Emergence and clonal expansion of *Vibrio aestuarianus* lineages pathogenic for oystersin Europe.

Aurélie Mesnil[1,2], Maude Jacquot[1], Céline Garcia[1], Delphine Tourbiez[1], Lydie Canier[1], Audrey Bidois[2], Lionel Dégremont[1], Deborah Cheslett[3], Michelle Geary[3], Alessia Vetri[4], Ana Roque[5], Dolors Furones[5], Alison Garden[6], Petya Orozova[7], Isabelle Arzul[1], Mathieu Sicard[8], Guillaume M. Charrière[2], Delphine Destoumieux-Garzón[2], Marie-Agnès Travers[2*]

1 ASIM, Ifremer, Station de La Tremblade, avenue de Mus du Loup, 17390 La Tremblade, France

2 IHPE, Université de Montpellier, CNRS, IFREMER, Univ Perpignan Via Domitia, Montpellier, France

3 Marine Institute MI, Rinville, Oranmore, Co. Galway, H91 R673, Ireland

4 Instituto Zooprofilattico Sperimentale del le Venezie (IZSVe), Viale del l'Università 10, 35020 Legnaro (PD), Italy

5 IRTA La Ràpita. Aquaculture Program. 43540 – San Carles de la Ràpita, Spain.

6 Marine Laboratory, Marine Scotland Science, Aberdeen, UK.

7 National Diagnostic and Research Veterinary Medical Institute, Sofia, Bulgaria.

8 ISEM, Université de Montpellier, CNRS, IRD, Montpellier, France.

* Corresponding author: marie.agnes.travers@ifremer.fr

Competing interests: The authors declare no conflict of interest.

Abstract

Crassostrea gigas oysters represent a significant global food source,with 4.7 million tons harvested per year. In 2001, the bacterium*Vibrioaestuarianus francensis*emerged as a pathogen that causes adult oyster mortality in France and Ireland. Its impact on oyster aquaculture has increased in Europe since its reemergence in 2012. To better understand the evolutionary mechanisms leading to the emergence and persistence over time of this pathogen, we conducted a survey of mollusk diseases through national reference laboratories (NRLs) across Europe. We analyzed54 new genomes of *Vibrio aestuarianus*(*Va*) isolated from multiple environmental compartments since 2001, in areas with and withoutbivalve mortalities. We used a combination of comparative genomics and population genetics approaches and show that*Va* has a classical epidemic population structure from which the pathogenic *Va francensis* subspecies emerged and clonally expanded.Furthermore, we identified a specific *cus-cop*-containing island conferring copper resistance to *Va francensis*whose acquisition may have favored the emergence of pathogenic lineages adapted and specialized to oysters.

Introduction

Molecular microbiology data offer opportunities to address crucial questions about infectious diseases epidemiology, with the aim to better understand mechanisms of emergence and transmission of pathogens. In particular, applying population genetics and comparative genomics approaches tothose data allow to get further insights into processes driving the evolution of pathogenic lineages. A better understanding of such key evolutionary features of infectious disease could allow effective prevention and control strategies.

Studies on bacterial pathogenic lineages showedthat their genomes cancarry low genetic diversity(1, 2).This is particularly true for those evolving clonally(3, 4).Low bacterial genetic diversity is often due to ecological isolation which is one of the main factors preventingthe

horizontal transfers and recombination events to occurwith other population(5, 6).Weakly divergent bacterial populations can also have either undergone a recent loss of genetic variation due to a demographic bottleneck or have been established by a very small founder population from a larger population. In addition, populations can also be subjected to either selection pressures that will lead to frequency increase and fixation of mutated alleles (selective sweep) along with surrounding loci by hitchhiking(7), or to elimination of mutated alleles and linked loci from the population (background selection, (8)). Founder events and selective sweepscan follow a favourable trait acquisition which can be achieved by mutation (9) or, more frequently, through horizontal transfer (through transformation, conjugation or transduction) of one or more genes favoring rapid adaptation to new niches and hosts (10). For example, among pathogenic bacterial lineages, some *Vibrio* species experienced trait acquisition that led to clonal expansion and pandemics. As for example,*Vibriocholerae* pandemic lineages emerged through genomic island acquisition conferring pathogenicity(11).

Althoughcultivated marine species are particularly threatened by infectious diseases, very little is known about the emergence and evolution of *Vibrio* bacteria responsible for marine mollusk diseases.It is yet crucial because aquaculture is practiced in an open and highly connected environment, with high population densities and animals being frequently transferred between production areas which increases the risk of pathogens introduction and favors their spatial dispersal (12, 13).*Vibrio aestuarianus* subsp. *francensis (Va francensis)*, affecting *Crassostrea gigas*, the most grownoysters worldwide (93% of the global oyster production) is of particular interest.Indeed, *Va francensis*has caused recurrent mortality of cultivated adult oysters since 2001(14), first in France, thenin several European countries (France, Italy, Ireland, and Spain)(15, 16). This spread of *Va francensis*has resulted in mortality rates of approximately 25% and is a major economic concernas itimpacts market-size oysters(17). Previous field studies found that pathogenic strains of *Vafrancensis* can colonize sentinel oysters, but are rare in samples fromthe local environment, *i.e.* sediment

and water column(18). Moreover, it has been proposed that *Va francensis* can persist in oyster tissues during winter(18). These data suggest that oysters are the preferred habitat of this pathogen andraise questions about how this host specialization has occurred.

Va francensis belongs to *Vibrio aestuarianus* species (*Va*), described in 1983 as a bacterium from estuarine environments(19), but whose genetic diversity has not been studied apart from the subspecies *Va francensis*(20). This lack of genomic data hinders a complete understanding of the mechanisms governing the emergence and evolutionary dynamics of *Vafrancensis.* In the present study, we therefore analyzed 54 genomes of isolates collected since 2001 across Europe. We combined phylogenetic, population genetics and comparative genomicsapproaches to decipher evolutionary processes associated with the emergence of this oyster pathogen. Our study shows that, like in other *Vibrio* species, *Va francensis*subspcies harbor an epidemic population structure, with clonal emergence and expansion of pathogenic lineages specialized to the oyster host.Our data suggest the acquisition of genes involved in copper resistance has been one mechanism promoting *Va*adaptation to oysters.

Material & methods

Isolate selection

The study collection was designed to represent the temporal, geographic and clinical heterogeneity of *Vibrioaestuarianus*(*Va*) origins. However, ascertainment bias could occur due to sampling strategy used. Indeed, strains were mainly isolated during bivalve mortalities. We maximized recovery of the genetic diversity of oyster pathogenic *Va* by collecting isolates from all European countries reporting mortalities related to *Va* or presence of the bacteria. We also included strains isolated during field surveys of sites without known mortality events. Strains were collected with the help of European NRLs responsible for the diagnosis and surveillance of mollusk diseases: Ifremer (France), MI (Ireland), IZSVE (Italy),

IRTA (Spain), Marine Scotland Science (Scotland), and NRLFMC (Bulgaria). Sequence comparisons of isolates from France, notably those isolated from moribund oysters, turned out to be very similar. In order to equilibrate our sampling, we excluded 17 isolates identical with included isolates. We selected approximately one strain per batch of moribund oysters (a batch represents a sampling point)(Table S1). The 54 isolates used in the study are summarized in Table S2.

Illumina, Ion torrent and PacBio whole-genome sequencing

Genome sequencing was performed between 2019 and 2021. Several sequencing technologies were used depending on the objectives. We used a long-read sequencing technology for some strains to improve assembly quality and study genomic features. DNA was isolated from 35 strains using the NucleoSpin Tissue Kit (Macherey-Nagel) following the manufacturer's guidelines. The concentration of each sample was adjusted to 0.2 ng/µl for library preparation using the Illumina Nextera XT DNA Library Preparation Kit according to the manufacturer's guidelines. For each library, paired-end reads of 150 bpwere produced on the Illumina NextSeq 550 platform (Bio Environnement, Perpignan, France). The quality of the resultant fastq files was assessed using FastQC(21). Primers, Illumina adapter sequences, low-quality bases at the 3′ end of reads with Phred-scaled quality scores less than 30 and reads with a length <50 bp were removed using Trim Galore (22). Quality-controlled reads were *de novo* assembled using the SPAdes genome assembler v3.13.1 (23) with the default kmer size and a coverage cutoff value of 20 reads (24).

For 2 strains (04_091_1T1 and U8), enzymatic shearing was performed on 500ng of genomic bacterial DNA using the NEBNext Fast DNA Library Prep kit (New EnglandsBiolabs) and following the manufacturer's recommendations. Samples were barcoded with individual IonCodes (Thermo). Purified libraries were then quantified by real-time PCR using the Ion Universal Library Quantification Kit (Thermo) prior to be diluted and pooled at 25µM each. The Ion Torrent S5 libraries were prepared using the "Ion 510™ & Ion

520™ & Ion 530™" for Ion Chef Kit and sequenced on the Ion Torrent S5 using an Ion 530 semi-conductor sequencing chip (Thermo).

DNA was isolated from other strains using a MagAttract HMW DNA Kit (Qiagen) following the manufacturer's instructions. Library preparation and sequencing on a PacBio Sequel II instrument were performed with the GENTYANE INRAe platform (Clermont-Ferrand, France). *De novo* genome assemblies were obtained using FLYE v2.8.3 (25). One round of polishing was performed using RACON v 1.4.3 (26).

The assemblies were evaluated with QUAST, and genome completeness were checked using the BUSCO v5.1.1 (27) and vibrionales_odb10 databases as lineage datasets, including 57 Vibrionales, and 1445 orthologous single copy genes known to be present in *Vibrio* genomes. BUSCO check for genomes completeness by checking these genes presence.

