Applied Microbiology And Biotechnology
2014, Volume 98 Issue 24 Pages 10165–1017
http://dx.doi.org/10.1007/s00253-014-6086-8
http://archimer.ifremer.fr/doc/00241/35244/

Genome sequence of Vibrio diabolicus and identification of the exopolysaccharide HE800 biosynthesis locus.

Goudenege David [2, 3, 4], Boursicot Vincent [1], Versigny Typhaine [1], Bonnetot Sandrine [1], Ratiskol Jacqueline [1], Sinquin Corinne [1], Lapointe Gisele [5], Le Roux Frederique [2, 3, 4, *], Delbarre-Ladrat Christine [1 *]

[1] Ifremer, Laboratoire EM3B (Ecosystèmes Marins et Molécules Marines pour les Biotechnologies), Rue de l'Ile d'Yeu, BP21105, F-44311, Nantes Cedex 3, France
[2] Ifremer, Unité Physiologie Fonctionnelle des Organismes Marins, ZI de la Pointe du Diable, CS 10070, F-29280, Plouzané, France
[3] Sorbonne Universités, UPMC Univ Paris 06, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff Cedex, France
[4] CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff Cedex, France
[5] Université Laval, Institute for Nutrition and Functional Foods, 2440 boul. Hochelaga, Québec, Québec, Canada

* Corresponding authors : Frédérique Le Roux, email address : frederique.le-roux@sb-roscoff.fr ; Christine Delbarre-Ladrat, email address : christine.delbarre.ladrat@ifremer.fr

Abstract :

Vibrio diabolicus, a marine bacterium originating from deep-sea hydrothermal vents, produces the HE800 exopolysaccharide with high value for biotechnological purposes, especially for human health. Its genome was sequenced and analyzed; phylogenetic analysis using the core genome revealed V. diabolicus is close to another deep-sea Vibrio sp. (Ex25) within the Harveyi clade and Alginolyticus group. A genetic locus homologous to the syp cluster from Vibrio fischeri was demonstrated to be involved in the HE800 production. However, few genetic particularities suggest that the regulation of syp expression may be different in V. diabolicus. The presence of several types of glycosyltransferases within the locus indicates a capacity to generate diversity in the glycosidic structure, which may confer an adaptability to environmental conditions. These results contribute to better understanding exopolysaccharide biosynthesis and for developing new efficient processes to produce this molecule for biotechnological applications.

Keywords : Genome, Vibrio, Exopolysaccharide, Biosynthesis

1. Introduction

Vibrionaceae are ubiquitous marine bacteria that are ecologically and metabolically diverse members of planktonic and animal associated microbial communities (Thompson et al. 2004). They encompass the ancient and well-studied human pathogen, *Vibrio cholerae*, as well as some less thoroughly characterized animal pathogens (Thompson et al. 2004). Perhaps less widely recognized are the species inhabiting extreme environments such as deep seas. The well-studied *Photobacterium profundum* SS9 has been adopted as a model organism to understand the molecular basis of piezophile adaptation (El-Hajj et al. 2010; Vezzi et al. 2005).

Vibrio diabolicus was isolated from a sample collected on the polychaete annelid *Alvinella pompejana* from a deep-sea hydrothermal vent on the East Pacific Rise. It is a facultative anaerobic, heterotrophic, and mesophilic organism that is surprisingly easily cultivable under classical laboratory conditions (Raguenes et al. 1997). Upon cultivation, this bacterium is able to produce an exopolysaccharide (EPS), designated HE800, composed of a tetrasaccharide repeating unit with two *N*-acetyl-hexosamine and two glucuronic acid residues: [3)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*A-(1→4)-β-D-Glc*p*A-(1→4)-α-D-Gal*p*NAc-(1→]$_n$ (Rougeaux et al. 1999). The HE800 structure shows similarities with hyaluronan and chondroitin sulfate that are glycosaminoglycans (GAGs) used in various medical applications (Senni et al. 2013).

All GAGs found in animal tissues (hyaluronic acid, heparan sulfate, chondroitin sulfate, dermatan sulfate and keratan sulfate) play critical biological roles in various pathological and physiological cellular and matrix processes as they can bind a variety of proteins involved in cell-cell interactions (for review see (Gandhi and Mancera 2008)). Therefore, these molecules present a great potential for the design and preparation of therapeutic drugs (Islam and Linhardt 2003). Interestingly, the bioactivity of HE800 in bone repair has been described, suggesting that this polymer in its native and derivative forms could find applications in skin and cartilage therapy (Senni et al. 2013; Zanchetta et al. 2003). The use of bacteria as a source of GAGs offers many advantages over plant, macroalgae and animal sources: EPSs are easily produced by fermentation in laboratory conditions allowing a stable supply independent of climatic, seasonal or physiologic variations, and are safer products with respect to prions/pathogen risks (De Angelis 2012). Deciphering the genetic basis of this bacterial EPS biosynthesis may enhance its biotechnological applications.

Bacterial polysaccharides include EPSs which are released to the surrounding medium, and two surface polysaccharides: lipopolysaccharides (LPS) with an *O*-antigen polysaccharide linked to the Lipid A core complex and capsular polysaccharides (CPS) with K-antigen. They are all composed of repeating sugar units and exhibit therefore similar biosynthesis mechanisms (Whitfield 2006). The biosynthesis begins with the production of activated sugars in the central metabolic pathway, they are used for other purposes and their biosynthetic pathways are encoded by housekeeping genes. The repeating unit is then synthesized by appropriate glycosyltransferases (GTs) (Rehm 2010; Whitfield 2006). After completion, the repeating unit is exported outside the cell and polymerized on the growing EPS chain. Based on biosynthesis and export mechanisms, three pathways have been described. In the first biosynthetic pathway, a multifunctional processive GT, the synthase, catalyzes both polymerisation and export (Weigel and De Angelis 2007). The second pathway relies on the ABC transporter for exportation of the biosynthesized polymer. The third mechanism depends on Wzx-Wzy proteins (Whitfield 2006). The Wzx-Wzy dependent mechanism has been widely studied in Gram negative bacteria especially for

heteropolysaccharide production. A particular GT involved in the initiation step, the phosphoglycosyltransferase (pGT), links the first osidic residue to a membrane lipid carrier through a phosphoryl bond (De Vuyst et al. 2001; Whitfield 2006). After completion of the repeating unit by successive GTs, it is exported outside the cell across the inner membrane by Wzx and subsequently polymerised by the Wzy protein through addition to the growing EPS chain on the outer face of the inner cell membrane (Whitfield 2006). The final translocation across the outer membrane involves a member of the outer membrane polysaccharide export protein family such as Wza (Reid and Whitfield 2005).

The present study aimed at determining which biosynthetic pathway is involved in HE800 production and to elucidate the distinctive characters of the deep sea *V. diabolicus* strain CNCM I-1629. To this end, we sequenced the genome, and compared it i) to one publicly available conspecific isolate (Ex25; Los Alamos National Laboratory), ii) to other members of the *Vibrionaceae* family and iii) to other bacteria inhabiting deep sea ecosystems. Polysaccharide biosynthesis was analysed by a computational approach and knock out mutants were generated to give functional evidence that an identified genetic locus is involved in HE800 EPS biosynthesis.

2. Material and methods

2.1. Bacterial strains, plasmids and media

The bacterial strains and plasmids used in this study are described in Table 1. *Escherichia coli* strains Π3813 and β3914 (Le Roux et al. 2007) were used for cloning and conjugation, respectively. *E. coli* strains were grown in Luria-Bertani (LB) medium incubated at 37°C. *V. diabolicus* strain CNCM I-1629 was grown at 30°C in LB-NaCl 0.5M, marine broth (MB) or marine agar (MA). Conjugations were performed at 37°C. All media were from BD Difco (Fisher Scientific, Illkirch, France). Antibiotics were used at the following concentrations: chloramphenicol (Cm), 12.5 μg mL^{-1}; erythromycin (Erm), 200 μg mL^{-1}; kanamycin (Km), 25 μg mL^{-1}; thymidine (dT) and diaminopimelate (DAP) were supplemented when necessary to a final concentration of 0.3 mM. Induction of *ccdB* expression under the control of the P_{BAD} promoter was achieved by adding 0.2% L-arabinose to the growth media and repressed by 1% D-glucose.

2.2. Vector construction and mutagenesis

The *V. diabolicus* CNCM I-1629 *ΔsypR* and *ΔsypK* derivatives (GV694 and GV708) were constructed by allelic exchange using the method described previously (Le Roux et al. 2007). Briefly, alleles carrying an internal deletion were generated in vitro using a two-step PCR construction method (Binesse et al. 2008) using primers 1 to 4 (Table S1, Online Resource) and cloned into pSW7848, a R6K γ-*ori*-based suicide vector that encodes the *ccdB* toxin gene under the control of an arabinose-inducible and glucose-repressible promoter, P_{BAD}. Matings between *E. coli* and *Vibrio* were performed at 30°C as described previously (Le Roux et al. 2007). Selection of the plasmid-borne drug marker resulted in the integration of the entire plasmid in the chromosome by a single crossover. Elimination of the plasmid backbone resulting from a second recombination step was selected by arabinose induction of the *ccdB* toxin gene. Mutants were screened by PCR using external primers 5 and 6 (Table S1, Online Resource). For the complementation experiment, the *sypK* gene was amplified using primers 7 and 8 and inserted in place of *gfp* in MRB (Le Roux et al. 2011). The plasmids (pMRB-P_{LAC} *gfp* or *sypK*) were transfered by conjugation to GV708 (*V. diabolicus*

CNCM I-1629 Δ*sypK*), leading to GV729 and GV735. The constitutive expression of *gfp*, under the control of the P_{LAC} promoter, was confirmed by epifluorescence (not shown).

