Environmental Microbiology
August 2009, Volume 11, Issue 8, Pages 1959 - 1970
http://dx.doi.org/10.1111/j.1462-2920.2009.01918.x

Genome sequence of *Vibrio splendidus*: an abundant planctonic marine species with a large genotypic diversity

Frédérique Le Roux[1, 2, *], Mohamed Zouine[1], Nesrine Chakroun[1], Johan Binesse[1], Denis Saulnier[2], Christiane Bouchier[3], Nora Zidane[3], Laurence Ma[3], Christophe Rusniok[4], Aurélie Lajus[5], Carmen Buchrieser[4], Claudine Médigue[5], Martin F. Polz[6] and Didier Mazel[1]

[1] Institut Pasteur, Unité Plasticité du Génome Bactérien, CNRS URA 2171, F-75015, Paris, France.
[2] IFREMER, Laboratoire de Génétique et Pathologie, La Tremblade, France.
[3] Institut Pasteur, PF1, Paris, France.
[4] Institut Pasteur, Unité de Génomique des Micro-Organismes Pathogènes, Paris, France.
[5] CEA, Institut de Génomique, Genoscope & CNRS UMR8030 Laboratoire de Génomique Comparative, Evry, France.
[6] Department of Civil and Environmental Engineering, MIT, Cambridge, MA, USA.

*: Corresponding author : F. Le Roux, Tel. (+33) 1 40 61 32 87; Fax (+33) 1 45 68 88 34, email address : fleroux@pasteur.fr

Abstract:

Vibrio splendidus is a dominant Vibrio species in seawater presenting a remarkable genetic diversity; several strains have been linked to invertebrate's mortality. We report the complete genome sequence of *V. splendidus* LGP32, an oyster pathogen, and its comparison with partial genome sequences from related strains. As is typical for the genus, *V. splendidus* LGP32 contains two chromosomes (3.29 and 1.67 Mb) and most essential cellular processes are encoded by chromosome 1. Comparison with two other V. splendidus partial genome sequences (strains 12B01 and Med222) confirms the previously suggested high genotypic diversity within this species and led to the identification of numerous strain-specific regions that could frequently not be assigned to a specific mechanisms of recombination. Surprisingly, the chromosomal integron, the most variable genetic element in all other *Vibrio* species analysed to date, is absent from 12B01 and inactivated by a mobile element in Med222, while in LGP32 it only contains a limited number of cassettes. Finally, we found that the LGP32 integron contains a new dfrA cassette, related to those found in resistance integrons of Gram-negative clinical isolates. Those results suggest that marine Vibrio can be a source of antibiotic resistance genes.

1. Introduction

Vibrios are ubiquitous marine bacteria, that are ecologically and metabolically diverse members of planktonic and animal associated microbial communities. In temperate ocean regions, *Vibrio splendidus* is the dominant vibrioplankton, but it is also abundant in and on many animal species (for a review see Le Roux and Austin, 2006). Analyses of the diversity and annual dynamics of *V. splendidus*-related strains isolated from seawater revealed an extreme genotypic diversity, with at least a thousand distinct genotypes co-occurring at extremely low environmental concentrations (Thompson et al., 2005). The adaptive significance of this extensive variation remains unknown but may be related to the ecological diversity within this group. It has recently been shown that *V. splendidus* constitutes several, closely related populations, which are engaged in fine-scale resource partitioning even when co-occurring in the water column. In contrast, most other vibrioplankton species formed ecologically cohesive populations (Hunt et al. 2008). The ecologically and genetically diverse *V. splendidus* populations may be useful models for understanding the genomic correlates of environmental adaptation (Hunt et al., 2008).
V. splendidus has also been shown to be associated with a wide range of animals, with interactions ranging from commensal to pathogenic, depending on the identities of both strain and host (Le Roux and Austin, 2006). Previously, we established a collection of *V. splendidus*- related strains associated with the "summer mortalities" syndrome in oysters, which has been responsible for high losses in commercial oyster farming in France since 1991 (Gay et al. 2004a). Experimental infection of oysters demonstrated that some strains were pathogenic individually while others displayed enhanced or repressed virulence when inoculated concomitantly with other strains. Thus vibriosis may result from both single infection and microbial interactions.
To investigate the origin of the remarkable genetic diversity of this species and identify homologs of genes that could play a role in virulence or resistance of *V. splendidus,* we sequenced the genome of strain LGP32 and compared it to two publicly available conspecific isolates (12B01 and Med222; J. Craig Venter Institute). LGP32 was isolated from oysters suffering "summer mortalities" syndrome and demonstrated to be virulent after injection of oysters and clams (Gay et al. 2004a, b). The strains 12BO1 and Med222 were obtained from the water column samples from the Plum Island Sound estuary (Northeastern Massachusetts) (Thompson et al. 2005) and the Mediterranean Sea (Jarone Pinhassi, unpublished data), respectively. Both strains were also demonstrated to be virulent towards oysters (F. Le Roux and D. Saulnier, unpublished data).

2. Results and discussion

General features of the LGP32 genome

The genome of *V. splendidus* strain LGP32 consists of two circular chromosomes of 3,299 (chromosome 1) and 1,675 (chromosome 2) Mb with an average percent G+C content of 44.03 and 43.64, respectively (Figure 1). The origins of replication for both chromosomes were identified by similarities to other *Vibrio* species by co-localization of genes (*myoC* on chromosome 1; *parA* on chromosome 2) and GC nucleotide skew (G-C/G+C). There are a total of 4,498 predicted open reading frames (ORFs) with 2,998 on chromosome 1 and 1,500 on chromosome 2 (identified gene are prefixed with VS_Ixxxx and VS_IIxxxx when carried on chromosome 1 and 2, respectively). Chromosome 1 and 2 contain 7 and 1 rRNA and 97 and 17 tRNA genes, respectively. Similar to other vibrios, there is a pronounced asymmetry in the distribution of functional classes of genes between the two chromosomes (Table 1) (Reen et al., 2006). Chromosome 1 contains 199/205 of the minimal gene set. A higher percentage of genes encoding cellular processes and signaling (18 vs. 6,7%), information

storage and processing (11,5 vs. 4,6%), metabolism (26,2 vs. 13,5%) and conserved hypothetical protein (12,5 % vs. 6,5%) are found on chromosome 1 vs. 2. However chromosome 2 contains a similar percentage of unknown genes (9,1 vs. 12,3%).
Relatively few mobile elements are part of the LGP32 genome; one RS1-like phage is present on chromosome 1 and there are 8 and 1 transposons on chromosome 1 and 2, respectively. All DNA repair enzymes are present, suggesting that the high genomic polymorphism is not directly caused by a missing repair pathway. A recent study showed that *V. cholerae* can acquire new genetic material by natural transformation during growth on chitin, a biopolymer that is abundant in aquatic habitats (e.g., from crustacean exoskeletons) (Meibom et al., 2004). Transformation competence was found to require several genes which are activated by chitin, all of which are present in LGP32 : *tfoX* (VS_I1829), a type IV pilus assembly complex (*pilA, B, C, Q* : VS_I2547, 2548, 2549, 2873) and a putative DNA binding protein (VS_I2199). The presence of these genes in LGP32 suggests that natural transformation maybe operative in at least some strains of *V. splendidus* and could be one of the sources of its genetic diversification.