Assembly features and GenBank accession numbers are summarized in Table S2.

Genome annotations

For each assembly, coding sequences (CDSs) were predicted using PROKKA v1.12 (28) with default databases and –compliant and –rfam flags and annotation of rRNAs, tRNAs and other ncRNAs enabled. Functional annotation of the detected CDSs was performed using EggNog-Mapper v2.1.0 (29, 30).The annotated genomes were then processed with Roary v3.13, using the following params "-e -n -I 90 -cd 100" to perform a pangenome analysis and identify the set of core and accessory genes (31). In order to determine if pangenomes are open or closed in *Va* clades, we used methodology developed by Tettelin et al (32). Briefly, we determined the number of new genes added to the pangenome based on the number of genomes analyzed. This number was normalized by the number of genes in the analyzed genome. Toconsider the effect of the order in which new genomes are added to the pangenome, we performed 50 genome order permutations per group. Then we calculated the median, and the 25^{th} and 75^{th} percentiles. Finally, we determined the equations of the regression curves according to the power law, to define the value of the exponent alpha.

Insertion sequences (IS) were detected in the genomesassembled from PacBio sequences using ISEScan v1.7 (33). Pseudogene detection was performed with the GenBank files released from PROKKA with PSEUDOFINDER v1.0 (34), using the diamond flag and NCBI nonredundant database. These later analyses were performed only on resolved genomes to avoid obtaining false negatives in IS detection and false positives in pseudogene detection at the intercontig locations.

Phylogeny and recombination analysis

Pairwise ANIs (%) were calculated using FastANI(35). Whole-genome alignments of the 54 isolate contigs were made using Progressive Mauve v13/02/2015 (36). Whole-genome alignment was used to infer the maximum likelihood midpoint-rootedphylogeny using RaxML v8.2.12 with the TVM model of nucleotide substitution (transversion model, AG=CT and unequal base frequencies) and 1000 bootstrap replicates.

To identify polymorphic sites, the haploid variant calling pipeline snippy v4.6(available at https://github.com/tseemann/snippy) was used with default parameters.To get further insights into population structure, we focused on allele frequency differences within genomes and applied the linkage model of Structurev2.3.4(37).

To assess the influence of various structuring factors on the observed genetic diversity, we performed AMOVA on the sequence dataset, with environmental origin, isolation context, sampling year, and country as structuring factors, using the R package *pegas*.

Recombination was evaluated across this core genome alignment using ClonalFrameML v1.12 (38). We used an extension that allows different recombination parameters to be inferred on different branches of the clonal genealogy. Linear regression analysis of the root-to-tip distances against sampling time was performed using TempEst v1.5.3 (39). The substitution rate was estimated under different demographic and clock models, using Beast v1.10.4, taking advantage of the sampling timeframe between 2001 and 2021. We tested both a strict and a relaxed molecular clock. Constant-size, exponential-size and GMRF

Bayesian skyride plot models (40, 41)were used. For each model, two independent chains were conducted for 200 million generations and convergence was assessed by checking effective sample size values for key parameters using Tracer v1.7.1 (42). For robust model selection, both path sampling and stepping-stone sampling approaches were applied to each BEAST analysis to estimate the marginal likelihood(43, 44). The best model selected by Bayes factor comparison was a relaxed molecular clock with a GMRF Bayesian skyride demographic model.For this model, triplicate runs were combined using LogCombiner v1.10.4, with removal of 10% burn-in.

Selection analysis

Tajima's D was calculated considering core genome alignment using MEGA 11 Tajima's test on neutrality. Core genome alignment were previously partitioned according to lineages or subspecies, resulting in *Va francensis* lineage A alignment (n = 17 sequences), *Va francensis* lineage B alignment (n = 8 sequences), *Vacardii* alignment (n = 17 sequences) and non-clinical*Va*alignment (n = 12 sequences). Since sample size is known to influence population genetic parameter estimations (45), we calculated Tajima's D on subsamples of eight sequences.For *Va francensis* lineage A, *Vacardii* and non-clinical*Va*isolates, Tajima's D was calculated on ten different alignments constructed with eight sequences randomly subsampled. For the *Va francensis* lineage B alignment, the eight sequences represented the entire sampled set. Tajima's D was also calculated for each core genes, considering the same partitions,employing the R package *pegas*.Tajima's D values follow a beta distribution under neutral evolution(46). Tajima's D p-value are calculated by comparison with this distribution. It has been observed that genes with p-values >0.05,have to be considered under neutral evolution, while the others are under positive or negative selection.

In order to detect SNPs that may explain *Va francensis* emergence, we performed a genome scan analysis using *pcadapt* package v.4.3.3 for R (47) that detect structure and clustering of individuals in a population. Number of principal components was set tosix. We used

LD.clumping option, with window size of 5 kb. LD.clumpinguses p-value to sort the SNPs by importance (*e.g.* keeping the most significant ones). It takes the first one (*e.g.* most significant SNP) and removes SNPs that are too correlated with this one in a window around it. We also removed SNP with minor allele frequency under 0.05. After computing the test statistic, we used Bonferroni adjustment to correct for multiple comparisons, and threshold of 0.001 was applied to identify outliers SNPs.

Bacterial virulence estimation by injection

Bacteria were grown with constant agitation at 22°C for 24 h in Zobell medium. Cultures were washed by centrifugation (10 min, 1500 g) and resuspended in artificial, sterile seawater to obtain OD_{600nm} = 0.1. Fifty microliters of bacterial suspensions (10^7cfu, checked by plating on Zobell agar) were injected intramuscularly into anesthetized specific pathogen-free (SPF) oysters produced and maintained in the IFREMER facility (48). After injection, the oysters were transferred to aquaria (10 oysters per 2.5 L aquarium, duplicate aquaria per bacterial strain) containing 2 L of aerated, UV-treated seawater at 22°C and kept under static conditions for 5 days.

Copper resistance phenotyping

Paper discs (6 mm) were loaded with $CuSO_4$ by soaking them overnight in 50 mM $CuSO_4$. Cells from stationary phase cultures at DO = 1 were washed twice with SSW before being spread onto agar plates containing 15 $g.L^{-1}$bactopeptone and 0.5 M NaCl. Using sterile tweezers, copper-impregnated disks were placed on the bacterial lawn before incubation at 20°C. After 5 days of incubation, the diameter of the inhibition zone around the disk was measured.

Statistic tests

Basic statistical analysis of all of the data were performed using R v4.1.2. Data were checked for normality and homogeneity of variance before analysis. Wilcoxon-Mann-Whitney was performed to evaluate if recombination ratios, IS number in genomes, virulence toward oysters and halo size (representing strains sensitivity to copper) differed among *Va francensis* lineages and between *Va francensis* lineages and other strains.

Results

European spread of *Vibrio aestuarianus* oyster pathogens

To assess the geographic distribution of pathogenic *Vibrio aestuarianus*(*Va*) in Europe, a questionnaire was sent in 2021 to all European laboratories in charge of the diagnosis and surveillance of mollusk diseases (Supplementary material). Eight out of the 12 respondent laboratories reported that*Va*was detected during adult oyster mortality events. The earliest detections were reported by France and Ireland in 2001, and the most recent detection was reported by Denmark in 2021 (case used DNA detection only) (Figure 1, A). The number of countries reporting detection has increasedsince 2011, with 8 European countries currently reporting *Va*. Importantlythe date of first detection may be under-estimated, since surveillance and testing for the pathogen has varied between countries and was only implemented in the majority of the laboratories following the development of a diagnostic toolin 2009 (49).

Vibrioaestuarianus diversity is structured according to environmental origin and isolation context.

To assessthe overall population structure of *Va* in Europe, we studied genetic diversity at the whole genome level using *Va* isolates froma broad variety of origins.54strains were sequenced and genomes were*de novo* assembled. We obtained 15 4.5-Mb complete

genomes (with 2 to 4 contigs) and 39 draft genomes (with up to 150 contigs) (Table S2). Whole genomes were aligned, and relationships between isolates were studied by building a maximum likelihood phylogeny (Figure 2). Overall, genomes were close to each other in the tree and showed highly similar nucleotide compositions, showing average nucleotide identities (ANIs) higher than 97.1% for all pairwise comparisons (Table S3).The tree showed a tightly clustered clade of oyster clinical isolates (*i.e*. isolates from moribund oysters) (n = 25) inside genetically diverse *Va* strains isolated outside of oyster mortalities(n = 29), encompassing non-clinical isolates (*i.e.* strains isolated from sites without mortalities, n = 12) and strains isolated during cockles mortalities, that belong to *Vacardii* subspecies (n = 17).All oyster clinical isolates areencompassed in*Va francensis*subspecies(50) andall *Va francensis* isolates were found in oysters, and showed higher virulence levels than *Vacardii* strains and non-clinical isolates when assayed throughinjection into oyster muscle (Wilcoxon test,p-value< 0.001). The reconstructed tree also confirmed that this subspecies is composed of two clades named A (n = 17) and B (n = 8)(20).To get further insights into *Va*population structure, we focused on allele frequency differences within genomes and applied the linkage model of Structureto our data and identified six ancestral populations in our samples. *Va francensis* strains were grouped into a unique population predicted to have descended from a unique ancestor (Figure S1).

Analysis of molecular variance (AMOVA) performed on the sequence dataset revealed that the isolation context (“oyster mortality” or not) and environmental origin(five host species, zooplankton and water) explained most significantly (p< 0.005) genetic variation (Table 1). AMOVA alsoindicated the absence of spatial and temporal delineation of the oyster pathogen lineages.Altogether, our data reveal*Va*lineageslinked European expansion ofthis oyster pathogen.

