Isolation of Harveyi clade Vibrio spp. collected in aquaculture farms: How can the identification issue be addressed?

Antoine Culot, Noel Grosset, Quentin Bruey, Michel Auzou, Jean-Christophe J.-C. Giard, Baptiste Favard, Alexandre Wakatsuki, Sandrine Baron, Stephane Frouel, Clarisse Techer, et al.

▶ **To cite this version:**

Antoine Culot, Noel Grosset, Quentin Bruey, Michel Auzou, Jean-Christophe J.-C. Giard, et al.. Isolation of Harveyi clade Vibrio spp. collected in aquaculture farms: How can the identification issue be addressed?. Journal of Microbiological Methods, 2021, 180, pp.106106. 10.1016/j.mimet.2020.106106 . hal-03099757

HAL Id: hal-03099757
https://hal.inrae.fr/hal-03099757

Submitted on 6 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Journal of Microbiological Methods 180 (2021) 106106

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Isolation of *Harveyi* clade *Vibrio* spp. collected in aquaculture farms: How can the identification issue be addressed?

Antoine Culot [a,b,**,1], Noel Grosset [a], Quentin Bruey [b], Michel Auzou [d], Jean-Christophe Giard [d], Baptiste Favard [b], Alexandre Wakatsuki [e], Sandrine Baron [c], Stephane Frouel [b], Clarisse Techer [b], Michel Gautier [a,*]

[a] *Microbiology Laboratory, Institut National de Recherche en Agriculture, Alimentation et Environnement, (INRAE), UMR 1253 Science and, Technology of Milk and Eggs (STLO), Rennes, France*
[b] *Mixscience SAS, Bruz, France*
[c] *Laboratoire de Ploufragan-Plouzane-Niort, ANSES, Ploufragan, France*
[d] *Service Microbiologie, Centre Hospitalier Universitaire de Caen, Caen, France*
[e] *Zoopeixes Fish Company, Jardim America, Goiânia, Brazil*

ARTICLE INFO

Keywords:
Vibrio harveyi
Aquaculture
Shrimp pathogen
Identification
Fur
Ldh
MLSA
MALDI-Tof

ABSTRACT

Aquaculture is a fast growing industry with its development hampered by bacterial diseases. Vibriosis caused by *Harveyi* clade strains is known for causing heavy loss especially in shrimp aquaculture farms. For farm treatment and pathogen spread management, veterinarians and researchers need reliable bacterial identification tools. A range of identification methods have been presented for *Vibrio* spp. in recent literature but little feedback on their performance have been made available to this day.

This study aims at comparing *Vibrio* spp. identification methods and providing guidance on their use.

Fifty farms were sampled and bacterial colonies were isolated using specific culture media before microscopic analysis and genomic profiling using ERIC-PCR. A preliminary identification step was carried out using MALDI-ToF mass spectrometry. Four methods were compared for strain identification on 14 newly isolated Harveyi clade *Vibrio* spp. strains: whole genome sequencing (digital DNA DNA Hybridization (dDDH)), 5 MLSA schemes, ferric uptake regulation (*fur*) and lecithin-dependent haemolysin (*ldh*) single gene based identification methods.

Apart from dDDH which is a reference method, no technique could identify all the isolates to the species level. The other tested techniques allowed a faster, cheaper but sub genus clade identification which can be interesting when absolute precision is not required. In this regard, MALDI-ToF and fur based identification seemed especially promising.

1. Introduction

According to the United Nations, the human population could reach 8.5 billion in 2030 and get to 9.7 billion by 2050 (United Nations Department of Economic and Social Affairs, 2019). This forthcoming population calls for an increase in food production of which aquaculture is believed to take the biggest part, as the fastest-growing food production industry. Currently, 45% of global fish production relies on aquaculture and this market share is predicted to grow to 52% before 2025 (FAO, 2018). As aquatic animal production loss due to infectious diseases are estimated at USD 6 billion per year, they are one of the main obstacles to the growth of the sector (Akazawa et al., 2014).

Species belonging to the *Vibrio* genus are Gram-negative straight or curved rods, with accessory flagella. This genus comprises more than a hundred species, which are facultative anaerobes and are found in fresh, brackish and sea water (Farmer and Michael Janda, 2015; Takemura et al., 2014). Some species can cause diseases and target crustaceans, finfishes and molluscs with mortality rates ranging from 45% to 100%

* Corresponding author.
** Corresponding author at: Microbiology Laboratory, Institut, National de Recherche en Agriculture, Alimentation et Environnement, (INRAE), UMR 1253 Science and, Technology of Milk and Eggs (STLO), Rennes, France.
E-mail addresses: a.culot@hotmail.com (A. Culot), michel.gautier@agrocampus-ouest.fr (M. Gautier).
[1] Present Address: Laboratoire de Microbiologie, Agrocampus Ouest, 65 Rue de Saint, Brieuc, 3500, Rennes, France

https://doi.org/10.1016/j.mimet.2020.106106
Received 21 September 2020; Received in revised form 14 November 2020; Accepted 14 November 2020
Available online 25 November 2020
0167-7012/© 2020 Elsevier B.V. All rights reserved.

(Defoirdt et al., 2007). In certain shrimp farmed species, namely *Penaeus monodon* and *Litopenaeus vanamei*, strains belonging to the *Harveyi* clade have caused severe impact on productions: in 2011 global loss were evaluated at USD 1.6 billion (Akazawa et al., 2014). One of the major bacterial diseases behind these losses has been identified and called Early Mortality Syndrome (EMS) or AHPND (Acute HepatoPancreatic Necrotic Disease). It has spread from Asia to Southern America (Akazawa et al., 2014; Gomez-Gil et al., 2014; Nunan et al., 2014).

The *Harveyi* clade is made of *Vibrio harveyi* and the following species: *Vibrio harveyi*, *Vibrio alginolyticus*, *Vibrio parahæmolyticus*, *Vibrio diabolicus*, *Vibrio rotiferianus*, *Vibrio owensii*, *Vibrio campbellii*, *Vibrio jasicida*, *Vibrio azureus*, *Vibrio sagamiensis*, *Vibrio natriegens*, and *Vibrio mytili* (Sawabe et al., 2013; Urbanczyk et al., 2013). Reliable identification of *Vibrio* spp. strains encountered in aquaculture farms is of prime importance since such identification is used to monitor disease spread and occurence. Unequivocal identification is also key for development and use of narrow spectrum treatments like phage therapy (Culot et al., 2019).

To properly identify a new strain, its DNA is compared to reference strains' DNA using DNA-DNA Hybridization (DDH) (Moore et al., 1987). This technique is, however, time-consuming and only available in few specialized laboratories (Goris et al., 2007; Raina et al., 2019). Rapid means of identification are needed, which led to the wide use of 16S rRNA for phylogeny and identification during the 2000s. As 16S rRNA gene interspecies identity varies between 97.60% and 99.99% and species-delineating threshold is usually 97.00%, 99.00% or 99.50% depending on studies, this technique is not accurate to the species level for *Vibrio* spp. (Gomez-Gil et al., 2014; Pascual et al., 2010; Woo et al., 2009). In recent scientific literature, species identification is often unreliable as numerous works rely on 16S rRNA, phenotypical and biochemical tests, which produce inaccurate results (Bauer et al., 2018). Several methods have been proposed to identify species in the *Harveyi* clade. As little feedback on their use is available, choosing between them is not a trivial task.

Part of our project to study vibriosis in Brazilian shrimp farms required to build a collection of *Vibrio* spp. After collecting water samples and eliminating duplicate isolates, the next step was to identify the species. To assess the best method for our collection, a first high-throughput but non species-specific identification step was performed using MALDI-ToF (Bauer et al., 2018). Then, 4 techniques were evaluated: whole genome sequencing and digital DNA DNA Hybridization (dDDH)(Meier-Kolthoff and Göker, 2019), 5 Multi-Locus Sequence Analysis (MLSA) schemes (Cano-Gomez et al., 2011; Pascual et al., 2010; Sawabe et al., 2007; Sawabe et al., 2013; Thompson et al., 2009), and 2 single gene based identifications (*fur* sequence analysis (Machado et al., 2017; Machado and Gram, 2015) and *ldh* detection (Eshik et al., 2018)). These methods are used in recent publications and were chosen for comparison because they feature different cost, practicality and delay.

The *ldh* detection method allows the detection of *V. parahæmolyticus* based on the exclusive presence of this gene in *V. parahæmolyticus* strains (Eshik et al., 2018). Since it is based on a single gene PCR amplification, this method is fast and cheap. It is, however, limited to the detection of *V. parahæmolyticus*. MALDI-ToF mass spectrometry is also a fast method, as colonies are directly used and results are produced in a few hours. The spectrometer creates a spectrum of the analyzed strain's protein content, which is then compared to type species spectra for identification. This technique is *a priori* able to identify a wide range of genus and species but it can be limited by the size of the type strain spectra database and the spectra comparison algorithm. The initial investment for a mass spectrometer can be amortized if it is used for a high number of strains, which puts the cost of a single analysis in the same price range as a PCR. *Fur* sequence analysis is based on the same principle as 16S rRNA gene sequence analysis: the sequence of the gene from the strain to be identified is aligned to its type strains homologs. If the alignment score reaches a given threshold, the species of the type strain is given to the strain to be identified. Although being more labor intensive than a single PCR or MALDI-ToF identification, this method is *a priori* able to deliver accurate results with little sequencing, which is usually the most expensive step. MLSA's principle is the same as *fur* sequence analysis. The difference lies in the fact that several (usually less than a dozen) genes are used for the analysis. This increase in genes allows a more precise identification (since more data is used in the analysis), but also comes with an increase in sequencing costs. Finally, dDDH is an *in silico* method for wet lab DDH, which is the golden standard for bacterial identification. However, DDH is very labor intensive and difficult to implement: it is only available in a handful of specialized laboratories, which makes it impractical. As dDDH simulates the results provided by DDH, it allows accurate species identification. This technique requires full genome sequencing, which makes it the most expensive and the lengthiest method addressed in this article. It was chosen as a reference method to compare all the other techniques.

2. Material & methods

2.1. Reference strains

Vibrio alginolyticus DSM104621 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Germany). The sequence of the type strains of the *Harveyi* clade as well as a reference strain of *Vibrio choler* were retrieved from NCBI Nucleotide database (National Library of Medicine (US), National Center for Biotechnology Information, 2019) and included in the bioinformatics analyzes:

- *Vibrio harveyi* DSM19623 T
- *Vibrio alginolyticus* DSM2171 T
- *Vibrio parahæmolyticus* DSM10027 T
- *Vibrio diabolicus* CNCM:I-1629 T
- *Vibrio rotiferianus* DSM17186 T
- *Vibrio owensii* DSM23055 T
- *Vibrio campbellii* DSM19270 T
- *Vibrio jasicida* DSM21061 T
- *Vibrio azureus* LC2-005 T
- *Vibrio sagamiensis* LC2-047 T
- *Vibrio natriegens* DSM759 T
- *Vibrio mytili* LMG19157 T
- *Vibrio choler* O1 biovar El Tor N16961

2.2. Bacterial strains obtention

2.2.1. Water sampling

Fifty antibiotic and probiotic-free Brazilian shrimp farms were selected for the sampling campaign (Fig. 1). Two hundred and fifty mL of water were taken at each sampling site. Water from the top and the bottom of the sampled ponds was collected and shipped to the laboratory (Agrocampus Ouest, Rennes, France) with ice packs in insulated boxes. The shipments took about two weeks and were stored at 4 °C upon arrival. The samples were still cold when they were taken out of the package but it is impossible to say how their temperature have changed during shipping.

