The ISME Journal
July 2015, Volume 9, Issue 7, Pages 1523-1531
http://dx.doi.org/10.1038/ismej.2014.233
http://archimer.ifremer.fr/doc/00244/35565/
© 2014 International Society for Microbial Ecology All rights reserved

Archimer
http://archimer.ifremer.fr

Populations, not clones, are the unit of vibrio pathogenesis in naturally infected oysters

Lemire Astrid [1, 2, 3], Goudenege David [1, 2, 3], Versigny Tiphaine [1, 2, 3], Petton Bruno [2], Calteau Alexandra [4, 5], Labreuche Yannick [1, 2, 3], Le Roux Frederique [1, 2, 3, *]

[1] Sorbonne Universités, UPMC Univ Paris 06, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, Roscoff, France
[2] Ifremer, Unité Physiologie Fonctionnelle des Organismes Marins, ZI de la Pointe du Diable, Plouzané, France
[3] CNRS UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, Roscoff, France
[4] Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Direction des Sciences du Vivant (DSV), Institut de Génomique (IG), Genoscope, Evry, France
[5] CNRS UMR 8030, Laboratoire d'Analyses Bioinformatiques en Génomique et Métabolisme (LABGeM), Evry, France

* Corresponding author : Frédérique Le Roux, email address : frederique.le-roux@sb-roscoff.fr

Abstract :

Disease in oysters has been steadily rising over the past decade, threatening the long-term survival of commercial and natural stocks. Our understanding and management of such diseases are of critical importance as aquaculture is an important aspect of dealing with the approaching worldwide food shortage. Although some bacteria of the Vibrio genus isolated from diseased oysters have been demonstrated to be pathogenic by experimental infection, direct causality has not been established. Little is known about the dynamics of how the bacterial population hosted by oysters changes during disease progression. Combining experimental ecology, a high-throughput infection assay and genome sequencing, we show that the onset of disease in oysters is associated with progressive replacement of diverse benign colonizers by members of a phylogenetically coherent virulent population. Although the virulent population is genetically diverse, all members of that population can cause disease. Comparative genomics across virulent and nonvirulent populations identified candidate virulence factors that were clustered in population-specific genomic regions. Genetic analyses revealed that one gene for a candidate virulent factor, a putative outer membrane protein, is necessary for infection of oysters. Finally, analyses of oyster mortality following experimental infection suggest that disease onset can be facilitated by the presence of nonvirulent strains. This is a new form of polymicrobial disease, in which nonpathogenic strains contribute to increase mortality.

Please note that this is an author-produced PDF of an article accepted for publication following peer review. The definitive publisher-authenticated version is available on the publisher Web site.

44 **INTRODUCTION**

45

46 Vibrios have been associated with successive mortality outbreaks of oyster beds (*Crassostrea gigas*) in
47 France that have resulted in losses up to 100% of production (Samain 2008). Given the near monoculture
48 of *C. gigas* in Europe, there is an urgent need to understand the epidemiology of these outbreaks,
49 particularly the role of vibrios in the diseases. To date, it has been difficult to determine whether vibrios
50 resident in diseased oysters are mutualistic, opportunistic or pathogenic, due to a lack of diagnostic tools
51 for distinguishing pathogenic from non-pathogenic strains, the fact that individual animals can harbor
52 multiple bacterial genotypes, and limitations inherent within the experimental systems available for
53 studying infection. Consequently, it has been difficult to conclusively identify bacterial genotypes or
54 genes that are linked to virulence in oysters.

55

56 In the last few years, significant progress has been made in understanding the population structure and
57 diversity of vibrios (Hunt et al 2008, Thompson et al 2005). Despite their enormous microdiversity, these
58 organisms fall into well-defined genetic clusters that have similar resource preferences. These clusters
59 have been hypothesized to correspond to populations that act as cohesive ecological units, i.e., ecological
60 populations (Hunt et al 2008). However, a link between ecological populations and pathogenicity has not
61 been demonstrated, and it is unclear whether pathogenicity is a trait primarily linked to clones, or to
62 populations comprising a large number of distinct genotypes.

63

64 Experimental infections of oysters, which has been performed for a limited number of bacterial strains,
65 have allowed identification of a few factors that contribute to virulence, namely a metalloprotease

66 (Labreuche et al 2010, Le Roux et al 2007) and the outer membrane protein OmpU (Duperthuy et al ,
67 Duperthuy et al 2011). However, knowledge of the absence/presence of these genes is not sufficient for
68 determination of a strain's pathogenicity (Saulnier et al 2010). Furthermore, the laboratory analyses
69 previously used to define virulence-linked loci do not capture the complexity of infection within the
70 natural environment. Oysters are typically injected with a single bacterial strain, whereas in their natural
71 environment animals are typically colonized by a diverse assemblage of vibrios (Gay et al 2004a,
72 Wendling et al 2014). This diversity may contribute to virulence; in fact, experimental infections have
73 demonstrated that some strains are moderately virulent when injected into animals individually, and
74 display heightened virulence in mixed experimental infections (Gay et al 2004b).
75
76 Oyster vibrioses studied in the laboratory also do not accurately model the natural infection process.
77 Numerous attempts to kill oysters by immersion in vibrio-contaminated seawater have proven
78 unsuccessful, and have necessitated reliance on infection via injection (Duperthuy et al 2011, Gay et al
79 2004a, Le Roux et al 2007). It is possible that infection in the wild is aided by vibrios' association with
80 and attachment to other organisms and particles; a recent study has reported that marine aggregates
81 facilitate retention of nanoparticles (including bacteria) by suspension-feeding bivalves (Froelich et al
82 2012). Thus, growth of oysters in an environment in which bacteria are not simply in a planktonic form
83 may yield a more accurate understanding of the factors that contribute to virulence.
84
85 In the present study, we investigate the oyster disease ecology of microdiverse *Vibrio* genotypes using a
86 new, field-based approach. We take advantage of recently developed specific-pathogen-free (SPF) spats
87 of *C. gigas,* which become naturally infected when placed in an oceanic environment (Petton et al 2013).
88 In addition, we use these standardized animals for high-throughput experimental infections. We show that
89 pathogenicity can be ascribed to a cluster of genetically related strains that coincides with a previously

90 defined ecologically cohesive population. Genes specific to this population likely reflect the selective
91 pressure associated with population specialization, and we demonstrate that one of them is required for
92 pathogenicity.
93
94
95 **MATERIALS AND METHODS**
96
97 **Strains, plasmids collections and culture conditions.** In May 2011, SPF oyster spats were transferred
98 to a farming area to allow infection (supplementary methods). An oyster set (designed as sentinel) was
99 maintained in the field to monitor the first mortality onset and determine the cumulative mortality rates
100 occurring naturally after one month. At the first mortality report, infected animals were reintroduced in
101 the laboratory to reveal the disease. Each day, from a pool of 10 living oysters, vibrios were isolated on
102 selective media (Thiosulfate-citrate-bile salts-sucrose agar, TCBS, Difco) and re-streaked two times
103 before genotyping using *gyrB* partial sequence (supplementary methods). The strains used for the
104 genomic analyses are described in Table 1. Other bacterial strains are described in Table S1. Vibrio
105 isolates were grown in Zobell or Zobell agar, Luria-Bertani (LB) or LB-agar (LBA) + NaCl 0.5M, at
106 20°C. *E. coli* strains were grown in LB or on LBA at 37°C. Chloramphenicol (12µg/ml), spectinomycin
107 (100µg/ml), thymidine (0.3 mM) and diaminopimelate (0.3 mM) were added as supplements when
108 necessary. Induction of the P_{BAD} promoter was achieved by the addition of 0.2% L-arabinose to the
109 growth media, and conversely, was repressed by the addition of 1% D-glucose.
110
111 **Genome sequencing, assembly and annotation.** Thirty-four strains (Table 1) were sequenced using the
112 Illumina HiSeq2000 technology with ~50-fold coverage (supplementary methods). Contigs were
113 assembled *de novo* using Velvet (Zerbino and Birney 2008) and genome assembly was improved by

114 contig mapping against the LGP32 reference genome (Le Roux et al 2009). Computational prediction of
115 coding sequences together with functional assignments were performed using the automated annotation
116 pipeline implemented in the MicroScope platform (supplementary methods) (Vallenet et al 2013).
117
118 ***In silico* analyses.** A dedicated precomputing repository (marshalling) was created to perform
119 comparative genomic and phylogenomic analyses. Orthologous proteins were defined as reciprocal best
120 hit proteins with 80% MaxLrap and a minimum of 60% identity cutoff (Daubin et al 2002). The nucleic
121 acid sequences were aligned using Muscle (Edgar 2004) and filtered by BMGE (Criscuolo and Gribaldo
122 2010). Phylogenetic trees were built using the parallel version of PhyML applied to Maximum-likelihood
123 algorithm and GTR model as parameters (Guindon et al 2010). A first phylogenetic analysis of
124 concatenated nucleic acid sequences derived from 3229 shared proteins from the 34 genome sequences
125 suggested the clonality of some isolates within our collection (Table 1). This was confirmed by an ANI
126 value >99,5% and an accessory genome <150 CDS between isolates. Consequently only 21/34 isolates
127 were considered as distinct strains.
128
129 **Vector construction and mutagenesis.** Alleles carrying an internal deletion were cloned into a suicide
130 vector using the Gibson method (New England Biolabs) (supplementary methods). The R6K γ-*ori*-based
131 suicide vector encodes the *ccdB* toxin gene under the control of an arabinose-inducible and glucose-
132 repressible promoter, P_{BAD} (Le Roux et al 2007). Matings between *E. coli* and *Vibrio* were performed at
133 30°C as described previously (Le Roux et al 2007) (supplementary methods). Selection of the plasmid-
134 borne drug marker (Cm^R) resulted in integration of the entire plasmid in the chromosome by a single
135 crossover. Elimination of the plasmid backbone resulting from a second recombination step was selected
136 by arabinose induction of the *ccdB* toxin gene. Mutants were screened by PCR and are described in
137 Table S1. For complementation experiments, the Gibson assembly method was used to clone the *R-5.7*

138 gene under a constitutive promoter (P_{LAC}) in pMRB plasmid known to be stable in vibrios (Le Roux et al
139 2011). This plasmid was then transferred to *Vibrio* by conjugation as described previously.
140
141 **Virulence studies using oysters.** Bacteria were grown under constant agitation at 20°C for 24 h in
142 Zobell media. One hundred microliters of the culture (10^6 cfu) pure or diluted were injected
143 intramuscularly into oysters. The bacterial concentration was confirmed by conventional dilution plating
144 on Zobell agar. After injection, the oysters were transferred to aquaria (20 oysters per aquarium of 2.5 l)
145 containing 1 liter of aerated 5 µm-filtered seawater at 20°C, kept under static conditions for 24 hours.
146
147
148 **RESULTS AND DISCUSSION**
149
150 **Disease is associated with progressive replacement of non-virulent vibrios by genetically related**
151 **virulent strains.** SPF oysters were exposed to natural seawater in the field during a mortality outbreak
152 and then returned to the laboratory after 15 days. On each subsequent day, 10 oysters were sacrificed and
153 bacteria were isolated from the tissue (supplementary methods). Mortalities started at day 3, reached 50%
154 at day 5 and then ceased (Fig. 1A, red bars). The cumulative mortality after 5 days in the laboratory was
155 similar to the extent of mortality observed for a subset of the same batch of oysters maintained in the field
156 for one month. We speculate that this mesocosm allows development of disease to proceed more rapidly
157 in the lab.
158
159 Roughly 30 of the bacterial isolates from each day were characterized by partial sequencing of a protein-
160 coding gene (*gyrB*). Phylogenetic analysis allowed the grouping of 162/173 isolates in 6 clades