Temporal emergence of oyster pathogen lineages

Next, we used a Bayesian approach to date the emergence and reconstruct the temporal evolution of oyster pathogens lineages. As a critical prerequisite, we checked whether the *Va francensis* subspeciesis a measurably evolving population. A temporal signal was observed with a correlation between genetic distances from a reconstructed common ancestor andstrains sampling time, yielding a R^2 of 0.4236(FigureS2A).Root-to-tip divergence shows a weak temporal signal. This can be the consequence of a limited representativeness of the genome collection, which is composed of strains collected since the emergence of the clone and does not represent isolates for periods prior to the observation of mortalities in Europe.The best-fitting evolutionary model for the *Va francensis*population was obtained under a GMRF Bayesian skyline population model with a relaxed clock, leading to a rate of 7.27×10^{-7} substitutions per position per year, or 3.2 mutations per genome per year. The mean time of emergence of the most recent common ancestor (TMRCA) of thelineageA was estimated to be 1956 (95% highest posterior density (HPD): 1919–1972). At this time, a divergenceoccurred between strain 01-151, which is non-virulent toward oysters, and other lineageA strains that are mostly pathogenic. The TMRCA of lineageB maps to the late 1970s, albeit with uncertainty (95% HPD: 1898–1993) (Figure S2B).

Genome stability in*Vibrio aestuarianus francensis* lineages

Recombination can drive genetic diversity and evolutionary dynamics in bacterial populations. We therefore quantified homologous recombination events using*Va*core genome alignment.

Recombination has little impact on oyster pathogen evolution. We found that recombination (r) has a lower impact on the core genome substitution rate in *Va francensis* lineages than the impact of mutation (m), with mean r/m = 0.36+/- 0.22 and 0.28+/- 0.09 in lineages A and B respectively (Figure S3). The ratios ofthese lineages are significantly (Student test, p-value < 0.005) smaller than in non-clinical *Va* isolates (*i.e.* strains isolated

fromsites without mortalities) (r/m = 0.91 +/- 0.45). Recombination in genomes can generate phylogenetic inconsistencies between taxa that can emerge in phylogenetic networks as the presence of alternative pathways(51). The network that we constructed from the core genome alignment (Figure S4) confirmed that recombination does not affect *Va francensis*isolaterelationships, with few ambiguitiesamong these taxa. One known consequence of the absence of recombination impact is strong linkage between loci in non-recombinant genomes(52). In *Vafrancensis* lineage A and, to a lesser extent, in lineage B, strong linkage between single nucleotide polymorphisms (*SNPs*) was observed (Figure S5). In contrast, the recombination impact detected in all other genomes was confirmed with phylogenetic networks, which showed ambiguities between taxa (Figure S4) and low linkage between loci in these genomes (Figure S5).

To further characterize genetic exchanges in *Va* populations, we analysed the pan-genomesof 54*Va* strains (17 *Vacardii*isolates, 12 non-clinical *Va*, 17*Va francensis* clade A and 8 *Va francensis* clade B isolates). First, analysis with Roary indicates that *Va francensis*hasa closed pan-genome, with the large majority of the genes being shared between the *Va francensis* isolates (n=2 899) and almost no gene specific to one strain (Figure S6,n= 1 to 115). On average, 81.3% of *Va francensis* genomes are present in *Va francensis* core genomes alignment, which is significantly higher (Wilcoxon test, p-value < 0.001) than *Vacardii* genomes in *Vacardii* core genome alignment (on average 65.5%) andsignificantly higher (Wilcoxon test, p-value < 0.001) than non-clinical *Va* genomes in non-clinical *Va*core genome alignment (on average 74.2%).Consistently, pan-genome rarefaction analyses support that *Va francensis* pan-genomes are closed, as indicated by α coefficient values being greater than 1(32)(Figure3A).To elucidate the importance of accessory genome and strain-specific genes in genomic architecture, we finally constructed the pan-genome and core-genome trees of *Va* and compared their topology(Figure S7).High congruence in the phylogenetic relationshipsdepicted bycore and pan-genome treesof*Va francensis* strains support that genetic exchanges with other *Va* populations are rare.

Core genome stability was observed in *Va francensis* lineages. Mutations also generated little genetic diversity among *Va francensis*. Indeed, apart from the recombinant segments, among the 19 775 variant genomic positions in the full dataset, 666 and 83 positions vary among *Va francensis*A and B lineages. Finally, low impact of recombination and mutation on *Va francensis*lineages results in core genome stability. We used Tajima's D to compare observedgenetic diversity in *Va* to that expected under neutral evolution and constant population size.Deviation from neutrality was largeramong the *Va francensis* lineages A (Tajima's D = -2.3) and B (-2.5) than in other *Va* groups (*Vacardii* and non-clinical *Va*)(respectively -0.45 and -0.62), with less polymorphism than expected under neutral evolution and constant population size (Figure4A). To determine whether the low genetic diversity was due to demographic effects (affecting whole genomes) or to selection effects (affecting particular genes), we adopted a gene-based approach to investigate selection pressures at the gene level. While most of the 1,974 core genes were under neutral evolution outside of the *Va francensis*subspecies, the vast majority were monomorphic within lineages A and B (Figure4B), which is consistent with demographic effects impacting whole genome rather than selectioneffects affecting particular genes.

Positive selection signatures of oyster pathogen lineages

We used *pcadapt,* which detects structure and clustering of individuals in a population,to detect potential signatures of positive selection affecting oyster pathogen lineages. 2,925 SNPs were selected with an adjusted p-value <0.001. Based on principal component analysis, principal component 1 (PC1) segregated the *Va francensis*lineages from the rest of *Va* strain collection (FigureS8). 46 outlier SNPs were associated with PC1 including 17non-synonymousand onestop-gained SNP potentially involved in diversifying selection.Indeed, these mutations positively selected in this group could have led to an increase in genetic diversity at the level of *Va* species by favouring the oyster pathogen group emergence.These

genes are listed in Table S4and notably include genes encoding flagellar proteins (*pomA*) or related to chemiotaxis.

Oyster-pathogenic *Vibrio aestuarianus* have genomic features of specialist pathogens

According to low rates of homologous recombination and the closed pan-genome, the two *Va francensis*lineagescan be considered as clonally evolving lineages. Furthermore, clonal evolution, as well as the small number of known habitats (18),suggest that these lineagesare oyster specialists. Pathogen populations have often undergone a change of niche (passage from a free-living lifestyle to an obligatory parasitic one), ultimately resulting in ecological isolation and thus inhibition of DNA transfer with the parental population, or even with the broader bacterial community (53, 54).

We further tested the hypothesis of specialization by analyzing genomic features known to be enriched during transition to specialist lifestyle: insertion sequences(IS) and pseudogene(55). We found a higher number of IS in the genomes of pathogenic oyster lineages A and B than in strains isolated from contexts without bivalve mortality (Wilcoxon test,p-value = 0.01 and 0.02, respectively). Interestingly, in groups containing strains isolated during cockle mortality events, IS were more abundant than in pathogenic oyster genomes (Wilcoxon test, p-value= 0.02 and 0.03, respectively) (Figure3B). IS families were conserved between oyster pathogen genomes, with most IS belonging to the IS66, IS982 and ISAS1 families (Figure S9). IS conservation in theses clades suggests that IS enrichment and expansion in genomes stopped after pathogen emergence. The dynamics of the IS appeared to be different in *Vacardii*. Indeed, the number and families of IS represented appeared to be less conserved than in *Va francensis* genomes.We suspect that IS flux is active in these isolates, with ongoing horizontal exchange and intragenomic expansion.

To check whether IS enrichment is involved in pseudogenization, we crossed IS and pseudogene genomic coordinates and determined the number of IS with pseudogenes

around them (from 2,000 bp before the IS start to 2,000 bp after the IS end) (Tables S1and S5). Regardless of the genome and chromosome considered, the vast majority (on average 86%) of IS had pseudogenes nearby. IS and pseudogene enrichment supports our hypothesis that the *Va francensis* lineages are specialized toward oyster.

Copper resistance acquisition: One potential step in the emergence of oyster-pathogenic clades

Trait acquisition is often the first step preceding clonal expansion of host-adapted pathogens. To identify traits that may have preceded oyster pathogen emergence, we studied the*Va* pan-genome and detected 139 genes that were shared exclusively by *Va francensis*genomes, 8 *Va francensis* clade A specific genes and 111 *Va francensis* clade B specific genes. Most of them (106 of the subspecies specific genes, all the clade A specific genes and 88 of the clade B specific genes) were annotated as hypothetical proteins. Among the *Va francensis* subspecies specific genes, 33 encode proteins with functions related to copper export, cell-bacteria interaction (LPS/antigen O, cell wall or peptidoglycan), chemotaxis and DNA repair. Among the clade B specific genes, some of them encode proteins with functions related to recombination.The sequences of these subspecies or clades specific genes were identical or well conserved between strains (Table S6).

Some oyster pathogen specific genes localize to a putative mobile genetic element (MGE), as determined by crossing their coordinates with IS coordinates in an oyster pathogen genome (02_041) (Figure 5A). This putative MGE contains Cus operon genes (*cusABCF,S/R)* surrounded by IS from the IS66 family (length = 15 kb), as well as the copper export genes *copA*and *copG* and the regulator *cueR*, located near tn7 transposon genes (tnA, B, C and E) (length = 10 kb), and is later referred to as a *cus-cop*-containing island.

The *Va francensis*ancestor could have acquired copper export genes by horizontal transfer. Presence in an MGE could have facilitated integration of these genes into the bacterial genome. To investigate the origin of *cus-cop*-containing islands, we used BLAST analyses of each operon and of the entire region in 127 reference genomes of different *Vibrio*species

(Table S7and Figure S10). The *cus-cop*-containing island presence is scattered along the *Vibrio* phylogeny and is not specific to a monophyletic group. It appeared highly similar to some *V. fluvialis*and *V. furnissii*genomic sequences with a nucleotide identity >95%. Moreover, this region was close to plasmidic elements identified in *V. crassostreae* and *Vibrio* sp. dgh with a nucleotide identity >89% (Figure 5B).

