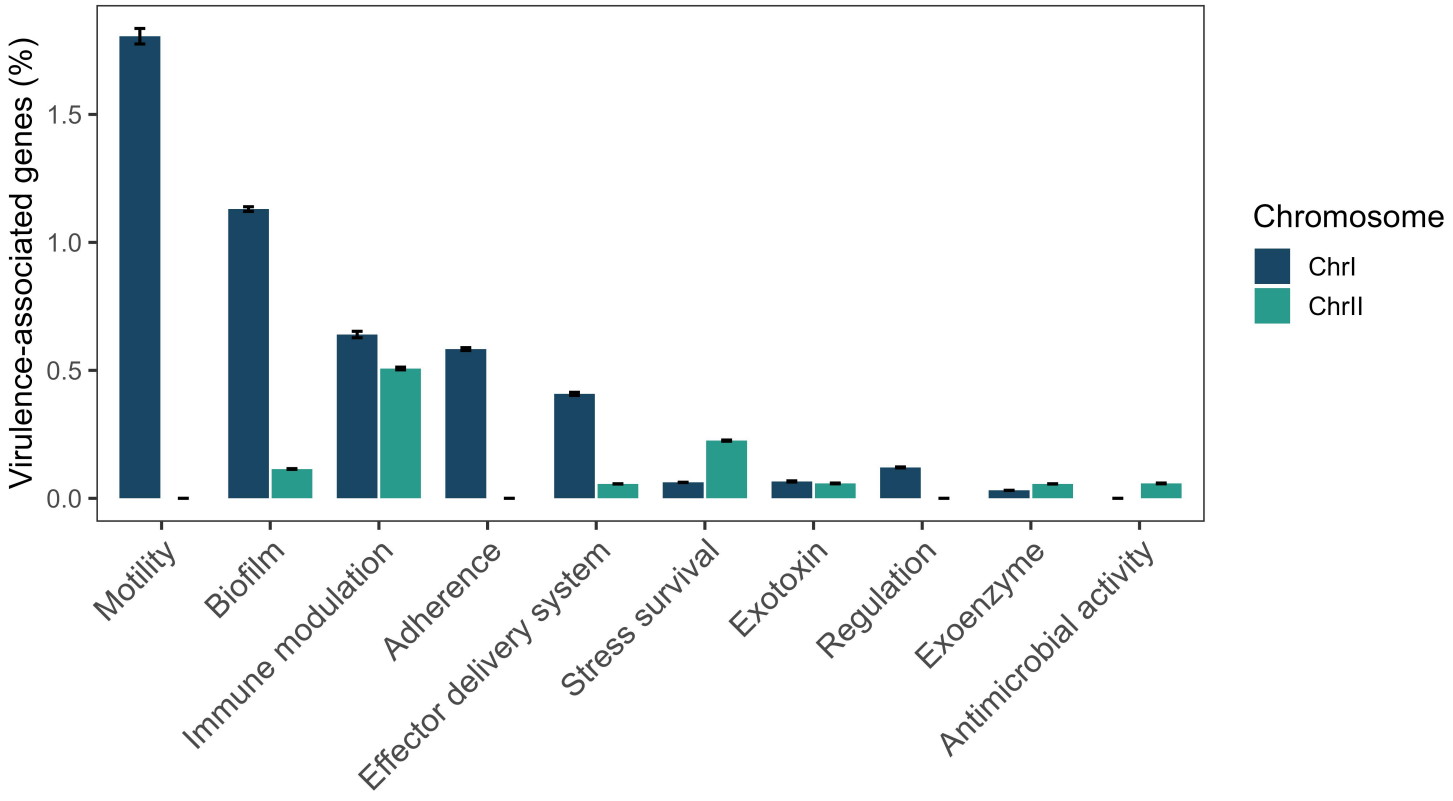


Figure 3

Distribution of virulence-associated genes between chromosome I (ChrI) **and chromosome II** (ChrII)**.** Relative abundance of genes assigned to different virulence-related functional categories in ChrI and ChrII, normalised to the total number of genes encoded by each replicon. Bars represent the mean percentage across the 41 genomes analysed and error bars indicate the standard error. Functional categories were assigned based on VFDB annotations.