161 (designated a to f) with a bootstrap value > 70% (Fig. 1B). These clades were matched with named
162 species using type strains *V. mediterranei* (a), *V. chagasii* (b), *V. lentus* (c), *V. splendidus* (d), *V.*
163 *cyclitrophicus* (e), *V. crassostreae* and *V. gigantis* (f). Clades (b) to (f) belong to the Splendidus super
164 clade (Sawabe et al 2013). Strains isolated at the beginning of the experiment were mainly related to
165 clades (d) (36%) and (f) (39%) (Fig. 1A). The clade (f) strains increased to as high as 77% when the
166 mortalities started but returned to their starting level by the final day. The prevalence of clade (d) strains
167 declined to 12%, while clade (a) strains increased from 0% to 32% when mortality reached 50%.
168
169 To address the pathogenic potential of individual strains, we used an injection model of infection, which
170 enables more reliable and rapid infection in the laboratory. Among the 173 isolates individually
171 introduced into SPF oysters, 143 (83%) were found to induce <20% mortality (Fig. 1B and Fig. S1).
172 Twenty strains induced >50% mortality and were consequently classified as virulent (vir+). The majority
173 of vir+ strains (75%) clustered into a subclade of (f) that contains the *V. crassostreae* type strain (LGP7^T)
174 (Fig. 1B) (Faury et al 2004). This type strain and other conspecifics were isolated from an oyster
175 mortality event in 2001 and demonstrated to be pathogenic for oysters (Gay et al 2004a).
176
177 **The genetic cluster of virulent strains coincides with an ecological population.** We investigated
178 whether the vir+/vir- subclades of (f) coincides with any of the ecological populations described by Polz
179 and collaborators (Hunt et al 2008). As the *Vibrio* ecological population structure analysis was based on
180 *hsp60* sequences, this genetic marker was used to compare the populations (Fig. S2). The vir+ and vir-
181 strains were found to be included in distinct clades corresponding to ecologically differentiated
182 populations. These two populations were reported to show preferences for either zooplankton or large
183 particles (Hunt et al 2008).
184

185 It is possible that the association between vir+ strains and particles/plankton is important for natural
186 infection of oysters by these pathogens. Recent studies have reported the central role of marine
187 aggregates in facilitating colonization of *C. virginica* oysters by *V. vulnificus* and showed that differences
188 in the ability to incorporate into these aggregates may play a role in the bacterial population disparity
189 observed within oysters (Froelich et al 2012). Thus, identifying the microhabitats of *C. gigas* pathogens
190 may facilitate development of an experimental infection model mimicking the natural route of infection,
191 e.g., through the use of adapted polymeric substrates. Such approach would also allow for a better
192 understanding of the mode of transmission and primary target tissues or organs for these pathogens.
193
194 **Comparative genomic analyses reveal limited gene clusters that distinguish vir+ and vir-**
195 **populations.** To identify genetic features that distinguish the vir+ population from vir- strains within
196 clade f, we used high throughput sequencing (HTS, supplementary methods) to analyze the genomes of
197 34 isolates (Fig. 1B and Table 1). Twenty-one of those isolates appeared to be distinct strains (Table 1): 9
198 virulent strains (6 from the present study and 3 from a 2001 mortality outbreak) and 12 avirulent strains.
199 The core genome phylogeny (3229 genes) shows that the strains split into two lineages consistent with
200 the pathogenicity status (Fig. 2). The average nucleotide identity (ANI) was 92-93% between
201 populations, whereas within vir+ and vir- lineages, the ANI values ranged between 97.8-99.9% and 96.6-
202 97.3%, respectively. Despite a strong clonal frame in the core genome, we detected extensive genetic
203 diversity in the flexible genome. Strain-specific gene numbers varied from 1000-1700 and 100-1200, at
204 respectively inter and intra-population levels, which is two times higher than the flexible genome
205 diversity observed in *V. cholerae* (Fig. S3).
206
207 We performed comparative analyses to identify sequences present in the vir+ and absent from the vir-
208 population. A total of 101 genes were found to be specific to vir+ strains, 53% of which were localized in

209 7 distinct regions (designated R-1 to 7) (Table S2). Four regions (R-2, 3, 4 and 5) are involved in drug
210 resistance and metabolic function, suggesting adaptation of strains to local competitive or environmental
211 pressure. The R-6 region encoding for arylsulfatase may have an important scavenging function in
212 removing sulfate groups from exogenous substrates such as macroalgal polysaccharides and providing
213 carbon sources (Cohen et al 2007, Mann et al 2013). The R-1 region is homologous to the widespread
214 colonization island (also named *tad* gene cluster), which encodes adhesive pili, and was demonstrated as
215 essential for biofilm formation, colonization and pathogenesis in numerous bacteria (Tomich et al 2007).
216 The R-1 region also encodes the PhoPQ two-component system, which controls a variety of processes
217 including resistance to antimicrobial peptides (AMPs) (Otto 2009). AMPs in concert with reactive
218 oxygen species (ROS) play a crucial role in the invertebrate immune system (Bachere et al 2004).
219 Interestingly, in the R-4 specific locus, we identified genes encoding a catalase and a superoxide
220 dismutase (SOD) putatively implicated in ROS resistance (Ibarra and Steele-Mortimer 2009). Finally, the
221 R-7 region carries genes encoding putative transposases and proteins of unknown function.
222
223 Reciprocally, we identified 193 genes present in all vir- strains but absent from all vir+ strains, half of
224 which were localized in 14 regions (Table S3). Two regions are putatively involved in phosphonate
225 transport (Yu et al 2013). Altogether, our data suggest that ecological specialization, possibly through
226 differential association with hosts and/or particulate material, results from gene acquisition conferring
227 function as scavenging, drug resistance, adhesion, and host immune response survival.
228
229 **Clade f vir+ strains encode a putative outer membrane protein that is necessary for virulence.** We
230 assessed the importance of vir+ specific loci for *V. crassostreae* virulence using a genetic knockout
231 approach. Deletion of regions R-1, 2, 4, 5 or 7 in strain J2-9 did not impair bacterial growth in culture
232 media, but deletion of R-5 resulted in a 3-fold decrease in mortalities induced after bacteria injection

233 (Fig. 3A). The importance of R-5 was confirmed using two additional strains (J5-5 and LGP8) belonging

234 to the vir+ population (Fig. 3A). Among the 32 genes localized in the R-5 region of J2-9, only 8 genes

235 were present in all vir+ strains and absent from all vir- strains (Fig. 3B and Table S2). Analyses of

236 mutants lacking individual genes revealed that the R-5.7 gene accounts for the contribution of R-5 to

237 virulence (Fig. 3C). When constitutively expressed in trans, R-5.7 was sufficient to restore the virulence

238 of the mutants *ΔR5.7* and *ΔR5* (Fig. 3D). Thus R-5.7 is the only gene necessary for the contribution of R-

239 5 to virulence. On another hand the expression of R-5.7 in trans in a vir- strain was not sufficient to

240 induce a virulent phenotype (Fig. 3D). These complementation experiments confirm that *R-5.7* gene is

241 necessary but not sufficient for the pathogenicity. The R-5.7 gene (labeled VRSK9J2v1_730268 in J2-9,

242 Table S2) is predicted to encode a 798 aa exported protein with a theoretical molecular mass of 89 kDa.

243 No functional domains within the protein could be identified using InterProScan, PFAM or Figfam, but

244 Psort predicted R-5.7 to encode an outer membrane protein. Blast analysis revealed the presence of R-5.7

245 orthologous genes in several genomes of vibrios belonging to Splendidus, Orientalis and Photobacterium

246 clades (Fig. S4), although not in LGP32, a *V. splendidus*-related strain previously demonstrated to be

247 pathogenic for oysters (Le Roux et al 2009).

248

249 It is notable that of the 81 genes analyzed by deletion only one was found to be necessary for *V.*

250 *crassostreae* pathogenicity. This finding suggests that the primary role of these population specific genes

251 may be unrelated to virulence. It is possible that these shared genes are simple a 'fossil' of the common

252 ancestry. On another hand, the vir+ strains may be specifically adapted to a particular environmental

253 niche where these genes are beneficial. In such a scheme, oysters may be considered as an alternative

254 habitat for the vir+ population rather than their principal environmental niche. This hypothesis is in

255 accordance with previous data demonstrating that *V. crassostreae* was associated with both algal detritus

256 and zooplankton (Preheim et al 2011, Szabo et al 2012). Further mapping of virulent strains onto *V.*

257 *crassostreae* ecological populations may enable determination of the microhabitats from which oyster
258 pathogens emerge and provide more insight into the nature of populations that serve as reservoirs of
259 pathogens.
260
261 **Non-virulent strains may facilitate the disease.** Naturally infected oysters initially contain a large
262 proportion of avirulent strains, but these are progressively replaced by a virulent population that
263 comprises ~50% of the bacterial isolates at the point of maximal mortalities (Fig. 1A). The low
264 prevalence of vir+ strains in the early infection process could reflect a contribution of the non-virulent
265 strains to the development of disease. To address this question, the vir+ strain J2-9 (*V. crassostreae*) was
266 injected into oysters at various doses, either alone or in combination with the non-virulent strains J2-8
267 (clade f) or J2-20 (*Shewanella* sp.) (Fig. 4). When injected alone, reduction of the injected dose of J2-9
268 from 10^6 to 4 x 10^4 cfu (via dilution in culture media) significantly reduced oyster mortality (~90% vs.
269 5% mortality), even if the infection was allowed to progress for a longer time. In contrast, when J2-9 was
270 injected at 4 x 10^4 cfu following serial dilutions with pure cultures of vir- strains (J2-8 or J2-20), so that
271 the final cfu/inoculum was 10^6, mortality rates were markedly higher than with J2-9 alone (cumulative
272 mortalities of 70% for J2-9 diluted in J2-8 and 60% for J2-9 diluted in J2-20). Thus, the presence of non-
273 virulent bacteria dramatically increases the virulence of low doses of J2-9, suggesting that there are
274 genotype-independent effects of bacterial density upon virulence.
275
276 Thus, although vir- strains are not sufficient for pathogenesis, they clearly have some features (as yet
277 undetermined) that contribute either directly or indirectly to virulence. One possibility is that vir- strains
278 provide resources lacking by the vir+, enabling the vir+ strains to act as “cheaters”, as seen in some
279 analyses of siderophore synthesis and utilization (Cordero et al 2012). An alternate role for the vir- strains
280 may be to generate a sufficient bacterial load, either to overcome host defenses or to induce expression of

281 virulence factors that are regulated by quorum sensing (Bassler 2002). Notably, autoinducer synthases
282 (CsqA, LuxM and LuxS), which initiate the quorum sensing signaling cascades, appear to be encoded by
283 both vir+ and vir- strains. In the future, we will investigate the importance of quorum sensing pathways in
284 virulence, as well as explore additional means by which vir- strains contribute to the disease in order to
285 better understand this process of density-dependent pathogenesis.
286
287
288 **CONCLUSION**
289
290 Our results demonstrate the consistency of the virulent population, which also correspond to a previously
291 identified ecologically cohesive genotypic cluster. In the future delineation of ecological populations
292 together with experimental infections should allow the determination of populations with high or low risk
293 of pathogenicity, the microhabitats from which oyster pathogens emerge and, consequently, which
294 populations serve as reservoirs of pathogens. Hence it may be possible to develop diagnostic tools at the
295 taxonomic level as soon as population specific genes are targeted.
296

296 **ACKNOWLEDGEMENTS**

297 We warmly thank Brigid Davis (HMS, Boston, USA), Marylise Duperthuy (Umea University, Sweden),
298 Otto X. Cordero (ETH Zurich, CH) and Martin F. Polz (MIT, Cambridge, USA) for fruitful discussions
299 and critically reading the manuscript. We thank the staff of the station Ifremer Argenton and Bouin, the
300 ABIMS and CRBM (Roscoff) and LABGeM (Evry) plateforms for technical support. FLR acknowledges
301 the VIBRIOGEN consortium (Annick Jacq, Delphine Destoumieux and Guillaume Charrière) for our
302 precious exchanges. The present study has been supported by the Region Bretagne (SAD Vibrigen 6633;
303 TV funding), the ANR blanc (11-BSV7-023-01 « VIBRIOGEN », DG funding), the EMBRC France
304 (AL funding) and Ifremer (DG and TV fundings).