Compartive genomics

The number and percentage of gene couples satisfying the bi-directional best hit (BBH) and ORF in synton between LGP32 and the other vibrio genomes reveals that the *V. splendidus* chromosome 1 appears closer to *V. vulnificus* chromosome 1 whereas the chromosome 2 appears closer to *V. harveyi* chromosome 2 (Table 1, supplementary data, and Figure 2). This analysis revealed a higher conservation of gene order in chromosome 1, confirming previous reports that chromosome 2 is subject to more frequent genomic rearrangements (Reen et al., 2006, Dryselius et al., 2007). Moreover, genes shared by all *Vibrionaceae* (core genome) are mainly located on chromosome 1, which contains 1725 of the shared genes as opposed to only 208 on chromosome 2 (Figure 1, third circle). Such results confirm previous observations that *Vibrionaceae* core and flexible genomes are largely segregated into two distinct chromosomes (Reen et al., 2006, Dryselius et al., 2007).
Sequences found in LGP32 but not in the majority of the *Vibrionaceae* are also distributed differently among the two chromosomes (Figure 1, third circle). On chromosome 1, these sequences constitute 25% of the chromosome and predominantly flank the *ter* site, whereas such regions constitute the majority of the chromosome 2 and appear more randomly distributed.
To assess strain-to-strain variation within *V. splendidus*, the LGP32 genome was compared to two other closely related strains (Med222 and 12B01) available as draft genome sequences. LGP32 and Med222 contain a closely related core genome, as indicated by an average nucleotide identity (ANI) (Goris et al., 2007) of 95.71%, while 12B01 is more distantly related to LGP32 (89.2%) and Med222 (89.6%). Many more genes were shared among *V. splendidus* strains than by LGP32 and non- *splendidus* vibrios. Of the 3644 genes shared by all 3 genomes (81% of LGP32 ORFs), 2508 (69%) were present on chromosome 1 and 1136 (31%) on chromosome 2 of LGP32 (Figure 3), a distribution roughly proportionally to chromosome size. Among these genes, 152 are only found in the 3 *V. splendidus* strains and not in other *Vibrionaceae.* The large majority of this *V. splendidus*-specific common gene set (65%), encodes proteins of unknown function. Interestingly, these genes are not divided between the two chromosomes in proportion to their sizes, as 67 are on chromosome 1 and 77 on chromosome 2.
We identified 409 in LGP32 that appear to be absent from both 12BO1 and Med222. The majority of these genes (161 in chromosome 1 and 114 in chromosome 2) were also absent from other *Vibrionaceae* and thus appear to be LGP32- specific regions. Annotation of the LGP32-specific genes revealed that 82% encode (conserved) hypothetical protein.
The mapping of strain specific genes onto the chromosomes (Figure 1) suggests that differences are more often colocalized for LGP32+/12B01- and LGP32+/Med222-. Thus

these regions may correspond to regions of genome plasticity instead of bias due to incomplete shotgun sequencing of Med222 and12B01.
The largest LGP32 region of genome plasticity (VS_II0950-VS_II1024, chromosome 2, position 10611438-1149922) has been found in other *Vibrio* species such as V. *parahaemolyticus* (Figure 1 in supplementary data) and *V. harveyi.* The presence of a gene encoding a putative transposase, and the GC% average of 36% (instead of 43.64% for the chromosome 2) indicate that it might be acquire through horizontal gene transfert (Table 2 in supplementary data). This region contains an haemolysin co-regulated protein gene (*hcp*) and the *vas* operon, which encodes a type VI secretion system implicated in virulence in several bacteria species (Figure 1 in supplementary data) (Folkesson et al., 2002; Gray et al., 2002; Rao et al. 2004; Pukatzki et al., 2006).
Several other LGP32 regions of genome plasticity contain genes commonly found in mobile elements suggesting that these regions were assembled by horizontal gene transfer: putative integrases (VS_I1387, VS_I1389), transposases (VS_I1869, VS_I1880, VS_II827), putative addiction modules (VS_I1444-1445, VS_II86-87), putative phage helix-destabilizing protein, rstB1 (VS_I1453) and putative plasmid related protein (VS_II104) (Table 2 in supplementary data). Several of these regions also display a GC% that could argue for an HGT origin. However in many cases, we did not find any such evidence, and nearly half of the LGP32 region of genome plasticity could not be classified as genomic islands.

Chromosomal integron

A surprising feature of the *V. splendidus* genomes is the small size of the chromosomal integron in LGP32 and Med222, and its complete absence in 12B01. The chromosomal integron is thought to be a universal feature of vibrios and has been shown to be the most variable part of the genomes of all human *Vibrio* pathogens studied to date (Chen et al., 2003; Labbate et al., 2007). Integron arrays can consist of up to two hundred cassettes, suggesting a highly active capture process and a strong adaptive role for these elements (Mazel, 2006). In LGP32, as in *V. fisheri* (Ruby et al., 2005), the chromosomal integron is split in two arrays (Figure 4). The functional platform, carrying the integrase gene *intIA* (VS_I1890) and the *attI* cassette integration site, is found in chromosome 1 and contains 20 cassettes (VS_I1869-1889; 15,5 kb). A second array containing only 5 cassettes, is present on chromosome 2, carrying genes VS_II0832 to VS_II0828. The presence of an abutting ISVisp2 suggest that the splitting of the array is due to recombination between identical ISs on the two chromosomes or to the formation of a compound transposon (Figure 4). This peculiar structure led us to look in more detail into the content and distribution of this locus among the other *V. splendidus* strains. Surprisingly, strain 12B01 does not contain an *intIA* ortholog. In Med222, this gene is inactivated by a 20 kb DNA insertion in its 3' part. PCR amplification of *intIA* failed in 12B01, while it confirmed the intIA disruption in Med222 (data not shown) indicating that these results are not due to the unfinished state of both genome projects or to mis-assembly processes. The contig AAND01000031, which carries the Med222 *intIA* gene, ended inside the first cassette. Another contig was found to carry 9 cassettes, identified through the conservation of their specific *attC* sites. Using PCR, we have been able to show that these two contigs are adjacent in the Med222 genome, the cassette array is thus located upstream of the inactivated *intIA* (data not shown). Therefore the chromosomal integron of Med222 is reduced to 9 cassettes, which are unrelated to those found in LGP32, while the C-term truncated integrase of Med222 is identical to the LGP32 one. We probed a collection 70 *V. splendidus* related isolates from various origins with *V. splendidus intIA* by colony-blot hybridization and demonstrate a lack of hybridization signal in 47 strains (data not shown). To further assess the functionality of this genetic element in the *V. splendidus* strain group, we determined the sequence of the chromosomal region carrying the *intIA* locus in *V. splendidus* LGP31, a second isolate from diseased oysters (Gay et al, 2004b), and in the closely related *V. tasmaniensis* LMG 20012T (Thompson et al 2003). The LGP31 locus is highly similar, except for the cassette array, both in organization and

sequence, to the LGP32 one, their IntIA sharing 98.5% identity (GenBank FJ529221). The LGP31 cassette array is different from the LGP32, as all PCR attempts using primers from the LGP32 cassettes failed. *V. tasmaniensis* is more distantly related to LGP32, their IntIA only share 95% identity and one of the gene (VS_I1892) is absent from the neighboring locus. We have however been able to amplify the first 3 cassettes inserted at its *attI* site, and found that they have no homology to those carried in LGP32 (GenBank FJ529220). These cassette variations indicate that the integron is functional in these 3 strains, and able to capture cassettes.
The inactivation of the chromosomal integron in most *V. splendidus* isolates suggests however that this element has not been advantageous during the diversification of this species, while remaining functional in other *Vibrio* species.