2.2.2. Isolation of bacterial colonies

The samples were filtered in sterile conditions using 0.45 μm polyethersulfone filters (Thermofisher Scientific, Waltham, USA). The filters were cut, placed in 50 mL tubes, and vortexed for 5 min. with phosphate buffered saline (8.0 g/L NaCl (VWR, Radnor, USA), 0.2 g/L KCl, 1.44 g/L Na_2HPO_4, 0.24 g/L KH_2PO_4 (Merck, Darmstadt, Germany)) to release the bacterial cells from the filters. The resulting bacterial suspension was serially diluted to 10–2 and 100 μL of each dilution was plated on four different culture media to allow culturing of more and different isolates. The media used were *Vibrio* CHROMagar (Chromagar, Paris, France), TCBS agar (Difco, Saint-Ferol, France), TCBS agar (Merck), and TCBS agar (VWR, Radnor, Pennsylvania). TCBS's selection spectrum is known

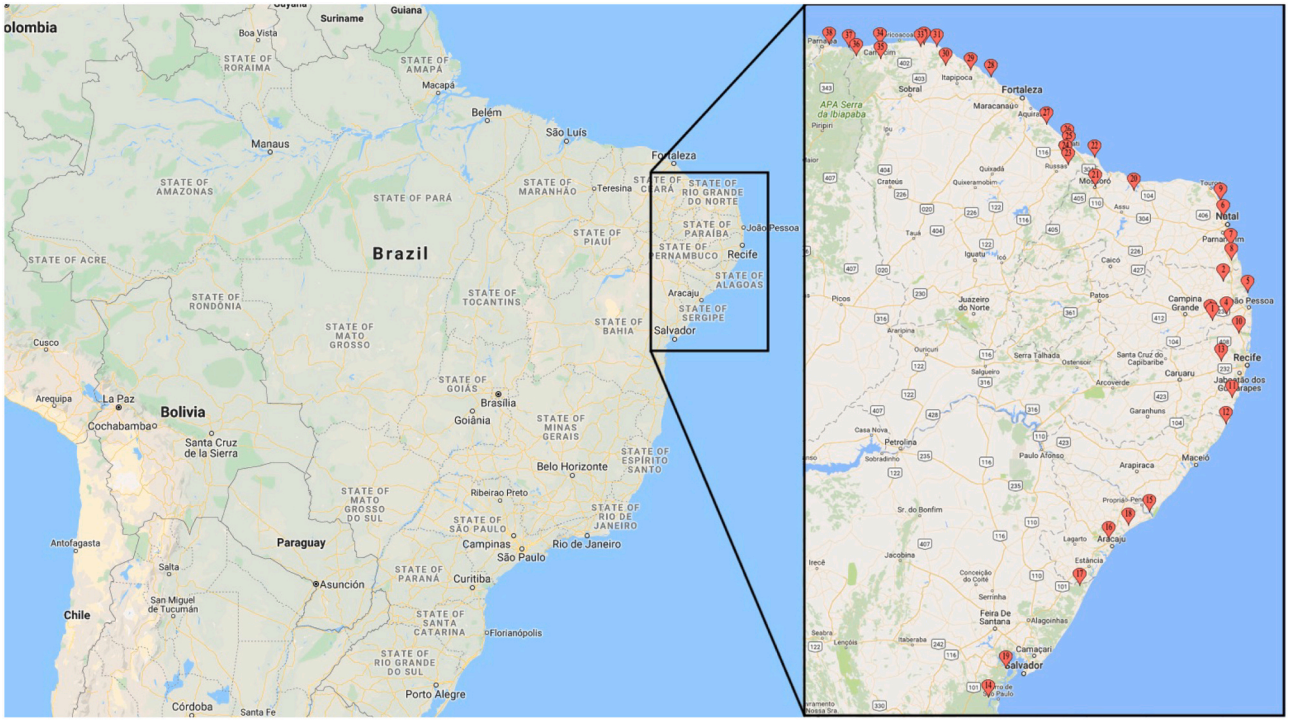

Fig. 1. Location of the 50 sampling zones in Brazil. (Data: Google maps).

for varying between manufacturer (World Health Organization, 2017), which is why 3 different TCBS media were used to increase isolated diversity. Ten colonies from each culture media and displaying different morphological phenotype were chosen and purified on marine agar (NaCl 19.4 g/L, MgCl 8.8 g/L, Na_2SO_4 3.24 g/L, CaCl 1.8 g/L, yeast extract 1 g/L, KCl 0.55 g/L, $NaHCO_3$ 0.16 g/L, ferric citrate 0.1 g/L, KBr 0.08 g/L, SrCl 0.034 g/L, H_3BO_3 0.022 g/L, Na_2Si_3 0.004 g/L, NaF 0.0024 g/L, NH_4NO_3 0.0016 g/L, Na_2HPO_4 0.008 g/L, pH = 7.6 ± 0.2, agar 15 g/L) (Carl Roth, Germany). Those were consequently considered as isolates.

Although being the standard media for *Vibrio* spp. isolation, TCBS are able to grow other bacteria such as *Escherichia spp*, *Bacillus spp*, *Proteus spp*, *Streptococcus* spp. and *Pseudomonas spp.* (Lotz et al., 1983 Jan-Feb). To reduce the number of isolates that entered the next steps and discard the non-*Vibrio* spp, the isolates were observed under a phase contrast microscope (Olympus, Shinjuku, Japan) and those corresponding to the *Vibrio* spp. morphology (1–3 μm motile rods, sometimes curved (Farmer and Michael Janda, 2015)) were kept frozen at −80 °C in marine broth supplemented with 30% glycerol (v/v).

2.3. Strain typing and preliminary identification

2.3.1. ERIC-PCR subtyping

Duplicate isolates were identified by strain subtyping using Enteric Repetitive Intergenic Consensus Polymerase Chain Reaction (ERIC-PCR). Bacterial colonies grown on marine agar were mixed in micro tubes with 1 mL ultrapure water (Thermofisher Scientific) and 100 μL of sterile suspension of chelex beads in ultrapure water (Sigma-Aldrich, Saint Louis, USA) (25% m/v). The tubes were centrifuged at 5200*g*, 4 °C for 7 min. using a 2 K15 centrifuge and a 383C rotor (Sigma-Aldrich) and 800 μL of supernatant were discarded and the pellet was resuspended. The tubes were heated at 100 °C for 10 min. and ice cooled. The samples were centrifuged again with the same parameters and 100 μL of DNA-containing supernatant was recovered. DNA concentration was measured using with a DS-11 Spectrophotometer (Denovix, Wilmington, USA). The spectrophotometer was used to assess DNA quality as well, using 260/280 and 260/230 absorbance ratio. DNA samples were kept if they exceeded 1.65 260/280 absorbance ratio and 1.00 260/320 ratio (Giron and Eric, 2019).

The ERIC-PCR assay was executed using a single primer (5'-ATG-TAAGCTCCTGGGGATTCAC-3') as described by Versalovic et al., 1991. The PCR mixture used is as follows for each sample: 10 μL of 50 ng/μL genomic DNA, 5 μL of 10 μmol/L primers (Sigma-Aldrich), 1.25 μL of 10 mmol/L dNTPs, 5 μL of 10× Taq buffer, 6 μL of 25 mmol/L $MgCl_2$, 0.5 μL of 5000 U/mL Taq polymerase, 22.5 μL of ultrapure water. The PCR cycle consists in the following sequence: 10 min. at 95 °C, 4× (5 min at 94 °C, 5 min. at 40 °C, 5 min. at 70 °C), 40× (1 min. at 94 °C, 1 min. at 55 °C, 2 min. at 72 °C) and 10 min. at 72 °C. Except for the primers, all the reagents were part of the OZYA001-1000D kit (Ozyme, Saint-Cyr-l'Ecole, France). A positive control was made using the DNA of strain DSM104621. A negative control was made using ultrapure water instead of DNA.

The PCR products were migrated alongside the ExactLadder DNA premix 100 pb plus (Ozyme) size marker in a Tris Acetate EDTA gel comprising 2% m/v agarose supplemented with 1 μL of ethidium bromide (Thermofisher Scientific) for 50 mL of total volume. The gel was migrated at 80 V for 2 h and was observed with a Transilluminator (Thermofisher Scientific). The pictures taken were analyzed with Bionumerics 6.6.1 (Applied Math, Sint-Martens-Latem, Belgium).

2.4. Matrix assisted laser desorption ionization - time of flight assay (MALDI - ToF)

All the isolates were preliminary identified using a MALDI - ToF mass spectrometer. A fraction of bacterial colony grown on marine agar was spread in duplicates on a MALDI plate and covered with alpha-cyano-4-hydroxycinnamic acid and subsequently analyzed with a Bruker Daltonik MALDI Biotyper MALDI-ToF mass spectrometer (Bruker, Billerica, USA). Species identification was retained when the two first propositions displayed a confidence score between 2 and the maximum score (3). When the identification score was comprised between 1.7 and 1.99 or above 2 but with different first two identification propositions, both proposition were kept. Isolates identified with a confidence score falling below 1.69 were considered not identified (Bauer et al., 2018). The

contents of the MALDI-ToF profiles database cannot be detailed here because it is covered by a non-disclosure agreement between the manufacturer and the users.

Isolates associated with different electrophoresis profiles or from different farms were regarded as distinct bacterial strains. As MALDI-ToF results exceed sub-genus clade precision for *Vibrio* spp. (Bauer et al., 2018), isolates belonging to different clades of the *Vibrio* genus (as defined by Sawabe et al., 2013) were also considered as discrete bacterial strains.

2.5. Identification method comparison

2.5.1. Whole genome sequencing and assembly

To generate WGS data required for dDDH, 14 strains isolated from shrimp farms were sequenced by Labofarm (Loudeac, France) using a Kingfisher Flex DNA extractor and an IonTorrent S5 Sequencer (Thermofisher Scientific). The 14 strains were selected because MALDI-ToF identification associated them with various species from the *Harveyi* clade; they are also used for other confidential projects in our laboratory. The reads' quality was assessed using FastQC 0.72 (Andrews, 2020) before trimming with PRINSEQ 0.20.4 (Schmieder and Edwards, 2011) and assembly with SPAdes 3.12.0 (Nurk et al., 2013). The assembly quality was subsequently evaluated with Bowtie2 2.3.4.3 (Langmead and Salzberg, 2012) and Samtools 2.02 (Li et al. Li et al., 2009), BUSCO 4.0.2 (Seppey et al., 2019) and Quast 5.0.1 (Mikheenko et al., 2018). Whole genome data analysis was conducted on INRA Migale's Galaxy instance (Jouy en Josas, France) (Afgan et al., 2018; Migale, 2019). Strains that had poor results (too few high quality reads (phred >20), that only allowed for an assembly broken in more than a thousand contigs) in the first sequencing run were sequenced again and data from both run was used for the bioinformatics analysis. The sequences (cleaned reads, assemblies, CDS used for analysis) were deposited on ENA-EMBL database under the accession number PRJEB39286.