305

306 **AUTHOR CONTRIBUTIONS**

307 AL, TV, BP, YL and FLR performed experiments. DG and AC performed the *in silico* analyses. YL and
308 FLR designed experiments, interpreted results, and wrote the paper. AL, DG and TV contribute equally
309 to this work.

310

311 Supplementary information is available at ISMEJ's website

312

313 The manuscript has been seen and approved by all of the authors. There is no conflict of interest. The
314 material represents an original result and has not been submitted for publication elsewhere.

315

316

317

317 **REFERENCES**

318

319 Bachere E, Gueguen Y, Gonzalez M, de Lorgeril J, Garnier J, Romestand B (2004). Insights into the anti-
320 microbial defense of marine invertebrates: the penaeid shrimps and the oyster *Crassostrea gigas.*
321 *Immunol Rev* **198:** 149-168.

322

323 Bassler BL (2002). Small talk. Cell-to-cell communication in bacteria. *Cell* **109:** 421-424.

324

325 Cohen AL, Oliver JD, DePaola A, Feil EJ, Boyd EF (2007). Emergence of a virulent clade of *Vibrio*
326 *vulnificus* and correlation with the presence of a 33-kilobase genomic island. *Applied and environmental*
327 *microbiology* **73:** 5553-5565.

328

329 Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F *et al* (2012). Ecological populations
330 of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* **337:** 1228-1231.

331

332 Criscuolo A, Gribaldo S (2010). BMGE (Block Mapping and Gathering with Entropy): a new software
333 for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10:**
334 210.

335

336 Daubin V, Gouy M, Perriere G (2002). A phylogenomic approach to bacterial phylogeny: evidence of a
337 core of genes sharing a common history. *Genome research* **12:** 1080-1090.

338

339 Duperthuy M, Binesse J, Le Roux F, Romestand B, Caro A, Got P *et al* The major outer membrane
340 protein OmpU of *Vibrio splendidus* contributes to host antimicrobial peptide resistance and is required
341 for virulence in the oyster *Crassostrea gigas. Environmental microbiology* **12:** 951-963.
342
343 Duperthuy M, Schmitt P, Garzon E, Caro A, Rosa RD, Le Roux F *et al* (2011). Use of OmpU porins for
344 attachment and invasion of *Crassostrea gigas* immune cells by the oyster pathogen *Vibrio splendidus.*
345 *Proc Natl Acad Sci U S A* **108:** 2993-2998.
346
347 Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput.
348 *Nucleic acids research* **32:** 1792-1797.
349
350 Faury N, Saulnier D, Thompson FL, Gay M, Swings J, Le Roux F (2004). *Vibrio crassostreae* sp. nov.,
351 isolated from the haemolymph of oysters (*Crassostrea gigas*). *International journal of systematic and*
352 *evolutionary microbiology* **54:** 2137-2140.
353
354 Froelich B, Ayrapetyan M, Oliver JD (2012). *Vibrio vulnificus* integration into marine aggregates and
355 subsequent uptake by the oyster, *Crassostrea virginica. Applied and environmental microbiology*.
356
357 Gay M, Berthe FC, Le Roux F (2004a). Screening of *Vibrio* isolates to develop an experimental infection
358 model in the Pacific oyster *Crassostrea gigas. Dis Aquat Organ* **59:** 49-56.
359
360 Gay M, Renault T, Pons AM, Le Roux F (2004b). Two *Vibrio splendidus* related strains collaborate to
361 kill *Crassostrea gigas*: taxonomy and host alterations. *Dis Aquat Organ* **62:** 65-74.
362

363 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and
364 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst*
365 *Biol* **59:** 307-321.

366

367 Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF (2008). Resource partitioning and
368 sympatric differentiation among closely related bacterioplankton. *Science* **320:** 1081-1085.

369

370 Ibarra JA, Steele-Mortimer O (2009). *Salmonella*--the ultimate insider. *Salmonella* virulence factors that
371 modulate intracellular survival. *Cell Microbiol* **11:** 1579-1586.

372

373 Labreuche Y, Le Roux F, Henry J, Zatylny C, Huvet A, Lambert C *et al* (2010). *Vibrio aestuarianus* zinc
374 metalloprotease causes lethality in the Pacific oyster *Crassostrea gigas* and impairs the host cellular
375 immune defenses. *Fish Shellfish Immunol* **29:** 753-758.

376

377 Le Roux F, Binesse J, Saulnier D, Mazel D (2007). Construction of a *Vibrio splendidus* mutant lacking
378 the metalloprotease gene *vsm* by use of a novel counterselectable suicide vector. *Applied and*
379 *environmental microbiology* **73:** 777-784.

380

381 Le Roux F, Zouine M, Chakroun N, Binesse J, Saulnier D, Bouchier C *et al* (2009). Genome sequence of
382 *Vibrio splendidus*: an abundant planctonic marine species with a large genotypic diversity.
383 *Environmental microbiology* **11:** 1959-1970.

384

385 Le Roux F, Davis BM, Waldor MK (2011). Conserved small RNAs govern replication and
386 incompatibility of a diverse new plasmid family from marine bacteria. *Nucleic acids research* **39:** 1004-
387 1013.
388
389 Mann AJ, Hahnke RL, Huang S, Werner J, Xing P, Barbeyron T *et al* (2013). The genome of the alga-
390 associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for
391 degradation of algal polysaccharides. *Applied and environmental microbiology* **79:** 6813-6822.
392
393 Otto M (2009). Bacterial sensing of antimicrobial peptides. *Contrib Microbiol* **16:** 136-149.
394
395 Petton B, Pernet F, Robert R, Boudry P (2013). Temperature influence on pathogen transmission
396 and subsequent mortalities in juvenile Pacific oysters *Crassostrea gigas. Aquacult Environ Interact* **3:**
397 257-273.
398
399 Preheim SP, Timberlake S, Polz MF (2011). Merging taxonomy with ecological population prediction in
400 a case study of *Vibrionaceae. Applied and environmental microbiology* **77:** 7195-7206.
401
402 Samain (2008). Summer mortality of Pacific oyster *Crassostrea gigas. The Morest Project, Editions*
403 *Quae, 379p*.
404
405 Saulnier D, De Decker S, Haffner P, Cobret L, Robert M, Garcia C (2010). A large-scale epidemiological
406 study to identify bacteria pathogenic to Pacific oyster *Crassostrea gigas* and correlation between
407 virulence and metalloprotease-like activity. *Microbial ecology* **59:** 787-798.
408

409 Sawabe T, Ogura Y, Matsumura Y, Feng G, Amin AR, Mino S *et al* (2013). Updating the *Vibrio* clades
410 defined by multilocus sequence phylogeny: proposal of eight new clades, and the description of *Vibrio*
411 *tritonius* sp. nov. *Front Microbiol* **4:** 414.
412
413 Szabo G, Preheim SP, Kauffman KM, David LA, Shapiro J, Alm EJ *et al* (2012). Reproducibility of
414 *Vibrionaceae* population structure in coastal bacterioplankton. *The ISME journal*.
415
416 Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt DE, Benoit J *et al* (2005). Genotypic
417 diversity within a natural coastal bacterioplankton population. *Science* **307:** 1311-1313.
418
419 Tomich M, Planet PJ, Figurski DH (2007). The *tad* locus: postcards from the widespread colonization
420 island. *Nat Rev Microbiol* **5:** 363-375.
421
422 Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A *et al* (2013). MicroScope--an
423 integrated microbial resource for the curation and comparative analysis of genomic and metabolic data.
424 *Nucleic acids research* **41:** D636-647.
425
426 Wendling CC, Batista FM, Wegner KM (2014). Persistence, seasonal dynamics and pathogenic potential
427 of *Vibrio* communities from pacific oyster hemolymph. *PloS one* **9:** e94256.
428
429 Yu X, Doroghazi JR, Janga SC, Zhang JK, Circello B, Griffin BM *et al* (2013). Diversity and abundance
430 of phosphonate biosynthetic genes in nature. *Proc Natl Acad Sci U S A* **110:** 20759-20764.
431

432 Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn
433 graphs. *Genome research* **18:** 821-829.
434
435
436
437

437 **TITLES AND LEGENDS TO FIGURES**

438

439 **Figure 1.** Bacterial population dynamics during oyster infection. **A-** Specific pathogen free oysters were
440 transferred to open seawater for 2 weeks and then transferred to the laboratory to study *Vibrio* dynamics
441 during the disease expression. Mortalities (pink bar) were recorded daily (D0 to 5) and expressed in
442 percentage (right y-axis). Lines indicate the percentage of strains (left y-axis) belonging to the
443 phylogenetic clades described in B. The black lines correspond to the 3 most represented clades (f: circle;
444 d: square; a: triangle) whereas grey lines (e: plain; c: large dots; b: small dots) correspond to less
445 abundant genotypes. The red line correspond to the strains inducing >50% mortality (right y-axis). **B-**
446 Date of isolation and virulence superimposed on the phylogeny of bacterial isolates inferred by maximum
447 likelihood analysis of partial *gyrB* gene sequences, with outer and inner rings indicating the % of
448 mortalities obtained 24H after oyster injection (brown bars >50%; red bars<50%) and the day of isolation
449 (D0 in white, D1-4 in grey gradient, D5 in black, as in A bellow the x-axis) respectively. Clades a, b, c,
450 d, e, f were obtained with a bootstrap value of 99, 94, 92, 100, 100 and 71% and contain the type strains
451 of *V. mediterranei,* CIP 10320^T (a), *V. chagasii,* R-3712^T (b), *V. lentus,* CECT 5110^T (c), *V. splendidus,*
452 LMG 4042^T (d), *V. cyclitrophicus,* LMG 21359^T (e), *V. crassostreae,* LGP7^T and *V. gigantis,* LGP13^T
453 (f). Asterisks indicate the strains sequenced in the present study.