2.3. Genomic studies

The genome sequence of *V. diabolicus* CNCM I-1629 was obtained using Illumina Genome Analyser II (GATC Biotech, Konstanz, Germany) from a 3 kb mate pair library. Assembly was performed by GATC Biotech using Genome Analyser Contigs. Computational prediction of coding sequences (CDSs) and other genome features (RNA encoding genes, ribosome binding sites, signal sequences, etc...), together with functional assignments were performed using the automated annotation pipeline implemented in the MicroScope platform (Vallenet et al. 2013). Some functional assignments were manually curated using InterPro, FigFam, PRIAM, COGs, PsortB, TMHMM and synteny groups computation. Expert annotation was also performed using Blast searches (Koriblast, Korilog, France). The genome sequences reported in this paper have been deposited in the EMBL database under the accession number PRJEB5898.

To investigate the core and flexible genomes, an all-versus-all BlastP search was performed using genomic sequences of 378 organisms: 377 *Vibrionaceae* and 1 *Shewanella baltica* (strain OS155, complete genome) available in GenBank (http://www.ncbi.nlm.nih.gov/genbank/). A dedicated precomputing repository (marshalling) was created to perform comparative genomic and phylogenomic analyses. Orthologous proteins were defined as reciprocal best hit proteins with 80% MaxLrap and a minimum of 30% identity and >75 amino acid length cutoff (Daubin et al. 2002). The nucleic acid sequences were aligned using Muscle (Edgar 2004) and filtered by BMGE (Criscuolo and Gribaldo 2010). Phylogenetic trees were built using FastTree2, an approximately maximum-likelihood inference tool, with GTR model as parameters (J-C distance, NNIs, SPRs) (Price et al. 2010). The Venn diagram was built using Vennerable R package (http://r-forge.r-project.org/projects/vennerable, Jonathan Swinton, 2009). CAZy family of glycosyltransferases was determined by B. Henrissat and P. Coutinho (Carbohydrate-Active enZYmes database, (Lombard et al. 2014), http://www.cazy.org/).

2.4. Polysaccharide preparation and analysis

The polymers were produced in MB medium supplemented with 30 g L^{-1} glucose in a bio-reactor as previously described (Rougeaux et al. 1999) or when necessary, in batch in Erlenmeyers during a 48 h-incubation. EPS were recovered from the supernatant, ultrafiltered using a 100 kDa membrane (OMEGA PES membrane, Pall Life Sciences, Saint-Germain-en-Laye, France) on a Pellicon 2 system (Millipore, Saint-Quentin en Yvelines, France) and a Masterflex pumping system (Cole-Parmer, Fisher Scientific, Illkirch, France). After freeze-drying, their osidic composition and molecular weight were determined.

EPS molecular weight (MW) was obtained from size exclusion chromatography combined with a multi-angle laser light scattering detector (SEC-MALS). The system consisted in a column (PL Aquagel-OH, Varian, Les Ulis, France), a pump and an injector (HPLC Prominence, Shimadzu, Marne La Vallée, France). Elution was performed at 1 mL min^{-1} with 0.1 M ammonium acetate filtered on a 0.1-µm membrane. A refractive index (RI) detector (Hitachi L2490, VWR, Fontenay sous Bois, France) and a MALS detector (Dawn Heleos II™; Wyatt Technology, Toulouse, France) were coupled on-line. Samples were centrifuged and filtered on a 0.45 µm membrane prior to injection (100 µL). Data for MW determinations and conformation were analyzed with Astra software (Wyatt Technology) based on a dn/dc of

0.145. The number-average molecular weight (M_n), the weight-average molecular weight (M_w), the polydispersity index (M_w/M_n), z-average radius of gyration (Rz) of the samples were determined.

To determine osidic composition, colorimetric assays were first performed: neutral oses by orcinol (Rimington 1931; Tillmans and Philippi 1929), uronic acids (Blumenkrantz and Asboe-Hansen 1973; Filisetti-Cozzi and Carpita 1991) and osamines including *N*-acetyl-osamines (Belcher et al. 1954). Osidic residues were then identified and quantified by gas chromatography after trimethylsylilation (Kamerling et al. 1975; Montreuil et al. 1986) using a CP-SIL 5CB column in 100% dimethylpolysiloxane. Proteins were assayed by BCA (bicinchoninic acid). Proteins and sugar residues were expressed as mass % (w/w).

3. Results

3.1. General features of the *Vibrio diabolicus* CNCM I-1629 genome

Illumina sequencing and assembly of the *V. diabolicus* CNCM I-1629 strain resulted in the generation of a draft genome sequence that contains a total of 29 contigs (5.13 Mb in size and 44,8% GC). 5161 coding DNA sequences (CDSs), 5 rRNAs operons and 67 tRNAs were predicted by MaGe annotation tools (Microbial Genome Annotation and Analysis Platform) (Vallenet et al. 2013). Because the genome assembly is not closed, some rRNA operons may have been missed in the assembly.

Based on 50 *Vibrionaceae* genome sequences and using *Shewanella baltica* as an outgroup, a total of 433 shared proteins was selected to study the phylogenetic relationship between *V. diabolicus* and other *Vibrionaceae* (Fig. 1). *V. diabolicus* strain CNCM I-1629 belongs to the Harveyi clade that contains *V. harveyi*, *V. parahaemolyticus* and *V. alginolyticus* type strains (Sawabe et al. 2013). Interestingly, its closest phylogenetic neighbor is the strain Ex25 that has also been isolated from samples collected from a deep-sea hydrothermal vent community along the East Pacific Rise. The strains *V. diabolicus* CNCM I-1629 and *Vibrio* sp. Ex25 contain a closely related core genome, as indicated by an average nucleotide identity (ANI) (Goris et al. 2007) of 98%, while *V. alginolyticus* strains are more distantly related (95%).

Comparative genomic analyses were performed between *V. diabolicus* CNCM I-1629, its two closely-related phylogenetic neighbors (*Vibrio* sp. Ex25 and the *V. alginolyticus* strain 12G01) and the more distantly related *Photobacterium profundum* SS9 (Vezzi et al. 2005), a piezophilic *Vibrionaceae* isolated from deep sea (Fig.2). A total of 423 CDSs were found in *V. diabolicus* CNCM I-1629 and not in the 3 other strains. Comparing to the other *Vibrionaceae* genomes available, 120 genes were unique to the *V. diabolicus* genome. A large part of these strain specific genes were located in 10 putative genomic islands (GIs), 5 from phage origin. In addition, the chromosomal integron (Mazel et al. 1998) contains the gene encoding the integrase (IntIA: VDIABv1_220004) and 122 cassettes (VDIABv1_160001 to 220003; VDIABv1_271228 to 290008) encoding membrane proteins (n=29), acetyl transferases (n=7) and a majority of hypothetical proteins. The predominance of secondary metabolism and cell surface modification genes in the integron has been associated previously with a role in competition and predation defence (Boucher et al. 2011).

A total of 148 genes shared by *V. diabolicus* CNCM I-1629 and *Vibrio* sp. Ex25 are absent from *V. alginolyticus* and *P. profundum* but are present in some other *Vibrionaceae* (Fig. 2).

We also identified 1072 CDSs (21% of the *V. diabolicus* genome) that are shared between *V. diabolicus* CNCM I-1629, *Vibrio* sp. Ex25 and *V. alginolyticus* and absent from *P. profundum* (Fig. 2). Finally 2337 CDSs (45% of the *V. diabolicus* genome) were shared with the three other strains (Fig. 2). As deep seas represent specific environmental conditions (for example, high pressure, oligotrophic nutrient conditions, low or high temperature with large gradients), adaptive traits may be highlighted by the comparison of *V. diabolicus* genome with other bacteria inhabiting such environments. We found that 161 CDS are shared by *V. diabolicus* CNCM I-1629, *Vibrio* sp. Ex25 and *P. profundum*, but all these genes were also found in other *Vibrionaceae* (Fig. 2). Thus, *V. diabolicus* might not be endemic to the deep-sea hydrothermal environment. A more generalist behavior is in accordance with its cultivability under lab conditions (Raguenes et al. 1997) and the isolation of other *V. diabolicus* strains from other habitats (Badhai et al. 2013; Klein et al. 2014).

3.2. Identification of polysaccharide biosynthesis loci in the *V. diabolicus* genome

As a first step toward analyzing *V. diabolicus* polysaccharide biosynthesis pathways, genes encoding carbohydrate-active enzymes (CAZYmes) were identified (Table S2, Online Ressource, lists GTs) (Coutinho et al. 2003). Among 31 GTs belonging to different CAZy families, 18 were colocalized in four loci (Fig. 3), 3 of these loci also contain gene(s) encoding for a pGT.

A first locus (Fig. 3 A) contains 3 GT encoding genes (5 GT2, 2 GT4 and 1 GT9). The GT9 family is specific to the biosynthesis of LPS Lipid A core. The GT2 family catalyses oligosaccharide synthesis. The last part of this locus (Fig. 3 A) contains 5 GT and 2 pGT genes which could be involved in oligosaccharide synthesis of the LPS. However, it is noteworthy that the two copies of *wzx* genes, the LPS acyltransferase and a GT encoding genes are fragmented, suggesting that the *O*-antigen biosynthesis is not functional. Therefore, this locus might not be associated with a complete LPS biosynthesis.

The second locus (Fig. 3 B) is similar to the *cps* locus of *Vibrio* sp. strain Ex25 and resembles capsular polysaccharide gene loci also found for group I capsules in *E. coli* (Keenleyside et al. 1992; Rahn et al. 1999). *cpsC* (VDIABv1_110287) and the fragmented *cpsD* (VDIABv1_110289) show similarity to *E. coli* K12 *wza* and *wzc* respectively; no *wzb*, usually associated with *cps* clusters, could be assigned (Drummelsmith and Whitfield 2000; Whitfield 2006); such spontaneous deletions have already been described without any influence on the biosynthesis of the CPS (Guo and Rowe-Magnus 2010). This locus also presents synteny with the VPA1403-1412 locus from *V. parahaemolyticus*, which has been designated as an exopolysaccharide gene locus related to the rugose colony phenotype (Chen et al. 2010). Therefore, this locus could be associated with capsular as well as exopolysaccharide biosynthesis. In *V. diabolicus*, this locus also contains *cpsG* (VDIABv1_110292) annotated as a Lipid A ligase suggesting the capsular polysaccharide is linked to Lipid A (K_{LPS}) (Rahn et al. 1999; Whitfield 2006). We suggest that this locus is a K_{LPS} cluster rather than a CPS or an EPS cluster.