Carriage of a trimethoprim-resistance cassette

Interestingly, among the cassettes present in the LGP32 integron, the sixth cassette encoded for a dihydrofolate reductase (DHFR) (VS_I1884) from the A family (Huovinen et al, 1995) phylogenetically related to the *dfr* genes from the same family identified in clinical isolate (Figure 5A). VS_I1884 shares 73% identity with the type VI DHFR (DfrA6) conferring resistance to trimethoprim ($Trim^R$), isolated in a class 1 mobile integron in *Proteus mirabilis* (Wylie and Koornhof, 1991) and 72 % with a DHFR, also carried in class 1 mobile integron, identified in a *V. cholerae* plasmid (strain 2004VN84; Gb accession AB200915). LGP32 is highly resistant to Trim (MIC>1 g/l), apparently due to the presence of VS_I1884, as disruption of this gene rendered the strain sensitive to this antibiotic (Figure 5B). Two additional strains within our collection of *V. splendidus* isolates were also $Trim^R$. The first strain, *V. splendidus* P52, had been isolated from an octopus (Farto et al., 2003), while the second, *V. splendidus* GR15, had been isolated from a sponge (Isabelle Coulon, unpublished results). PCR amplification using VS_I1884 specific primers revealed that GR15 carried a copy of the VS_I1884 cassette (99.7% identity) but chromosome walking showed that this cassette was located in a different genetic context: between an ISVisp6 copy inserted inside the *attC* site of its preceding cassette and a non cassette-borne chromosomal gene (data not shown). Interestingly, according to our hybridization probing (see above), this strain, like the other $Trim^R$ P52 strain, apparently lacked the *intIA* gene. For strain P52, as VS_I1884 amplification attempts failed, we constructed an EcoRI genomic library in *E. coli* and selected the $Trim^R$ clones. Sequencing of the unique insert carried in the $Trim^R$ clones, led to the identification a novel *dfr* gene, dfr_{P52}, related to VS_I1884 and to *dfr6* (ca 70% identity with both of these), but the analysis of the surrounding sequences did not reveal any of the cassette structural signatures (ie *attC* specific features). Figure 5A shows the phylogenetic relationships between the different Class A DHFRs and indicate whether they are embedded in integron cassettes (black star) or not (white star). Interestingly, all the integron DHFRs are found in two related sub-groups, and are clearly distinct from the non-cassette borne $Trim^R$ DHFRs. The single exception is the P52 DHFR, which branches among the integron DHFRs, suggesting that this gene was originally cassette-borne, and became fixed in this strain by losing its cassette structural elements. It is also striking to see that three vibrios $Trim^R$ DHFRs group together and are slightly divergent from the DHFR cassettes identified in the common proteobacteria of clinical importance (*E. coli*, *Salmonella*,…). The DfrA6 isolated in *P. mirabilis* is the unique exception, as it is more related to the vibrio DHFR, suggesting that this cassette may have been recruited from a *Vibrio* integron. Another cassette, *dfrA12*, has also been identified in *V. cholerae* but is clearly coming from the enterobacteria pool of *dfr* cassettes, all highly related which form the first subgroup of the *dfr* cassettes (see Figure 5).
The *V. splendidus dfr* cassettes are not the first resistance cassettes discovered inside vibrio characteristic chromosomal integron. The one hosted by different *V. cholerae* isolates has already been occasionally found to carry several such cassettes (*catB9* (Rowe-Magnus et al., 2002), *carb7* and *carb9* (Melano et al., 2002; Petroni et al., 2004) and more recently a *qnr*

cassette (Fonseca et al, 2008), but this was less surprising as these antibiotics are used to cure people from *V. cholerae* infections. In the case of these environmental *V. splendidus* isolates resistance this is more unexpected. Moreover, all 3 sequenced *V. splendidus* genomes have been found to carry the *qnrVS* gene (VS_II0063), which has been recently demonstrated to confer quinolone resistance (Cattoir et al., 2007).

Virulence related genome features

The genome sequence of LGP32 revealed homologs of genes that have been associated with virulence in other organisms ((Table 2 in supplementary data). This includes potential toxins such as haemolysins (Zhang and Austin, 2005), proteases (Miyoshi and Shinoda, 2000), a type VI secretion system (Pukatzki et al., 2006), genes for siderophore transport and utilization (Crosa and Walsh, 2002) and adhesins (Kachlany et al., 2001). Other homologs of genes that have been implicated in resistance to anti microbial peptides (AMPs) have also been identified, i.e. *phoQ* and *phoP* encoding a two-component system (Groissman, 2001) and ompU (Mathur and Waldor, 2004).
The distribution of putative virulence genes varies among the three strains, several being shared by all three, such as the metalloprotease Vsm, the only confirmed virulence factor to date (Le Roux et al., 2007, Binesse et al., 2008). Others putative virulence genes could be shared by two strains, such as the InA-like metalloprotease, and several seem to be specific of one strain, such as the *vas* operon or some siderophores. Such a variation in the carriage of virulence determinants has been reported for other Vibrio species, most notably the genomes and genotypes of V. cholerae strains (Rahman et al., 2008). These data suggest that each *V. splendidus* strain contains a variable set of virulence genes, explaining their observed variable virulence for aquatic animals when tested independantly. On the other hand, the fact that complex *V. splendidus* populations are commonly found in moribund animals (Gay et al 2004 a), and that agonisms between strains have been observed in oyster experimental challenge (Gay et al., 2004b), could indicate that *V. splendidus* vibriosis may result from microbe interactions.

3. Conclusion

V. splendidus represents the most successful vibrio of temperate waters and appears to have relatively recently diversified into a range of ecologically defined populations, including some that result in complex host interaction (Thompson et al., 2005). Such versatility could indicate a large (flexible) gene pool that, if shared efficiently by horizontal gene transfer, can give rise to large numbers of phenotypes including diverse virulence traits.
Surprisingly our analysis revealed relatively few mobile genetic elements in these genomes and notably the absence, or its reduction to a small size, of the chromosomal integron, which is a common element in other vibrios and which has been shown to be the most variable part of the genome in other *Vibrio* species (Chen et al, 2004, Labbate et al 2007). Moreover, although numerous LGP32 regions of genome plasticity were identified, they could rarely be associated to a specific mechanism of recombination, suggesting that they are not of recent origin, or that their mechanism of acquisition is not known yet. Thus the mechanisms at the origin of the high genotypic and phenotypic diversity (as habitat, life style or virulence) within *V. splendidus* are not obvious from our genomic analysis. Further genome sequencing of numerous *V. splendidus* strains may be necessary to understand the basis for this diversity, and will at least help to assess the pan genome of this versatile species.
Finally, the finding of antibiotic resistance genes in the genome of several of these environmental strains points to the major role of the marine bacteria as source of resistance determinants, be they carried in vibrio characteristic integron cassettes or in an apparently

immobile structure such as the *qnrVS* (VS_II0063). Quinolones, such as flumequine, and trimethoprim, are approved in EU and USA for fish treatment in case of infection in aquaculture settings, but they are not approved for mollusk and crustacean farming. These observations suggest that if selection for quinolone and Trim resistance can easily occur through their utilization for fish farming, the circulation of the vibrios environmental species, such as *V. splendidus* may ultimately spread these resistance determinants in seas and oceans, ending in the colonizing flora of likely untreated mollusks and crustaceans. These animals will be handled and ingested by man, and as such able to transfer these genes to human pathogens as already demonstrated for the fish pathogen *Aeromonas salmonicida* (Kruse and Sørum, 1994).