454

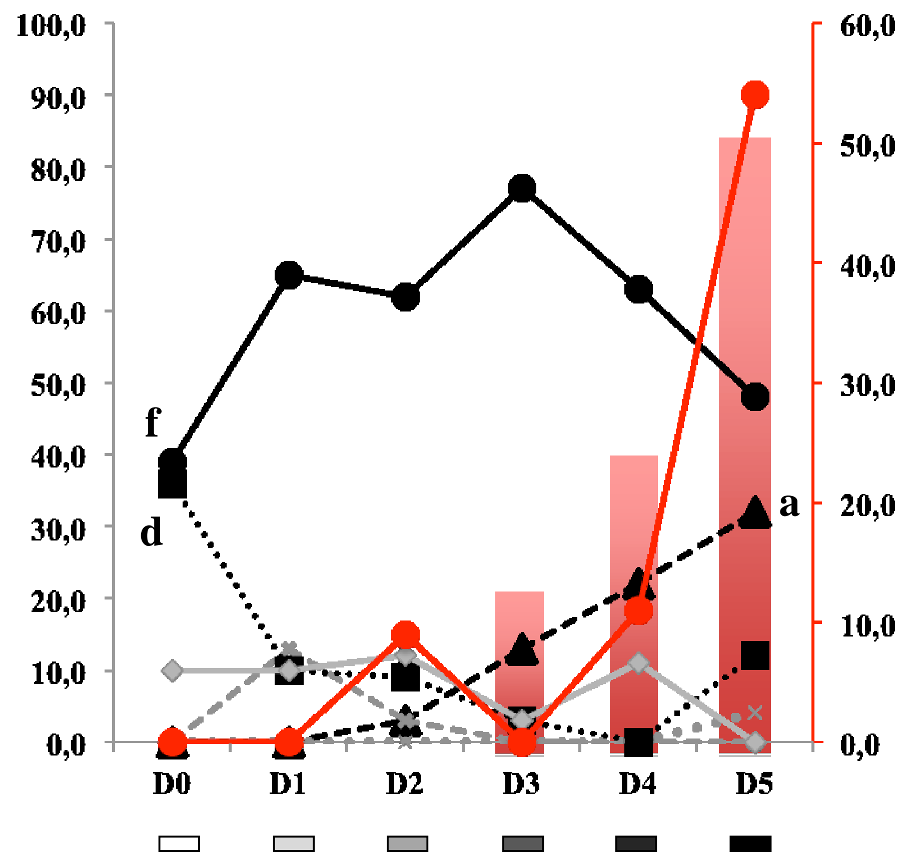
455 **Figure 2.** Phylogenetic analysis based on concatenated alignments of nucleic acid sequences of core
456 genes of strains sequenced in the present study. J2-x and J5-x strains were isolated in the present study at
457 day 2 and 5 respectively. LGPx strains were isolated in 2001. LGP7^T is the type strain of *V. crassostreae*
458 species. The strain LGP32 was used as an outgroup. Trees were built by the Maximum-Likelihood
459 method (GTR substitution model, NNIs, γ4, invariant site) based on sequences aligned using Muscle and
460 filtered with BMGE. Branch lengths are drawn to scale and are proportional to the number of nucleotide

461 changes. Numbers at each node represent the percentage value given by bootstrap analysis of 100
462 replicates. Strains in which the deletions of the R-5 region were made are underlined.
463
464 **Figure 3.** Oyster mortality in response to experimental infection with *V. crassostreae* wild type strains
465 and derivatives. 10^6 CFU of the tested strains (wt : wild type ; **A-** ΔR-1 to 7 for deleted regions 1 to 7
466 respectively ; **C-** ΔR-5.1 to 8 for deleted genes 5.1 to 8 in the region R-5; **D-** J2-9 and derivatives ΔR-5,
467 ΔR-5.7 or J2-8 carrying (indicated with a star) or not the expression vector pMRB-P_{LAC}*R-5.7*) was
468 intramuscularly injected into oysters (n= 20, in duplicate). Mortality (%) was assessed after 24 hours. **B-**
469 Organization of the genomic region R-5. Genes in red are found in all vir+ strains and absent in all vir-
470 strains (plain: exported unknown protein; hatched: regulators). Genes in white were also found in vir-
471 strains. Genes in black indicate the region boundaries that have been targeted for the deletion of the entire
472 region (locus tag of the J2-9 strain: VRSK9J2v1_ 73037 and 730270). Arrows indicate gene targeted for
473 single-gene mutagenesis; we were unable to established a deletion in R-5.4 (grey, italics).
474
475 **Figure 4.** Oyster mortality in response to experimental infection with vir+ strain in the presence of vir-
476 strains cultures. The J2-9 (*V. crassostreae*, vir+), was intramuscularly injected into oysters (two aquaria
477 of n= 20) in pure culture (10^6 CFUs/animal), diluted (4 x 10^4 CFUs/animal) with culture media or with a
478 pure culture of vir- strains J2-8 (clade f, vir-) or J2-20 (*Shewanella* sp.). As a negative control a pure
479 culture (10^6 CFUs/animal) of J2-8 or J2-20 was also injected. Mortality (%) was assessed after 24 hours.
480

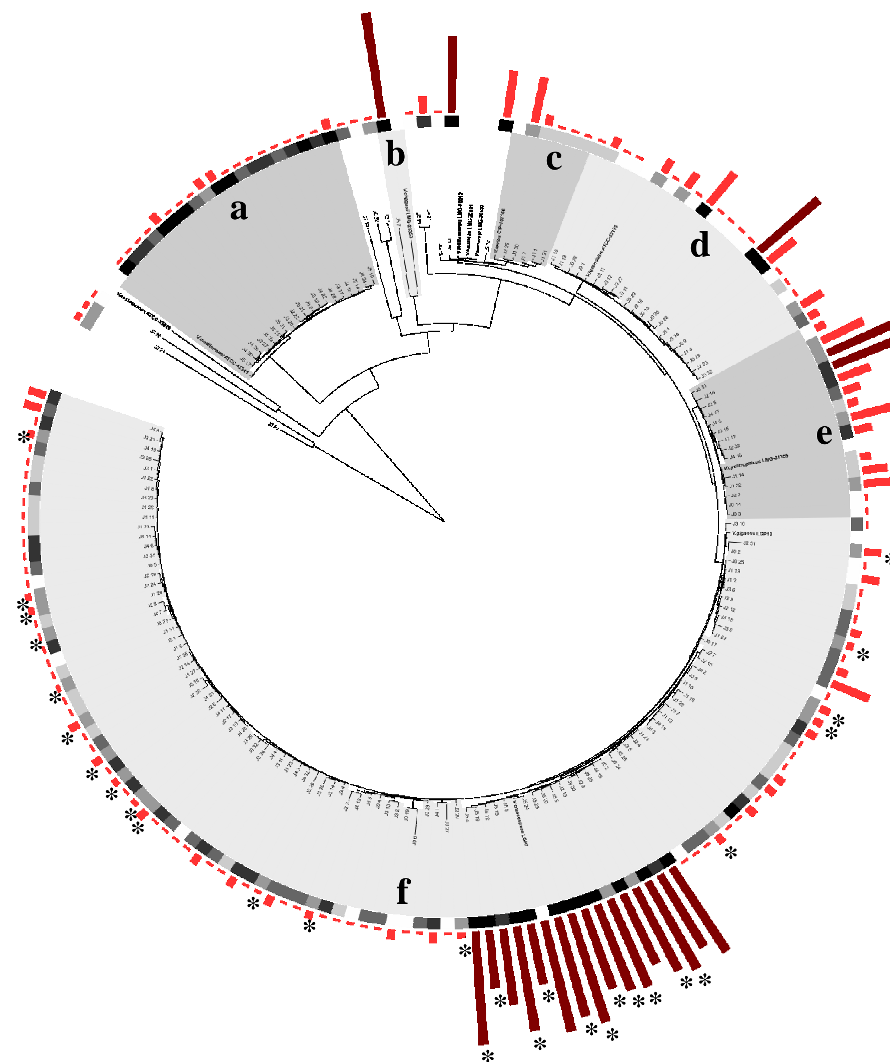
Virulence	Strain	Origin	Contigs number	Genome size (Mb)	CDSs	Accession number	Apparent clonality
vir+	J2-9	This study	174	5.79	5495	PRJEB5876	J2-13
	J5-4	This study	189	5.86	5589	PRJEB5877	
	J5-5	This study	165	5.81	5543	PRJEB5878	
	J5-15	This study	92	5.65	5345	PRJEB5879	
	J5-19	This study	145	5.61	5255	PRJEB5880	
	J5-20	This study	145	5.74	5462	PRJEB5882	J5-6
	LGP7T	Oyster mortalities, 2001	122	5.64	5413	PRJEB5883	J5-23, J5-24, J5-28, LGP15
	LGP8	Oyster mortalities, 2001	177	5.58	5427	PRJEB5884	
	LGP107	Oyster mortalities, 2001	86	5.49	5249	PRJEB5885	LGP108
vir-	J2-1	This study	69	5.36	4941	PRJEB5886	J2-24, J2-28
	J2-3	This study	54	5.52	5063	PRJEB5887	
	J2-4	This study	84	5.67	5285	PRJEB5888	
	J2-6	This study	64	5.38	4934	PRJEB5889	J2-18
	J2-8	This study	98	5.39	4969	PRJEB5890	
	J2-12	This study	158	5.55	5177	PRJEB5891	
	J2-14	This study	65	5.67	5244	PRJEB5892	
	J2-15	This study	74	5.57	5158	PRJEB5893	J2-7
	J2-17	This study	95	5.42	4977	PRJEB5894	J2-10
	J2-26	This study	74	5.42	4981	PRJEB5895	J2-30
	J2-29	This study	117	5.66	5301	PRJEB5896	
	J2-31	This study	255	5.73	5437	PRJEB5897	

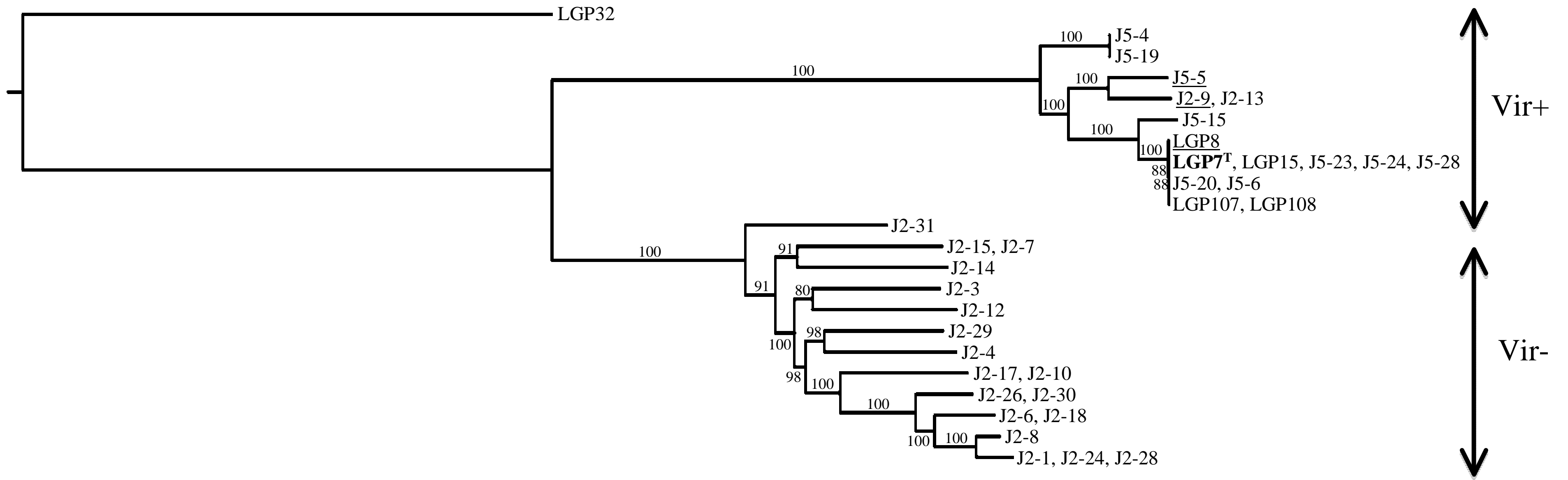
Table 1 : Strains sequenced by HTS in the present study.