To test whether the genomic differences between the oyster pathogens and otherstrains could lead to phenotypic differences, we tested copper resistance in 27 *Va*strains using an agar diffusion assay. Supporting our hypothesis, all tested oyster pathogens (8 strains from lineage A and 5 from lineage B) appeared significantly (Wilcoxon test,p-value < 0.001) more resistant than other strains to copper in radial diffusion assays (Figure 5C).

Discussion

Here, using a wide range of approaches, we describe a genomic analysis of *Vibrio aestuarianus*(*Va*) isolated from multiple European countries, in different epidemic contexts and from multiple environments (five host species, zooplankton and water) since 2001. We report the expansion and clonal evolution of two oyster-pathogenic lineages from the*Va francensis* subspecies(14).

Many pathogenic bacterial clones emerged from recombining background populations, leading to “epidemic" population structures,where clones, in the short term, resist to the homogenising effect of recombination(56).*Vibrio* species such as*V. parahaemolyticus* and *V. cholerae*have been showed to display such epidemic population structures(57-60).Here, we provide several evidencesupporting the hypothesis that*Va*harbors a similar specific population structure.

First, our results suggest clonal evolutionin *Va francensis*pathogenic lineages.Homologous recombination has low impact on genetic diversity. The r/m ratio, ~0.3 in both lineages, is

similar to clonally evolving pathogenic bacteria like the *V.cholerae* 7 pandemic lineage (r/m = 0.1 ; (61)),*M. tuberculosis* complex (between 0,426 and 0,565, (62)) and lineage 1 of *Salmonella enterica* (r/m = 0.2 ; (63)). On the contrary, homologous recombination impact appears more important in environmental isolates, which can belong to the recombining background population. However, the impact of recombination in the evolution of *Va* environmental strains appears to be lower than what has been described for *V. parahaemolyticus* where an r/m ratio of 39.8 was calculated (64), which shows that overall, the recombination plays a much smaller role in the evolution of the *Va*species.We also found that *Va francensis* has a closed pangenome whereas environmental isolates have a more open pangenome, suggesting that horizontal gene transfer do not occur in *Va francensis* populations whereas it affects environmental populations.The low impact of homologous recombination, compare to mutations, and the closed pangenome of *Va francensis* suggest that genetic exchanges with other microbial populations are scarce, which leads to the resistance of this population to the homogenizing effect of recombination.

The epidemic population structure can explain the observed monomorphism in *Va francensis* core genes since a particular clone probably emerged from one genotype of the background population.We propose that thissuccessful pathogenic clone isfavoured in a particular ecological niche, with oysters as a single habitat. Indeed, *Va francensis* appears ecologically isolated and could be anoyster-specialist pathogen that do not exchange genes with other bacterial populations. This hypothesis is also supported by enrichment of insertion sequences (IS) and pseudogenizationin both oyster-pathogenic lineages. During specialization, pathogens transition from a free-living lifestyle to a permanent association with a host. Through host association, pathogens can acquire compounds of intermediate metabolism that allow genes associated with related biosynthetic pathways to become facultative or redundant (51). Studies on pathogen specialization have largely focused on highly specialized intracellular bacteria, which can undergo an extreme genome rationalization process (52). This process implies pseudogenizationof non-essential genes

that accumulate mutations or IS and can ultimately result in gene loss (55, 65, 66). In our data, genome size reduction was not observed, probably because *Vafrancensis* pathogenic lineages are not obligate intracellular parasites. While they have a restricted habitat, they can still replicate in the absence of their host, at least in rich medium. Moreover, *Va francensis* can colonize diverse host tissues, including hemolymph, gills and digestive connective tissues (18). Consequently, *Va francensis*is adapted to diverse tissular micro-environments, consistent with being less specialized than obligate intracellular bacteria. Further studies comparing environmental *Va* and *Va francensis*capacities to colonize, growth or persist within oyster tissues may help to identify specific adaptative traits of this specialized pathogen. Moreover, new studies on oyster pathogen microhabitats may help to elucidate their complete intra-host life cycle, and the functions of specific genes linked to adaptation. Whether clades A and B occupy similar niches in infected oysters and whether they have the potential to cooperate remains unknown. Neither epidemiological parameter estimation (67) nor in-depth analysis of their genomes identified lineage-specific functional traits.

Based on the time-scaled phylogeneticreconstruction, we inferred that *Va francensis* emerged recently and more precisely during the second half of 20th century. Our hypothesis is that the environmental conditions became favorable to the selected genotype at the origin of the *Va francensis*subspecies due to the great increase of *C. gigas*oysters production in Europe (68).In addition, oyster culture is practiced intensively with high population densities and many transfers of live animals, which favours the introduction and spread of pathogens on a large geographical scale.

Unfortunately, because of the absence of isolates prior to the detection of mortalities in Europe, we cannot reconstruct the evolution and identified ancestral states connecting recent genomes with common ancestors that would allow to distinguish bottleneck from clonal emergence and expansion. To determine if a purge of genetic diversity in the ancestral population of *Va francensis* has occurred, or if this entire population is derived from

the expansion of an ancestral genotype, it would be interesting to sample *Va* in other regions to access to this ancestral state.Several genomic modifications may help pathogenic clone emergence and expansion (56, 64), including core genome substitutions and gene acquisition. We found non-synonymous substitutions that were fixed in pathogenic oyster lineages.These occurred ingenes involved in host interaction or virulence, such as the *pomA*gene which encodes a flagellum motor protein andthe gene *luxP*linked to quorum-sensing. The functional consequences of these substitutions remain to be explored.Acquisition of copper resistance traits may also have contributed to the emergence of the oyster-pathogenic lineages. Our data indicate that the*Vafrancensis* ancestor acquired a genomic island carrying copper detoxification genes. Indeed, copper is a heavy metal found at high levels in coastal waters because of agricultural and industrial activities(69-71). Copper can accumulate in oyster tissues (up to 18,000 µg/g in China(72)), particularly at infection sites (73)where it supportsantimicrobial defense as copper can kill invading microorganisms(73-75). Bacteria have, in turn, developed copper detoxification mechanisms which utilize the *cop* and *cus* genes(76). *Cop* genes support initial Cu^{+} translocation from the cytoplasm to the periplasm. *Cus* genes encode aCusABC system that exports periplasmic copper into the extracellular medium(77). Importantly, *copA* is essential for copper resistance, oyster colonization and virulence in *V. tasmaniensis* LGP32, an oyster-pathogenic strain associated withthe Pacific Oyster Mortality Syndrome(78). Similarly, *copA* deletion in*V. crassostreae*also attenuates virulence in oysters (79). Our data further support copper resistance as a shared trait thatallowspathogens to adapt to oysters as a host.

Beyond their emergence, the global spread of pathogens is a recurrent problem in agriculture and aquaculture and is often favored by anthropogenic introductions via infected hosts (80). The many transfers taking place during the lifetime of an oyster probably explain most of the spread of pathogenic lineages of *Va francensis* across Europe. Preventing the spread of pathogens during transfers is particularly complicated in the case of *Va francensis* since the animals may be asymptomatic (18). There is therefore an urgent need to develop

sensitive detection assays to monitor the spread of pathogens: the specific *Va francensis* genes revealed in this study, such as the cusS/R genomic island, may provide new diagnostic tools, which may help to monitor and prevent future mortality events.

Acknowledgments

We first thank the NRLs for mollusk diseases that contributed to the questionnaire survey (BojanAdžić, Raquel Aranguren, Giuseppe Arcangeli, Deborah Cheslett, Marc Engelsma, Mark Fordyce, Ana Grade, Michael Gubbins, Lone Madsen, PetyaOrozova, and EwaPaździor). We thank Nicolas Bierne for helpful discussion and Viviane Boulo, Jean-Michel Escoubas, Nicole Faury and GaelleCourtay for their valuable technical assistance. We also thank Bruno Petton and staff of IFREMER Bouin, who contributed to animal production and delivery. We thank the Bio-Environment platform (University of Perpignan Via Domitia) and Jean-François Allienne for support in library preparation and sequencing. We are grateful to the Ifremer Bioinformatics Core Facility (SeBiMER;https://ifremer-bioinformatics.github.io/) for providing technical assistance and scientific support in bioinformatics analysis. We also acknowledge the Pôle de Calculet de Données Marines (https://wwz.ifremer.fr/en/Research-Technology/Research-Infrastructures/Digital-infrastructures/Computation-Centre) for providing DATARMOR computing and storage resources. This work was funded by Europe DG-Sante (through European Union Reference Laboratory for mollusk diseases funding) and French Ministry DPMA (MIRAGE project) and DGAl through NRL funding. Moreover, A.M., PhD, was supported by IFREMER and Région Nouvelle Aquitaine. This study is part of the « Laboratoired'Excellence (LabEx) » TULIP (ANR-10-LABX-41) framework.