4. Material And Methods

Bacterial strains and media

The *V. splendidus* related strains used in this study have been described previously (Gay et al., 2004a, Thompson et al. 2005). The strains Med222, P52 and Gr15, isolated respectively from the water column, *Octopus vulgaris* and *Grantia compressa*, have been provided graciously by Dr Jarone Pinhassi (Kalmarsundslab, Sweeden), Dr Farto (Facultad de Ciencias, Universidad de Vigo, Spain) and Dr Domart-Coulon (Muséum National d'Histoire Naturelle, Paris). Strains were grown in LB-NaCl 0.5M, marine broth (MB) or marine agar (MA) at 20°C. *Escherichia coli* strains, used in *dfr* knock out experiments, were grown in Luria-Bertani (LB) or, in case of π1, Mueller-Hinton (MH) broth, at 37°C (Demarre et al., 2005). All media were from Difco. Antibiotics were used at the following concentrations: ampicillin (Ap), 100 μg/ml ; chloramphenicol (Cm), 12,5 μg/ml or trimethoprim (Trim), 100 μg/ml. Thymidine (dT) and diaminopimelate (DAP) were supplemented when necessary to a final concentration of 0.3mM.

Cloning, sequencing, assembly and annotation

Genome sequencing was performed using the whole-genome shotgun strategy (Fleishman et al., 1995) as described previously (Frangeul et al., 1999, 2004). A library (1-2 kb inserts) was generated by random mechanical shearing of genomic DNA and cloning into pcDNA-2.1 (Invitrogen). A BAC library was constructed as previously described (Buchrieser et al., 1999). Assembly and annotation are detailed in supplementary data. The transfer of the data in the MaGe software (Magnifying Genome, Vallenet *et al*., 2006) allows dynamic and permanent update of the genome annotations (http://www.genoscope.cns.fr/agc/mage).

In silico comparative genomic

Orthologous genes shared by pairs of *V. splendidus* LGP32, *V. parahaemoliticus*, *V. vulnificus*, *V. cholerae* and *Photobacterium profundum* species were determined through high-stringency sequence homology matching using the program InParanoid (Remm et al., 2001). To identify genes that are present in all these organisms, orthologs family assembly was computed by MultiPARANOID program (Alexeyenko et al., 2006).
Conserved genes between LGP32, *Vibrio splendidus* 12B01 and *Vibrio sp*. MED222 were determined using BLASTN (Basic Local Alignment Search Tool for Nucleotide; http://blast.ncbi.nlm.nih.gov/). LGP32 genomic sequences were used as query on genomic sequences of 12B01 and MED222 strains (accession numbers: NZ_AAMR00000000 and NZ_AAMR00000000 respectively).

Cloning of the V. splendidus LGP31 and V. tasmaniensis LMG 20012T integron platform locus.

We designed a set of antiparallel primers from each of the different genes, VS_I1891 to VS_I1894, located upstream of the LGP32 intIA (VS_I1890), and one primer for each of the *attC* sites associated with each of the six first cassettes of LGP32 in orientation 5'-3' toward the *attI* site. Then we tested the different combination of antiparallel primers, and we cloned the largest amplicon specifically obtained in pTOPO (Invitrogen) according to manufacturer's instructions, and used insert walking to determine the complete insert sequence. This strategy was previously successfully used to characterize the integron and their very first cassettes in different bacterial species (Rowe-Magnus et al , 2001). For LGP31, the largest amplicon was obtained with primers VspORF14 (GTTCAGTCAAATTAGACTGCGG) and LGPattIr2 (ATGAAGTTAACCATACCCCG), and fully sequenced (Gbk FJ529221). Unfortunately, none of the PCR made with combination of the attC primers and antiparallel intIA primers lead to a notable amplicon, suggesting that the LGP31 cassettes are different. For *V. tasmianensis* LMG20012T, primers VspORF13 (CGATTCGACGTTGGAAGATGC) and VspAttc2 (GGAATCACTCTTAAATTGTTTGTTA) gave the largest amplicon, which was then fully sequenced (Gbk FJ529220).

Cloning of the Trim^R DHFR encoding loci from V. splendidus GR15 and P52.

VS_I1884 specific primers (DHFR32s GCCCGAATTCATAAGGAGATAAATTTGAAATTATCATTGATGGC and DHFR32as GCCCGGATCCATTATCAGCCCTTTTGCCAAATTTGGTACG) were used in PCR attempts with low annealing temperature (52°C), and lead to amplification with GR15 genomic DNA. This amplicon sequence revealed 99.7% identity when compared to the LGP32 gene. This PCR attempt failed with strain P52. We then made two genomic libraries for each of the two strains, after digestion of their genomic DNA with HindIII and EcoRI independently and cloning in pUC18. Top10 cells (Invitrogen) carrying the different libraries were then plated on MH Ap Trim medium, and Trim^R clones were reisolated. The inserts carried by the different TrimR plasmids were then digested by different restriction enzymes, subcloned in appropriately digested pUC18, and tested for the TrimR phenotype. The smallest inserts were then sequenced. The sequence of the *dfrA* locus from GR15 has been deposited in Genbank under accession number FJ529219 and the one from P52 under accession number FJ529218.

dfrA inactivation

A 300 bp long internal region of the cassette carrying *dfrA* (VS_I1884) was PCR amplified using primers forward (5'-GCCCGAATTCTGAGAAAATGGAGTAATTGGCTC-3') and reverse (5'-GCCCGGATCCTCAGTCCACTTGCGTAATGAGTG-3'), digested with EcoRI and BamHI and cloned in pSW23T (Demarre et al., 2005). Inactivation of the chromosomal target was performed through conjugation as described previously (Le Roux et al., 2007). Trim sensitivity was established by growth comparison on MH supplemented with 0.5 N NaCl and with or without 100μg/mlTrimwith 0.5 N NaCl, Once dried, a Whatman paper disk carrying 3 mg of trimethoprim was deposited on each plates, they were then incubated at 25°C for 30 hours, before photography.

The genome sequences reported in this paper have been deposited in the EMBL databases under the accession numbers FM954972 and FM954973.

Acknowledgements

The authors acknowledge Dr Isabelle Domart-Coulon, Dr Rosa Farto and Dr Jarone Pinhassi for providing strains and plasmid constructions. We wish to thank Dr Bridget Davis, Dr Céline Loot and Dr Eric Duchaud for critical reading of the manuscript. We are indebted to Dr Dirk Gevers for helpful discussion during the course of this work. This study was carried out with financial assistance from the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS-URA 2171), the Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER) and the Institut de Génomique Marine (contrat Ministère de la Recherche n°0425).