2.6. Digital DNA-DNA hybridization identification

Digital DNA DNA Hybridization (dDDH) identification was performed using TYGS (Type Strain Genome Server) to compare the newly isolated strains to type strains from the up to date prokaryotic nomenclature database. According to scientific literature, the 70% DDH/dDDH score was used as the species-delineating threshold for this analysis (Meier-Kolthoff et al., 2013; Meier-Kolthoff and Göker, 2019).

A core-genome alignment was performed to produce a phylogenetic tree of all the newly isolated strains along with the type strains of the *Harveyi* clade species. This tree was used to determine if strains identified by TYGS as belonging to new species were clustered together. To decide whether new species strains that are grouped together belonged to the same new species or to 2 different new species, their genomes were compared using the algorithm used by TYGS to compute dDDH score (GGDC (Genome to Genome Distance Calculator)).

Core genome phylogeny was obtained using Roary 3.13.0 (Page et al., 2015) and its dependencies (Bedtools 2.29.2 (Quinlan and Hall, 2010), Cdhit 4.8.1 (Fu et al., 2012; Niu et al., 2010), Blast+2.9.0 (Camacho et al., 2009; Li et al., 2009), MCL 12–068 (Dongen, 2000, Stijn 2000), GNU parallel 20,191,122 (Tange, 2018), PRANK 170427 (Löytynoja, 2014), MAFFT 7.455 (Katoh and Standley, 2013) and FastTree 2.1.11 (Price et al., 2010). Roary was called with 90% blastp percentage identity and using the options: -e –mafft -i 90. The maximum likelihood tree where generated with FastTree 2.1.1 using the general time reversible model and visualized with FigTree 1.4.4 (Rambaut, A. 2018).

2.7. Multi locus sequence analysis

As some genes involved in MLSA can be difficult to amplify using PCR (Sawabe et al., 2013) and the strains had already been sequenced, we used WGS data to extract genes to build MLSA schemes. To build the gene sequence database required for MLSA analysis, the genomes were annotated using Prokka 1.14.5 (Seemann, 2014) and its dependencies (Uniprot KB database (The UniProt Consortium, 2019), NCBI Bacterial Antimicrobial Resistance Reference Gene Database (National Library of Medicine (US), National Center for Biotechnology Information, 2019), Barrnap 0.9 (Seemann, 2019), Prodigal 2.6.3 (Hyatt et al., 2010) and EggNOG mapper 2.0.0 (Huerta-Cepas et al., 2017) using EggNOG 4.5.1 database (Huerta-Cepas et al., 2019). Prokka was used for gene calling using Prodigal for gff file generation and for 16S rRNA prediction using Barrnap. The predicted CDS were annotated using EggNOG mapper and the gene sequences were automatically recovered using a R 3.6.0 script (R Core Team, 2019). EggNOG sequence annotation was also validated by querying the CDS to the NCBI Nucleotide database using Blast.

All the sequences of each gene were aligned using Clustal Omega 1.2.3 (Sievers et al., 2011) and trimmed with trimAl 1.2 (Capella-Gutierrez et al., 2009) using the -nogaps option. The trimmed sequences where concatenated using a R script before comparison with Blast. Part of the MLSA analysis were run on GenOuest bioinformatics' cluster (Genouest, 2019). The following already published MLSA schemes were tested:

- MLSA 1: *ftsZ*, *gapA*, *gyrB*, *mreB*, *pyrH*, *recA*, *rpoA*, and *topA* with a species-delineating threshold of 98% identity (Sawabe et al., 2013).
- MLSA 2: *rpoD*, *rctB*, and *toxR* with a 90,3% identity threshold (Pascual et al., 2010).
- MLSA 3: *rpoA*, *pyrH*, *topA*, *ftsZ*, and *mreB* with a 94,5% identity threshold (Cano-Gomez et al., 2011).
- MLSA 4: *ftsZ*, *gapA*, *gyrB*, *mreB*, *pyrH*, *recA*, *rpoA*, *topA*, and the 16S rRNA gene with a 95% identity threshold (Sawabe et al., 2007).
- MLSA 5: *rpoA*, *recA*, *pyrH*, *ftsZ*, *topA*, *mreB*, *gyrB*, and *gapA* with a 95% identity threshold (Thompson et al., 2009).

2.8. Single gene based identifications

PCR amplification of *fur* with various PCR mix and cycle parameters (DNA concentration, MgCl concentration, elongation time, number of cycles, annealing temperature) were tested for our strains in our laboratory and by an independent service provider (Labofarm). Tests were also conducted using the ready-made PCR mix recommended by the authors of the *fur* identification method (TEMPase Hot Start Master Mix Blue (Ampliqon A/S, Odense, Denmark)). To test if *fur* sequence was amplified after PCR assay, a PCR amplification targeted at a fragment of *fur* was conducted as recommended by Machado and Gram, 2015, which turned out negative except for 2 of the 25 strains used for testing. During this testing steps, negative controls where DNA was replaced by ultrapure water were executed. Positive controls using the strains that allowed PCR amplification of *fur* were also executed.

Our assay and the assays conducted by Labofarm concluded independently that the primers described in the literature were too degenerate, as most of the strains were resistant to *fur* PCR amplification. It was therefore decided to use WGS data to extract *fur*. *Fur* based identification was performed using FurIOS webserver (Machado et al., 2017). For *ldh* identification, the gene was also retrieved after genome annotation.

3. Results

3.1. Construction of the bacterial collection

From 50 water samples, 2000 isolates were obtained. One thousand three hundred and thirty were discarded upon microscopical observation.

ERIC PCR and MALDI-ToF results allowed us to detect 296 duplicate isolates, which were discarded. All the samples were positive for *Vibrio* spp. As illustrated by Figure 2 various profiles were obtained: the

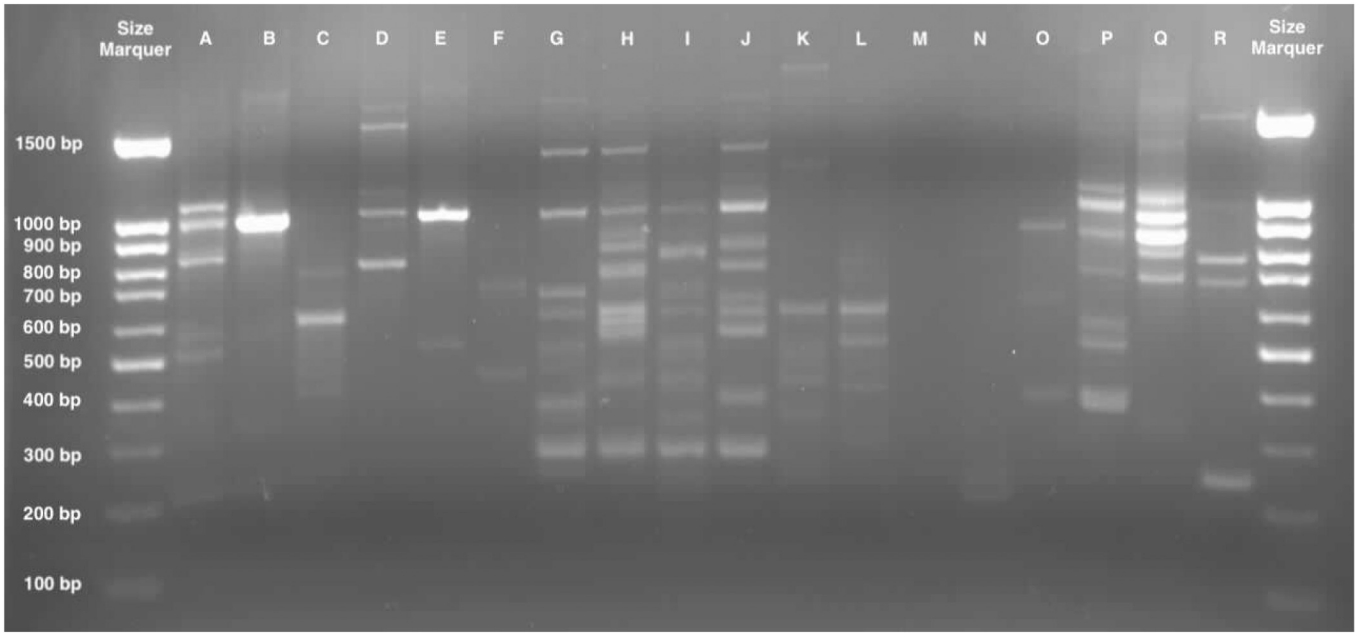

Fig. 2. Example of ERIC PCR result. Tracks A to J and K to R are from two distinct farms. The number of bands per strain and sample varied from 0 to a dozen between strains from the same sample.

number of bands per strain varied between 0 and a dozen. Two hundred and forty six different band profiles were found. Identical profiles coming from different sampling sites were also obtained.

Fig. 3 shows an overview of the species diversity observed in our research. All water samples were positive for *Vibrio* spp. detection. To give the best estimate of this diversity the results were not filtered by MALDI-ToF score to build the graph, which allows isolates that are not identified precisely to the species level by this technique to appear nonetheless. The data presented here shows that more than a third (30%) of the strains could not be identified by mass spectrometry, and that *Harveyi* clade members represented the majority (72%) of collected *Vibrio* spp. strains. Isolates belonging to other *Vibrio* clades represented 26% of the identified stains. *Vibrio* spp. represented 234 isolates.