Bibliographic references

1. Charlesworth D, Charlesworth B.1995. Quantitative Genetics in Plants: The Effect of the Breeding System on Genetic Variability. Evolution 49:911-920.
2. Tibayrenc M, Ayala FJ.2012. Reproductive clonality of pathogens: A perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. Proceedings of the National Academy of Sciences 109:E3305-E3313.
3. Anderson RM, Spratt BG, Maiden MCJ.1999. Bacterial population genetics, evolution and epidemiology. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 354:701-710.
4. Levin BR.1981. Periodic selection, infectious gene exchange and the genetic strucutre of E. coli populations. Genetics 99:1-23.
5. Cohan FM.2002. Sexual isolation and speciation in bacteria. Genetics of Mate Choice: From Sexual Selection to Sexual Isolation:359-370.
6. Polz MF, Alm EJ, Hanage WP.2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet 29:170-5.
7. Smith JM, Haigh J.2009. The hitch-hiking effect of a favourable gene. Genetical Research 23:23-35.
8. Charlesworth D, Charlesworth B, Morgan MT.1995. The pattern of neutral molecular variation under the background selection model. Genetics 141:1619-32.
9. Sheppard SK, Guttman DS, Fitzgerald JR.2018. Population genomics of bacterial host adaptation. Nature Reviews Genetics 19:549-565.
10. Martínez JL.2013. Bacterial pathogens: from natural ecosystems to human hosts. Environmental Microbiology 15:325-333.
11. Davis BM, Waldor MK.2003. Filamentous phages linked to virulence of Vibrio cholerae. Current Opinion in Microbiology 6:35-42.
12. Lafferty KD, Harvell CD, Conrad JM, Friedman CS, Kent ML, Kuris AM, Powell EN, Rondeau D, Saksida SM.2015. Infectious Diseases Affect Marine Fisheries and Aquaculture Economics. Annual Review of Marine Science 7:471-496.
13. Murray AG, Peeler EJ.2005. A framework for understanding the potential for emerging diseases in aquaculture. Preventive Veterinary Medicine 67:223-235.
14. Garnier M, Labreuche Y, Nicolas JL.2008. Molecular and phenotypic characterization of *Vibrio aestuarianus* subsp. *francensis* subsp. nov., a pathogen of the oyster *Crassostrea gigas*. Syst Appl Microbiol 31:358-65.
15. Lasa A, di Cesare A, Tassistro G, Borello A, Gualdi S, Furones D, Carrasco N, Cheslett D, Brechon A, Paillard C, Bidault A, Pernet F, Canesi L, Edomi P, Pallavicini A, Pruzzo C, Vezzulli L.2019. Dynamics of the Pacific oyster pathobiota during mortality episodes in Europe assessed by 16S rRNA gene profiling and a new target enrichment next-generation sequencing strategy. Environ Microbiol 21:4548-4562.
16. Mandas D, Salati F, Polinas M, Sanna MA, Zobba R, Burrai GP, Alberti A, Antuofermo E.2020. Histopathological and Molecular Study of Pacific Oyster Tissues Provides Insights into *V. aestuarianus* Infection Related to Oyster Mortality. Pathogens 9.
17. Dégremont L, Azéma P, Maurouard E, Travers M-A.2020. Enhancing resistance to *Vibrio aestuarianus* in *Crassostrea gigas* by selection. Aquaculture 526:735429.
18. Parizadeh L, Tourbiez D, Garcia C, Haffner P, Dégremont L, Le Roux F, Travers M-A.2018. Ecologically realistic model of infection for exploring the host damage caused by *Vibrio aestuarianus*. Environmental Microbiology 20:4343-4355.
19. Tison DL, Seidler RJ.1983. *Vibrio aestuarianus*: a New Species from Estuarine Waters and Shellfish†. International Journal of Systematic and Evolutionary Microbiology 33:699-702.
20. Goudenège D, Travers MA, Lemire A, Petton B, Haffner P, Labreuche Y, Tourbiez D, Mangenot S, Calteau A, Mazel D, Nicolas JL, Jacq A, Le roux F.2015. A single

regulatory gene is sufficient to alter *Vibrio aestuarianus* pathogenicity in oysters. Environ Microbiol 17:4189-99.
21. Wingett SW, Andrews S.2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res 7:1338.
22. Bush SJ.2020. Read trimming has minimal effect on bacterial SNP-calling accuracy. Microbial Genomics 6.
23. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology 19:455-477.
24. Gurevich A, Saveliev V, Vyahhi N, Tesler G.2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-5.
25. Kolmogorov M, Yuan J, Lin Y, Pevzner PA.2019. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology 37:540-546.
26. Vaser R, Sovic I, Nagarajan N, Sikic M.2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Research doi:10.1101/gr.214270.116.
27. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM.2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212.
28. Seemann T.2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.
29. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J.2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38:5825-5829.
30. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen Lars J, von Mering C, Bork P.2018. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research 47:D309-D314.
31. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J.2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691-3693.
32. Tettelin H, Riley D, Cattuto C, Medini D.2008. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 11:472-7.
33. Xie Z, Tang H.2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics 33:3340-3347.
34. Syberg-Olsen MJ, Garber AI, Keeling PJ, McCutcheon JP, Husnik F.2021. Pseudofinder: detection of pseudogenes in prokaryotic genomes. bioRxiv doi:10.1101/2021.10.07.463580:2021.10.07.463580.
35. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM.2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81-91.
36. Darling AE, Mau B, Perna NT.2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLOS ONE 5:e11147.
37. Falush D, Stephens M, Pritchard JK.2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics 164:1567-1587.
38. Didelot X, Wilson DJ.2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. PLOS Computational Biology 11:e1004041.
39. Rambaut A, Lam TT, Max Carvalho L, Pybus OG.2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution 2.

40. Minin VN, Bloomquist EW, Suchard MA.2008. Smooth Skyride through a Rough Skyline: Bayesian Coalescent-Based Inference of Population Dynamics. Molecular Biology and Evolution 25:1459-1471.
41. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W.2002. Estimating Mutation Parameters, Population History and Genealogy Simultaneously From Temporally Spaced Sequence Data. Genetics 161:1307-1320.
42. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA.2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology 67:901-904.
43. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV.2012. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. Molecular Biology and Evolution 29:2157-2167.
44. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P.2012. Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics. Molecular Biology and Evolution 30:239-243.
45. Subramanian S.2016. The effects of sample size on population genomic analyses – implications for the tests of neutrality. BMC Genomics 17:123.
46. Tajima F.1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-95.
47. Paradis E.2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26:419-420.
48. Petton B, Pernet F, Robert R, Boudry P.2013. Temperature influence on pathogen transmission and subsequent mortalities in juvenile Pacific oysters *Crassostrea gigas*. Aquaculture Environment Interactions 3:257-273.
49. Saulnier D, De Decker S, Haffner P.2009. Real-time PCR assay for rapid detection and quantification of Vibrio aestuarianus in oyster and seawater: A useful tool for epidemiologic studies. Journal of Microbiological Methods 77:191-197.
50. Garnier M, Labreuche Y, Garcia C, Robert M, Nicolas JL.2007. Evidence for the involvement of pathogenic bacteria in summer mortalities of the Pacific oyster Crassostrea gigas. Microb Ecol 53:187-96.
51. Moran NA.2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108:583-6.
52. Floriano AM, Castelli M, Krenek S, Berendonk TU, Bazzocchi C, Petroni G, Sassera D.2018. The Genome Sequence of "Candidatus *Fokinia solitaria*": Insights on Reductive Evolution in Rickettsiales. Genome Biol Evol 10:1120-1126.
53. Hauck S, Maiden MCJ. 2018. Clonally Evolving Pathogenic Bacteria, p 307-325. *In* Rampelotto PH (ed), Molecular Mechanisms of Microbial Evolution doi:10.1007/978-3-319-69078-0_12. Springer International Publishing, Cham.
54. Bergstrom CT, McElhany P, Real LA.1999. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. Proceedings of the National Academy of Sciences 96:5095-5100.
55. Bäumler A, Fang FC.2013. Host Specificity of Bacterial Pathogens. Cold Spring Harbor Perspectives in Medicine 3.
56. Smith JM, Feil EJ, Smith NH.2000. Population structure and evolutionary dynamics of pathogenic bacteria. BioEssays 22:1115-1122.
57. Bhuiyan NA, Ansaruzzaman M, Kamruzzaman M, Alam K, Chowdhury NR, Nishibuchi M, Faruque SM, Sack DA, Takeda Y, Nair GB.2002. Prevalence of the Pandemic Genotype of Vibrio parahaemolyticus in Dhaka, Bangladesh, and Significance of Its Distribution across Different Serotypes. Journal of Clinical Microbiology 40:284-286.
58. González-Escalona N, Martinez-Urtaza J, Romero J, Espejo RT, Jaykus L-A, DePaola A.2008. Determination of Molecular Phylogenetics of Vibrio parahaemolyticus Strains by Multilocus Sequence Typing. Journal of Bacteriology 190:2831-2840.