References

Alexeyenko, A., Tamas, I., Liu, G. and Sonnhammer, E.L. (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* **22**:9-15.

Binesse J. , Delsert, C., Champomier-Vergès, M-C., Saulnier, D., Zagorec, M., Mazel, D. and Le Roux, F. (2008) The Metalloprotease Vsm is the Main Toxic Factor for *Vibrio splendidus* Secretome ; *Appl Environ Microbiol.* **74** :7108-7117.

Buchrieser, C., Rusniok, C., Frangeul, L., Couve, E., Billault, A., Kunst, F., Carniel, E., and Glaser, P. (1999) The 102-kilobase pgm locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. *Infect Immun* **67:**4851-61.

Cattoir, V., Poirel, L., Mazel, D., Soussy, C.J. and Nordmann, P. (2007) *Vibrio splendidus* as the source of plasmid-mediated QnrS-like quinolone resistance determinants. *Antimicrob Agents Chemother.* **51**(7):2650-1.

Chen, C. Y., Wu, K. M., Chang, Y. C., Chang, C. H., Tsai, H. C. , Liao, T. L., et al. (2003). Comparative genome analysis of *Vibrio vulnificus,* a marine pathogen. *Genome Res* **13:**2577-87.

Crosa, J. H., and Walsh, C. T. (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* **66:**223-49.

Demarre, G., Guerout, A. M., Matsumoto-Mashimo, C., Rowe-Magnus, D. A., Marlière, P. , and Mazel, D. (2005) A new family of mobilizable suicide plasmids based on the broad host range R388 plasmid (IncW) or RP4 plasmid (IncPa) conjugative machineries and their cognate *E.coli* host strains. *Research in Microbiology* **156:**245-55.

Dryselius, R., Kurokawa, K., Iida, T. (2007) *Vibrionaceae*, a versatile bacterial family with evolutionarily conserved variability. *Res Microbiol.* **158**(6):479-86.

Farto, R., Armada, S.P., Montes, M., Guisande, J.A., Pérez, M.J. and Nieto, T.P. (2003) *Vibrio lentus* associated with diseased wild octopus *(Octopus vulgaris).* J Invertebr Pathol. **83**(2):149-56.

Fleischmann, R. D., Adams, M. D. , White, O. , Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., and et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269:**496-512.

Folkesson, A., Lofdahl, S., and Normark, S. (2002) The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. *Res Microbiol* **153:**537-45.

Fonseca, E.L., Dos Santos Freitas, F., Vieira, V.V. and Vicente, A.C. (2008) New *qnr* gene cassettes associated with superintegron repeats in *Vibrio cholerae* O1. *Emerg Infect Dis* **14**(7):1129-31.

Frangeul, L., Glaser, P., Rusniok, C., Buchrieser, C., Duchaud, E., Dehoux, P., and Kunst, F. (2004) CAAT-Box, Contigs-Assembly and Annotation Tool-Box for genome sequencing projects. *Bioinformatics* **20:**790-7.
Frangeul, L., Nelson, K. E. , Buchrieser, C., Danchin, A., Glaser, P. , and Kunst, F. (1999) Cloning and assembly strategies in microbial genome projects. *Microbiology* **145:**2625-34.
Gay, M., Berthe, F. C., and Le Roux, F. (2004a) Screening of *Vibrio* isolates to develop an experimental infection model in the Pacific oyster *Crassostrea gigas*. *Dis Aquat Organ* **59**:49-56.
Gay, M., Renault, T., Pons, A. M., and Le Roux, F. (2004b) Two *Vibrio splendidus* related strains collaborate to kill *Crassostrea gigas*: taxonomy and host alterations. *Dis Aquat Organ* **62**:65-74.
Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P. and Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol.* **57**:81-91.
Gray, C. G., Cowley, S. C., Cheung, K. K., and Nano, F. E. (2002) The identification of five genetic loci of *Francisella novicida* associated with intracellular growth. *FEMS Microbiol Lett* **215**:53-6.
Groisman, E. A. (2001) The pleiotropic two-component regulatory system PhoP-PhoQ. *J Bacteriol* **183:**1835-42.
Huovinen, P., Sundström, L., Swedberg, G. and Sköld, O. (1995) Trimethoprim and sulfonamide resistance. *Antimicrob. Agents Chemother.* **39**: 279–289.
Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J. and Polz, M.F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**:1081-5.
Kachlany, S. C., Planet, P. J., DeSalle, R., Fine, D. H., and Figurski, D. H. (2001) Genes for tight adherence of *Actinobacillus actinomycetemcomitans*: from plaque to plague to pond scum. *Trends Microbiol* **9:**429-37.
Kruse, H.E., Sørum, H. (1994) Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. *Appl Environ Microbiol;* **60** (11):4015-21.
Labbate, M., Boucher, Y., Joss, M. J., Michael, C. A., Gillings, M. R., and Stokes H. W. (2007) Use of chromosomal integron arrays as a phylogenetic typing system for *Vibrio cholerae* pandemic strains. *Microbiology* **153:**1488-98.
Le Roux, F., and Austin, B. (2006) *Vibrio splendidus*, *In* F. L. Thompson, B. Austin, and J. Swings (ed.), The Vibrios. ASM Press, Washington, DC. Chapter 21p. 285-296.
Le Roux, F., Binesse, J., Saulnier, D., and Mazel D. (2007) Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene *vsm* by use of a novel counterselectable suicide vector. *Appl Environ Microbiol* **73:**777-84.
Mathur, J., and Waldor, M. K. (2004) The *Vibrio cholerae* ToxR-regulated porin OmpU confers resistance to antimicrobial peptides. *Infect Immun* **72:**3577-83.
Mazel, D. (2006) Integrons: agents of bacterial evolution. *Nat Rev Microbiol* **4:**608-20.
Meibom, K. L., Li, X. B., Nielsen, A. T., Wu, C. Y., Roseman, S., and Schoolnik, G. K. (2004) The *Vibrio cholerae* chitin utilization program. *Proc Natl Acad Sci U S A* **101**:2524-9.
Melano, R., Petroni A., Garutti, A., Saka, H.A., Mange, L., Pasterán, F., Rapoport, M., Rossi, A. and Galas, M. (2002) New carbenicillin-hydrolyzing beta-lactamase (CARB-7) from *Vibrio cholerae* non-O1, non-O139 strains encoded by the VCR region of the *V. cholerae* genome. *Antimicrob Agents Chemother.* **46**(7):2162-8.
Miyoshi, S., and Shinoda, S. (2000) Microbial metalloproteases and pathogenesis. *Microbes Infect* **2:**91-8.
Petroni, A., Melano, R.G., Saka, H.A., Garutti, A., Mange, L., Pasterán, F., Rapoport, M., Miranda, M., Faccone, D., Rossi, A., Hoffman, P.S. and Galas, M.F. (2004) CARB-9, a carbenicillinase encoded in the VCR region of Vibrio cholerae non-O1, non-O139 belongs to a family of cassette-encoded beta-lactamases. *Antimicrob Agents Chemother* **48**(10):4042-6.
Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., Heidelberg, J.F. and Mekalanos, J.J. (2006) Identification of a conserved bacterial protein secretion