The third C locus comprises genes involved in the $(Kdo)_2$-Lipid A biosynthesis and is similar to the Lipid A core biosynthesis locus of *V. vulnificus* ATCC 27562 (Nakhamchik et al. 2007). Neither *wzx*-like gene for the translocation of *O*-antigen nor the *waaL* gene to ligate the *O*-antigen to the Lipid A core was identified. Therefore, this locus is most probably involved in the biosynthesis of the Lipid A core part of LPS.

The fourth and last locus (Fig. 3 D) is similar to the symbiosis exopolysaccharide biosynthetic genes (*syp*) of *V. fischeri* (Yildiz and Visick 2009; Yip et al. 2005). *V. fischeri* produces a specific biofilm to promote colonization of its eukaryotic host, the squid *Euprymna scolopes* (Morris and Visick 2010; Yip et al. 2005). Formation of this biofilm requires the *syp* locus containing genes involved in polysaccharide biosynthesis and encoding regulatory proteins (Table 2). Among genes involved in polysaccharide biosynthesis, the *V. diabolicus syp*-like locus contains one pGT (*syp*R), 5 GT4 (*syp*H, I, J, N) and 1 GT2 (*syp*Q). *syp*Q has been associated with poly-*N*-acetylglucosamine biosynthesis in *V. parahaemolyticus* (Ye et al. 2014). Genes encoding for a periplasmic protein involved in polysaccharide export (Wza), an oligosaccharide translocase (Wzx), a polymerase (Wzy) and a putative polysaccharide biosynthesis chain length regulator (Wzc) indicate that the EPS biosynthesis and export depends on the Wzx-Wzy mechanism (Whitfield 2006).

In addition to the genes involved in polysaccharide biosynthesis, three genes were predicted to encode regulatory proteins such as an anti-sigma factor antagonist (*sypA*), a response regulator with a σ54 interaction domain (*sypG*) and a hybrid sensor kinase (*sypF*). The SypF kinase, involved in *V. fischeri* biofilm formation (Visick 2009), is truncated in *V. diabolicus.* In addition, the *sypM* and *sypE* genes present in the *V. fischeri* cluster are absent in *V. diabolicus. SypM* encodes an *O*-acetyltransferase (Shibata et al. 2012). *SypE* encodes a multidomain response regulator and has been demonstrated to inhibit biofilm formation by *V. fischeri* when induced by overexpression of SypG (Morris and Visick 2013).

Altogether, our analysis suggests that among the polysaccharide biosynthesis loci identified in the *V. diabolicus* genome, only the *syp*-like cluster may be involved in the HE800 EPS biosynthesis under laboratory conditions.

3.3. The *syp* cluster controls HE800 production in *V. diabolicus*

Because genome analysis suggested that the *syp*-like cluster may be involved in HE800 production, allelic exchange was used to disrupt potential biosynthetic genes as described previously (Le Roux et al. 2007). The genes *sypR* and *sypK*, coding for a pGT and an oligosaccharide translocase respectively, were deleted in *V. diabolicus*.

EPS productions by the wild type (wt) and *ΔsypR* were studied in fermenter. The yield, respectively 1.18 g $L^{-1}$and 1.12 g L^{-1}, was not affected by the deletion. The molecular weight (MW) profile of the polymers determined by SEC-MALS showed one peak for the wt (1.6 x 10^6 g mol^{-1}) and three for *ΔsypR* (1.5 x 10^6, 6.5 x 10^5 and 2.1 x 10^6 g mol^{-1}) suggesting that *sypR* deletions altered the polymer MW and homogeneity. The osidic composition and protein content of the *ΔsypR* polymer showed a reduced amount of sugars and an increased protein content (20.2%). The osidic composition was 7.7, 2.9 and 3.3 % of glucuronic acid, *N*-acetyl-glucosamine and *N*-acetyl-galactosamine, thus a 2:1:1 ratio similar to that of the wild type (Rougeaux et al. 1999). This result suggests that the *sypR* deletion may be complemented by other pGT encoding genes present in the genome (Table S2, Online Resource).

sypK gene deletion influence on the polysaccharide produced was studied together with complementation experiments. The *sypK* gene was cloned under the control of the constitutive P_{LAC} promoter in a MRB plasmid previously shown to be stable in *Vibrionaceae* (Le Roux et al. 2011) and the construct was transferred to the wild type and the *ΔsypK* mutant. As a control, the GFP encoding gene was cloned in the same expression plasmid. In

order to perform in parallel the study of the wt, the *ΔsypK* mutant and the complemented strains, the experiments were carried out in Erlenmeyers. wt and wt complemented strains produced the biopolymer at similar yield while Δ*sypK* and Δ*sypK* pMRB-P_{LAC} *gfp* showed a high decrease in the production yield (Fig. 4). Neither pMRB-P_{LAC} *gfp* nor pMRB-P_{LAC} *sypK* affected the osidic composition of the produced polymers by the wt complemented strains with respect to ratios of glucuronic acid, *N*-acetyl-glucosamine and *N*-acetyl-galactosamine which were similar to the 2:1:1 ratio of the wt. However, the polymer produced by the Δ*sypK* strain contained neither GlcNAc nor GalNAc; addition in the Δ*sypK* strain of plasmid pMRB-P_{LAC} *gfp* confirmed the absence of GlcNAc. A peak corresponding to GalNAc in gas chromatography was measured although we can not explain it from the addition of GFP. In the Δ*sypK* pMRB-P_{LAC} *sypK* strain, the production yield is higher. Nonetheless, the production level of the wt strains was not restored; this could be explained by the fact that a gene complementation outside the original biosynthetic genetic locus is not as effective as the wild type situation. Both *N*-acetyl hexosamine residues are found within the extracted polymer from the *sypK* complemented strain (Fig. 4) showing that *sypK* is able to translocate efficiently oligosaccharides containing GlcA, GlcNAc and GalNAc out of the cell. We can conclude that the complementation by *sypK* allowed the recovery of a polysaccharide with an osidic composition similar to the wild type one and that this gene is involved in HE800 production.

HPSEC-MALS gave molecular weights of $2.027\ 10^6$ ±0.319% for the wt, $2.465\ 10^6$ ±0.526% for the wt pMRB-P_{LAC} *gfp*, $2.757\ 10^6$ ±0.381% for the wt pMRB-P_{LAC} *sypK,* $3.580\ 10^6$ ±0.262% for the Δ*sypK* strain, $3.201\ 10^6$ ±0.205% for Δ*sypK* pMRB-P_{LAC} *gfp*, $5.671\ 10^6$ ±0.202% for the Δ*sypK* pMRB-P_{LAC} *sypK.* From a general point of view, the wt and complemented wt strains produce polysaccharides with lower molecular weight than the Δ*sypK* strains series. The lower the yield, the higher the molecular weight.

Altogether, these results show that *sypK* is required for HE800 production by *V. diabolicus* and highlight the role of the *syp*-like locus in the synthesis of this polymer.

4. Discussion

Vibrionaceae are widely distributed over the ocean. Global Ocean Sampling Expeditions revealed that within 16SrRNA sequences, ribotypes (sequences exhibiting more than 97% identity) distribution was not homogeneous all over the ocean and different microbial communities could be distinguished with direct adaptive implications (Rusch et al. 2007). In the same way, a recent comparative MLSA study has shown *Alteromonas* strains can be clustered in deep-sea and surface groups (Ivars-Martinez et al. 2008). On the contrary, our results show that *V. diabolicus* does not possess any genes specific to deep-sea life although it has been first isolated from deep sea. This suggests a widespread biogeographical distribution of this species.

Four gene clusters have been assigned to polysaccharide biosynthesis in *V. diabolicus*: two to LPS biosynthesis, most probably involved in the Lipid A core part of LPS biosynthesis, one to K_{LPS} biosynthesis (a capsular or exo-polysaccharide that remains linked to Lipid A) and an exopolysaccharide cluster homologous to the *syp V. fischeri* cluster.

No complete LPS biosynthesis cluster which would include the whole set of genes for Lipid A, inner and outer core oligosaccharide, together with *O*-antigen biosyntheses was identified in *V. diabolicus*. Two Lipid A core biosynthesis clusters have been identified and we can not rule out the complementation of the fragmented genes, in particular for the *O*-antigen export and ligation (*wzx* and *waaL*-like genes), by homologs located elsewhere on the genome. On the other hand, LPS can exist as a molecule composed solely of Lipid A and the core oligosaccharide without the *O*-antigen oligosaccharide. Some colonies of *E. coli* and *Salmonella enterica* have been shown to exhibit a rough or a smooth appearance depending on the LPS structure: rough LPS is composed of Lipid A and the core oligosaccharide, whereas a smooth one includes also the *O*-antigen component (Caroff and Karibian 2003). LPSs in the outer membrane of Gram-negative bacteria provide cell integrity especially needed in harsh oceanic conditions and are important for the growth of the bacteria (Anwar and Choi 2014; Dong et al. 2014). They are also related to pathogenicity and virulence (Faruque and Mekalanos 2003), although in the case of marine Gram-negative bacteria, LPS often show low virulence (Leone et al. 2007). These molecules display great structural diversity depending on the bacterial species (Anwar and Choi 2014). Virulence depends on the core oligosaccharide and *O*-antigen domains whereas Lipid A induces inflammatory response and is known as endotoxin; therefore, core oligosaccharide and *O*-antigen parts of the LPS may encounter a high level of genetic mutations resulting in variation in the structure and biological activity (Anwar and Choi 2014). The assumptions that *V. diabolicus* LPSs may be only composed of Lipid A and the core would need experimental evidence. The determination of the structure of LPS is a challenge because of the complexity and heterogeneity of these molecules (Anwar and Choi 2014) and could be envisioned as a future research work.