A

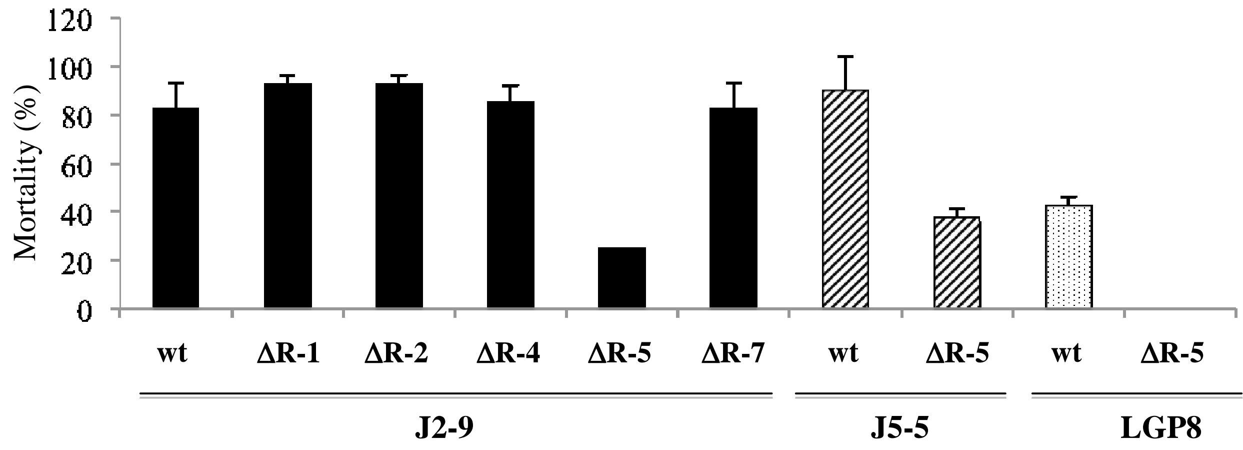


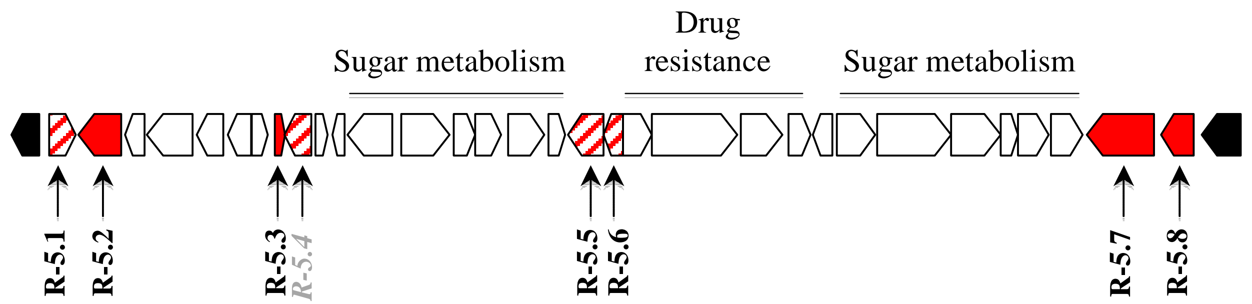
B

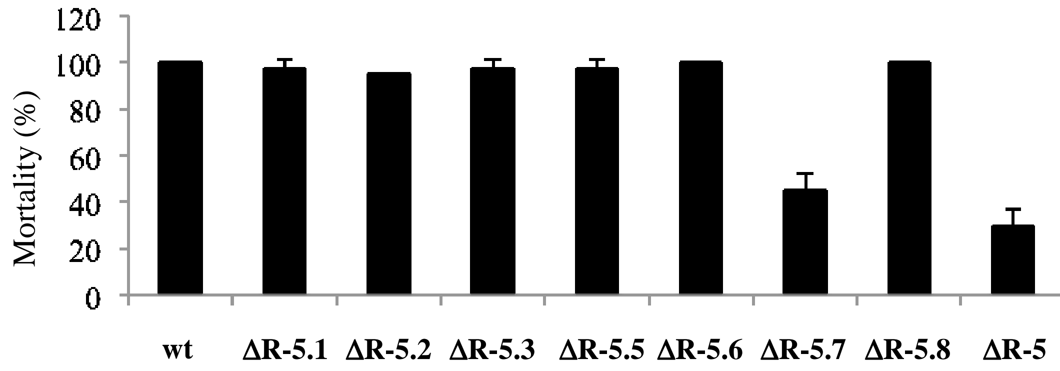


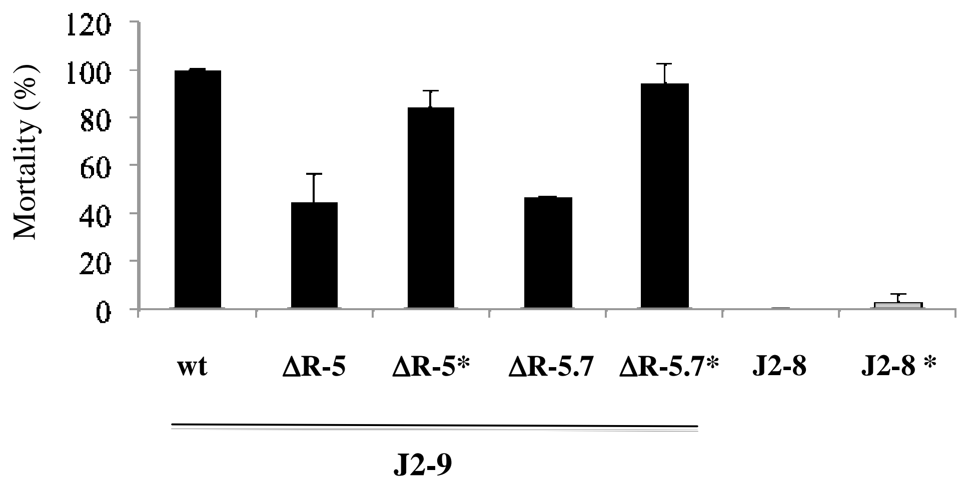


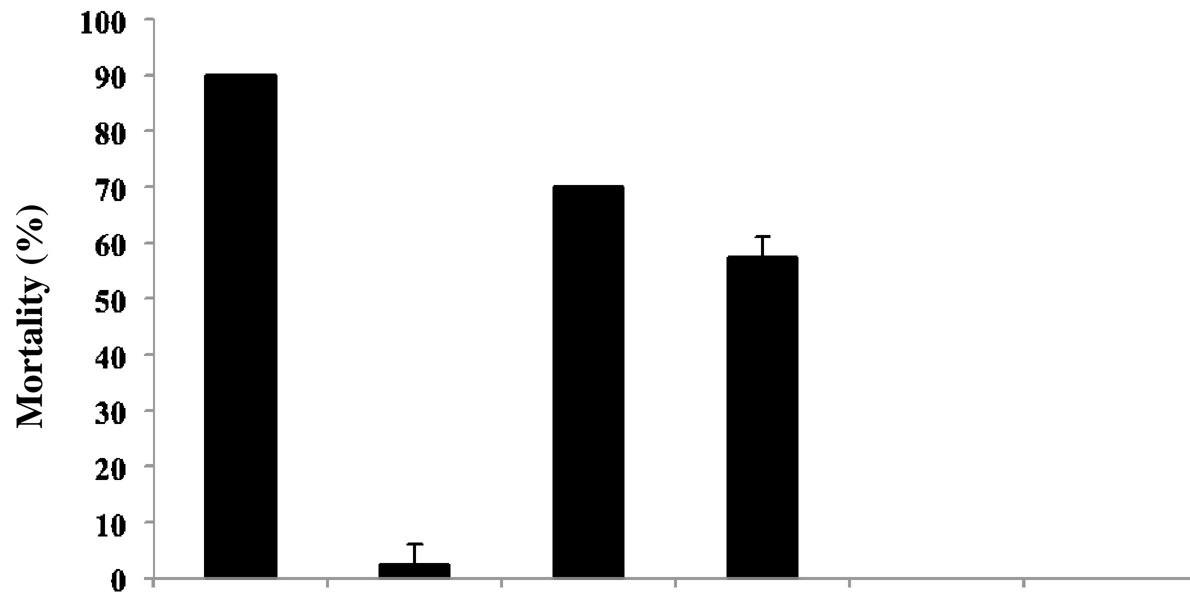
0.01
LGP32
100
J5-4
J5-19
J5-5
J2-9, J2-13
J5-15
LGP8
LGP7T, LGP15, J5-23, J5-24, J5-28
J5-20, J5-6
LGP107, LGP108
Vir+
J2-31
J2-15, J2-7
J2-14
J2-3
J2-12
J2-29
J2-4
J2-17, J2-10
J2-26, J2-30
J2-6, J2-18
J2-8
J2-1, J2-24, J2-28
Vir-


A
Mortality (%)
120
100
80
60
40
20
0
wt
ΔR-1
ΔR-2
ΔR-4
ΔR-5
ΔR-7
wt
ΔR-5
wt
ΔR-5
J2-9
J5-5
LGP8


B
Sugar metabolism
Drug resistance
Sugar metabolism
R-5.1
R-5.2
R-5.3
R-5.4
R-5.5
R-5.6
R-5.7
R-5.8


C
Mortality (%)
120
100
80
60
40
20
0
wt
ΔR-5.1
ΔR-5.2
ΔR-5.3
ΔR-5.5
ΔR-5.6
ΔR-5.7
ΔR-5.8
ΔR-5


D
Mortality (%)
120
100
80
60
40
20
0
wt
ΔR-5
ΔR-5*
ΔR-5.7
ΔR-5.7*
J2-8
J2-8 *
J2-9



J2-9 (CFU/animal	10^6	4.10^6	4.10^4	4.10^4	0	0
Dilution in (CFU/animal)		Medium	J2-8 (10^6)	J2-20 (10^6)	J2-8 (10^6)	J2-20 (10^6)

1 **SUPPLEMENTARY INFORMATIONS**

2

3 **MATERIAL AND METHODS**

4

5 **Production of "pathogen free" oysters (Specific Pathogen Free, SPF)**

6 A pool of 40 genitors (wild animals collected in Fouras Bay, Marennes- Oléron, France) was transferred
7 to the Ifremer facility located at Argenton (Brittany, France) for maturation conditioning as described
8 previously(Petton et al 2013). After gamete stripping and fertilization, obtained larvae then spat were
9 reared under controlled conditions up to 5 months. PCR detection of Herpes was performed to confirm
10 the negative status of oyster spats(Petton et al 2013). Vibrios isolation on selective culture medium
11 (Thiosulfate-citrate-bile salts-sucrose agar, TCBS) confirmed a low vibrio presence (~10 cfu/gr tissus).

12

13 **Infection on field**

14 The 30 May 2011, SPF oysters spat (n=600) were transferred to a farming area (Bay of Brest at Pointe du
15 Château ; 48° 20' 06.19" N, 4° 19' 06.37" W) to allow the development of infection. Mean seawater
16 temperature in the field was 17.1°C. An oyster set (designed as sentinel, n=300) was maintained in the
17 field to monitor the first mortality onset and determine the cumulative mortality rates occurring naturally
18 after one month. At the first mortality reports (15th of June), infected animals (n=300) were reintroduced
19 in the laboratory for 10 days to reveal the disease (21°C). Mortalities started at day 3, reached 50% at day
20 5 and then ceased.