system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc Natl Acad Sci U S A.* **103**(5):1528-33.
Rahman, M.H., Biswas, K., Hossain, M.A., Sack, R.B., Mekalanos, J.J., Faruque, S.M. (2008) Distribution of genes for virulence and ecological fitness among diverse Vibrio cholerae population in a cholera endemic area: tracking the evolution of pathogenic strains. DNA Cell Biol. **27**(7):347-55.
Reen, F. J., Almagro-Moreno, S., Ussery, D. and Boyd E. F. (2006) The genomic code: inferring *Vibrionaceae* niche specialization. *Nat Rev Microbiol* **4**:697-704.
Remm, M., Storm, C.E. and Sonnhammer, E.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol.* 314(5):1041-52.
Rowe-Magnus, D.A., Guerout, A.M. and Mazel, D. (2002) Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Mol Microbiol.* **43**(6):1657-69.
Rowe-Magnus, D.A., Guerout, A.M., Ploncard, P., Dychinco, B., Davies, J. and Mazel, D. (2001) The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc. Natl. Acad. Sci. U.S.A.* **98** (2), 652-657.
Ruby, E. G., Urbanowski, M., Campbell, J., Dunn, A., Faini, M., Gunsalus, R., et al. (2005) Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. *Proc Natl Acad Sci U S A* **102:** 3004-9.
Thompson, F.L., Thompson, C.C., Swings, J. (2003) Vibrio tasmaniensis sp. nov., isolated from Atlantic salmon (Salmo salar L.) *Syst Appl Microbiol.* **26**(1):65-9.
Thompson, J. R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D. E., Benoit, J., Sarma-Rupavtarm, R., Distel, D. L., and Polz, M. F. (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**:1311-3.
Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., Lajus, A., Pascal, G., Scarpelli, C., Médigue, C. (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* **34**(1):53-65.
Wylie, B.A. and Koornhof, H.J. (1991) Nucleotide sequence of dihydrofolate reductase type VI. *J Led Microbiol* **35** :214-8.
Zhang, X.H., Austin, B. (2005) Haemolysins in Vibrio species. *J Appl Microbiol.* **98**(5):1011-9.

Tables

Table 1: The distribution of functional classes of genes (Cluster of orthologous genes, COG), in percentage of total predicted ORFs, between the two chromosomes and species of *Vibrionaceae* family: *Vibrio splendidus* (VS), *Vibrio vulnificus* (VV), *Vibrio parahaemolyticus* (VP), *Vibrio cholerae* (VC), *Vibrio harveyi* (VH), *Vibrio fisheri* (VF), *Photobacterium profundum* (PP).

	Chromosome 1							**Chromosome 2**						
	VS	**VV**	**VP**	**VC**	**VH**	**VF**	**PP**	**VS**	**VV**	**VP**	**VC**	**VH**	**VF**	**PP**
Cellular process and signaling	18	17,3	17	19,8	15,9	18,6	16,2	6,7	8,7	8,8	5,9	7,9	8,5	6,5
Information storage and processing	11,5	11,7	11,1	12,3	14,6	11,9	12,3	4,6	4,9	4,6	3,8	7,4	4,4	7,6
Metabolism	26,2	24,3	24,9	29,1	23	27,3	25,4	13,5	14,2	14,4	10,5	13,2	11,4	13,9
Conserved hypothetical protein	12,5	11,8	11,7	12,5	11,1	12,4	11,7	6,5	6,8	7	5,7	6,6	5,3	6
Unknown	12,2	13,4	15,6	14,6	15,8	11,8	13,3	9,1	8,6	12	9,7	15,9	8,9	13

Figures

Figure 1

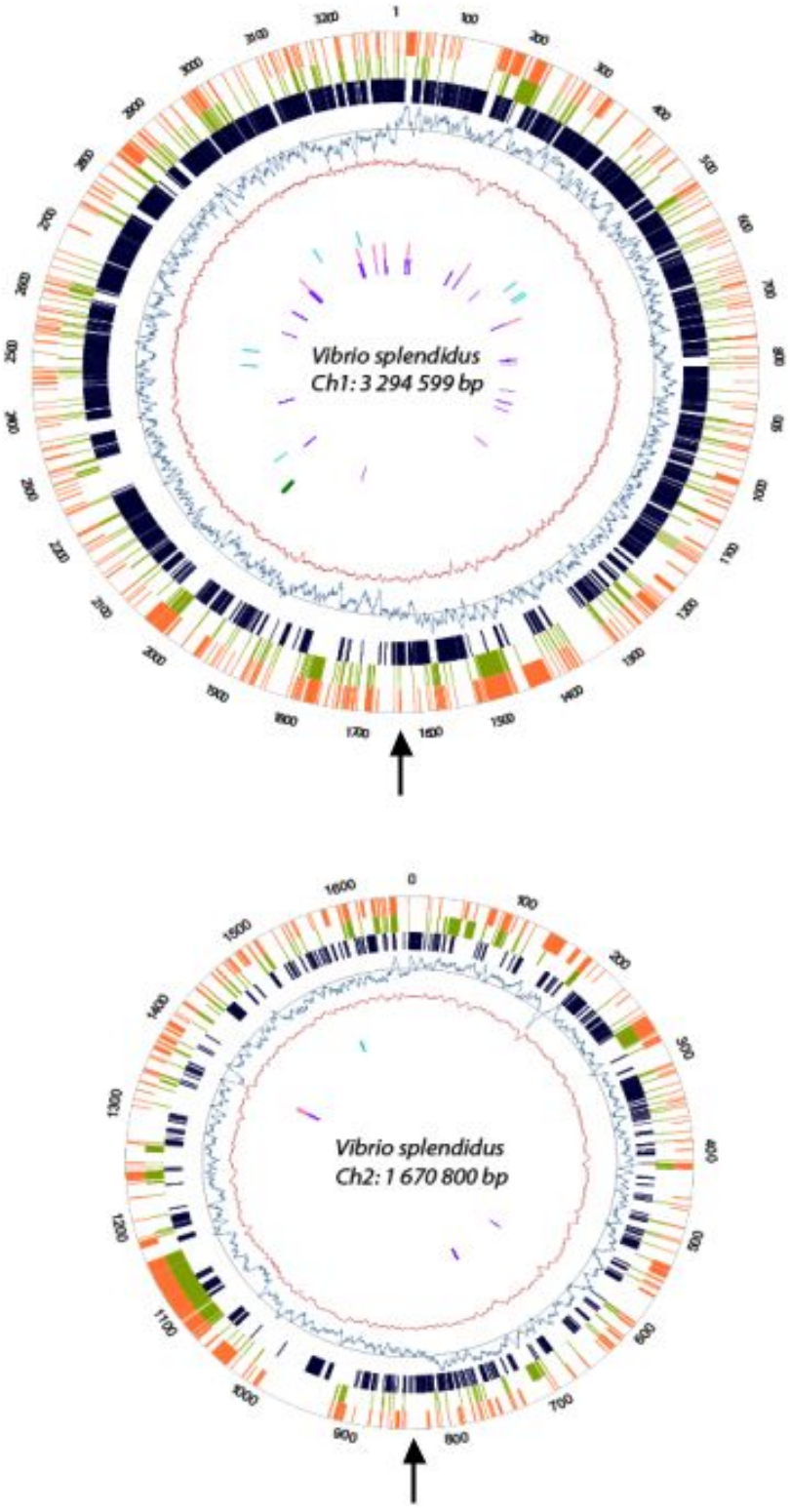

Figure 1: Circular representation of the *V. splendidus* LGP32 genome with comparative genomic data. From the outside inwards: the 1rd circle (orange bars) shows LGP32 ORFs that are absent from the partial genome sequences of 12B01, the 2th circle (green bars) shows LGP32 ORFs that are absent from the partial genome sequences of Med222. The 3th circle (blue bars) corresponds to ORFs common to all sequenced *Vibrio* species. The 4th circle shows GC skew. The 5th circle shows percentage G+C in relation to mean G+C for the genome. The 6th, 7th and 8th circles show insertion sequences (IS), rRNA and tRNA, respectively. The ter site is indicated by an arrow.