Altogether, our analysis suggests that among the polysaccharide biosynthesis loci identified in the *V. diabolicus* genome, only the *syp*-like cluster may be involved in the HE800 EPS biosynthesis under laboratory conditions. This has been confirmed by inactivated mutant construction. In *V. fischeri*, the *syp* cluster has been shown to have a major role in biofilm formation and colonization (Morris and Visick 2010). By analogy to *V. fischeri,* a role of the *syp* genes in *V. diabolicus* biofilm formation can be hypothesized, with distinct regulation due to the absence of specific regulators such as SypF and SypE. Some other not shown preliminary data suggest an antagonist effect of HE800 polymer on adhesion capability in vitro. However, *V. diabolicus* is a poor biofilm former when tested in laboratory conditions (unpublished results). Thus, optimization of the culture conditions controling EPS expression and composition is a prerequisite towards investigating biofilm formation by *V. diabolicus*. This strain has been collected from polychaete annelid dorsal teguments (Raguenes et al. 1997) and it may therefore have some biofilm formation capability. It is highly probable that the HE800 EPS could be involved in attachment to surfaces, along with protection of the bacterial cells from toxic compounds in its original deep sea environment. In the deep sea hydrothermal ecosystems, biofilms provide homeostasis in the face of fluctuating and harsh conditions (extreme temperatures, pH and exposure to ultraviolet (UV) light) (Jannasch and Taylor 1984).

Wzy-dependent mechanism for heteropolysaccharide biosynthesis is widely distributed in Gram-negative bacteria and more generally across all the bacteria (Whitfield 2010; Whitfield and Paiment 2003). However, the *V. diabolicus syp*-like gene cluster comprises an excess of GT genes. Indeed, from the structure of the EPS HE800 repeating unit, it was reasonable to expect that 3 GT genes at the most were necessary. The structure of the EPS produced by *V. fischeri* has not been characterized, but osidic composition analysis showed the presence of glucose, galactose, mannose, rhamnose and arabinose (Rodrigues and Bhosle 1991); therefore, in *V. fischeri*, a high number of GT genes should be probably necessary for *syp* polysaccharide biosynthesis. On the other hand, such an excess number has already been

described in the teichuronic acid operon of *Bacillus subtilis* 168 (Soldo et al. 1999) and this could probably account for EPS versatility resulting in adaptability, as it has also been described for *O*-antigens (Lerouge and Vanderleyden 2002). The *syp* cluster in *V. fischeri* is also important for colonization (Yip et al. 2005) and its expression could be variable, depending upon changes of the environment, host, as well as surface nature resulting in a high level of adaptability. SypI GT has been shown to be less critical for the production of the molecule, suggesting it only adds a structure, such as a branch, to the polysaccharide (Shibata et al. 2012). In *Bacteroides fragilis*, 8 CPS gene clusters exist and are differentially expressed by promoter switching depending of growth conditions; in the same way, some in vivo conditions induce the DNA inversion of the promoter of a high-molecular-weight EPS genes (Chatzidaki-Livanis et al. 2008). Therefore, if six GT genes are present in *V. diabolicus syp*-like cluster, GT genes may not have the same importance; in particular, some of them may not have any critical effect on polysaccharide production and may allow some versatility.

In addition to the initiating SypR, 6 other GT genes were predicted in the *V. diabolicus syp* cluster. One GT of the *V. diabolicus syp* cluster is a member of family 2 and the five others are GTs from family 4. Both families are polyspecific. GT2 enzymes catalyse β glycosidic linkages by inversion (http://www.cazy.org). Therefore, we suggest that the GT2 SypQ transfers the two β-linked glucuronic acid residues to the growing repeating unit. The third β bond of the *N*-acetyl-glucosamine residue would actually be the result of the polymerization of the repeating units on the outer face of the membrane. Therefore, the *N*-acetyl-glucosamine residue would be the first to be transferred by *V. diabolicus* SypR, as priming glycosyltransferases are able to transfer the sugar-1-phosphate to the lipid carrier, and do not form a glycosidic linkage. GT4 are retaining enzymes catalysing α glycosidic bonds (http://www.cazy.org); it is proposed that the closest GT4 to SypQ, SypP, is involved in the transfer of the *N*-acetyl-galactosamine residue to the repeating unit of *V. diabolicus* EPS because these three genes are probably part of the same operon within the *syp* cluster (Yip et al. 2005) (Fig. 5). The proposed scenario is the only one explaining the three beta linkages while only one inverting GT is present in *V. diabolicus syp* cluster. However, the contribution of the other GT4-coding genes in the *V. diabolicus syp* cluster cannot be ruled out at this time. Subsequently, the SypK translocase and the SypL polymerase would respectively transport and transfer the repeating unit to the growing EPS chain on the outer face of the membrane.

In summary, we have shown that the *syp* cluster can be assigned to HE800 EPS biosynthesis in *V. diabolicus*. The availability of the genomic sequence for *V. diabolicus* contributes to our understanding of the EPS biosynthetic machinery including the regulatory network. This would help us in obtaining a better production yield as well as in developing in cell production of molecules harboring targeted modifications; these issues would be of high value for biotechnological applications (Colliec-Jouault and Delbarre-Ladrat 2014).

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank B. Henrissat and P. Coutinho for annotation of CAZYmes. We acknowledge the LABGeM and the National Infrastructure France Genomique for the Microscope platform tools and Yannick Labreuche for critical reading of the manuscript. This work was supported through a grant awarded to V. Boursicot, co-funding was from Ifremer and the Pays de la Loire Region. D. Goudenège acknowledges Ifremer and the ANR blanc project (11-BSV7-023-01 « VIBRIOGEN ») for post doctoral funding.

References

Anwar M, Choi S (2014) Gram-negative marine bacteria: structural features of lipopolysaccharides and their relevance for economically important diseases. Mar Drugs 12:2485-2514. doi:10.3390/md12052485

Badhai J, Kumari P, Krishnan P, Ramamurthy T, Das SK (2013) Presence of SXT integrating conjugative element in marine bacteria isolated from the mucus of the coral *Fungia echinata* from Andaman Sea. FEMS Microbiol Lett 338:118-123. doi:10.1111/1574-6968.12033

Belcher R, Nutten AJ, Sambrook CM (1954) The determination of glucosamine. Analyst 79:201-208. doi:10.1039/AN9547900201

Binesse J, Delsert C, Saulnier D, Champomier-Verges M-C, Zagorec M, Munier-Lehmann H, Mazel D, Le Roux F (2008) Metalloprotease Vsm is the major determinant of toxicity for extracellular products of *Vibrio splendidus*. Appl Environ Microbiol 74:7108-7117. doi:10.1128/aem.01261-08

Blumenkrantz N, Asboe-Hansen G (1973) New method for quantative determination of uronic acids. Anal Biochem 54:484-489. doi:10.1016/0003-2697(73)90377-1

Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Bapteste E, Lopez P, Tarr CL, Polz MF (2011) Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. mBio 2:e00335-10. doi:10.1128/mBio.00335-10

Caroff M, Karibian D (2003) Structure of bacterial lipopolysaccharides. Carbohydr Res 338:2431-2447. doi:10.1016/j.carres.2003.07.010

Chatzidaki-Livanis M, Coyne MJ, Roche-Hakansson H, Comstock LE (2008) Expression of a uniquely regulated extracellular polysaccharide confers a large-capsule phenotype to *Bacteroides fragilis*. J Bacteriol 190:1020-1026. doi:10.1128/JB.01519-07

Chen Y, Dai J, Morris JG, Johnson J (2010) Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. BMC Microbiol 10:274. doi:10.1186/1471-2180-10-274

Colliec-Jouault S, Delbarre-Ladrat C (2014) Marine-derived bioactive polysaccharides from microorganisms. In: Brahmachari G (ed) Natural bioactive molecules. Impacts and prospects. Narosa Publishing House, New Dehli, India, pp 5.1-5.21

Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. J Mol Biol 328:307-317. doi:10.1016/S0022-2836(03)00307-3

Criscuolo A, Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210. doi:10.1186/1471-2148-10-210

Daubin V, Gouy M, Perriere G (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. Genome Res 12:1080-1090. doi:10.1101/gr.187002

De Angelis PL (2012) Glycosaminoglycan polysaccharide biosynthesis and production : today and tomorrow. Appl Microbiol Biotechnol 94:295-305. doi:10.1007/s00253-011-