59. Moore S, Thomson N, Mutreja A, Piarroux R.2014. Widespread epidemic cholera caused by a restricted subset of Vibrio cholerae clones. Clinical Microbiology and Infection 20:373-379.
60. Salim A, Lan R, Reeves PR.2005. Vibrio cholerae pathogenic clones. Emerg Infect Dis 11:1758-60.
61. Didelot X, Pang B, Zhou Z, McCann A, Ni P, Li D, Achtman M, Kan B.2015. The Role of China in the Global Spread of the Current Cholera Pandemic. PLOS Genetics 11:e1005072.
62. Namouchi A, Didelot X, Schöck U, Gicquel B, Rocha EPC.2012. After the bottleneck: Genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. Genome Research 22:721-734.
63. Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D, Donnelly P.2011. Recombination and Population Structure in Salmonella enterica. PLOS Genetics 7:e1002191.
64. Martinez-Urtaza J, Aerle Rv, Abanto M, Haendiges J, Myers RA, Trinanes J, Baker-Austin C, Gonzalez-Escalona N.2017. Genomic Variation and Evolution of Vibrio parahaemolyticus ST36 over the Course of a Transcontinental Epidemic Expansion. mBio 8:e01425-17.
65. Siguier P, Gourbeyre E, Chandler M.2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiology Reviews 38:865-891.
66. Touchon M, Rocha EPC.2007. Causes of Insertion Sequences Abundance in Prokaryotic Genomes. Molecular Biology and Evolution 24:969-981.
67. Travers MA, Tourbiez D, Parizadeh L, Haffner P, Kozic-Djellouli A, Aboubaker M, Koken M, Degremont L, Lupo C.2017. Several strains, one disease: experimental investigation of *Vibrio aestuarianus* infection parameters in the Pacific oyster, *Crassostrea gigas*. Vet Res 48:32.
68. FranceAgriMer.2021. https://www.franceagrimer.fr/. https://www.franceagrimer.fr/. Accessed
69. Claisse D, Alzieu C.1993. Copper contamination as a result of antifouling paint regulations? Marine Pollution Bulletin 26:395-397.
70. Hall LW, Anderson RD.1999. A Deterministic Ecological Risk Assessment for Copper in European Saltwater Environments. Marine Pollution Bulletin 38:207-218.
71. Rossi N, Jamet J-L.2008. *In situ* heavy metals (copper, lead and cadmium) in different plankton compartments and suspended particulate matter in two coupled Mediterranean coastal ecosystems (Toulon Bay, France). Marine Pollution Bulletin 56:1862-1870.
72. Weng N, Wang W-X.2014. Variations of trace metals in two estuarine environments with contrasting pollution histories. Science of The Total Environment 485-486:604-614.
73. Shi B, Wang T, Zeng Z, Zhou L, You W, Ke C.2019. The role of copper and zinc accumulation in defense against bacterial pathogen in the fujian oyster (*Crassostrea angulata*). Fish Shellfish Immunol 92:72-82.
74. Destoumieux-Garzón D, Canesi L, Oyanedel D, Travers M-A, Charrière GM, Pruzzo C, Vezzulli L.2020. Vibrio–bivalve interactions in health and disease. Environmental Microbiology 22:4323-4341.
75. Yeh S-T, Liu C-H, Chen J-C.2004. Effect of copper sulfate on the immune response and susceptibility to *Vibrio alginolyticus* in the white shrimp *Litopenaeus vannamei*. Fish & Shellfish Immunology 17:437-446.
76. Bondarczuk K, Piotrowska-Seget Z.2013. Molecular basis of active copper resistance mechanisms in Gram-negative bacteria. Cell Biol Toxicol 29:397-405.
77. Argüello JM, Raimunda D, Padilla-Benavides T.2013. Mechanisms of copper homeostasis in bacteria. Front Cell Infect Microbiol 3:73.
78. Vanhove AS, Duperthuy M, Charriere GM, Le Roux F, Goudenege D, Gourbal B, Kieffer-Jaquinod S, Coute Y, Wai SN, Destoumieux-Garzon D.2015. Outer

membrane vesicles are vehicles for the delivery of Vibrio tasmaniensis virulence factors to oyster immune cells. Environ Microbiol 17:1152-65.
79. Bruto M, James A, Petton B, Labreuche Y, Chenivesse S, Alunno-Bruscia M, Polz MF, Le Roux F.2017. *Vibrio crassostreae,* a benign oyster colonizer turned into a pathogen after plasmid acquisition. The ISME Journal 11:1043-1052.
80. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P.2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends in Ecology & Evolution 19:535-544.

Data accessibility and Benefit-sharing

Genome assemblies are deposited in Genbank under the following accession numbers JABANJ000000000.1, JABANK000000000.1, JAAZTV000000000.1, JAKNBK000000000, JAAZTU010000000, JAKNBJ000000000, JABANH000000000.1, CP091995-CP091996, JAKNBI000000000, CP091993-CP091994, JAKNBH000000000, JAKNBG000000000, JAKNYS000000000, JAKOGL000000000, JAAKZN000000000.1, JAKNAO000000000, JAKOAX000000000, JAAKZK000000000.1, JAKOAW000000000, JAAKZL000000000.1, JAKNAP000000000, JAKNBF000000000, JAKNBE000000000, JAKOGK000000000, JAKOGJ000000000, JAKNBD000000000, JAKNBC000000000, JAKNBB000000000, JAKNBA000000000, JAKNAZ000000000, JAKNAY000000000, JAKNAX000000000, JAKNAW000000000, JAKNAV000000000, JAKNAU000000000, JAKNAT000000000, JAKNAS000000000, JAKNAR000000000, JAKNAQ000000000, JAKNAN000000000, JAKNAM000000000, JAKNAL000000000, JAKNAK000000000, JAKNAJ000000000, JAKNAI000000000, JAKNAH000000000, JAKNAG000000000, JAKNAF000000000, JAKNYR000000000, JAKNYQ000000000, JAKNYP000000000, JAKNYO000000000
Benefits Generated: Benefits from this research accrue from the sharing of our data and results on public databases as described above

Authors contribution

AM, MJ, GC, DDG and MAT designed research, AM, MJ, CG, DT, LC, AB, LD, DC, MG, AV, AR, DF, AG, PO performed research, AM, MJ, CG, LC, IA, MS, GC, DDG and MAT analyzed data and AM, MJ, GC, DDG and MAT wrote the paper

Tables and figures

Table 1: **Analysis of molecular variance** supporting genetic partitioning according to environmental origin and isolation context

Factor	*Variability (%)*	*p-value*	*Phi stat*
Environmentalorigin	**17.8**	**0.002**	**0.178**
Isolation context	**16.0**	**0.001**	**0.160**
Country of isolation	7.1	0.074	0.071
Year of isolation	3.4	0.25	0.034

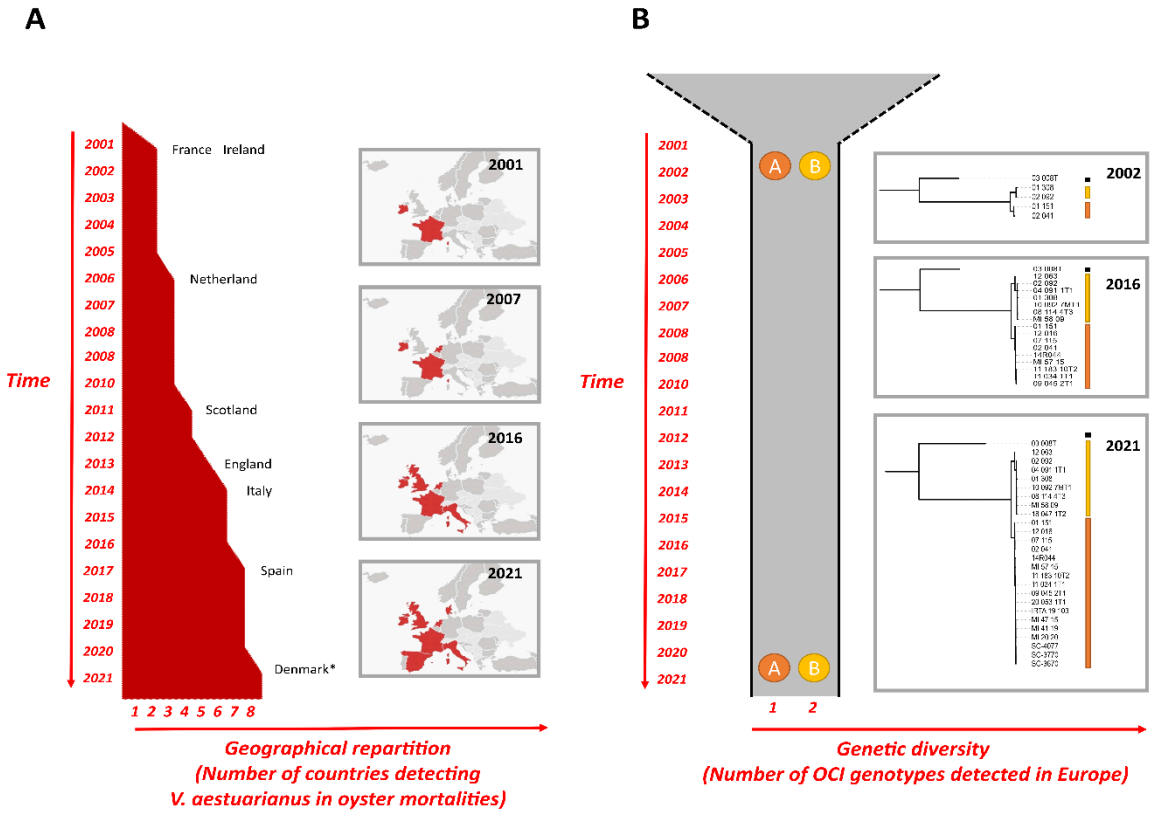


Figure 1: **European expansion of *Vibrioaestuarianus francensis***. While the geographic distribution of *Vibrioaestuarianus* strains involved in *Crassostreagigas* mortalities expands, genetic diversity remains stable over time. **A) The number of countries reporting the detection of *Vibrio aestuarianus***was assessed through a questionnaire sent to European national reference laboratories (NRLs). *Note: Denmark case report corresponds to a suspected case without bacterial isolation. **B) The two lineage expansions**: twogenotypes were determined based on a maximum likelihood phylogenetic tree from complete genome alignments (GTR substitution model).