Figure 2

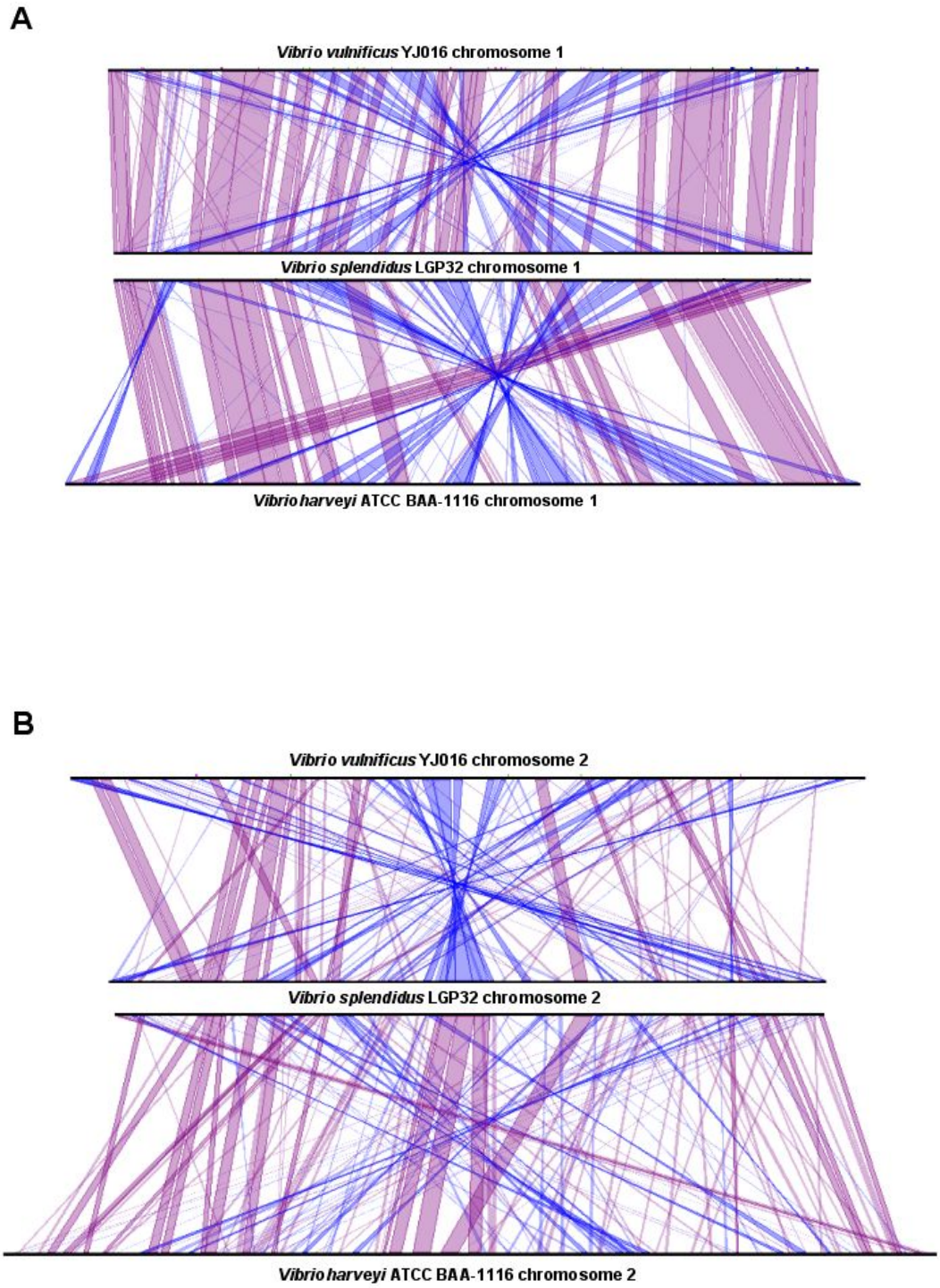

Figure 2: Synteny plot between *V. splendidus* LGP32, *V. vulnificus* YJO16 and *V. harveyi* ATCC BAA-1116 genomes. Putative orthologous relations between LGP32 and other vibrio genomes were defined as gene couples satisfying the bi-directional best hit (BBH) criterion or a blastP alignment threshold, a minimum of 35% sequence identity on 80% of the length

of the smallest protein. These relations were subsequently used to search for conserved gene clusters, e.g. synteny groups (syntons). All possible kinds of chromosomal rearrangements are allowed (inversion, insertion, and deletion). A gap parameter, representing the maximum number of consecutive genes which are not involved in a synteny group, is set to five genes. In purple : synteny groups are organized on the same strand; in blue : synteny groups are organized on two opposite strands. Comparative analysis were analyzed using the MaGe interface in the VibrioScope database (https://www.genoscope.cns.fr/agc/mage).
A. Synteny plot between the chromosome 1 of *V. splendidus* LGP32, *V. vulnificus* YJO16 and *V. harveyi* ATCC BAA-1116. **B.** Synteny plot between the chromosome 2 of *V. splendidus* LGP32, *V. vulnificus* YJO16 and *V. harveyi* ATCC BAA-1116.

Figure 3

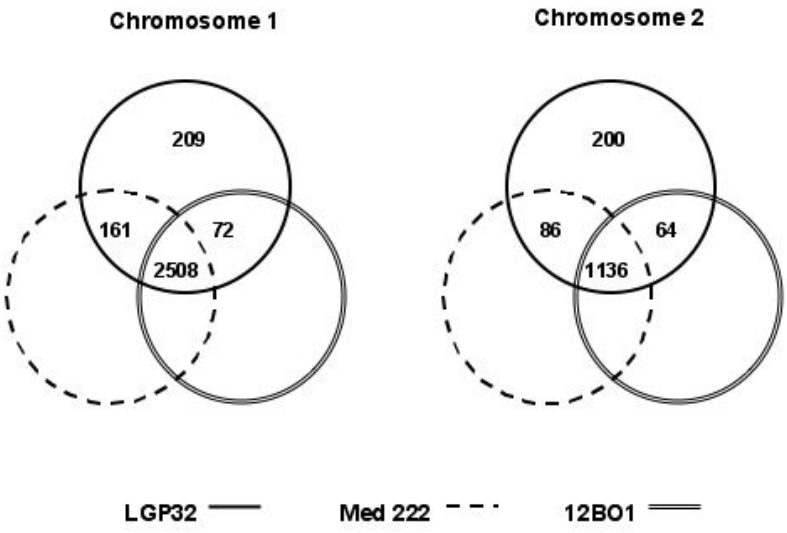

Figure 3: Strain-to-strain variation within *V. splendidus*. Distribution between chromosome 1 and 2 of genes that are specific of LGP32, shared by LGP32 and Med222 or 12B01 and shared by the three strains.