4. Whole genome sequencing and assembly

First sequencing run results quality were poor for strains B1ASS3, B1HAN19, and B1REV9 (too few high quality reads (phred >20), that only allowed for an assembly broken in more than a thousand contigs) so they were sequenced again. Assembly was made using data from both sequencing runs. Reference strains genomes were included in Table 1 alongside to the genomes sequenced for this study to provide a reference

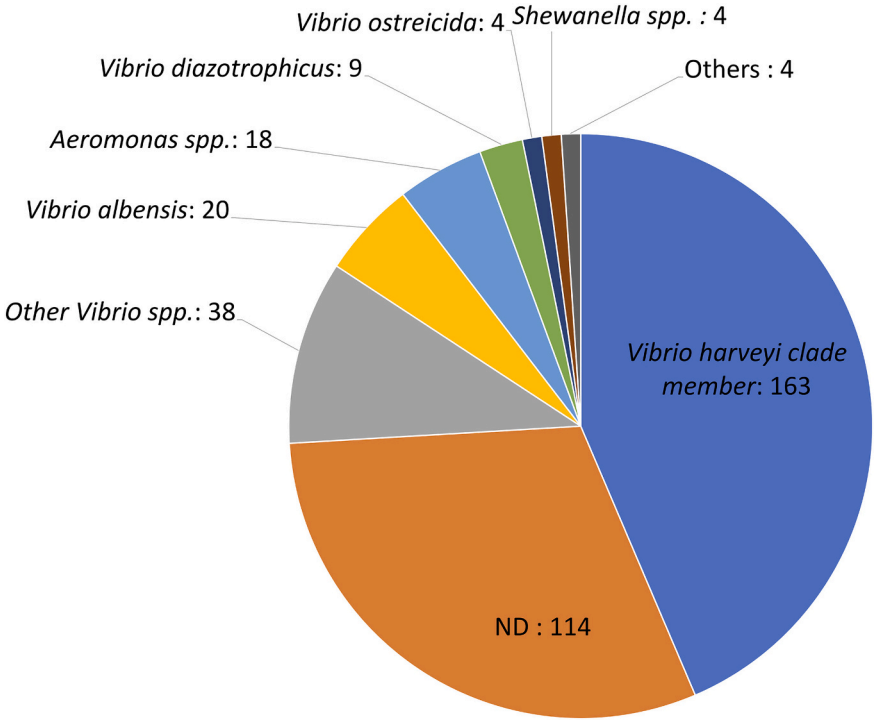

Fig. 3. MALDI-ToF identification results on isolates. *Vibrio harveyi*, *Vibrio alginolyticus*, *Vibrio mytili*, *Vibrio parahæmolyticus* and *Vibrio rotiferianus* are grouped in "*Vibrio harveyi clade member*". *Aeromonas spp.* (*Aeromonas enteropelogenes*, *Aeromonas hydrophila*, *Aeromonas veronii*, unidentified *Aeromonas spp.*) are represented in the eponymous group and so are *Shewanella* spp. strains (*Shewanella fidelis*, *Shewanella putrefaciens*). *Vibrio* species represented by less than 3 strains (namely *Vibrio vulnificus*, *Vibrio ponticus*, *Vibrio pelagius*, *Vibrio navarrensis*, *Vibrio hispanicus*, *Vibrio fortis*, *Vibrio mediterranei*, *Vibrio fluvialis*, *Vibrio shilonii*, *Vibrio brasiliensis*, unidentified *Vibrio* spp.) are grouped in "Other *Vibrio* spp.". The "Other" group contains *Exiguobacterium auranticum* and *Bacillus flexus*. The "ND" section of the graph contains isolates that could not be identified by this technique. *V. albensis* is a synonym of *V. choler* (Reichelt et al., 1976).

Table 1
Whole genome assembly statistics.

Strain	Assembly size (bp)	N50	L50	Coverage	BUSCO Score	Accession number
B1ASS3	6,119,842	59,651	33	53	99.7	ERZ1466925
B1BET6	5,161,640	258,264	7	40	98.7	ERZ1466913
B1BET14	5,127,072	178,539	9	29	99.4	ERZ1466915
B1CBC12	5,427,723	216,559	6	37	99.0	ERZ1466916
B1DRD10	5,172,845	229,171	8	31	99.3	ERZ1466917
B1FIG11	5,721,605	201,665	11	25	99.2	ERZ1466918
B1FLJ16	4 725 811	553 918	4	33	99.2	ERZ1466919
B1FVB17	5 409 824	66,859	23	21	98.7	ERZ1466920
B1HAN19	5,173,310	492,004	4	36	99.5	ERZ1466921
B1JAC2	5,301,427	264,954	8	20	98.1	ERZ1466910
B1REV17	5,301,427	274,790	8	26	99.1	ERZ1466923
B1STR6	5 117 716	175,115	8	27	99.4	ERZ1466924
B1BNVF9	5,148,292	260 208	7	48	97.5	ERZ1466914
B1REV9	5,134,388	67,137	24	23	97.5	ERZ1466922
DSM19623 T	5,583,643	92,314	20	NA	99.9	BCUF01000001.1
DSM2171 T	5,146,637	3,334,467	1	NA	99.5	GCF_00354175.2
DSM10027 T	5,067,729	366,570	5	NA	100	LATW01000001.1
CNCM:I-1629 T	5,117,716	776,801	3	NA	96.2	GCF_01048675.1
DSM17186 T	5 368 584	188 275	10	NA	97.9	APHW01000001.1
DSM23055 T	6 342 356	89,583	23	NA	100	BAOH01000001.1
DSM19270 T	5 094 940	73,528	21	NA	99.7	BAOF01000001.1
DSM21061 T	6 045 622	109 122	17	NA	99.7	BAOG01000001.1
LC2-005 T	4,749,542	25,467	54	NA	98.7	BAOB01000001
LC2-047 T	4,534,101	50,295	29	NA	99.0	BAOJ01000001.1
LMG19157 T	4 610 819	122 372	12	NA	91.9	JXOK01000001.1
DSM759 T	5 113 814	62,844	26	NA	96.9	ATWU01000001.1

point for comparison. The L50 statistic is the smallest number of contigs whose length sum makes up half of genome size. N50 statistics is the size of the smallest contig used to compute L50 (International Human Genome Sequencing Consortium, 2001). Coverage (which is the mean number of reads containing a given nucleotide) was not calculated for reference strains as the diversity of sequencing techniques used for these assemblies did not allow a relevant comparison between them or with our assay. The BUSCO score is the ratio of the number of genes that were found in the assembly out of the 1445 *Vibrionales* core genes that are searched by the BUSCO software. This score gives an estimate of assembly completeness.

Assembly coverage, was high (superior to 20) for all our assemblies. This indicates that the contigs are supported by a significant number of reads. For L50 and N50 statistics, lower values and higher values are respectively preferable as such values are linked to less fragmented assemblies with larger contigs. Despite variations in those metrics, all of the assemblies were of higher quality than the lower quality reference assemblies.

L50 and N50 statistics varied from inferior values for B1ASS3 ($L50_{B1ASS3}$ = 33, $N50_{B1ASS3}$ = 59,651) to better values for B1FLJ16 ($L50_{B1FLJ16}$ = 4, $N50_{B1FLJ16}$ = 553,918). Sequencing of the strains allowed us to obtain assemblies containing nearly all the DNA sequence of our strains, as shown by the BUSCO scores. The scores were high and some of them were superior to reference assemblies. All the assembly statistics of the newly sequenced strains were similar to those of reference strains, which validates our sequencing runs and bioinformatics analysis. Interestingly, the BUSCO scores were (except for two reference assemblies) below 100. The BUSCO database is made of genes that are found in 90% of *Vibrionales*, which implies that scores below 100% are not necessarily linked to issues in sequencing and data analysis as some of the searched genes could be absent of the studied strains. The good quality (high coverage, low L50, high N50, high BUSCO score) of the assemblies allowed us to carry on with the bioinformatics analyses.

4.1. Sequence-based identification

All the genes were obtained for each MLSA scheme, using whole genome sequencing data. The five schemes tested in this study yielded the following alignment sizes:

- MLSA 1: 7555 bp.
- MLSA 2: 3330 bp.
- MLSA 3: 4698 bp.
- MLSA 4: 8316 bp.
- MLSA 5: 7555 bp.

As shown by Table 2, all the *Vibrio parahæmolyticus* strains where positive for *ldh* detection as well as *Vibrio diabolicus*, *Vibrio harveyi*, *Vibrio mytili*, *Vibrio natriegens*, *Vibrio owensii*, *Vibrio rotiferianus* strains. Three newly isolated strains where negative and so were *Vibrio choler*, *Vibrio alginolyticus*, *Vibrio azureus*, *Vibrio campbellii*, *Vibrio diabolicus*, and *Vibrio*

Table 2
Comparison of *ldh* presence test and dDDH results.

Strain name	dDDH identification	ldh presence
DSM10027 T	*Vibrio parahæmolyticus*	+
B1BET6	*Vibrio parahæmolyticus*	+
B1BET14	*Vibrio parahæmolyticus*	+
B1DRD10	*Vibrio parahæmolyticus*	+
B1FIG 11	No significant match	–
B1FLJ16	No significant match	–
B1FVB17	*Vibrio alginolyticus*	–
DSM2171 T	*Vibrio alginolyticus*	–
LC2-005 T	*Vibrio azureus*	–
DSM19270 T	*Vibrio campbellii*	–
N16961	*Vibrio choler*	–
DSM21061 T	*Vibrio jasicida*	–
LC2-047 T	*Vibrio sagamiensis*	–
B1ASS3	No significant match	+
B1REV17	No significant match	+
B1HAN19	*Vibrio diabolicus*	+
B1JAC2	*Vibrio diabolicus*	+
B1STR6	*Vibrio diabolicus*	+
DSMZ104621	*Vibrio diabolicus*	+
CNCM:I-169 T	*Vibrio diabolicus*	+
DSM19623 T	*Vibrio harveyi*	+
LMG19157 T	*Vibrio mytili*	+
DSM759 T	*Vibrio natriegens*	+
DSM23055 T	*Vibrio owensii*	+
B1CBC12	*Vibrio rotiferianus*	+
DSM17186 T	*Vibrio rotiferianus*	+

jasicida strains.

Table 3 presents the identification results for dDDH, MLSA, *fur* and MALDI-ToF based identifications. In this table, for dDDH, MLSA, and *fur*, only the relevant results, *e.g.* those associated with a higher score than the species-delineating threshold are displayed. All 3 *Vibrio parahæmolyticus* strains were properly identified by every method, which was not the case for the 4 *Vibrio diabolicus* strains which were only correctly identified by the MLSA 1 and 2 technique. Also, the *Vibrio alginolyticus* strain was identified according to dDDH only by MLSA 1 and 2 and MALDI-ToF mass spectrometry. Four strains isolated in this study did not have any match in digital DNA-DNA Hybridization analysis and where not identified in most of the methods, except for MALDI-ToF. The strains belonging to new species that were clustered together after core-genome analysis, namely.

B1FLJ16 and B1REV17 (Fig. 4), were compared using GGDC to decide if they belonged to the same species. The dDDH result was: 80.60%, with a confidence interval of [77.7–83.2]. Raw values obtained during dDDH, *fur*, and MLSA analysis are available in supplementary data.

The tested methods are compared in Fig. 5, where each criterion was evaluated out of five points and weighted as follows: Result accuracy: 3, Ease of implementation: 1, Cost: 1, time: 1, Breadth of use: 0.5. "Ease of implementation" represents the amount of work required to set up the method. "Breadth of use" represents the quantity of information brought by the use of the method, and the diversity of utilization that can be made of that information.

5. Discussion

5.1. Outcome of the sampling campaign

One hundred and sixty three strains belonged to our target clade, *Vibrio harveyi*. The taxa identified using MALDI-ToF are not surprising to be found in marine environments (El-Barbary, 2017; Evangelista-Barreto et al., 2010; Subramanian and Maruthamuthu, 2019; Vishnivetskaya et al., 2009).