21

22 **Strain isolation and identification**

23 Each day 10 living oysters were pooled, grounded in sterile seawater (10 ml/gr of tissues) and serial
24 dilutions were spread on TCBS (Difco). TCBS is advocated as the main selective medium for vibrio

25 isolation from estuarine and marine waters (Pfeffer & Oliver 2003). It has been used extensively in the
26 past to recover a large range of *Vibrio* spp. and provides standard approach for comparison with other
27 studies (Turner et al., 2009).Randomly selected colonies (~30/day for 6 days) were re-streaked two times
28 on TCBS, cultivated in Zobell media (4 g/l bactopeptone and 1 g/l yeast extract in artificial sea water,
29 pH7.6) and stored at -80°C. For DNA sequencing, purified isolates were grown in Zobell overnight and
30 DNA extracted using a DNA extraction kit (Wizard, Promega). The partial *gyrB* or *hsp60* gene was
31 amplified for all isolates as described previously(Faury et al 2004, Hunt et al 2008). The PCR conditions
32 were: 3 min at 94°C followed by 30 cycles of 30 sec each at 94°C, 55° and 72°C with a final step of 6
33 min at 72°C. Genes were sequenced using the reverse primer and sequencing was performed at the
34 Station Biologique of de Roscoff, France.
35 The partial *gyrB* or *hsp60*gene sequences were aligned using Muscle(Edgar 2004) and filtered by
36 BMGE(Criscuolo and Gribaldo 2010). Phylogenetic trees were built using the parallel version of PhyML
37 applied to Maximum-likelihood algorithm and GTR model as parameters (NNIs, γ4, invariant
38 site)(Guindon et al 2010). Reliability was assessed by the bootstrap method with 100 replicates. Circular
39 tree figures were drawn using the online iTOL software package(Letunic and Bork 2011).
40
41 **Genome sequencing, assembly and annotation**
42 A total of 34 strains were sequenced using the Illumina HiSeq2000 technology with ~50-fold coverage.
43 Contigs were assembled *de novo* using Velvet(Zerbino and Birney 2008) and genome assembly was
44 improved by contig mapping against the LGP32 reference genome(Le Roux et al 2009).Computational
45 prediction of coding sequences and other genome features (RNA encoding genes, ribosome binding sites,
46 peptide signal sequences, etc...), together with functional assignments were performed using the
47 automated annotation pipeline implemented in the MicroScope platform
48 http://www.genoscope.cns.fr/agc/mage(Vallenet et al 2013). An extensive manual curation of specific

49 genes, which includes correction of the start codon position and of the functional assignments, was

50 performed. This expert procedure was supported by functional analysis results [e.g., InterPro, FigFam,

51 PRIAM, COGs (Clusters of Orthologous Groups), PsortB], which can be queried using an exploration

52 interface, and by synteny groups computation visualized in cartographic maps to facilitate genome

53 comparison.

54

55 **Cloning of deleted alleles into a suicide vector**

56 The cloning was performed using the Gibson assembly method according to the manufacturer's

57 instructions (New England Biolabs, NEB). For region's deletion, 3 independent PCR amplifications

58 were performed using primers R-y-1 to 6 (Table S4). Primer pairs R-y-1+ 2 and R-y-5+ 6 were used to

59 amplified the 500 bp fragment located upstream and downstream respectively of the region to be deleted.

60 Primer pair R-y-3+ 4 led to the amplification of a spectinomycin resistance cassette to be cloned between

61 the two previous 500 bp fragments. For single gene deletion, independent PCR amplifications of the

62 regions (500 bp) encompassing the gene to delete were performed using two primer pairs (R-5.x-1 and 2

63 and R-5.x-3 and 4) (Table S4).A PCR inside out was performed using PSW7848T suicide vector

64 DNA(Val et al 2012) and primer pair (SW-F and SW-R) (Table S4). After purification and

65 quantification, 100 ng of the PCR products were mixed with Gibson assembly Master Mix and incubated

66 for 60 minutes at 50°C. Samples were diluted at 1/3 before *E. coli* transformation.Strains Π3813 and

67 β3914 were used as a plasmid host for cloning and conjugation, respectively(Le Roux et al 2007).

68 Plasmids and strains used and established in the present study are presented in Table S1.

69

70 **Conjugation**

71 Overnight cultures of donor and recipient were diluted at 1:100 in culture media without antibiotic and

72 grown at 30°C to an OD_{600nm} of 0.3. The different conjugation experiments were done by a filter mating

73 procedure described previously(Le Roux et al 2007) with a donor/recipient ratio of 1ml/10ml.
74 Conjugations were performed overnight on filters incubated on LBA + NaCl 0.5N + diaminopimelic acid
75 (DAP) plates at 30°C. Counter-selection of *dapA* donor was done by plating on a medium devoid of
76 DAP, supplemented with chloramphenicol and 1% glucose. For mutagenesis, Cm^R resistant colonies
77 were grown in LB + NaCl 0.5N up to late logarithmic phase and spread on plates containing 0.2%
78 arabinose. Mutants were screened by PCR using primers 7+8 flanking the different genes targeted (Table
79 S4).
80
81

81
82 **REFERENCES**

83

84 Criscuolo A, Gribaldo S (2010). BMGE (Block Mapping and Gathering with Entropy): a new software
85 for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10:**
86 210.

87

88 Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput.
89 *Nucleic acids research* **32:** 1792-1797.

90

91 Faury N, Saulnier D, Thompson FL, Gay M, Swings J, Le Roux F (2004). Vibrio crassostreae sp. nov.,
92 isolated from the haemolymph of oysters (Crassostrea gigas). *International journal of systematic and*
93 *evolutionary microbiology* **54:** 2137-2140.

94

95 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and
96 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst*
97 *Biol* **59:** 307-321.

98

99 Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF (2008). Resource partitioning and
100 sympatric differentiation among closely related bacterioplankton. *Science* **320:** 1081-1085.

101

102 Le Roux F, Binesse J, Saulnier D, Mazel D (2007). Construction of a Vibrio splendidus mutant lacking
103 the metalloprotease gene vsm by use of a novel counterselectable suicide vector. *Applied and*
104 *environmental microbiology* **73:** 777-784.

105

106 Le Roux F, Zouine M, Chakroun N, Binesse J, Saulnier D, Bouchier C *et al* (2009). Genome sequence of
107 Vibrio splendidus: an abundant planctonic marine species with a large genotypic diversity.
108 *Environmental microbiology***11:** 1959-1970.

109

110 Letunic I, Bork P (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees
111 made easy. *Nucleic acids research***39:** W475-478.

112

113 Petton B, Pernet F, Robert R, Boudry P (2013). Temperature influence on pathogen transmission
114 and subsequent mortalities in juvenile Pacific oysters
115 Crassostrea gigas. *Aquacult Environ Interact***3:** 257-273.

116

117 Pfeffer C, Oliver JD (2003). A comparison of thiosulphate-citrate-bile salts-sucrose (TCBS) agar and
118 thiosulphate-chloride-iodide (TCI) agar for the isolation of Vibrio species from estuarine environments.
119 *Lett Appl Microbiol***36**:150-151.

120

121 Turner JW, Good B, Cole D, Lipp EK (2009). Plankton composition and environmental factors
122 contribute to Vibrio seasonality. ISME J. **3**:1082-1092

123

124 Val ME, Skovgaard O, Ducos-Galand M, Bland MJ, Mazel D (2012). Genome engineering in Vibrio
125 cholerae: a feasible approach to address biological issues. *PLoS Genet***8:** e1002472.

126

127 Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A *et al* (2013). MicroScope--an
128 integrated microbial resource for the curation and comparative analysis of genomic and metabolic data.
129 *Nucleic acids research* **41:** D636-647.
130
131 Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn
132 graphs. *Genome research* **18:** 821-829.
133
134
135
136

136

137 **Table S1:** Strains and plasmids used and constructed in this study

138

Strain	**Description**	**Reference**
Π3813	B462 *ΔthyA::(erm-pir116)* [Erm^R]	13
β3914	β2163 *gyrA462, zei298::Tn10* [Km^R Em^R Tc^R]	13
GV973	J2-9 *ΔR-1*	This study
GV964	J2-9 *ΔR-2*	This study
GV940	J2-9 *ΔR-4*	This study
GV953	J2-9 *ΔR-5*	This study
GV968	J2-9 *ΔR-7*	This study
GV976	J5-5 *ΔR-5*	This study
GV977	LGP8 *ΔR-5*	This study
GV1018	J2-9 *ΔR-5.1*	This study
GV1019	J2-9 *ΔR-5.2*	This study
GV1020	J2-9 *ΔR-5.3*	This study
GV1021	J2-9 *ΔR-5.5*	This study
GV1022	J2-9 *ΔR-5.6*	This study
GV1023	J2-9 *ΔR-5.7*	This study
GV1025	J2-9 *ΔR-5.8*	This study
GV1139	J2-8 + pMRB-P_{LAC}*R-5.7*	This study
GV1140	J2-9 *ΔR-5* + pMRB-P_{LAC}*R-5.7*	This study
GV1141	J2-9 *ΔR-5.7* + pMRB-P_{LAC}*R-5.7*	This study
Plasmid	**Description**	**Reference**
pSW7848T	*oriV*$_{R6K\gamma}$; *oriT*$_{RP4}$; *araC*-P_{BAD}*ccdB;* [Cm^R]	13
pSWδR-1	pSW7848T; *Δ R1*	This study
pSWδR-2	pSW8742T; *Δ R2*	This study
pSWδR-4	pSW8742T; *Δ R4*	This study
pSWδR-5	pSW8742T; *Δ R5*	This study
pSWδR-7	pSW8742T; *Δ R7*	This study
pSWδR-5.1	pSW8742T; *Δ R5-1*	This study
pSWδR-5.2	pSW8742T; *Δ R5-2*	This study
pSWδR-5.3	pSW8742T; *Δ R5-3*	This study
pSWδR-5.5	pSW8742T; *Δ R5-5*	This study
pSWδR-5.6	pSW8742T; *Δ R5-6*	This study
pSWδR-5.7	pSW8742T; *Δ R5-7*	This study
pSWδR-5.8	pSW8742T; *Δ R5-8*	This study
pMRB-P_{LAC}*R-5.7*	*oriV*$_{R6K\gamma}$; *oriT*$_{RP4}$; *oriV*$_{pB1067}$; P_{LAC}*R-5.7*[Cm^R]	This study

139

140

140 **Table S2:** Gene annotation of the regions R-1 to 7 in J2-9 strain. The genes present in all vir+ and absent

141 in all vir- strains are grey shaded.

142

Region	Locus tag in J2-9 VRSK9J2v1_	Product	Function
R-1	130064	Putative pilin Flp	Tad operon PhoQ/P
	130065	Putative prepilin leader peptidase TadV/CpaA	
	130066	Putative flp pilus assembly protein RcpC/CpaB	
	130067	Putative flp pilus assembly protein, RcpA/CpaC	
	130068	Putative flp pilus lipoprotein RcpB/CpaD	
	130069	Putative flp pilus assembly atpase TadZ/CpaE	
	130070	Putative flp pilus assembly protein, TadA/CpaF	
	130071	Putative flp pilus assembly protein TadB	
	130072	Putative flp pilus assembly protein TadC	
	130073	Putative flp pilus assembly protein TadD	
	130074	Putative flp pilus assembly pseudopilin TadE	
	130075	Putative flp pilus assembly pseudopilin TadF	
	130076	Putative flp pilus assembly protein TadG	
	130077	Outer membrane protein A	
	130078	Sensor histidine kinase PhoQ	
	130079	Transcriptional regulatory protein PhoP	
R-2	150070	Putative hydrolase (beta-lactamase)	Drug resistance
	150071	Putative transcriptional regulator, lysR family	
	150072	Putative lysR family transcriptional regulator	
	150073	Putative exported metal-dependent hydrolase	
	150074	Exported protein of unknown function	
	150075	Conserved exported protein of unknown function	
	150077	Protein of unknown function	
	150078	Putative porin	
	150079	Putative transcriptional regulator, lysR family	
	150080	Putative exported amidohydrolase	
	150081	Multidrug efflux pump membrane transporter	
	150082	Putative multidrug efflux membrane fusion protein	
	150083	Putative multidrug efflux outer membrane protein	
	150084	Putative cold shock-like protein	
	150085	Conserved protein of unknown function	
R-3	720235	Putative magnesium transporter	Drug resistance
	720236	Putative multiple antibiotic resistance protein	
	720237	Putative acriflavin resistance protein	