Figure 4

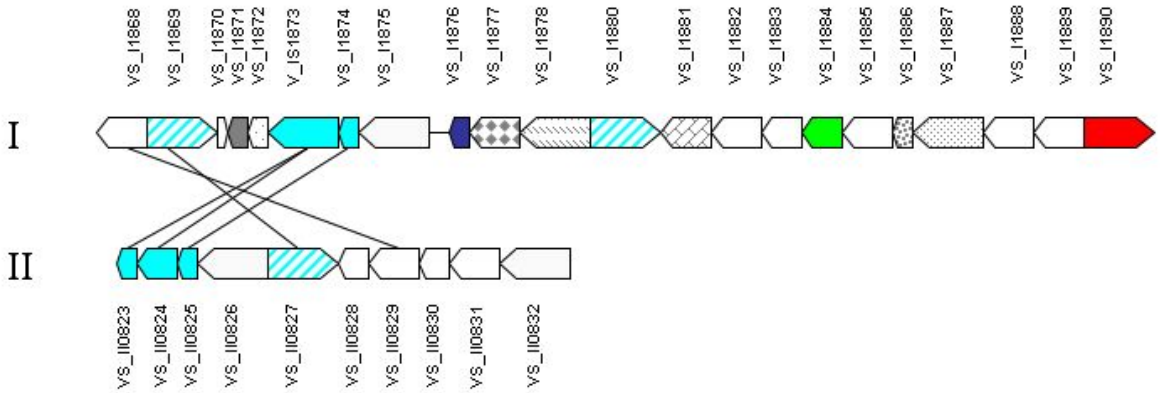

Figure 4: Organization of the *V. splendidus* LGP32 chromosomal integron. The chromosomal integron is split in two arrays. The functional platform, carrying the integrase gene *intIA* (red arrow) and the *attI* cassette integration site, is found in Chromosome 1 and contains genes VS_I1869 to VS_I1889. A second array containing only 5 genes, VS_II0832 to VS_II0828*,* is present on chromosome 2. The presence of transposase genes (blue arrows) suggest that the splitting of the array is due to recombination between identical ISs on the two chromosomes.
ORFs were annotated as: hypothetical protein (VS_I1870, 1875, 1879, 1882, 1883, 1885, 1888, 1889 ; VS_II826, 827, 828, 829, 830, 831); conserved hypothetical protein (VS_I1868, 1871, 1872, 1877, 1878, 1881, 1886, 1887) ; transposase of insertion sequence ISVisp2, IS110 family subgroup IS1111 (VS_I1869, 1880, VS_II827); transposase (orfB) of insertion sequence ISVisp1, IS3 family subgroup IS3 (VS_I1873, VS_II823, 824); transposase (orfA) of insertion sequence ISVisp1, IS3 family subgroup IS3 (VS_I1874, VS_II825) ; putative MutT/nudix hydrolase family protein (VS_I1876) ; dihydrofolate reductase type I (VS_I1884) and ABC transporter, ATP binding protein (VS_II832).

Figure 5

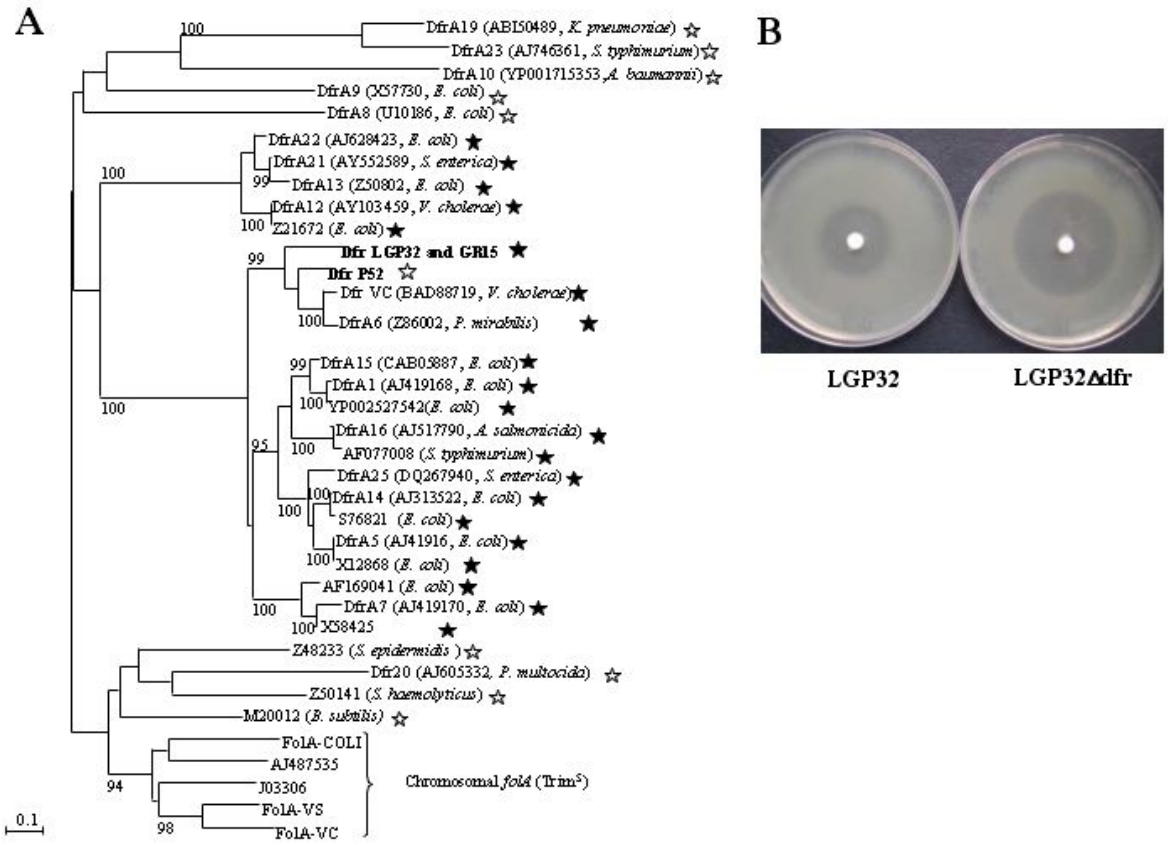

Fig. 5. A. Rooted phylogenetic tree of various DHFR proteins. The tree was built by the Neighbour-joining method based on sequence alignments using clustal x. Branch lengths are drawn to scale and proportional to the number of amino acid changes. Number at each node represents the percentage value given by bootstrap analysis of 1000 replicates. GenBank accession numbers of sequences are indicated. Genes that are present in an integron cassette are marked by a black star and genes that are not in an integron cassette by a white star.
B. Comparative trimethoprim sensitivity of LGP32 and its Δdfr derivative. The larger is the diameter of growth inhibition around the disk, the more sensitive to trimethoprim is the strain.