3801-6
De Vuyst L, De Vin F, Vaningelgem F, Degeest B (2001) Recent developments in the biosynthesis and applications of heteropolysaccharides from lactic acid bacteria. Int Dairy J 11:687-707. doi:10.1016/S0958-6946(01)00114-5
Dong H, Xiang Q, Gu Y, Wang Z, Paterson NG, Stansfeld PJ, He C, Zhang Y, Wang W, Dong C (2014) Structural basis for outer membrane lipopolysaccharide insertion. Nature 511:52-56. doi:10.1038/nature13464
Drummelsmith J, Whitfield C (2000) Translocation of group 1 capsular polysaccharide to the surface of *Escherichia coli* requires a multimeric complex in the outer membrane. EMBO J 19:57-66
Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797. doi:10.1093/nar/gkh340
El-Hajj ZW, Allcock D, Tryfona T, Lauro FM, Sawyer L, Bartlett DH, Ferguson GP (2010) Insights into piezophily from genetic studies on the deep-sea bacterium, *Photobacterium profundum* SS9. Ann N Y Acad Sci 1189:143-148. doi:10.1111/j.1749-6632.2009.05178.x
Faruque SM, Mekalanos JJ (2003) Pathogenicity islands and phages in *Vibrio cholerae* evolution. Trends Microbiol 11:505-510. doi:10.1016/j.tim.2003.09.003
Filisetti-Cozzi TMCC, Carpita NC (1991) Measurement of uronic acids without interference from neutral sugars. Anal Biochem 197:157-162. doi:10.1016/0003-2697(91)90372-Z
Gandhi NS, Mancera RL (2008) The structure of glycosaminoglycans and their interactions with proteins. Chem Biol Drug Des 72:455-482. doi:10.1111/j.1747-0285.2008.00741.x
Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81-91. doi:10.1099/ijs.0.64483-0
Guo Y, Rowe-Magnus D (2010) Identification of a c-di-GMP-regulated polysaccharide locus governing stress resistance and biofilm and rugose colony formation in *Vibrio vulnificus*. Infect Immun 78:1390-1402. doi:10.1128/IAI.01188-09
Islam T, Linhardt RJ (2003) Chemistry, biochemistry, and pharmaceutical potentials of glycosaminoglycans and related saccharides. In: Wong C-H (ed) Carbohydrate-based drug discovery. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp 407-433
Ivars-Martinez E, D'Auria G, Rodriguez-Valera F, Sanchez-Porro C, Ventosa A, Joint IAN, Mühling M (2008) Biogeography of the ubiquitous marine bacterium *Alteromonas macleodii* determined by multilocus sequence analysis. Mol Ecol 17:4092-4106. doi:10.1111/j.1365-294X.2008.03883.x
Jannasch HW, Taylor CD (1984) Deep-sea microbiology. Annu Rev Microbiol 38:487-487. doi:10.1146/annurev.mi.38.100184.002415
Kamerling JP, Gerwig GJ, Vliegenthart JFG, Clamp JR (1975) Characterization by gas-liquid chromatography mass spectrometry of permethylsilylglycosides obtained in the methanolysis of glycoproteins and glycolipids. Biochem J 151:491-495
Keenleyside WJ, Jayaratne P, MacLachlan PR, Whitfield C (1992) The *rcsA* gene of *Escherichia coli* O9:K30:H12 is involved in the expression of the serotype-specific group I K (capsular) antigen. J Bacteriol 174:8-16
Klein SL, Gutierrez West CK, Mejia DM, Lovell CR (2014) Genes similar to the *Vibrio parahaemolyticus* virulence-related genes *tdh*, *tlh*, and *vscC2* occur in other Vibrionaceae species isolated from a pristine estuary. Appl Environ Microbiol 80:595-602. doi:10.1128/aem.02895-13
Le Roux F, Binesse J, Saulnier D, 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 73:777-784. doi:10.1128/aem.02147-06
Le Roux F, Davis BM, Waldor MK (2011) Conserved small RNAs govern replication and incompatibility of a diverse new plasmid family from marine bacteria. Nucleic Acids Res 39:1004-1013. doi:10.1093/nar/gkq852

Leone S, Silipo A, Nazarenko E, Lanzetta R, Parrilli M, Molinaro A (2007) Molecular structure of endotoxins from Gram-negative marine bacteria: an update. Mar Drugs 5:85-112. doi:10.3390/md503085

Lerouge I, Vanderleyden J (2002) *O*-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions. FEMS Microbiol Rev 26:17-47. doi:10.1111/j.1574-6976.2002.tb00597.x

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490-D495. doi:10.1093/nar/gkt1178

Mazel D, Dychinco B, Webb VA, Davies J (1998) A distinctive class of integron in the *Vibrio cholerae* genome. Science 280:605-608. doi:10.1126/science.280.5363.605

Montreuil J, Bouquelet S, Debray H, Fournet B, Spik G, Strecker G (1986) Glycoproteins. In: Chaplin MF, Kennedy JK (eds) Carbohydrate analysis: a practical approach. IRL Press, Oxford, UK, pp 143-204

Morris AR, Visick KL (2010) Control of biofilm formation and colonization in *Vibrio fischeri*: a role for partner switching? Environ Microbiol 12:2051-2059. doi:10.1111/j.1462-2920.2010.02269.x

Morris AR, Visick KL (2013) The response regulator SypE controls biofilm formation and colonization through phosphorylation of the *syp*-encoded regulator SypA in *Vibrio fischeri*. Mol Microbiol 87:509-525. doi:10.1111/mmi.12109

Nakhamchik A, Wilde C, Rowe-Magnus DA (2007) Identification of a Wzy polymerase required for group IV capsular polysaccharide and lipopolysaccharide biosynthesis in *Vibrio vulnificus*. Infect Immun 75:5550-5558. doi:10.1128/IAI.00932-07

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS ONE 5:e9490. doi:10.1371/journal.pone.0009490

Raguenes G, Christen R, Guezennec J, Pignet P, Barbier G (1997) *Vibrio diabolicus* sp. nov., a new polysaccharide-secreting organism isolated from a deep-sea hydrothermal vent polychaete annelid, *Alvinella pompejana*. Int J Syst Bacteriol 47:989-995. doi:10.1099/00207713-47-4-989

Rahn A, Drummelsmith J, Whitfield C (1999) Conserved organization in the *cps* gene clusters for expression of *Escherichia coli g*roup 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps g*enes from *Klebsiella pneumoniae*. J Bacteriol 181:2307-2313

Rehm BHA (2010) Bacterial polymers: biosynthesis, modifications and applications. Nat Rev Micro 8:578-592. doi:10.1038/nrmicro2354

Reid AN, Whitfield C (2005) Functional analysis of conserved gene products involved in assembly of *Escherichia coli* capsules and exopolysaccharides: evidence for molecular recognition between Wza and Wzc for colanic acid biosynthesis. J Bacteriol 187:5470-5481. doi:10.1128/jb.187.15.5470-5481.2005

Rimington C (1931) The carbohydrate complex of the serum proteins improved method for isolation and re-determination of structure. Isolation of glucosaminodimannose from proteins of ox blood. Biochem J 25:1062-1071

Rodrigues C, Bhosle NB (1991) Exopolysaccharide production by *Vibrio fischeri*, a fouling marine bacterium. Biofouling 4:301-308. doi:10.1080/08927019109378220

Rougeaux H, Kervarec N, Pichon R, Guezennec J (1999) Structure of the exopolysaccharide of *Vibrio diabolicus* isolated from a deep-sea hydrothermal vent. Carbohydr Res 322:40-45. doi:10.1016/S0008-6215(99)00214-1

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers Y-H, Falcon LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC (2007) The *Sorcerer II* Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol 5:e77. doi:10.1371/journal.pbio.0050077

Sawabe T, Ogura Y, Matsumura Y, Gao F, Amin AR, Mino S, Nakagawa S, Sawabe T, Kumar R, Fukui Y, Satomi M, Matsushima R, Thompson FL, Gomez-Gil B, Christen R, Maruyama F, Kurokawa K, Hayashi T (2013) Updating the *Vibrio* clades defined by multilocus sequence phylogeny: proposal of eight new clades, and the description of *Vibrio tritonius* sp. nov. Front Microbiol 4:414. doi:10.3389/fmicb.2013.00414

Senni K, Gueniche F, Changotade S, Septier D, Sinquin C, Ratiskol J, Lutomski D, Godeau G, Guezennec J, Colliec-Jouault S (2013) Unusual glycosaminoglycans from a deep sea hydrothermal bacterium improve fibrillar collagen structuring and fibroblast activities in engineered connective tissues. Mar Drugs 11:1351-1369. doi:10.3390/md11041351

Shibata S, Yip ES, Quirke KP, Ondrey JM, Visick KL (2012) Roles of the structural symbiosis polysaccharide (*syp*) genes in host colonization, biofilm formation, and polysaccharide biosynthesis in *Vibrio fischeri*. J Bacteriol 194:6736-6747. doi:10.1128/jb.00707-12

Soldo B, Lazarevic V, Pagni M, Karamata D (1999) Teichuronic acid operon of *Bacillus subtilis* 168. Mol Microbiol 31:795-805. doi:10.1046/j.1365-2958.1999.01218.x

Thompson FL, Iida T, Swings J (2004) Biodiversity of Vibrios. Microbiol Mol Biol Rev 68:403-431. doi:10.1128/mmbr.68.3.403-431.2004

Tillmans J, Philippi K (1929) Über den Gehalt der wichtigsten Proteine der Nahrungsmittel an Kohlenhydrat und über ein kolorimetrisches Verfahren zur quantitativen Bestimmung von stickstofffreiem Zucker im Eiweiß. Biochem Z 215:36-60

Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fèvre F, Longin C, Mornico D, Roche D, Rouy Z, Salvignol G, Scarpelli C, Thil Smith AA, Weiman M, Médigue C (2013) MicroScope-an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res 41:D636-D647. doi:10.1093/nar/gks1194

Vezzi A, Campanaro S, D'Angelo M, Simonato F, Vitulo N, Lauro FM, Cestaro A, Malacrida G, Simionati B, Cannata N, Romualdi C, Bartlett DH, Valle G (2005) Life at depth: *Photobacterium profundum* genome sequence and expression analysis. Science 307:1459-1461. doi:10.1126/science.1103341

Visick KL (2009) An intricate network of regulators controls biofilm formation and colonization by *Vibrio fischeri*. Mol Microbiol 74:782-789. doi:10.1111/j.1365-2958.2009.06899.x

Weigel PH, De Angelis PL (2007) Hyaluronan synthases: A decade-plus of novel glycosyltransferases. J Biol Chem 282:36777-36781. doi:10.1074/jbc.R700036200

Whitfield C (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. Annu Rev Biochem 75:39-68. doi:10.1146/annurev.biochem.75.103004.142545

Whitfield C (2010) Polymerases: Glycan chain-length control. Nat Chem Biol 6:403-404. doi:10.1038/nchembio.376

Whitfield C, Paiment A (2003) Biosynthesis and assembly of Group 1 capsular polysaccharides in *Escherichia coli* and related extracellular polysaccharides in other bacteria. Carbohydr Res 338:2491-2502. doi:10.1016/j.carres.2003.08.010

Ye L, Zheng X, Zheng H (2014) Effect of *sypQ* gene on poly-*N*-acetylglucosamine biosynthesis in *Vibrio parahaemolyticus* and its role in infection process. Glycobiology 24:351-358. doi:10.1093/glycob/cwu001

Yildiz FH, Visick KL (2009) *Vibrio* biofilms: so much the same yet so different. Trends Microbiol 17:109-118. doi:10.1016/j.tim.2008.12.004

Yip ES, Grublesky BT, Hussa EA, Visick KL (2005) A novel, conserved cluster of genes promotes symbiotic colonization and σ54-dependent biofilm formation by *Vibrio fischeri*. Mol Microbiol 57:1485-1498. doi:10.1111/j.1365-2958.2005.04784.x

Zanchetta P, Lagarde N, Guezennec J (2003) A new bone-healing material: A hyaluronic acid-like bacterial exopolysaccharide. Calcif Tissue Int 72:74-79. doi:10.1007/s00223-001-2091-x

Tables

Table 1: Bacterial strains and mutants constructed as well as plasmids used in this study.