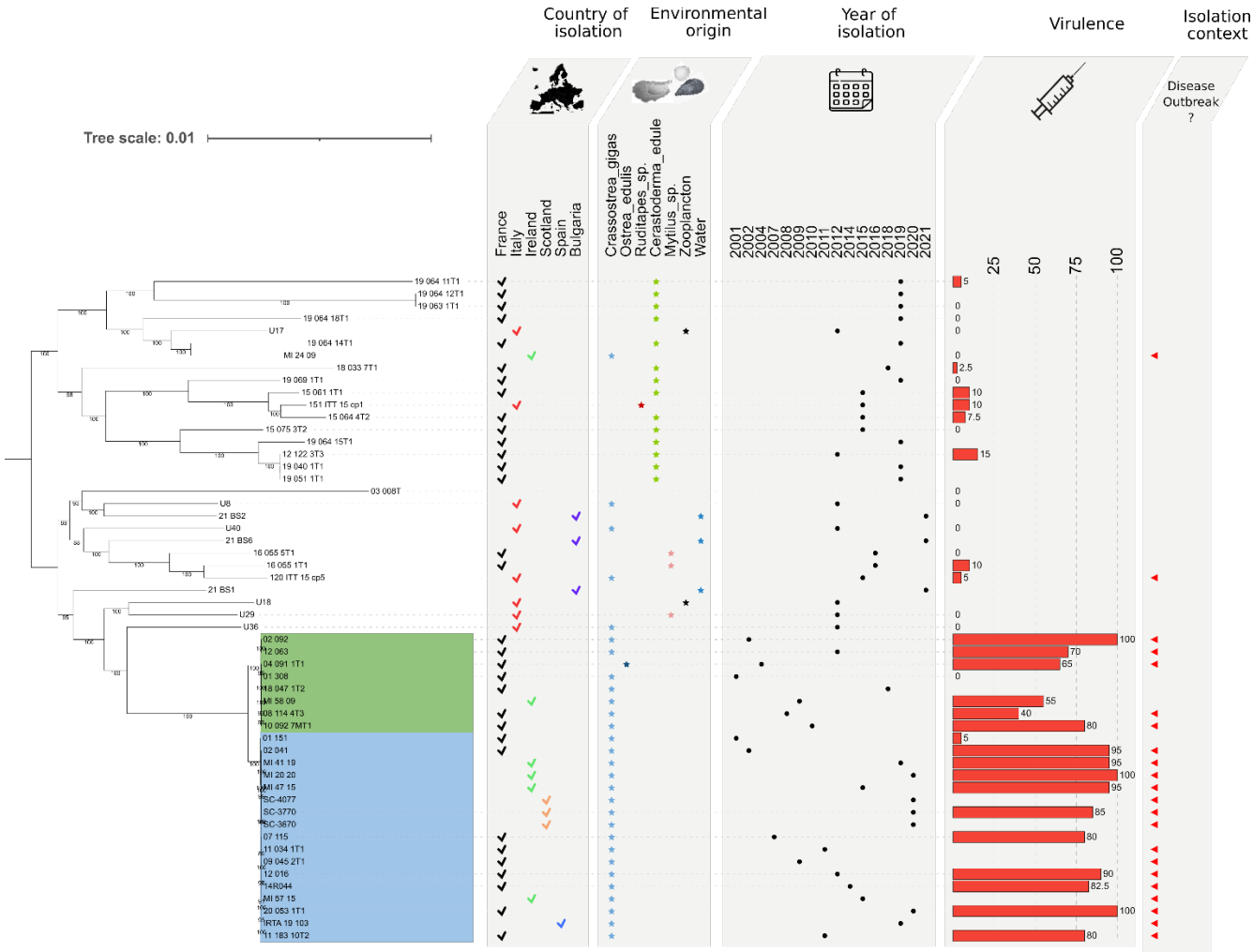


Figure 2: ***Vibrioaestuarianus* diversity is structured according to host species and isolation context.** Phylogenetic tree of *Vibrioaestuarianus*(*Va*) species calculated with RAxML using a GTR substitution model and rooted with the midpoint rooting option from whole-genome alignments. Several layers of annotations are added: 'Country of isolation' (check mark), 'Environmental origin' (star), 'Year of isolation' (point), 'Virulence' test in the laboratory by injection into oysters (10^7 UFC/animal, duplicates of 10 animals) indicated with a bar plot on the right side representing mortality percentages, and 'Isolation context' (strains isolated during oyster disease outbreaks are indicated with a red arrow on the right).*Va francensis* lineages A and B are highlighted in blue and green, respectively. Their virulence is significantly higher compared to non-clinical isolates (*i.e.* strains isolated on sites without mortalities) by Wilcoxon-Mann-Whitney test (*p*-value < 0.001). The absence of virulence data means the absence of a test for this strain.

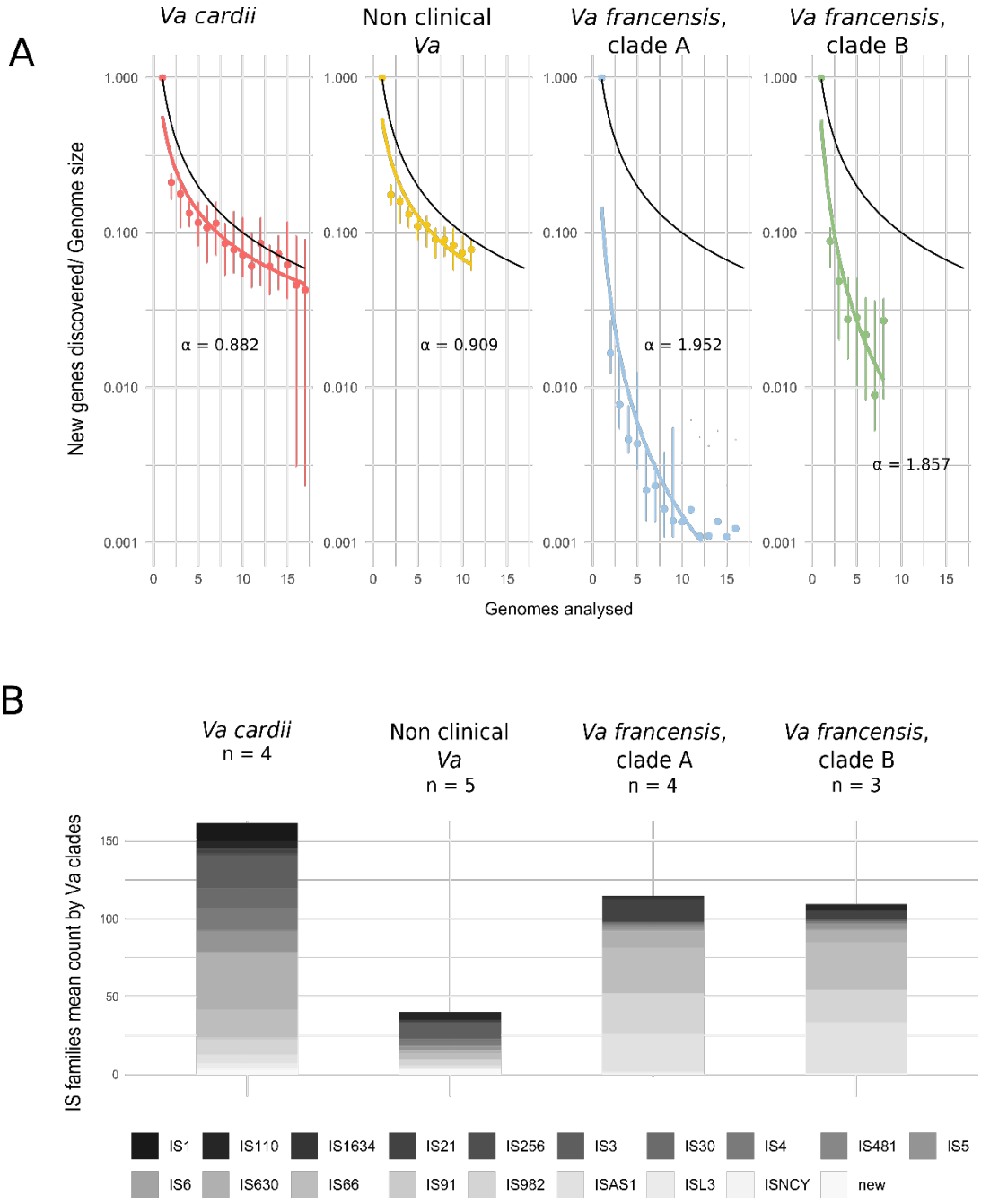


Figure 3: ***Va francensis*pangenomes are closed and enriched in Insertion Sequences (IS).A)Pan genomes of *Vibrio aestuarianus francensis* lineages are closed**. Power law regression for *Va* groups pangenomes. The medians of the number n of new genes discovered for increasing values of the number N of genomes analysed, normalized to the average genome size of the species, are displayed along with their 25–75 percentile intervals. In each box a black guide to the eye shows the borderline power law n~N-α with α=1, to facilitate the comparison of slopes. α values for *Vacardii* and non-clinical *Va* pangenome around 1 denote that pan-genome size follows a logarithmic trend, that is, grows very slowly but it is still technically unbounded. α values for the two francensis clades, which are higher than 1, denote that pangenome are closed.**B)Differentialinsertion sequence enrichment is observed in *Vibrio aestuarianus francensis*and cockle associated isolates genomes.** Distribution of the number of Insertion Sequence (IS) elements (complete and partial) in 16 complete genomes (chromosomes A and B). Grey shades

represent IS families predicted by ISEScan. *Vibrio aestuarianus* (*Va*) genomes are those of 4 strains isolated during cockle mortality events (*Vacardii*), 5 non-clinical strains (isolated on sites without mortalities), 4 *Va francensis* lineage A strains, 3 *Va francensis* lineage B strains. The conservation of IS number and families between genomes of *Va francensis* from lineage A or clade B indicates an ancestral origin of IS. The genomes that contain the most IS are those of strains isolated during cockle mortality events (*Vacardii*). The diversity of IS quantities and families represented suggests that IS enrichment is in progress. The number of IS in groups was significantly different between *Va francensis* and other strains from *Vacardii* and Non-clinical *Va* (**: Kruskal test p value < 0.05).