	720238	Putative efflux transporter	
	720239	Putative mechanosensitive ion channel	
	720240	Exported protein of unknown function	
	720241	Putative protein translocase subunit SecF	
	720242	Protein translocase subunit SecD	
	720243	Exported protein of unknown function	
R-4	730104	Putative transcriptional repressor	**Sugar metabolism ROS**
	730105	Putative sucrose-6-phosphate hydrolase	
	730106	Fructose-bisphosphate aldolase 2	
	730107	Putative sugar kinase	
	730108	Putative fructose transport system kinase	
	730109	Putative sugar-proton symporter	
	730110	Catalase	
	730111	Putative AnkB protein	
	730112	Superoxide dismutase	
	730113	Putative transcriptional regulator	
	730114	Conserved hypothetical protein	
R-5	730238	Putative transcriptional regulator, lysR family	**Sugar metabolism Drug resistance**
	730239	Conserved exported protein of unknown function	
	730240	Conserved exported protein of unknown function	
	730241	Conserved lipoprotein of unknown function	
	730242	Putative transcriptional regulator, lysR family	
	730243	Putative permease	
	730244	Putative transcriptional regulator	
	730245	Conserved exported protein of unknown function	
	730246	Putative transcriptional regulator	
	730247	Putative quinol oxidase subunit	
	730248	Conserved protein of unknown function	
	730249	Malate synthase A	
	730250	Glyoxylate carboligase	
	730251	Hydroxypyruvate isomerase	
	730252	Tartronate semialdehyde reductase	
	730253	Putative hydroxypyruvate reductase	
	730254	Putative transcriptional regulator, TetR family	
	730255	Putative two-component sensor	
	730256	Putative two-component regulator	
	730257	Putative multidrug resistance membrane protein	
	730258	Putative multidrug resistance protein	
	730259	Putative chlorohydrolase/deaminase family protein	
	730260	Putative outer membrane protein	
	730261	Putative transcriptional regulator	

	730262	Putative phosphotransferase system	
	730263	Putative alpha-mannosidase	
	730264	Putative sucrose phosphorylase	
	730265	Putative Pts fructose-specific enzyme	
	730266	Glycerate kinase	
	730267	Mannose-6-phosphate isomerase	
	730268	Conserved exported protein of unknown function	
	730269	Conserved exported protein of unknown function	
R-6	1350053	Putative transcriptional regulators, lysR family	**Arylsulfatase**
	1350054	Putative short chain dehydrogenase	
	1350055	Conserved membrane protein of unknown function	
	1350056	Conserved exported protein of unknown function	
	1350057	Putative arylsulfatase A	
	1350058	Putative arylsulfatase regulator	
	1350059	Protein of unknown function	
	1350060	Putative cytochrome BD ubiquinol oxidase	
	1350061	Putative transcriptional regulator, lysR family	
	1350062	Putative arylsulfatase	
R-7	1640006	Conserved protein of unknown function	**Transposon**
	1640007	Conserved protein of unknown function	
	1640009	Conserved protein of unknown function	
	1640010	Conserved protein of unknown function	
	1640011	Putative transposase	
	1640012	Putative transposase	
	1640013	Putative transposase	

143

144

145

145 **Table S3:** Gene annotation of the regions R-1 to 14 in J2-1 strain

146

Region	Locus Tag VRSK1J2v1_	Product	Function
R-1	20095	Conserved exported protein of unknown function	**Unknown**
	20099	Conserved exported protein of unknown function	
	20100	Conserved membrane protein of unknown function	
	20101	Transcriptional regulator	
	20102	Conserved exported protein of unknown function	
	20103	Conserved protein of unknown function	
	20106	Putative Sensor histidine kinase	
	20109	Major facilitator superfamily MFS_1	
	20110	Transcriptional Regulator, LysR family protein	
	20114	Conserved protein of unknown function	
R-2	20186	Conserved exported protein of unknown function	**Unknown**
	20187	Conserved protein of unknown function	
	20188	Putative D-aminoacylase	
	20189	Conserved exported protein of unknown function	
	20190	Protein of unknown function	
R-3	20208	GlcNAc-binding protein A	**Cytochrome d**
	20210	Conserved membrane protein of unknown function	
	20211	Putative Bacterial regulatory protein GntR	
	20212	Cytochrome d ubiquinol oxidase subunit 1	
	20213	Cytochrome d ubiquinol oxidase subunit 2	
R-4	20411	Cytochrome d ubiquinol oxidase subunit 2	**Cytochrome d**
	20412	Cytochrome d ubiquinol oxidase subunit 1	
	20413	DoxD-like family protein	
	20414	Conserved protein of unknown function	
	20415	Conserved protein of unknown function	
	20416	Conserved exported protein of unknown function	
	20417	Arylsulfatase	
R-5	90009	Conserved exported protein of unknown function	**Phosphanate**
	90010	Putative PhnE phosphonate ABC transporter, permease	
	90011	Putative PhnE phosphonate ABC transporter, permease	
	90012	Putative phnC phosphonate ABC transporter, ATPase	
	90013	Putative Phosphonate ABC transporter, periplasmic	
	90014	Putative transcriptional regulator	
	90015	Putative sodium/glucose cotransporter	
	90016	Conserved protein of unknown function	
	90020	Putative alginate lyase	

	90024	Putative alginate lyase	
	90028	3-ketoacyl-CoA reductase PhaB	
R-6	190007	Phosphonate import ATP-binding protein PhnC	**Phosphanate**
	190008	Phosphonate-binding periplasmic protein	
	190009	Putative phosphonate transport system phnE	
	190010	Putative Phosphonate C-P lyase system PhnF	
	190011	Phosphonate metabolism protein PhnG	
	190012	Carbon-phosphorus lyase complex subunit	
	190013	Carbon-phosphorus lyase complex subunit	
	190014	Phosphonate metabolism PhnJ	
	190015	Phosphonates transport ATP-binding protein phnK	
	190016	Phosphonates transport ATP-binding protein phnL	
	190017	Carbon-phosphorus lyase complex subunit	
	190018	Ribose 1,5-bisphosphokinase	
	190019	Carbon-phosphorus lyase complex accessory protein	
	190020	Putative acetyltransferase	
R-7	190292	Aspartate aminotransferase	**Transporter**
	190293	Uncharacterized transporter FTT_0829c	
	190294	Conserved protein of unknown function	
	190295	Efflux transporter, RND family, MFP subunit	
	190296	Uncharacterized transporter HI_0895	
	190297	Conserved protein of unknown function	
	190299	Transcriptional regulator CadC	
	190300	Putative Dehydrogenase	
R-8	200063	Phosphate transporter (ABC permease)	**Phosphate**
	200064	Phosphate transporter subunit membrane component	
	200065	Phosphate transporter subunit membrane component	
	200066	Phosphate import ATP-binding	
	200071	Protein of unknown function	
	200072	Conserved protein of unknown function	
	200073	Two-component sensor PilS	
R-9	200469	Putative Transcriptional regulator, LysR family	**Unknown**
	200470	Na+-driven multidrug efflux pump	
	200471	Enoyl-CoA hydratase/isomerase family protein	
	200473	Putative oxidase	
	200474	Transcriptional regulatory protein	
	200475	Prolyl endopeptidase	
	200476	Conserved hypothetical protein	
R-10	400123	Putative DNA-binding transcriptional regulator	**Unknown**
	400124	Putative oxalate formate antiporter	
	400125	Putative kinase	
	400126	Putative (phospho)hydrolase	

	400127	Putative transporter MFS superfamily	
R-11	400166 400167 400169 400170 400171 400172 400175	Conserved exported protein of unknown function Exported protein of unknown function Conserved exported protein of unknown function Conserved exported protein of unknown function Conserved exported protein of unknown function Short-chain fatty acids transporter Putative alkaline phosphatase	**Unknown**
R-12	450066 450067 450068 450069 450070 450071 450072 450073 450074	Putative lipoprotein Putative MoxR-like ATPase Conserved protein of unknown function Conserved protein of unknown function Conserved hypothetical protein Transporter Conserved protein of unknown function Conserved exported protein of unknown function Transcriptional regulator, LysR family protein	**Unknown**
R-13	590236 590237 590238 590241 590243 590245 590246 590247	Putative outer membrane protein Putative Thioredoxin family protein Putative Thiol:disulfide interchange protein Putative amidase Conserved exported protein of unknown function Putative Multidrug resistance efflux transporter EmrE Conserved exported protein of unknown function Transcriptional Regulator, LysR family	**Unknown**
R-14	660001 660002 660005 660006 660007 660008 660009 660010 660011 660012 660014	Putative metal-dependent hydrolase Putative Transcriptional regulator, LysR family protein Hypothetical protein Transcriptional regulators, LysR family Hypothetical protein Hypothetical protein Hypothetical protein Hypothetical protein Outer membrane protein Putative multidrug efflux membrane fusion protein Conserved hypothetical protein	**Unknown**

147

148

148
149 **Table S4:** Primers used in this study

150

Primer	Sequence 5'-3'
R-1-1	GTATCGATAAGCTTGATATCGAATTCAAGTAATCGAGCATGGCAGG
R-1-2	CGTCACAGGTATTTATTCGGCGTCAATCGCAGTAACACCACG
R-1-3	CGTGGTGTTACTGCGATTGACGCCGAATAAATACCTGTGACG
R-1-4	CTCGTCTCAACTTGGTGATCGATATCGTCGCAGACCAAAACG
R-1-5	CGTTTTGGTCTGCGACGATATCGATCACCAAGTTGAGACGAG
R-1-6	CCCCCGGGCTGCAGGAATTCGAGTTATCTCTCGTCTCACG
R-1-7	GCGATGAGTTCTCCACTGGC
R-1-8	GCTGATTACCTTCGCTACGC
R-2-1	GTATCGATAAGCTTGATATCGAATTCGTCCAGTTCCAGCAACACAA
R-2-2	CGTCACAGGTATTTATTCGGCGTGCTGATACGAGTGCGAATG
R-2-3	CATTCGCACTCGTATCAGCACGCCGAATAAATACCTGTGACG
R-2-4	AGTATGTCGCTCTGACTCTGGATATCGTCGCAGACCAAAACG
R-2-5	CGTTTTGGTCTGCGACGATATCCAGAGTCAGAGCGACATACT
R-2-6	CCCCCGGGCTGCAGGAATTCACTTGTACATACGTAGGCGC
R-2-7	GCAAGAGCACGCGTAATTGC
R-2-8	CGTGGTTCTGTTCACAACAC
R-4-1	GTATCGATAAGCTTGATATCGAATTCGTCGGTTCTGATGGGTAAAG
R-4-2	CGTCACAGGTATTTATTCGGCGTAAAGCCACCAACACACCTG
R-4-3	CAGGTGTGTTGGTGGCTTTACGCCGAATAAATACCTGTGACG
R-4-4	AGTGCAAAGCGATTCCAAGCGATATCGTCGCAGACCAAAACG
R-4-5	CGTTTTGGTCTGCGACGATATCGCTTGGAATCGCTTTGCACT
R-4-6	CCCCCGGGCTGCAGGAATTCGTAGACCAAGGTTAGTGGTG
R-4-7	GTTAACCATCATGTGTGGCG
R-4-8	GTTCCAAGGCAAGAACGCAT
R-5-1	GTATCGATAAGCTTGATATCGAATTCATGCCACTTCCATAGCGAAC
R-5-2	CGTCACAGGTATTTATTCGGCGTTCGCCTGAGATTGTTGCTC
R-5-3	GAGCAACAATCTCAGGCGAACGCCGAATAAATACCTGTGACG
R-5-4	TCTGCACATCAACCTCAGTGGATATCGTCGCAGACCAAAACG
R-5-5	CGTTTTGGTCTGCGACGATATCGCGTACCTGGACCAACAATT
R-5-6	CCCCCGGGCTGCAGGAATTCGCCCAGAATAATGCCGACAT
R-5-7	GTTGGTTGAACACTTTCGCC
R-5-8	GACTTCACCTCACTGCTTGT
R-7-1	GTATCGATAAGCTTGATATCGAATTCCCGTAATCCCCAGATATTGC
R-7-2	CGTCACAGGTATTTATTCGGCGCGTCAGCTATCACGTATAGC
R-7-3	GCTATACGTGATAGCTGACGCGCCGAATAAATACCTGTGACG
R-7-4	GTAGTCGTAGAGGATAAGGCGATATCGTCGCAGACCAAAACG
R-7-5	CGTTTTGGTCTGCGACGATATCGCCTTATCCTCTACGACTAC