	Description	Reference
Bacterial strains		
Vibrio diabolicus CNCM I-1629	Wild Type, wt	(Raguenes et al. 1997)
Π3813	B462 Δ*thyA*::(*erm-pir-116*) [Erm^R]	(Le Roux et al. 2007)
β3914	β2163 *gyrA462*, *zei-298*::Tn10 [Km^R, Em^R, Tc^R]	(Le Roux et al. 2007)
GV694	*V. diabolicus* CNCM I-1629 Δ*sypR*	This study
GV708	*V. diabolicus* CNCM I-1629 Δ*sypK*	This study
GV729	GV708 pMRB-P_{LAC} *gfp*	This study
GV735	GV708 pMRB-P_{LAC} *sypK*	This study
Plasmids		
pSW7848	*oriV* $_{R6K\gamma}$; *oriT* $_{RP4}$; *ara*C-P_{BAD} *ccdB* [Cm^R]	(Le Roux et al. 2007)
pSW Δ*sypR*	pSW7848, Δ*sypR*	This study
pSW Δ*sypK*	pSW7848, Δ*sypK*	This study
MRB	*oriV* $_{R6K\gamma}$; *oriT* $_{RP4}$; *oriV* $_{pB1067}$; [Cm^R]	(Le Roux et al. 2011)
pMRB-P_{LAC} *gfp*	pMRB-P_{LAC} *gfp*	This study
pMRB-P_{LAC} *sypK*	pMRB-P_{LAC} *sypK*	This study

Table 2: Polysaccharide biosynthetic gene cluster C in *V. diabolicus* strain CNCM I-1629 (locus tag VDIABv1_) compared to the *syp* gene cluster in *V. fischeri* (locus tag VFA) (Yip et al. 2005).

VDIABv1_	VFA	*syp*	Function
270274	1020	A	Anti-sigma factor antagonist
270273	1021	B	Outer membrane protein
270272	1022	C	Periplasmic protein involved in polysaccharide export (Wza)
270271	1023	D	Putative protein with a nucleotide domain
none	1024	E	Response regulator
270270	1025	F	Hybrid sensor kinase (truncated gene)
270269	1026	G	Response regulator with σ54 interaction domain
270268	1027	H	Glycosyltransferase (GT4)
270267	1028	I	Glycosyltransferase (GT4)
270266	1029	J	Glycosyltransferase (GT4)
270265	1030	K	Oligosaccharide translocase (Wzx)
270264	1031	L	Polymerase (Wzy)
none	1032	M	Acetyltransferase
270263	1033	N	Glycosyltransferase (GT4)
270262	1034	O	Putative polysaccharide biosynthesis chain length regulator (Wzc)
270261	1035	P	Glycosyltransferase (GT4)
270260	1036	Q	Glycosyltransferase (GT2)
270259	1037	R	Phosphoglycosyltransferase (pGT)

Figures

Fig. 1: Phylogenetic analysis based on concatenated alignments of nucleic acid sequences of 433 core genes from 50 *Vibrionaceae* strains and *Shewanella baltica* as outgroup. Phylogenetic trees were built using FastTree2, an approximately maximum-likelihood inference tool, with GTR model as parameters (J-C distance, NNIs, SPRs) (Price et al. 2010). Branch lengths are drawn to scale and are proportional to the number of nucleotide changes.

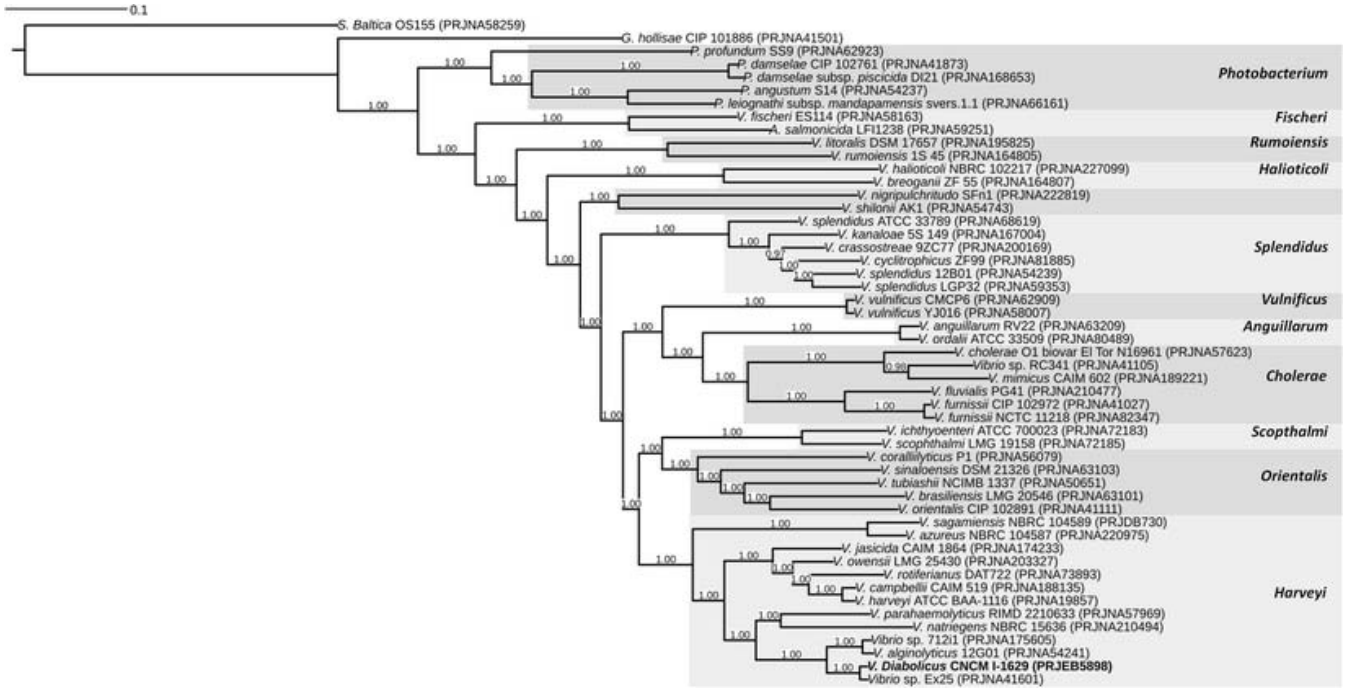


Fig. 2: Strain-to-strain variation between *V. diabolicus* strain CNCM I-1629, strain Ex25, *V. alginolyticus* strain 12G01 and *Photobacterium profundum* strain SS9. The Venn diagram was built using the Vennerable R package.