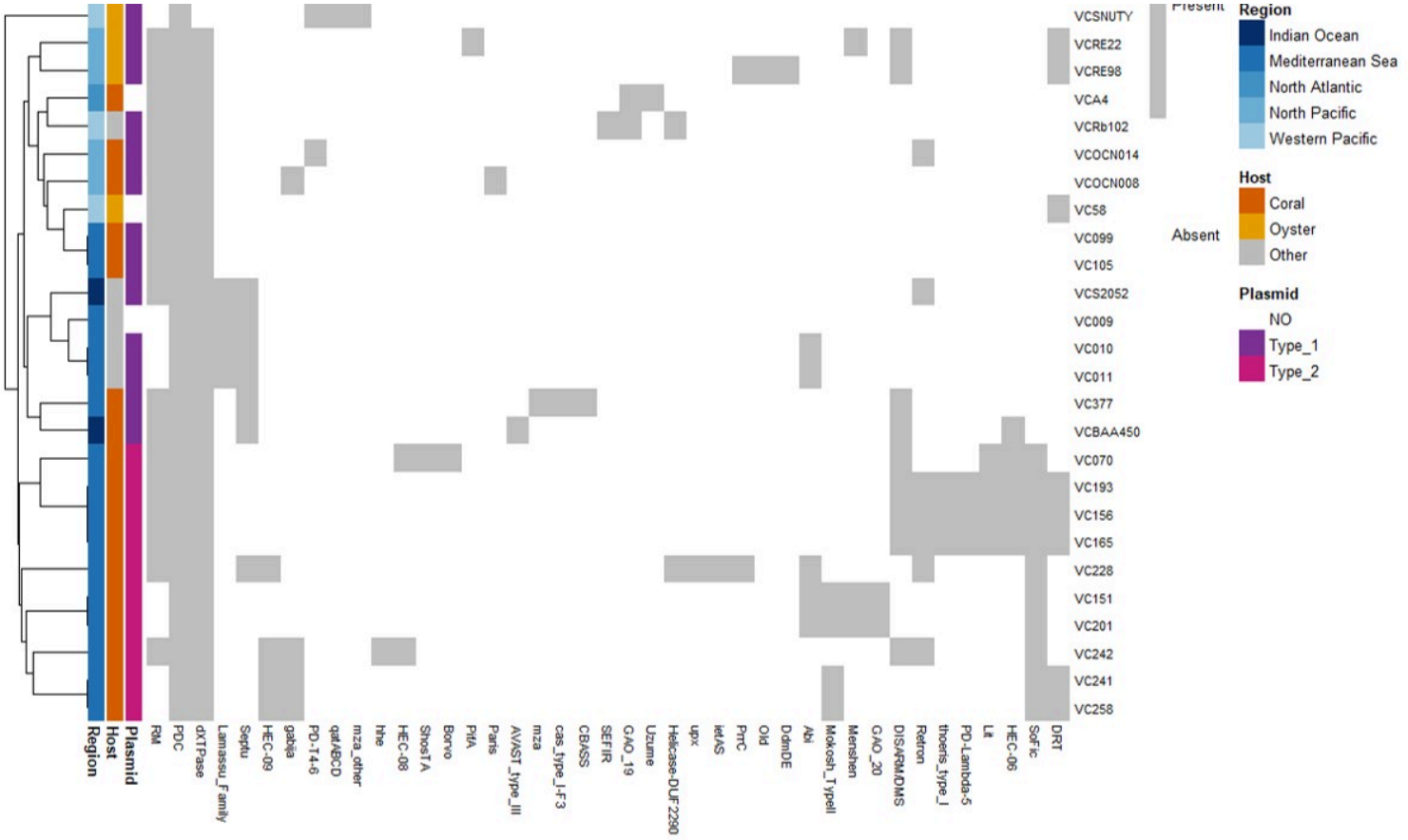


Figure 4

Distribution of defence-system repertoires across *V. coralliilyticus* genomes. Hierarchical clustering of 41 *V. coralliilyticus* genomes based on the presence and absence of defence-system types identified with PADLOC. Rows represent genomes and columns represent defence-system categories. Grey cells indicate the presence of a defence system and white cells its absence. Coloured annotations indicate the geographic region of isolation, the host or sample source, and the megaplasmid lineage.