Although a majority of isolates belonged to *Vibrio* spp. (Fig. 3), the methods used are not free of bias and this prevents the collection to be actually representative of the species that can be found in shrimp farms. The shipment time and conditions most likely selected a part of the sampled bacterial diversity. Also, culture-based methods are known to select a small fraction of the living cells, which can be even more of an issue with *Vibrio* spp. as members of this genus are known for their ability to switch to viable but not cultivable state (Coutard et al., 2007).

Although widely used, TCBS media's selectiveness is known to be tolerant, as they can allow *Escherichia spp*, *Bacillus spp*, *Proteus spp*, *Streptococcus* spp. and *Pseudomonas* spp. to grow (Lotz et al., 1983 Jan-Feb). Given the high number of isolates, it was necessary to carry out a first filtering step to remove these non-*Vibrio* spp. isolates. Detecting *Vibrio* spp. phenotype under a microscope is difficult, as some strains are more motile than others and viable but non cultivable cells can switch to a coccoid morphology (Fernández-Delgado et al., 2015 Jan-Feb), which deviates from the expected curved and motile rod phenotype. This step allowed us to save time by discarding 1330 isolates but might have biased the selection of species, since it is not an accurate method for bacterial identification. The microscope observation step could therefore impact the final *Vibrio* spp. diversity.

Finally, ERIC-PCR has a limited resolution and while this technique allows to reliably establish that two isolates are different, it cannot be used to tell if two isolates are the same. Therefore, isolates that are actually different might have been discarded after the ERIC-PCR assay.

5.2. Precision of the different identification methods

5.2.1. Digital DNA DNA hybridization

Since dDDH is the reference method for bacterial identification, it was chosen as the comparison point for all the tested techniques. For dDDH, MLSA and *fur*, the lack of significant match does not mean that the method failed to identify the isolate. This result actually means that based on the comparison method (*e.g.* dDDH, MLSA, *fur*) no type strain was close enough to the analyzed strain to yield a score higher than the identification threshold. The absence of a type strain matching significantly with an isolate means that, considering the method's "point of view", no type strain was close enough to the isolate, hence it belongs to an undescribed species. dDDH results suggest that 5 newly isolated strains (B1ASS3, B1FIG 11, B1REV17, B1FLJ16, B1REV9) belong to new species, as no type strain scored more than the 70% threshold when compared to these strains. Although TYGS stated that these strains did not match with existing species, caution should be used, as B1ASS3's, B1FIG11's and B1REV9's best hit's dDDH scores confidence intervals included the 70% mark (see supplementary data). To confirm these strains as part of new species, hybrid genome sequencing using long read technology could be performed to improve the quality of the assembly (Vasudevan et al., 2020). Wet lab DDH could also be used to confirm the results.

B1FLJ16 and B1REV17 had best hit scores far below 70% (25.2 and 25.3 respectively) and could therefore be reliably considered as new species. Since strains B1FLJ16 and B1REV17 were clustered in the same phylogenic group (Fig. 4), a dDDH comparison using GGDC was conducted to determine if these strains belonged to different species. The result (80.60%) was above species delineation threshold of 70% which indicates that B1FLJ16 and B1REV17 strains belong to the same species.

It can therefore be concluded that 4 new species might have been discovered here. It is interesting to note that strain DSM104621 was assigned to different species than previously published. This result is not completely unexpected, as DSM104621 was identified using 16S rRNA and phenotype-based methods (Kokkari et al., 2018).

5.3. Multi locus sequence analysis

Using WGS data instead of wet lab PCR amplification can arguably yield different results when performing MLSA. However, as underlined by Sawabe et al., 2013, it is difficult to design a truly universal primer for every gene in an MLSA scheme and WGS can be necessary for some *Vibrio* species. Our strains being resistant to *fur* PCR amplification, WGS was a convenient way to get the gene sequence and perform the analysis.

MLSA 1 and 2 were the best alternative to dDDH, as they allowed to identify nearly all our isolates. These two schemes yielded good results for all strains associated to known taxa but failed to identify B1FIG 11 and B1ASS3 as new species. Strains B1REV9, B1REV17 et B1FLJ16 belonged to new species based on dDDH results and were identified accordingly by MLSA 1 and 2. MLSA 3, 4, 5 where not able to reliably identify all the species in the dataset but the results were consistently in the phylogenic vicinity of the correct identification. Identification of *Vibrio diabolicus* was for example inaccurate even though the correct species was proposed among others.

Identification errors in MLSA can be caused by poor choice of genes or species-delineating identity threshold. If the genome loci used do not provide enough information about the proximity and differences of two compared strains from different species, the identity percentage will not be lower than the species-delineating identity threshold, which can prevent the analysis from differentiating distinct species. In that case, the method is limited by the quantity of available information, which is linked to the choice and number of genes. MLSA schemes can be enhanced in two ways:

- By using genes bearing more information, that is to say genes which evolution is more representative of the whole genome's evolution.
- By adding more genes to the analysis to increase the amount of information.

Adding genes to the analysis should however be limited to a few

Table 3
Result comparison of the different identification techniques. Cells containing results coherent with dDDH identification are in green. When the coherent result was proposed among others by an identification method, the cell were colored in orange. Non coherent results where colored in red. The identification score (Blast identity percentage) are displayed for MLSA schemes yielding more than one result, and are systematically given for MALDI-ToF (confidence score out of 3). For dDDH, MLSA, *fur*, only results associated with a higher score than the species-delineating threshold are displayed. In this table, "no significant match" does not mean that the method failed, but that no type strain was close enough to the tested train for the analysis to yield a higher score than the threshold.

Strain name	**dDDH**	MLSA 1	MLSA 2	fur	MLSA 3	MLSA 5	MLSA 4	Maldi -Tof
B1ASS3	No significant match	*V. owensii (98.607)*	*V. owensii (97.896)*	*V. jasicida*	*V. owensii (98.930)* *V. campbellii (95.141)*	*V. owensii (98.607)* *V. campbellii (95.557)*	*V. owensii (98.719)* *V. campbellii (95.894)*	*V. harveyi (2.12 / 2.07)*
B1BET6	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus (2.30 / 2.26)*
B1BET14	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus (2.41 / 2.33)*
B1CBC12	*V. rotiferianus*	*V. rotiferianus*	*V. rotiferianus*	*V. rotiferianus*	*V. rotiferianus*	*V. rotiferianus*	*V. rotiferianus (99.259)* *V. campbellii (95.068)*	*V. rotiferianus (2.13)* *V. harveyi (1.91)*
B1DRD10	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus*	*V. parahaemolyticus (2.04/2.04)*
B1FIG11	No significant match	*V. campbellii (98.112)*	*V. campbellii (98,146)*	*V. campbellii*	*V. campbellii (98.785)* *V. owensii (95.322)*	*V. campbellii (98,112)* *V. owensii (96.155)*	*V. campbellii (98.253)* *V. owensii (96.454)* *V. rotiferianus (95.405)*	*V. harveyi (2.06/ 1.95)*
B1FLJ16	No significant match	No significant match	No significant match	No significant match	No significant match	No significant match	No significant match	*V. parahaemolyticus (1.91)* *V. mytili (1.83)*
B1FVB17	*V. alginolyticus*	*V. alginolyticus*	*V. alginolyticus*	*V. alginolyticus*	*V. alginolyticus (99.456)* *V. diabolicus (97.953)*	*V. alginolyticus (99.459)* *V. diabolicus (96.166)*	*V. alginolyticus (99.492)* *V. diabolicus (96.475)*	*V. alginolyticus (2.11 /2.05)*
B1HAN19	*V. diabolicus*	*V. diabolicus*	*V. diabolicus*	*V. alginolyticus*	*V. diabolicus (98.839)* *V. alginolyticus (96.917)*	*V. diabolicus (98.803)* *V. alginolyticus (95.960)*	*V. diabolicus (98.900)* *V. alginolyticus (96.276)*	*V. alginolyticus (2,17 /1,99)*
B1JAC2	*V. diabolicus*	*V. diabolicus*	*V. diabolicus*	*V. alginolyticus*	*V. diabolicus (99.166)* *V. alginolyticus (96.844)*	*V. diabolicus (99.252)* *V. alginolyticus (96.120)*	*V. diabolicus (99.312)* *V. alginolyticus (96.422)*	*V. alginolyticus (2.09 / 2.03)*
B1REV17	No significant match	No significant match	No significant match	No significant match	No significant match	No significant match	No significant match	*V. parahaemolyticus (1.91/1.89)*
B1STR6	*V. diabolicus*	*V. diabolicus*	V. diabolicus	*V. alginolyticus*	*V. diabolicus (98.912)* *V. alginolyticus (96.989)*	*V. diabolicus (99.137)* *V. alginolyticus (96.143)*	*V. diabolicus (99.206)* *V. alginolyticus (96.443)*	*V. alginolyticus (2.18 / 2.17)*
DSMZ104621	*V. diabolicus*	*V. diabolicus*	V. diabolicus (98,949)	*V. alginolyticus*	*V. diabolicus (99.220)* *V. alginolyticus (97.026)*	*V. diabolicus (99.263)* *V. alginolyticus (96.247)*	*V. diabolicus (99.323)* *V. alginolyticus (96.539)*	*V. alginolyticus (2.15 / 2.02)*
LMG10946	No significant match	*V. campbellii (98.192)*	*V. campbellii (98.192)*	*V. campbellii*	*V. campbellii (99.093)* *V. owensii (95.177)*	*V. campbellii (98.377)* *V. owensii (95.913)*	*V. campbelli (98.497)* *V. owensii (96.232)* *V. rotiferianus (95.247)*	*V. harveyi (2.10)* *V. parahaemolyticus (2.08)*
B1REV9	No significant match	No significant match	No significant match	No significant match	No significant match	No significant match	No significant match	*V. harveyi (1.78/1.74)*
B1BNVF9	*V. alginolyticus*	*V. alginolyticus*	*V. alginolyticus*	*V. alginolyticus*	*V. alginolyticus (99.383)* *V. diabolicus (96.899)*	*V. alginolyticus (99.528)* *V. diabolicus (95.166)*	*V. alginolyticus (99.555)* *V. diabolicus (95.475)*	*V. alginolyticus (2.24 /2.22)*