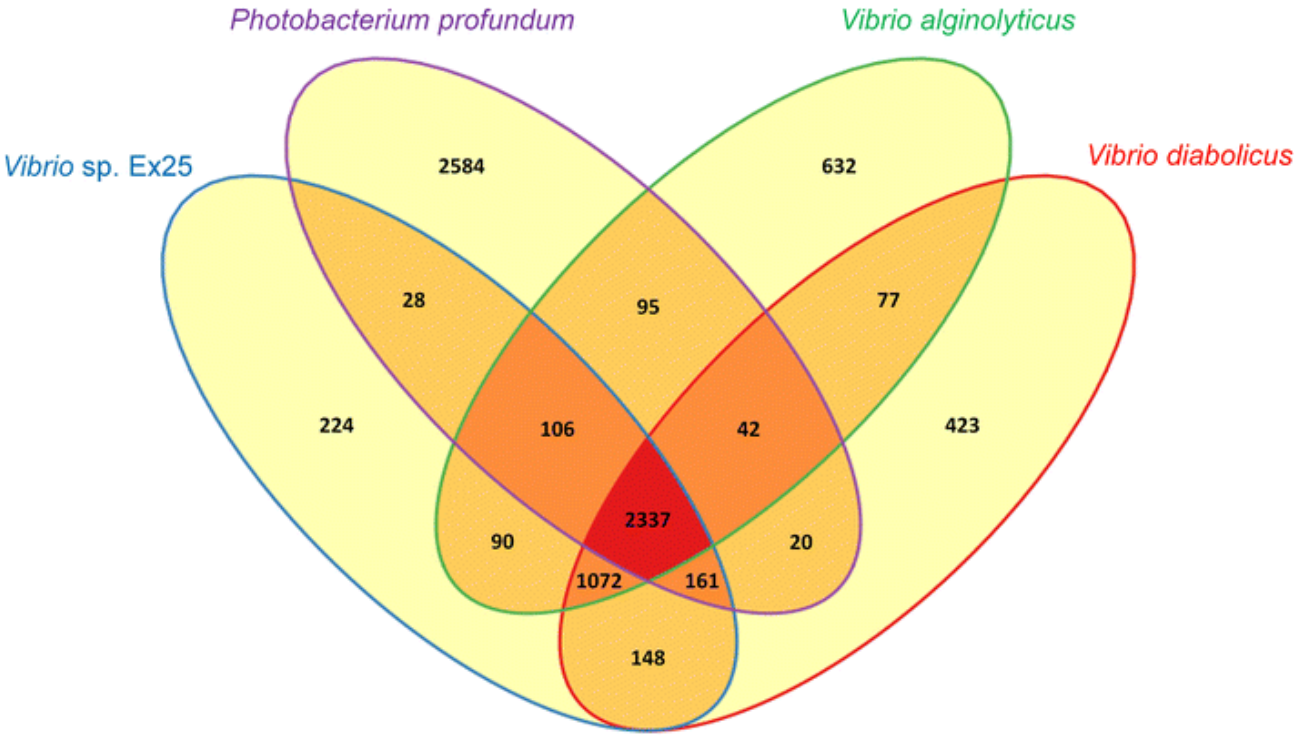


Fig. 3: Structure of the four polysaccharide biosynthesis loci in the *V. diabolicus* genome. **A)** The locus contains genes VDIABv1_10178 to VDIABv1_30021 assigned as Lipid A core biosynthesis and non functional *O*-antigen genes; **B)** The locus contains genes VDIABv1_110283 to VDIABv1_110298 (K_{LPS} biosynthesis); **C)** The locus contains genes VDIABv1_270001 to VDIABv1_270019 (Lipid A core biosynthesis); **D)** The locus contains genes VDIABv1_270957 to VDIABv1_270976 and is homologous to the *V. fischeri syp* cluster. Dark grey, light grey and hatched arrows correspond to genes encoding for GT with CAZy family GT number indicated, pGT and unknown protein respectively. Other CDS are indicated in white. Fragmented CDSs are indicated rectangles with points.

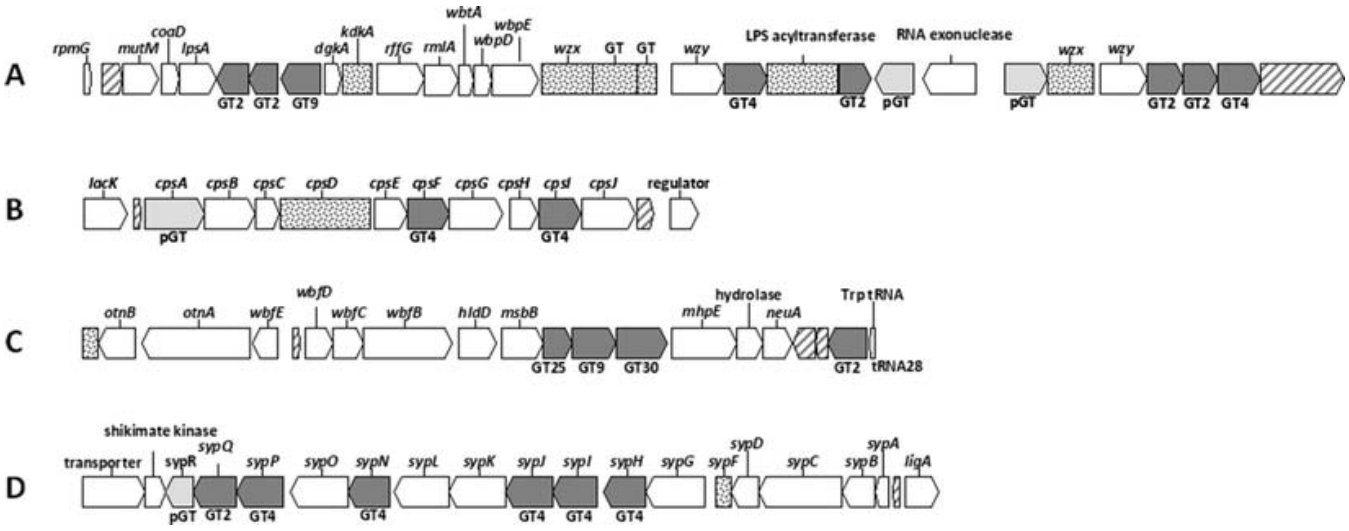


Fig. 4 : Analysis of the osidic composition in the obtained polymers after allele replacements and complementation in *V. diabolicus,* strain CNCM I-1629. Osidic composition in percentage (w/w) btained from the assay was expressed as mg L^{-1} of culture supernatant according to the dry weight yield of polymers.

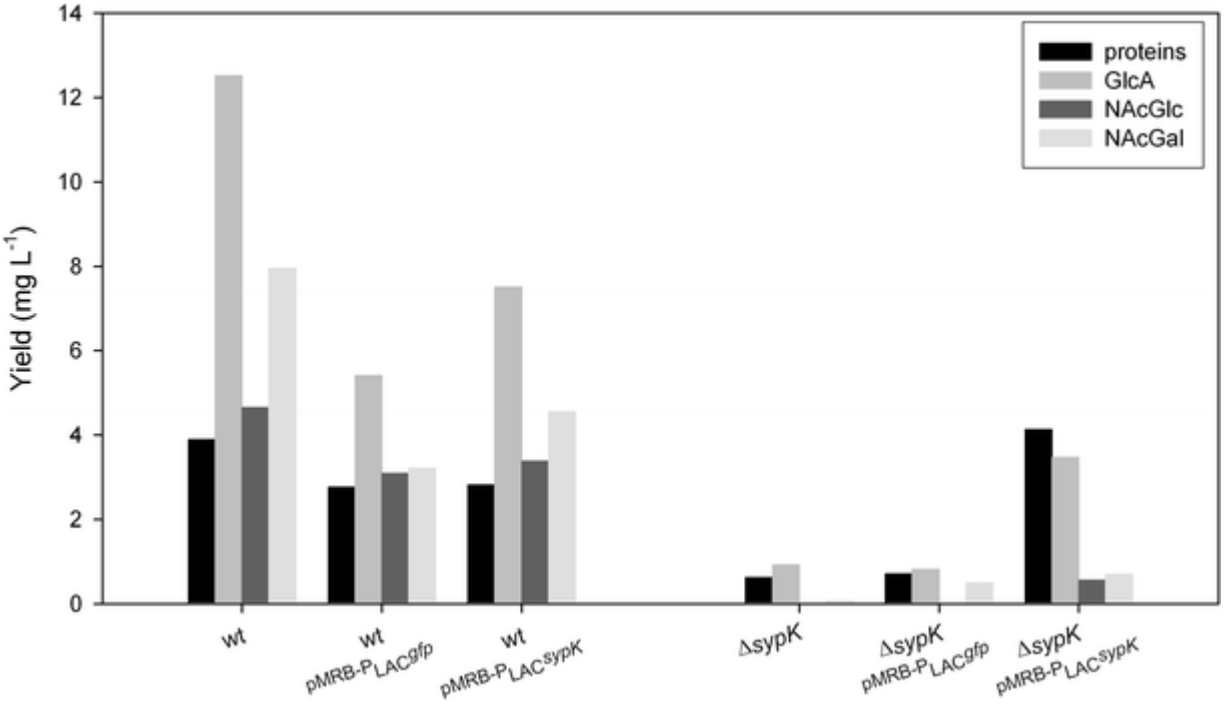
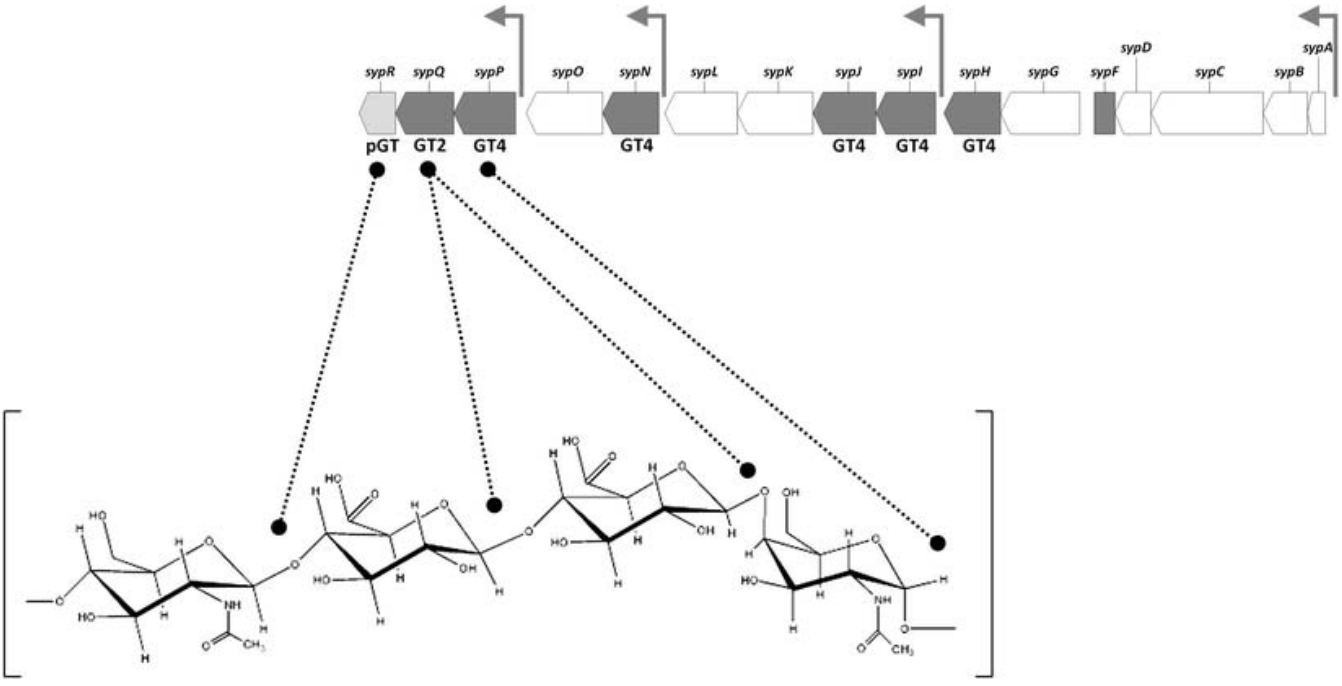


Fig. 5 : Probable GTs involved in HE800 linkage formation. Putative promoters are indicated by grey line arrows from similarity with *V. fischeri* (Yip et al. 2005). The box with vertical lines is the truncated *sypF* gene.