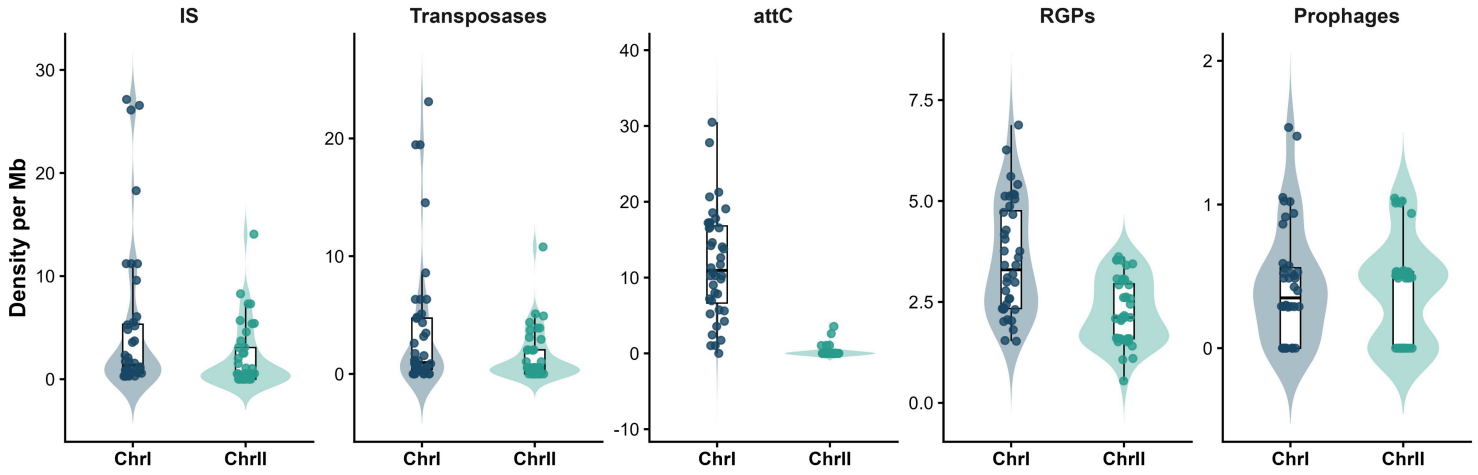


Figure 5

Enrichment of genome plasticity-associated features in chromosome I of *Vibrio coralliilyticus*. Violin plots showing the density per megabase (Mb) of insertion sequences (IS), transposases, integron-associated *attC* sites, regions of genomic plasticity (RGPs), and prophages in ChrI and ChrII. Points represent individual genomes and boxplots indicate the median and interquartile range.

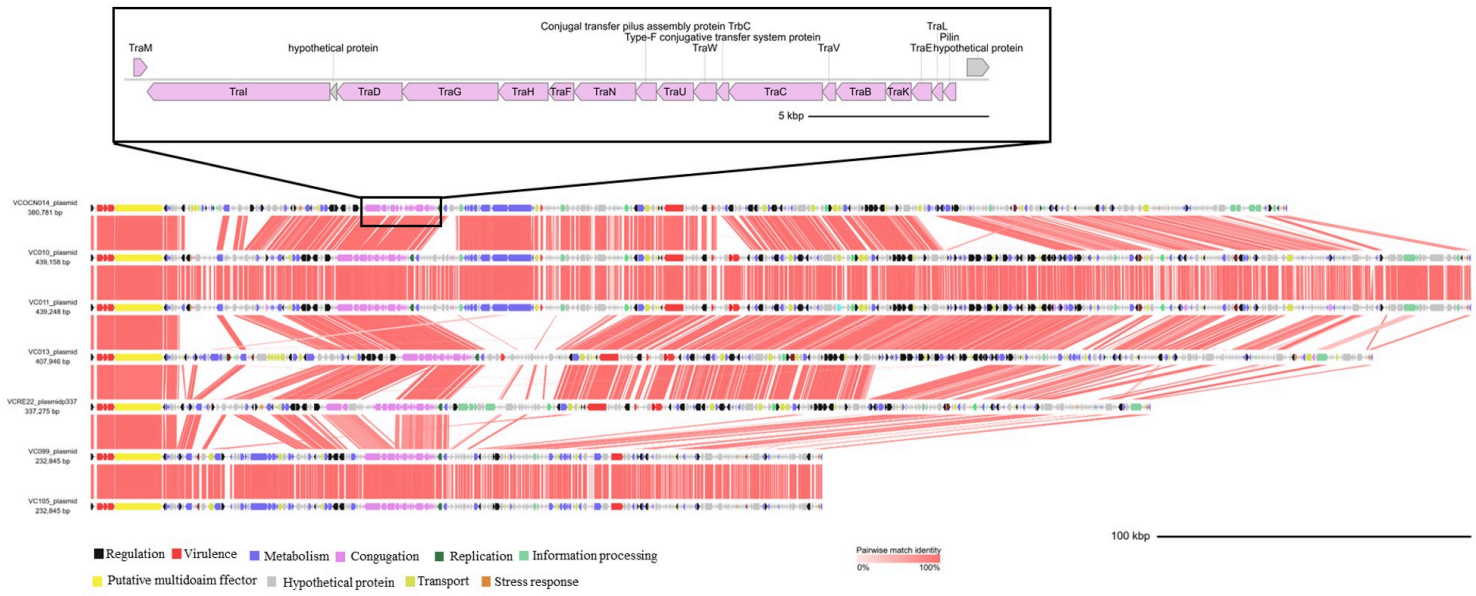


Figure 6

Comparative genomics of complete Type1 megaplasmids. Pairwise genomic comparison of the seven complete Type1 megaplasmids identified in *V. coralliilyticus*. Genes are coloured according to their predicted functional category, and shaded connections indicate regions of nucleotide similarity between plasmids, with colour intensity proportional to pairwise identity. The inset shows the organisation of the conjugative transfer region, with *tra*and *trb* genes labelled. Plasmid names and sizes are indicated on the left; scale bars correspond to 5 kb (inset) and 100 kb (main panel).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure6.jpg
- SupplementaryFigure4.jpg
- SupplementaryFigure8.jpg
- SupplementaryTable2.docx
- SupplementaryFigure5.jpg
- SupplementaryFigure3.jpg
- SupplementaryTable1.docx
- SupplementaryTable4.docx

- SupplementaryFigure9.jpg
- SupplementaryFigure7.jpg
- SupplementaryTable3.docx
- SupplementaryFigure1.jpg
- SupplementaryTable5.docx
- SupplementaryFigure2.jpg
- Supplementary.docx