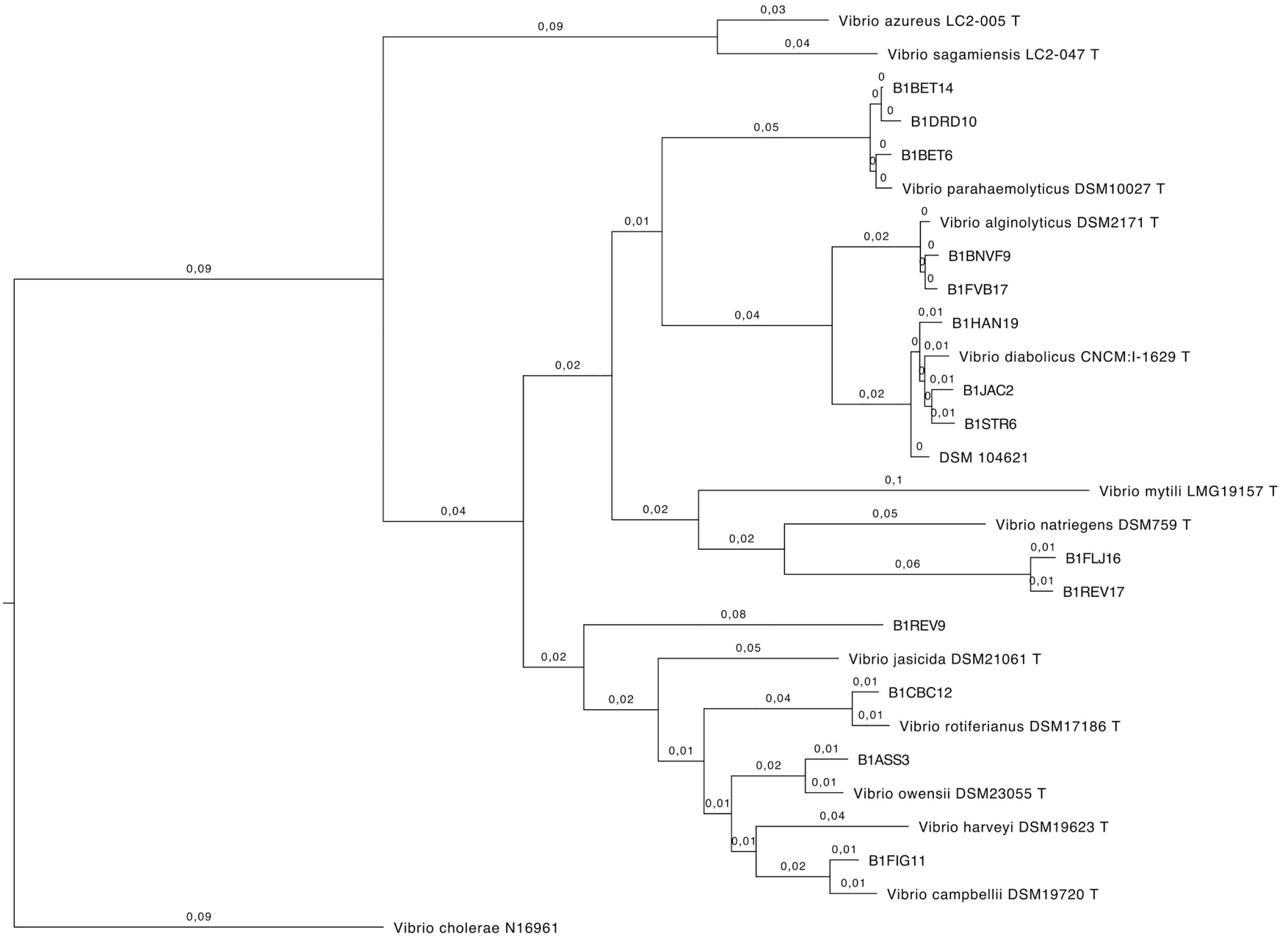

Fig. 4. Phylogenetic tree based on core genome alignment. The tree was rerooted using *Vibrio cholerae* O1 biovar El Tor N16961 as the outgroup. The values displayed at branches are *p*-values computed from Shimodaira-Hasegawa test rounded to the 2nd decimal.

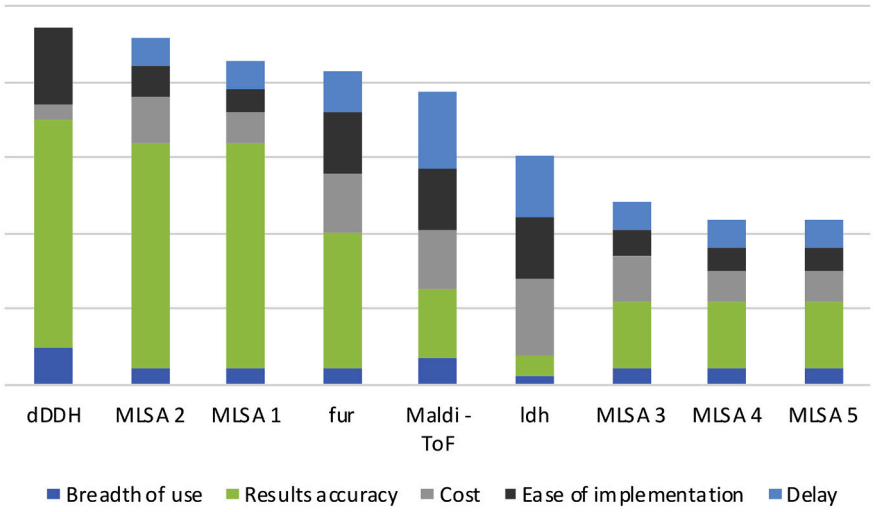

Fig. 5. Ranking of tested identification methods for *Vibrio* spp. identification. For all the evaluated criteria, a higher score is better. "Ease of implementation" represents the amount of work required to set up the method. "Breadth of use" represents the quantity of information brought by the use of the method, and the diversity of utilization that can be made of that information.

sequences: MLSA's principal advantage is to avoid sequencing whole genomes, thus limiting sequencing costs. This advantage is reduced as the number of sequenced genes increases. The majority of identification errors of MLSA 3, 4 and 5 seem to be caused by too low species-delineating threshold: identification of *V. diabolicus* is for example not precise but the correct species was among the proposed identifications. Despite these results being close to the "right answer" they are not usable beyond species group identification. It is not possible to choose the taxa associated with the highest identity percentage as an answer. Raising the identity threshold could be considered, but it requires to completely rebuild the MLSA scheme, taking into account new entries in the prokaryotic taxonomy (De Vos, 2011).

The issues of gene choice and species-delineating threshold discussed here are not linked to mistakes possibly made by the authors of the different MLSA schemes, but rather to the increasing number of known species. Genes used in a MLSA scheme are chosen to differentiate known species at the time it is created. As it is not possible to predict in advance what genes will be sufficient to distinguish species that will be discovered in the future, it is inherent to MLSA's method to be limited by new species. MLSA schemes need to be updated regularly to take the evolution of prokaryotic taxonomy into account.

It should be noted that although the result given by the analysis is that "no significant match were found" for strains belonging to new species, it is possible for dDDH and MLSA to ignore the identity threshold and have a look at the best hit to locate the strain in the prokaryotic taxonomy. This result can also be obtained by building a tree with MLSA data, which provides a similar result as Fig. 4.

5.4. *Fur* and *ldh* genes

Fur identification was more accurate than MLSA 3, 4, 5, which is interesting since it is based on a single gene, whereas MLSA schemes comprise up to 9 genes.

As *ldh* detection is a PCR-based method for detection of *Vibrio parahæmolyticus*, it provides result in a matter of hours which is interesting to get fast diagnosis. However, result showed many false positives in this study. It is demonstrated here that the gene is not species specific, which implies that the detection test could not be reliable. It is important to note that this assay was based on whole genome sequencing and annotation whereas the original method used PCR amplification and it is possible that the provided primers are species specific. The original publication of the method however states that it is also positive to the detection of *Vibrio hollisae* which was reclassed a decade ago as *Grimontia hollisae* (Nishibuchi et al., 1985; F. L. Thompson et al., 2003). It is therefore hard to recommend this method for its accuracy.

6. There is more to it than species identification accuracy

Accuracy is the most important criterion to evaluate an identification method. However other criteria have to be taken into account such as cost, time before result obtention, amount of yielded information and ease of implementation. Whole genome sequencing and dDDH cost hundreds of US dollar and take a week to provide results but provides high amount of information which can be used for further investigation of the strains (antibiotic resistance genes, pathogenicity genes for example). WGS requires bioinformatics skills to produce contigs but the dDDH analysis is trivial since TYGS runs a whole identification pipeline which provides a complete analysis (Meier-Kolthoff and Göker, 2019).

PCR-based methods, such as *ldh* detection, are fast and cheap methods as they only require a PCR which can be executed in half a day, but could not potentially give any more information than the presence or the absence of a species. *Fur* identification is a bit more expensive and lengthy method since sequencing is required but provides more accurate results, in a day. Also, it brings more information than PCR-based detection methods, as it can identify multiple species and assess their phylogeny. This technique is very affordable compared to WGS, as the cost for *fur* identification is less than USD 5. However, the PCR primers described in the original publication (Machado et al., 2017) are very degenerate and it did not work reliably on every tested strain (data not shown), which is why it was chosen to evaluate this method using whole genome sequencing data. The design of the primers should be optimized.

MLSA sits between *fur* identification and dDDH as it is cheaper than dDDH its cost lives in the USD 30–40 range. MLSA takes a comparable amount of time to *fur* method but can be complicated to implement. The primers described in the literature are indeed not always able to amplify the necessary genes (Sawabe et al., 2013). It also requires bioinformatics analyzes which can be tedious to set up. It is also important to note that MLSA 1, 4 and 5 are related, as MLSA 4 is MLSA 1 + 16S rRNA gene and MLSA 1 and 5 are the same with different strain delineating identity cutoff. MLSA 1 and 2 produce similar results but are set apart by the number of genes they require to be sequenced. MLSA 2 necessitates twice less genes than MLSA 1, which makes it easier to use and cheaper.

MALDI-ToF gave inaccurate results to the species level but has numerous advantages over the other methods:

- It is able to identify *a priori* any genus and can be used as a preliminary step to identify the genus of the newly isolated strains.
- Even though the identification is not accurate at the species level, it is accurate to the species group level (Fig. 4).
- It is the fastest method tested in this study, as the colonies are directly used for the analysis and the results are produced in less than an hour.
- It is high throughput as up to 96 strains could be analyzed at the same time.
- The price per isolate is small, as it is neighboring USD 1 (taking amortization of the equipment into account).
- Its performance can be greatly enhanced with software and database revisions and is likely to improve over time (Florio et al., 2018).

7. Conclusion

Species belonging to *Vibrio* genus are important part of aquatic ecosystems (Farmer and Michael Janda, 2015; Takemura et al., 2014) and especially aquaculture ecosystems, as they can be beneficial or cause disease to the fish (Defoirdt et al., 2007; J. Thompson et al., 2010). Identification by standard techniques such as 16S rRNA sequencing of species belonging to this genus is known to be unreliable and different techniques have been proposed in the literature. By comparing several of them, our objective was to provide insight on their up and downsides and to encourage researchers to use relevant identification techniques for *Vibrio* spp. identification.