Online resource of the article published in Applied Microbiology and Biotechnology

Genome sequence of *Vibrio diabolicus* and identification of the exopolysaccharide HE800 biosynthesis locus

David Goudenège[2,3,4a] and Vincent Boursicot[1a], Typhaine Versigny[1], Sandrine Bonnetot[1], Jacqueline Ratiskol[1], Corinne Sinquin[1], Gisèle LaPointe[5], Frédérique Le Roux[2,3,4b], Christine Delbarre-Ladrat[1b]

[1]Ifremer, Laboratoire EM[3]B (Ecosystèmes Marins et Molécules Marines pour les Biotechnologies), Rue de l'Ile d'Yeu, BP21105, F-44311, Nantes Cedex 3, France

[2]Ifremer, Unité Physiologie Fonctionnelle des Organismes Marins, ZI de la Pointe du Diable, CS 10070, F-29280 Plouzané, France

[3]Sorbonne Universités, UPMC Univ Paris 06, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff Cedex, France

[4]CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff Cedex, France

[5]Université Laval, Institute for Nutrition and Functional Foods, 2440 boul. Hochelaga, Québec, Québec, Canada, G1V 0A6

[a] Equally contributed to the work

[b]**Corresponding authors:**
Christine Delbarre-Ladrat, Ifremer, Laboratoire EM[3]B (Ecosystèmes Marins et Molécules Marines pour les Biotechnologies), Rue de l'Ile d'Yeu, BP21105, F-44311, Nantes Cedex 3, France. Tel +33 (0)2 40 37 40 57, Fax +33 (0)2 40 37 40 71, christine.delbarre.ladrat@ifremer.fr
Frédérique Le Roux, Equipe « Génomique des Vibrio », Station Biologique de Roscoff, Place Georges Tessier, 29680 Roscoff, France. Tel +33 (0)2 98 29 56 47, Fax +33 (0)2 98 29 frederique.le-roux@sb-roscoff.fr

Table S1: Primers used in this study

Primer	5'-3' sequence
sypR-1	gcccATCGATATGGACGTTGGAAACGCGTC
sypR-2	CCCGGTAATACGCCATATGCGCAGTTGACGATAGATCAC
sypR-3	GTGATCTATCGTCAACTGCGCATATGGCGTATTACCGGG
sypR-4	gcccGAATTCACTGACCTCGTCCGTCAATC
sypR-5	GCCCGGATCCGGTTTGTTCACTACTACTGGC
sypR-6	GCCCTCTAGATGCGACGCCAGTGGCAAGAG
sypK-1	gcccATCGATCGAGTATCTCTTCACCTGTC
sypK-2	CACCTTTGGCGTTGTCACAATTCACCTAAGTCTGATGGT
sypK-3	AACCATCAGACTTAGGTGAATTGTGACAACGCCAAAGGTG
sypK-4	gcccGAATTCGCACCGAATGGCGATTGTAA
sypK-5	AATGTCACGACTGGAAGCAA
sypK-6	ATACGGCCGAACGTTGTTAT
sypK-7	gcccGGGCCCATGGCGATTGTAAACTCAAAACA
sypK-8	gcccCTCGAGTCATAACGAGTATCTCTTACC

Table S2: Glycosyltransferase (GT) and glycosyl-phosphotransferase (pGT) locus tag of the genes identified within the *Vibrio diabolicus* genome. GT families (1 to 51) were determined using the CAZy annotation pipeline (Coutinho et al., 2003).

Classification	Activities	Mechanism	Locus tag
GT1	UDP-glucuronosyltransferase (EC 2.4.1.17); 2-hydroxyacylsphingosine 1-β-galactosyltransferase (EC 2.4.1.45); N-acylsphingosine galactosyltransferase (EC 2.4.1.47); flavonol 3-O-glucosyltransferase (EC 2.4.1.91); indole-3-acetate β-glucosyltransferase (EC 2.4.1.121); sterol glucosyltransferase (EC 2.4.1.173); ecdysteroid UDP-glucosyltransferase (EC 2.4.1.-); zeaxanthin glucosyltransferase (EC 2.4.1.-); zeatin O-β-glucosyltransferase (EC 2.4.1.203); zeatin O-β-xylosyltransferase (EC 2.4.2.40); limonoid glucosyltransferase (EC 2.4.1.210); sinapate 1-glucosyltransferase (EC 2.4.1.120); anthocyanin 3-O-galactosyltransferase (EC 2.4.1.-); anthocyanin 5-O-glucosyltransferase (EC 2.4.1.-); anthocyanidin 3-O-glucosyltransferase (EC 2.4.1.115); dTDP-β-2-deoxy-L-fucose: α-L-2-deoxyfucosyltransferase (EC 2.4.1.-); UDP-β-L-rhamnose: a-L-rhamnosyltransferase (EC 2.4.1.-); UDP-glucose: 4-hydroxybenzoate 4-O-β-glucosyltransferase (EC 2.4.1.194); flavonol L-rhamnosyltransferase (EC 2.4.1.159); salicylic acid β-glucosyltransferase (EC 2.4.1.-)	Inverting	VDIABv1_30137 VDIABv1_30138
GT2	Cellulose synthase (EC 2.4.1.12); chitin synthase (EC 2.4.1.16); dolichyl-phosphate b-D-mannosyltransferase (EC 2.4.1.83); dolichyl-phosphate b-glucosyltransferase (EC 2.4.1.117); N-acetylglucosaminyltransferase (EC 2.4.1.-); N-acetylgalactosaminyltransferase (EC 2.4.1.-); hyaluronan synthase (EC 2.4.1.212); chitin oligosaccharide synthase (EC 2.4.1.-); β-1,3-glucan synthase (EC 2.4.1.34); β-1,4-mannan synthase (EC 2.4.1.-); β-mannosylphosphodecaprenolï¿½mannooligosaccharide α-1,6-mannosyltransferase (EC 2.4.1.199); α-1,3-L-rhamnosyltransferase (EC 2.4.1.-)	Inverting	VDIABv1_10183 VDIABv1_10184 VDIABv1_30011 VDIABv1_30018 VDIABv1_30019 VDIABv1_100272 VDIABv1_100273 VDIABv1_110575 VDIABv1_270019 VDIABv1_270373 VDIABv1_270960
GT4	Sucrose synthase (EC 2.4.1.13); sucrose-phosphate synthase (EC 2.4.1.14); α-glucosyltransferase (EC 2.4.1.52); lipopolysaccharide N-acetylglucosaminyltransferase (EC 2.4.1.56); GDP-Man α-mannosyltransferase (EC 2.4.1.-); 1,2-diacylglycerol 3-glucosyltransferase (EC 2.4.1.157); diglucosyl diacylglycerol synthase (EC 2.4.1.208); digalactosyldiacylglycerol synthase (EC 2.4.1.141); trehalose phosphorylase (EC 2.4.1.231); phosphatidylinositol α-mannosyltransferase (EC 2.4.1.57); UDP-Gal α-galactosyltransferase (EC 2.4.1.-); UDP-Xyl α-	Retaining	VDIABv1_30008 VDIABv1_30020 VDIABv1_110291 VDIABv1_110294

	xylosyltransferase (EC 2.4.2.-)		VDIABv1_270961 VDIABv1_270963 VDIABv1_270966 VDIABv1_270967 VDIABv1_270968
GT5	UDP-Glc: glycogen glucosyltransferase (EC 2.4.1.11); ADP-Glc: starch glucosyltransferase (EC 2.4.1.21); NDP-Glc: starch glucosyltransferase (EC 2.4.1.242); UDP-Glc: α-1,3-glucan synthase (EC 2.4.1.183) UDP-Glc: α-1,4-glucan synthase (EC 2.4.1.-)	Retaining	VDIABv1_270523
GT9	Lipopolysaccharide N-acetylglucosaminyltransferase (EC 2.4.1.56); heptosyltransferase (EC 2.4.-.-).	Inverting	VDIABv1_10185 VDIABv1_270012
GT19	Lipid-A-disaccharide synthase (EC 2.4.1.182).	Inverting	VDIABv1_250050
GT25	Lipopolysaccharide β-1,4-galactosyltransferase (EC 2.4.1.-); β-1,3-glucosyltransferase (EC 2.4.1.-); β-1,2-glucosyltransferase (EC 2.4.1.-); β-1,2-galactosyltransferase (EC 2.4.1.-)	Inverting	VDIABv1_270011
GT28	1,2-diacylglycerol 3-β-galactosyltransferase (EC 2.4.1.46); 1,2-diacylglycerol 3-β-glucosyltransferase (EC 2.4.1.157); UDP-GlcNAc: Und-PP-MurAc-pentapeptide β-N-acetylglucosaminyltransferase (EC 2.4.1.227).	Inverting	VDIABv1_30226
GT30	α-1,6-mannosyltransferase (EC 2.4.1.-); α-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.-); α-1,4-N-acetylgalactosaminyltransferase (EC 2.4.1.-); GDP-Man: inositol-phosphorylceramide transferase (EC 2.4.1.-); UDP-Gal: β-galactoside a-1,4-galactosyltransferase (EC 2.4.1.-); UDP-Gal: lactose/N-acetyl-lactosamine α-1,4-galactosyltransferase (EC 2.4.1.-)	Inverting	VDIABv1_270013
GT35	Glycogen or starch phosphorylase (EC 2.4.1.1).	Retaining	VDIABv1_110520
GT51	Murein polymerase (EC 2.4.1.129).	Inverting	VDIABv1_30484 VDIABv1_110516 VDIABv1_250263
pGT	pGT		VDIABv1_30012 VDIABv1_30014 VDIABv1_110285 VDIABv1_270959