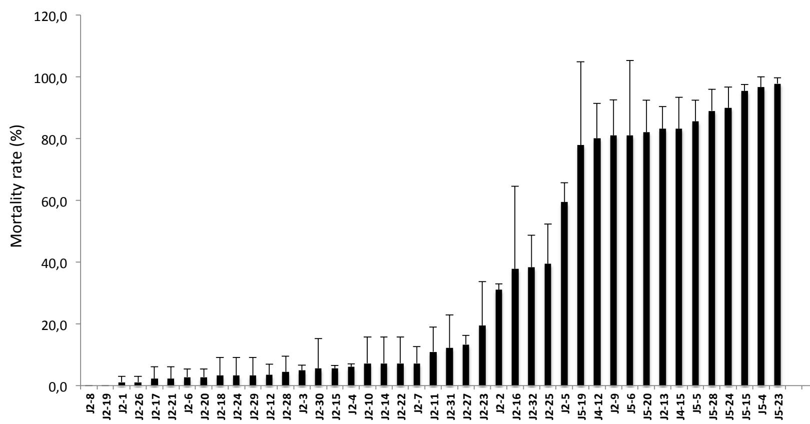
R-7-6	CCCCCGGGCTGCAGGAATTCCCCAACGATTAATAGTGGCG
R-7-7	GGAGTACCTTCTGGAACTTC
R-7-8	CTGTGCTCTAACATTGGTGC
R-5.1-1	GTATCGATAAGCTTGATATCGAATTCCATAGGCTTGTAGCGCCTTA
R-5.1-2	GACAGTAACTAAAACGCGCCCCTCATTTCCATGCTTTGCC
R-5.1-3	GGCAAAGCATGGAAATGAGGGGCGCGTTTTAGTTACTGTC
R-5.1-4	CCCCCGGGCTGCAGGAATTCGATGGCGTTGAAGCTTATCG
R-5.2-1	GTATCGATAAGCTTGATATCGAATTCGCTTATGTGGCCGATACACA
R-5.2-2	CTGCCGCGTATCAGATATGTACTAAAACGCGCCTTCATCG
R-5.2-3	CGATGAAGGCGCGTTTTAGTACATATCTGATACGCGGCAG
R-5.2-4	CCCCCGGGCTGCAGGAATTCGCGCGATGTTTCTTTATCCC
R-5.3-1	GTATCGATAAGCTTGATATCGAATTCCGCCTTGATAAACTGGAAGC
R-5.3-2	GGCTCAAACTCACTAGTTCGCTGATGGGTGGCATCTATTG
R-5.3-3	CAATAGATGCCACCCATCAGCGAACTAGTGAGTTTGAGCC
R-5.3-4	CCCCCGGGCTGCAGGAATTCTTCTCATCTCATCGATGCCC
R-5.5-1	GTATCGATAAGCTTGATATCGAATTCCCGATTGAAGAACTCAGCGA
R-5.5-2	ACCGTTCGACTTTCCAATGCGGACATCGTGTGATGTCCTT
R-5.5-3	AAGGACATCACACGATGTCCGCATTGGAAAGTCGAACGGT
R-5.5-4	CCCCCGGGCTGCAGGAATTCGCTAACTGCATTGAATGGCC
R-5.6-1	GTATCGATAAGCTTGATATCGAATTCAGTACACCCCAACTATCGTC
R-5.6-2	GCTACATTGTTGATAGGGGCATCGAGGATGTCAGAATCCG
R-5.6-3	CGGATTCTGACATCCTCGATGCCCCTATCAACAATGTAGC
R-5.6-4	CCCCCGGGCTGCAGGAATTCGTTGCTTTTCGTAGCGCAAG
R-5.7-1	GTATCGATAAGCTTGATATCGAATTCGAAACGACGGATATCGAAGG
R-5.7-2	AAGGTGTTGGCTTCTGCTACGTTAACCGAGCGTTGCTTTC
R-5.7-3	GAAAGCAACGCTCGGTTAACGTAGCAGAAGCCAACACCTT
R-5.7-4	CCCCCGGGCTGCAGGAATTCCTCTTATCATCGCTCTTCAG
R-5.8-1	GTATCGATAAGCTTGATATCGAATTCGAAGGGTTGGTGACGTATAG
R-5.8-2	AATGCCATCGAATGGTAGGGCATCTCGAAGAAATAGCGCG
R-5.8-3	CGCGCTATTTCTTCGAGATGCCCTACCATTCGATGGCATT
R-5.8-4	CCCCCGGGCTGCAGGAATTCCCTGAACATTATCTGGCACA
SW-F	GAATTCCTGCAGCCCGGGGG
SW-R	GAATTCGATATCAAGCTTATCGATAC
R-5.7-F	GTGAGCGGATAACAAAGGAAGGGCCCATGAGACTTCTACCTATTATTATTTC
R-5.7-R	CTCGAGCTGCAGACGCGTCGCTAATTAGAGCAGCTATTGCC
pMRB-F	CTCGAGCTGCAGACGCGTCG
pMRB-R	GGGCCCTTCCTTTGTTATCCGCTCAC

151

152

153

153 **Fig. S1:** Virulence of range of selected isolates. Strains isolated at day 2 (J2-1 to 32) and additional vir+
154 strains isolated at day 5 were injected intramuscularly to 20 oysters. Mortalities were recorded at 24H.
155 Experiments were performed in triplicate. Strains were classified as virulent when >50% mortality was
156 obtained.
157



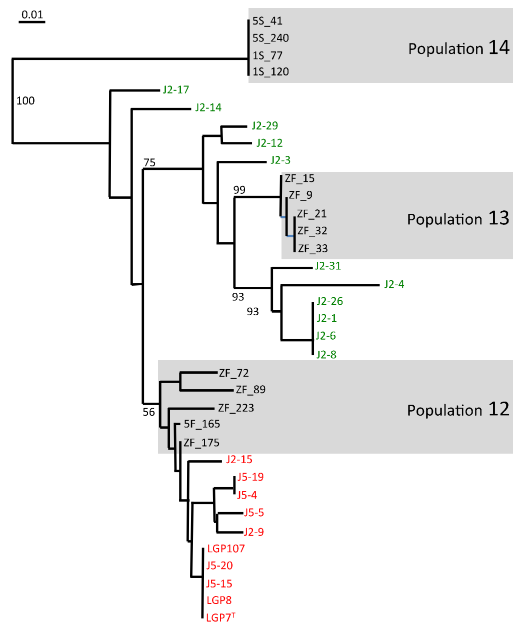
158
159

159 **Fig. S2:** Phylogenetic tree of *hsp60* genes built by the Maximum-Likelihood method. Strains belonging

160 to ecological populations 12, 13 and 14 were isolated from fractionated seawater(Hunt et al 2008) vir+

161 (in red) and vir- (in green) strains were isolated in the present study.

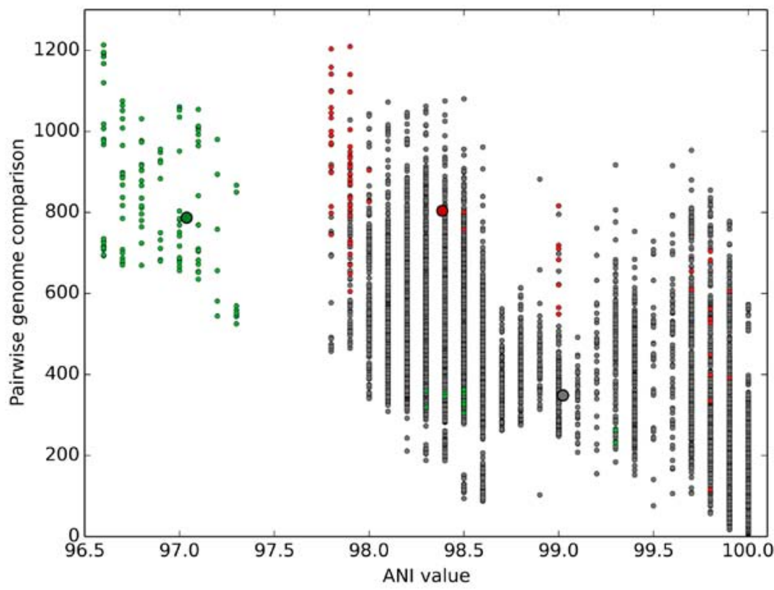
162



163

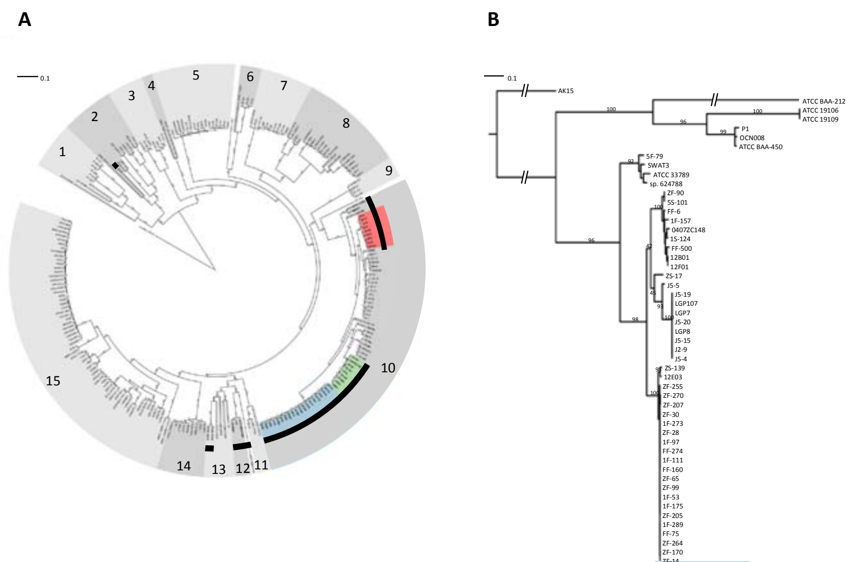
164

164 **Fig. S3:** Strain specific gene numbers according to the ANI value determined by pairwise genome
165 comparison at intra population level (green: 12 strains from clade f, population vir-; red: 9 strains from
166 clade f population vir+) or within *V. cholerae* (grey, 161 genomes). Larger circles indicate the centroids.
167
168



169
170

170 **Fig. S4:** Identification of R-5.7 orthologous genes in *Vibrionaceae* genomes. **A**- *Vibrionaceae* phylogeny
171 was built using FastTree and GTR model, based on concatenated nucleic alignment (muscle) of 42 core
172 genes. Clades are numbered by 1: Salinivibrio-Grimontia-Enterovibrio ; 2: Photobacterium ; 3: Fischeri ;
173 4: Rumoiensis ; 5: Nigripulchritudo ; 6: Halioticoli ; 7: Anguillarum ; 8: Cholerae ; 9: F10 ; 10:
174 Splendidus ; 11: Scophthalmi ; 12: Coralliilyticus ; 13: Orientalis ; 14: Vulnificus ; 15: Harveyi. Black
175 line indicates the presence of R-5.7 orthologous genes. **B** – Phylogenetic tree of R-5.7 orthologous genes
176 built by the Maximum-Likelihood method. Within the Splendidus clade, *V. crassostreae*, *V. splendidus*
177 and *V. cyclitrophicus* are colored in red, green and blue respectively.



178