MALDI-ToF identification can be used as a preliminary step to identify the genus of the isolated strains and *fur* identification can be used for a fast accuracy improvement, bringing the identification closer to the species level (Fig. 6). These two methods do not provide accurate identification but they nonetheless yield sub-genus clade identification and more importantly for routine diagnosis, they deliver results in a day which is essential to limit loss in infected farms. 16S rRNA identification could also be used for genus level identification but it requires more time to complete than MALDI-ToF.

If time and money are less of a constraint, MLSA 1 or 2 should be used for species level identification. To make up for the flaws of the tested MLSA schemes, new ones can be developed but they will also probably need new updates as new species are discovered, which makes them hard to rely on when absolute precision is required. For research works on a specific species, dDDH should be used as it proved to be the only method to reach species-level precision in our dataset.

In the small subset of the collection that was sequenced, 4 potentially new species were discovered and it is possible that more could be highlighted by sequencing the entire collection. Future steps could include the analysis of the geographic diversity of *Vibrio* spp. isolates in Brazilian shrimp farms, based on WGS data. (See Fig. 7.)

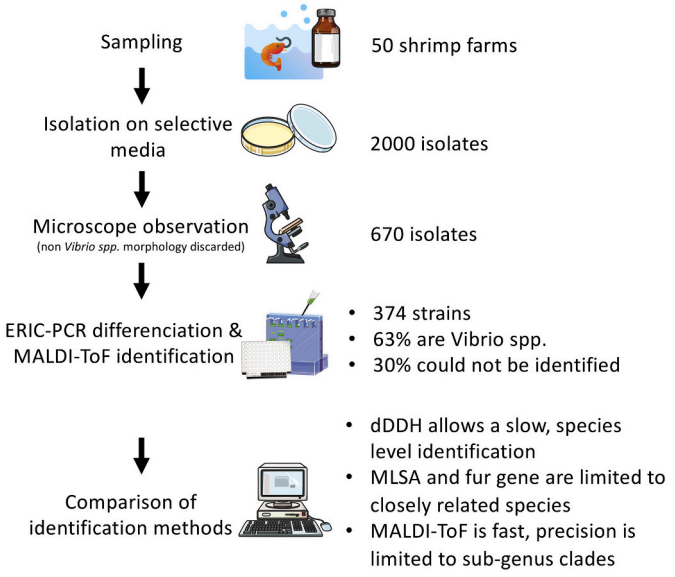

Fig. 7. Summary of the works presented in this paper. Icons are from Servier (Servier, 2020) and Becris, and are used under CC BY 3.0 license.

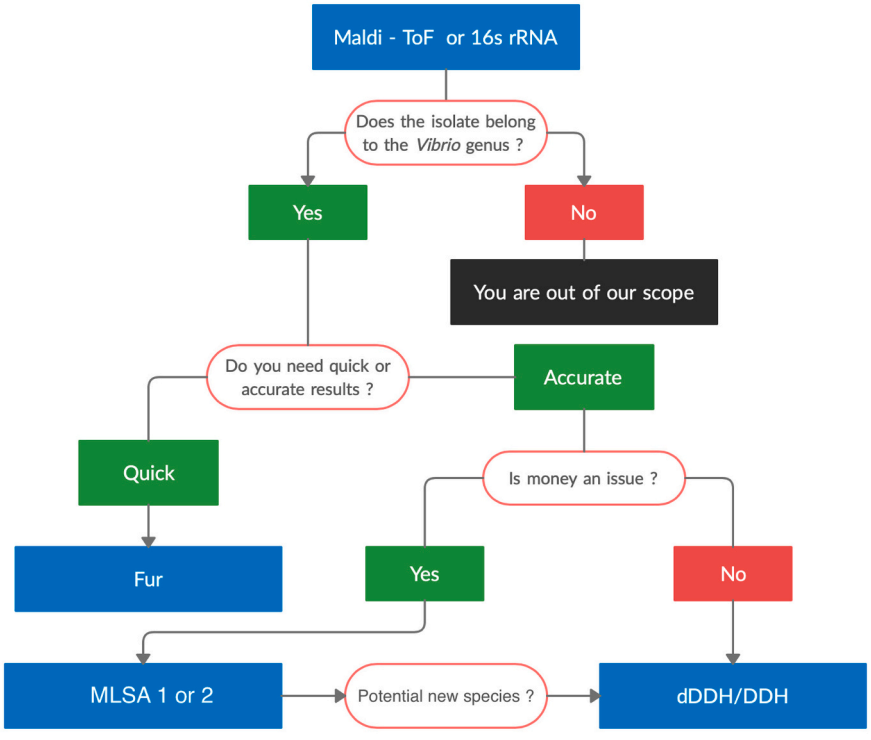

Fig. 6. Decision tree.

Author contributions

Conceptualization : Michel Gautier

Data curation, Formal analysis, Software, Visualization, Writing – original draft : Antoine Culot

Funding acquisition : Michel Gautier, Clarisse Techer, Stéphane Frouël, Jean-Christophe Giard

Investigation : Antoine Culot, Noel Grosset, Quentin Bruey, Michel Auzou, Baptiste Favard, Alexandre Wakatsuki

Methodology : Antoine Culot, Michel Gautier

Project administration, Supervision, Writing – review & editing : Michel Gautier, Clarisse Techer, Stéphane Frouël

Resources : Alexandre Wakatsuki

Validation : Michel Gautier

Data availability statement

The sequences (cleaned reads, assemblies, CDS used for analysis) were deposited on ENA-EMBL database under the accession number PRJEB39286.

Declaration of competing interest

The authors declare no potential conflict of interests.

Acknowledgments

This work was supported by the Association Nationale de la Recherche et de la Technologie, Paris, France (Grant N° 2017/0726).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mimet.2020.106106.

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Blankenberg, D., 2018, July. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research 46 (W1), W537–W544. https://doi.org/10.1093/nar/gky379.

Akazawa, N., Alvial, A., Blanc, Pierre-Philippe, Burgos, J.M., Chamberlain, G.W., Forster, J, Villarreal, M, 2014. Reducing disease risk in aquaculture - World Bank Report Number 88257-GLB. World Bank Rep. https://doi.org/10.13140/RG.2.1.4525.5529.

Andrews, S., 2020. S-Andrews/FastQC. January, Retrieved 2020-01-08, from. https://github.com/s-andrews/FastQc.

Bauer, J., Teitge, F., Neffe, L., Adamek, M., Jung, A., Peppler, C., Jung-Schroers, V., 2018, December. Recommendations for identifying pathogenic *Vibrio* spp. as part of disease surveillance programmes in recirculating aquaculture systems for Pacific white shrimps (*Litopenaeus vannamei*). J Fish Dis 41 (12), 1877–1897. https://doi.org/10.1111/jfd.12897.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10 (1), 421. https://doi.org/10.1186/1471-2105-10-421.

Cano-Gomez, A., Høj, L., Owens, L., Andreakis, N., 2011, December. Multilocus sequence analysis provides basis for fast and reliable identification of *Vibrio harveyi*-related species and reveals previous misidentification of important marine pathogens. Syst. Appl. Microbiol. 34 (8), 561–565. https://doi.org/10.1016/j.syapm.2011.09.001.

Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009, August. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25 (15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Coutard, F., Crassous, P., Droguet, M., Gobin, E., Colwell, R.R., Pommepuy, M., Hervio-Heath, D., 2007, June. Recovery in culture of viable but nonculturable *Vibrio parahæmolyticus*: regrowth or resuscitation? ISME J 1 (2), 111–120. https://doi.org/10.1038/ismej.2007.1.

Culot, A., Grosset, N., Gautier, M., 2019, November. Overcoming the challenges of phage therapy for industrial aquaculture: A review. Aquaculture 513, 734423. https://doi.org/10.1016/j.aquaculture.2019.734423.

De Vos, P., 2011. Multilocus sequence determination and analysis. In: Methods in Microbiology, 38. Elsevier, pp. 385–407. https://doi.org/10.1016/B9780-12-387730-7.00017-6.

Defoirdt, T., Boon, N., Sorgeloos, P., Verstraete, W., Bossier, P., 2007, October. Alternatives to antibiotics to control bacterial infections: luminescent vibriosis in aquaculture as an example. Trends Biotechnol. 25 (10), 472–479. https://doi.org/10.1016/j.tibtech.2007.08.001.

Dongen, Stijn, 2000. *Graph Clustering by Flow Simulation* (Unpublished Doctoral Dissertation).

El-Barbary, M.I., 2017, March. First recording of *Shewanella putrefaciens* in cultured *Oreochromis niloticus* and its identification by 16Sr RNA in Egypt. Egypt. J. Aquatic Res. 43 (1), 101–107. https://doi.org/10.1016/j.ejar.2017.01.002.

Eshik, M.M.E., Abedin, M.M., Punom, N.J., Begum, M.K., Rahman, M.S., 2018, January. Molecular identification of AHPND positive *Vibrio Parahæmolyticus* causing an outbreak in south-west shrimp farming regions of Bangladesh. J. Bangladesh Acad. Sci. 41 (2), 127–135. https://doi.org/10.3329/jbas.v41i2.35492.

Evangelista-Barreto, N.S., de Carvalho, F.C.T., Vieira, R.H.S.D.F., dos Reis, C.M.F., Macrae, A., Rodrigues, D.D.P., 2010, June. Characterization of *Aeromonas* species isolated from an estuarine environment. Braz. J. Microbiol. 41 (2), 452–460. https://doi.org/10.1590/S151783822010000200027.

FAO (Ed.), 2018. *Meeting the Sustainable Development Goals* (No. 2018). Rome.

Farmer, J., Michael Janda, J., 2015, September. Vibrionaceae. In: Whitman, W.B., et al. (Eds.), Bergey's Manual of Systematics of Archaea and Bacteria. John Wiley & Sons, Ltd., Chichester, UK, pp. 1–7. https://doi.org/10.1002/9781118960608.fbm00212.

Fernández-Delgado, M., García-Amado, M.A., Contreras, M., Incani, R.N., Chirinos, H., Rojas, H., Suárez, P., 2015. Survival, induction and resuscitation of *Vibrio cholerae* from the viable but non-culturable state in the southern Caribbean Sea. Rev. Inst. Med. Trop. Sao Paulo 57 (1), 21–26. https://doi.org/10.1590/S0036-46652015000100003 (Jan-Feb).

Florio, W., Tavanti, A., Barnini, S., Ghelardi, E., Lupetti, A., 2018, May. Recent advances and ongoing challenges in the diagnosis of microbial infections by MALDI-TOF mass spectrometry. Front. Microbiol. 9, 1097. https://doi.org/10.3389/fmicb.2018.01097.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012, December. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28 (23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Genouest, 2019. GenOuest Bioinformatics – Development, Expertise and Resources for Bioinformatics. Retrieved 2019-09-06, from. https://ww$_w$.genouest.org/.

Giron, K., Eric, C., 2019. A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers (New England Biolabs).