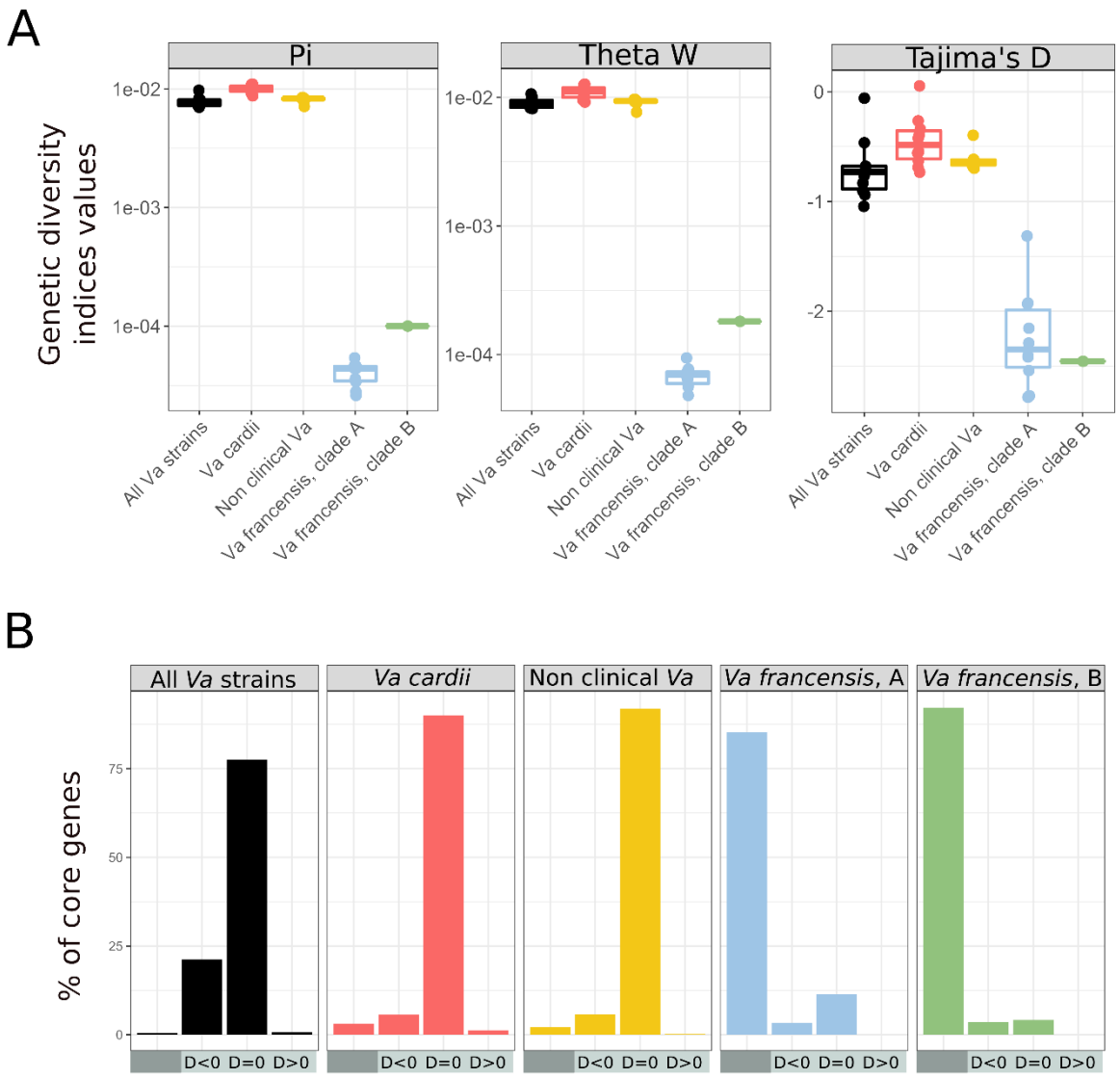
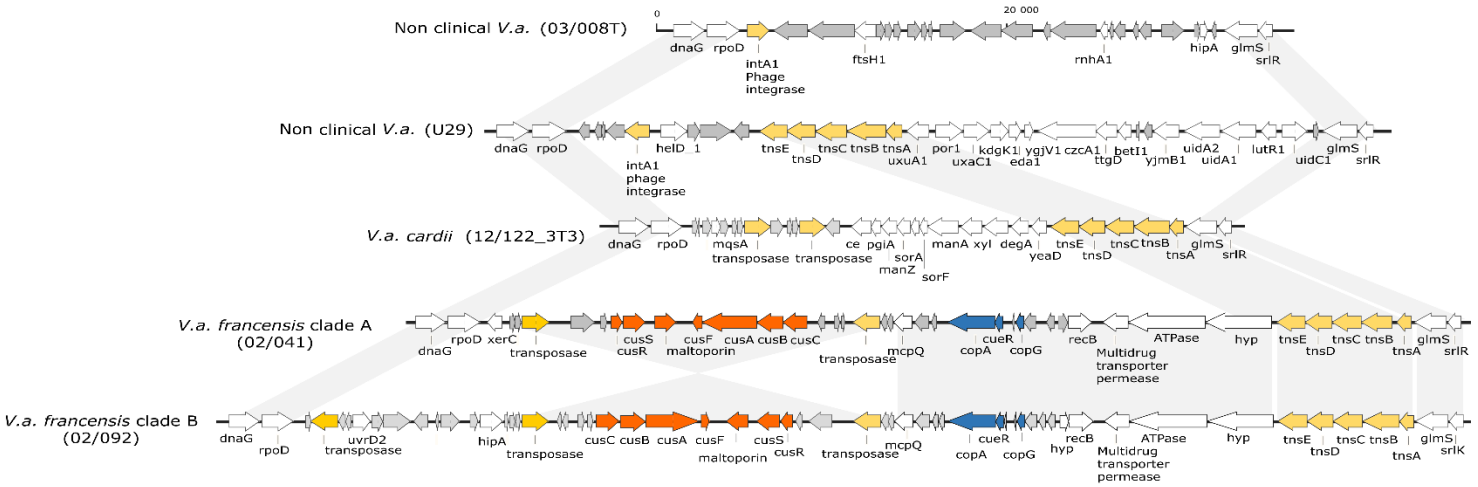
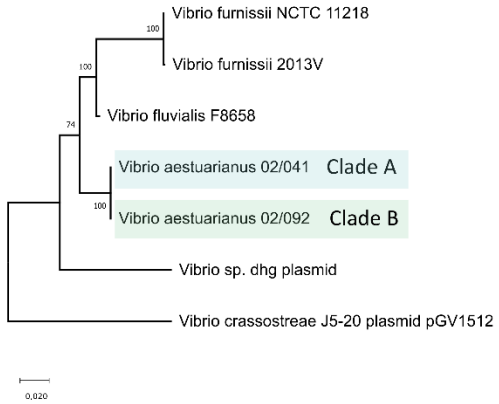


Figure 4: **Support for a recent clonal expansionof*Vibrio aestuarianus francensis* lineages.** Genetic statistics in all*Vibrio aestuarianus*(*Va*)strains (black, n=8), *Vacardii (red, n=8),* Non clinical*Va*(yellow, n=8), *Va francensis*, clade A (blue, n=8) and clade B (green, n=8).**A) Whole genome scale.**Expected genetic diversity under neutral evolution and a stable population size in *Va francensis* is low (Watterson's theta), but the observed diversity (Pi) is still lower, as shown by the Tajima's D value below -2.Pi: observed genetic diversity; Theta W: Watterson's Theta estimate of expected genetic diversity. The Y axis in the Pi and Watterson's theta plots is log10 scaled.**B) Gene scale.** Core genes show different selection patterns among groups of strains and among core genes within groups. While most core genes are under neutral evolution in isolates other than *Va francensis*, most genes are monomorphic in *Va francensis* lineage A and B. D>0 and D<0: significantly negative or positive Tajima's D value ($p < 0.05$) and D=0 ($p > 0.05$). A negative Tajima's D is indicative of purifying selection or recent population expansion, while a positive Tajima's D is indicative of diversifying selection or population contraction. A zero value is indicative of neutral evolution.

A *V. a. francensis* specific locus containing copper resistance genes



B Evolutionary history of cus/cop locus



C Copper resistance phenotyping

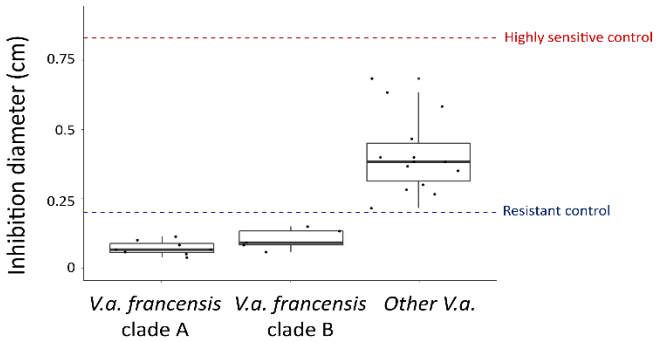


Figure 5: A) *Vibrio aestuarianus francensis*specific locus containing copper resistance genes. Transposase and integrase are indicated in yellow, cus genes in orange, and cop genes in blue. **B) Evolutionary history of the cus/cop locus.** NJ, bootstrap test (500 replicates). The evolutionary distances were computed using the Kimura 2-parameter method and are shown as the number of base substitutions per site. This analysis involved 8 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 29 555 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [4]. **C) Agar plate assay**: bacterial growth inhibition diameter around a pellet impregnated with 50 mM $CuSO_4$ solution. *Vibrio aestuarianus francensis* linages A and B are significantly more resistant to copper that other *Vibrioaestuarianus* strains (Wilcoxon-Mann-Whitney test p-value < 0.001).