Gomez-Gil, B., Soto-Rodríguez, S., Lozano, R., Betancourt-Lozano, M., 2014, May. Draft genome sequence of *Vibrio parahæmolyticus* strain M0605, which causes severe mortalities of shrimps in Mexico. Genome Announc. 2 (2) https://doi.org/10.1128/genomeA.00055-14.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J. M., 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57 (1), 81–91. https://doi.org/10.1099/ijs.0.64483-0.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., Bork, P., 2017. Fast genome-wide functional annotation through Orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34 (8), 2115–2122. https://doi.org/10.1093/molbev/msx148. January.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Bork, P., 2019, January. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research 47 (D1), D309–D314. https://doi.org/10.1093/nar/gky1085.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119. https://doi.org/10.1186/1471-2105-11-119.

International Human Genome Sequencing Consortium, 2001, February. Initial sequencing and analysis of the human genome. Nature 409 (6822), 860–921. https://doi.org/10.1038/35057062.

Katoh, K., Standley, D.M., 2013, April. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Kokkari, C., Sarropoulou, E., Bastias, R., Mandalakis, M., Katharios, P., 2018, January. Isolation and characterization of a novel bacteriophage infecting *Vibrio alginolyticus*. Arch. Microbiol. https://doi.org/10.1007/s00203-018-1480-8.

Langmead, B., Salzberg, S.L., 2012, March. Fast gapped-read alignment with bowtie 2. Nat. Methods 9 (4), 357–359. https://doi.org/10.1038/nmeth.1923.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N, 1000 Genome Project Data Processing Subgroup, 2009, August. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lotz, M.J., Tamplin, M.L., Rodrick, G.E., 1983. Thiosulfate-citrate-bile salts-sucrose agar and its selectivity for clinical and marine *vibrio* organisms. Ann. Clin. Lab. Sci. 13 (1), 45–48 (Jan-Feb).

Löytynoja, A., 2014. Phylogeny-aware alignment with PRANK. Methods Mol. Biol. 1079, 155–170. https://doi.org/10.1007/978-1-62703-646-7_10.

Machado, H., Gram, L., 2015. The fur Gene as a new phylogenetic marker for *Vibrionaceae* species identification. Appl. Environ. Microbiol. 81 (8), 2745–2752. https://doi.org/10.1128/AEM.00058-15 (April).

Machado, H., Cardoso, J., Giubergia, S., Rapacki, K., Gram, L., 2017. FurIOS: A web-based tool for identification of *Vibrionaceae* species using the fur Gene. Front. Microbiol. 8 https://doi.org/10.3389/fmicb.2017.00414. March.

Meier-Kolthoff, J.P., Göker, M., 2019, December. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat. Commun. 10 (1), 2182. https://doi.org/10.1038/s41467-019-10210-3.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14 (1), 60. https://doi.org/10.1186/1471-2105-14-60.

Migale, 2019. Migale platform | Migale. Retrieved 2020-01-13, from. https://migale.inra.fr/.

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., Gurevich, A., 2018, July. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34 (13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266.

Moore, W.E.C., Stackebrandt, E., Kandler, O., Colwell, R.R., Krichevsky, M.I., Truper, H. G., Moore, L.H., 1987, October. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. International Journal of Systematic and Evolutionary Microbiology 37 (4), 463–464. https://doi.org/10.1099/00207713-37-4-463.

National Library of Medicine (US), National Center for Biotechnology Information, 2019. Bacterial Antimicrobial Resistance Reference Gene … (ID 313047) - BioProject - NCBI. Retrieved 2020-01-08, from. https://www.ncbi.nlm.nih.gov/bioproject/313047.

National Library of Medicine (US), National Center for Biotechnology Information, 2019. Nucleotide. Retrieved 2020-01-08, from. https://www.ncbi.nlm.nih.gov/nucleotide/.

Nishibuchi, M., Ishibashi, M., Takeda, Y., Kaper, J.B., 1985, September. Detection of the thermostable direct hemolysin gene and related DNA sequences in *Vibrio parahæmolyticus* and other *vibrio* species by the DNA colony hybridization test. Infect. Immun. 49 (3), 481–486. Retrieved 2020-01-15, from. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC261186/.

Niu, B., Fu, L., Sun, S., Li, W., 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. BMC Bioinformatics 11 (1), 187. https://doi.org/10.1186/1471-2105-11-187.

Nunan, L., Lightner, D., Pantoja, C., Gomez-Jimenez, S., 2014. Detection of acute hepatopancreatic necrosis disease (AHPND) in Mexico. Dis. Aquat. Org. 111 (1), 81–86. https://doi.org/10.3354/dao02776 (August).

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Pevzner, P.A., 2013. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: Deng, M., Jiang, R., Sun, F., Zhang, X. (Eds.), Research in Computational Molecular Biology. Springer, Berlin, Heidelberg, pp. 158–170. https://doi.org/10.1007/978-3-642-37195-0_13.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Parkhill, J., 2015, November. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 31 (22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421.

Pascual, J., Macian, M.C., Arahal, D.R., Garay, E., Pujalte, M.J., 2010, January. Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, recA, pyrH, rpoD, gyrB, rctB and toxR genes. Int. J. Syst. Evol. Microbiol. 60 (1), 154–165. https://doi.org/10.1099/ijs.0.010702-0.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010, March. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 5 (3), e9490. https://doi.org/10.1371/journal.pone.0009490.

Quinlan, A.R., Hall, I.M., 2010, March. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26 (6), 841–842. https://doi.org/10.1093/bioinformatics/btq033.

R Core Team, 2019. R: A Language and Environment for Statistical Computing.

Raina, V., Nayak, T., Ray, L., Kumari, K., Suar, M., 2019. A Polyphasic taxonomic approach for designation and description of novel microbial species. In: Microbial Diversity in the Genomic Era. Elsevier, pp. 137–152 (doi: 10.1016/B978-0-12-814849-5.00009-5 Rambaut, A. (2018). FIgTree).

Reichelt, J.L., Baumann, P., Baumann, L., 1976. Study of genetic relationships among marine species of the genera *Beneckea* and *Photobacterium* by means of in vitro DNA/DNA hybridization. Arch. Microbiol. 110 (1), 101–120. https://doi.org/10.1007/BF00416975.

Sawabe, T., Kita-Tsukamoto, K., Thompson, F.L., 2007, November. Inferring the evolutionary history of *Vibrios* by means of multilocus sequence analysis. J. Bacteriol. 189 (21), 7932–7936. https://doi.org/10.1128/JB.00693-07.

Sawabe, T., Ogura, Y., Matsumura, Y., Feng, G., Amin, A.R., Mino, S., 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* https://doi.org/10.3389/fmicb.2013.00414.

Schmieder, R., Edwards, R., 2011, March. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27 (6), 863–864. Retrieved 2020-01-08, from. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3051327/.

Seemann, T., 2014, July. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 (14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Seemann, T., 2019. December. In: Tseemann/barrnap. Retrieved 2020-01-08, from. https://github.com/tseemann/barrnap.

Seppey, M., Manni, M., Zdobnov, E.M., 2019. BUSCO: Assessing genome assembly and annotation completeness. In: Kollmar, M. (Ed.), Gene Prediction, vol. 1962. Springer New York, New York, NY, pp. 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.

Servier, 2020. SMART. Retrieved 2020-05-28, from. https://smart.servier.com/.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Higgins, D.G., 2011, January. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7 (1), 539. https://doi.org/10.1038/msb.2011.75.

Subramanian, M., Maruthamuthu, M., 2019, February. Draft genome sequences of two *Bacillus* spp. and an *Oceanobacillus* sp. strain isolated from marine macroalgae. Microbiol. Resour. Announc. 8 (7) https://doi.org/10.1128/MRA.01417-18 e01417-18, /mra/8/7/MRA.01417-18.atom.

Takemura, A.F., Chien, D.M., Polz, M.F., 2014. Associations and dynamics of *Vibrionaceae* in the environment, from the genus to the population level. Front. Microbiol. 5 https://doi.org/10.3389/fmicb.2014.00038.

Tange, O., 2018. Gnu parallel 2018. Zenodo. https://doi.org/10.5281/ZENODO.1146014.

The UniProt Consortium, 2019, January. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47 (D1), D506–D515. https://doi.org/10.1093/nar/gky1049.

Thompson, F.L., Hoste, B., Vandemeulebroecke, K., Swings, J., 2003. Reclassification of *Vibrio hollisae* as *Grimontia hollisae* gen. nov., comb. nov. Int. J. Syst. Evol. Microbiol. 53 (5), 1615–1617. https://doi.org/10.1099/ijs.0.02660-0.

Thompson, C.C., Vicente, A.C.P., Souza, R.C., Vasconcelos, A.T.R., Vesth, T., Alves, N., Thompson, F.L., 2009, October. Genomic taxonomy of *vibrios*. BMC Evol Biol 9, 258. https://doi.org/10.1186/1471-2148-9-258.

Thompson, J., Gregory, S., Plummer, S., Shields, R., Rowley, A., 2010, October. An in vitro and in vivo assessment of the potential of *Vibrio* spp. as probiotics for the Pacific white shrimp, *Litopenaeus vannamei*. J. Appl. Microbiol. 109 (4), 1177–1187. https://doi.org/10.1111/j.1365-2672.2010.04743.x.

United Nations Department of Economic and Social Affairs, 2019. *World Population Prospects 2019* (Tech. Rep.). Retrieved from. https://population.un.org/wpp/.

Urbanczyk, H., Ogura, Y., Hayashi, T., 2013. Taxonomic revision of *Harveyi* clade bacteria (family *Vibrionaceae*) based on analysis of whole genome sequences. Int. J. Syst. Evol. Microbiol. 63 (7), 2742–2751. https://doi.org/10.1099/ijs.0.051110-0.

Vasudevan, K., Devanga Ragupathi, N.K., Jacob, J.J., Veeraraghavan, B., 2020, January. Highly accurate-single chromosomal complete genomes using IonTorrent and MinION sequencing of clinical pathogens. Genomics 112 (1), 545–551. https://doi.org/10.1016/j.ygeno.2019.04.006.

Versalovic, J., Koeuth, T., Lupski, R., 1991. Distribution of repetitive DNA sequences in eubacteria and application to finerprinting of bacterial enomes. Nucleic Acids Res. 19 (24), 6823–6831.

Vishnivetskaya, T.A., Kathariou, S., Tiedje, J.M., 2009, May. The *Exiguobacterium* genus: biodiversity and biogeography. Extremophiles 13 (3), 541–555. https://doi.org/10.1007/s00792-009-0243-5.

Woo, P.C.Y., Teng, J.L.L., Wu, J.K.L., Leung, F.P.S., Tse, H., Fung, A.M.Y., Yuen, K.-y., 2009, August. Guidelines for interpretation of 16S rRNA gene sequence-based results for identification of medically important aerobic Gram-positive bacteria. Journal of Medical Microbiology 58 (8), 1030–1036. https://doi.org/10.1099/jmm.0.008615-0.

World Health Organization, 2017. *Selection and Application of Methods for the Detection and Enumeration of Humanpathogenic … Halophilic Vibrio SPP.* in Seafood: Guidance. Place of publication not identified: World Health Organization.