Comparative Genomics of *Tenacibaculum dicentrarchi* and “*Tenacibaculum finnmarkense*” Highlights Intricate Evolution of Fish-Pathogenic Species

Sébastien Bridel\(^1,2,3\), Anne-Berit Olsen\(^4\), Hanne Nilsen\(^4\), Jean-François Bernardet\(^1\), Guillaume Achaz\(^5\), Ruben Avendaño-Herrera\(^6,7\), and Eric Duchaud\(^1\)*

\(^1\)VIM, INRA, Université Paris-Saclay, Jouy-en-Josas, France  
\(^2\)Labofarm, Finalab, Louédac, France  
\(^3\)Université de Versailles Saint-Quentin-En-Yvelines, Montigny-Le-Bretonneux, France  
\(^4\)Norwegian Veterinary Institute, Bergen, Norway  
\(^5\)Atelier de Bioinformatique, UMR 7205 ISyEB, MNHN-UPMC-CNRS-EPHE, Muséum National d’Histoire Naturelle, Paris, France  
\(^6\)Laboratorio de Patología de Organismos Acuáticos y Biotecnología Acuícola, Departamento de Ciencias Biológicas, Facultad de Ciencias Biológicas, Universidad Andrés Bello, Viña del Mar, Chile  
\(^7\)Interdisciplinary Center for Aquaculture Research (INCAR), Concepción, Chile

*Corresponding author: E-mail: eric.duchaud@inra.fr.

Accepted: January 18, 2018

**Data deposition:** All genome sequencing data have been deposited at the European Nucleotide Archive under the accessions GCA_900239455, GCA_900239305, GCA_900239485, GCA_900239185, GCA_900239505, GCA_900239345, and GCA_900239495.

**Abstract**

The genus *Tenacibaculum* encompasses several species pathogenic for marine fish. *Tenacibaculum dicentrarchi* and “*Tenacibaculum finnmarkense*” (Quotation marks denote species that have not been validly named.) were retrieved from skin lesions of farmed fish such as European sea bass or Atlantic salmon. They cause a condition referred to as tenacibaculosis and severe outbreaks and important fish losses have been reported in Spanish, Norwegian, and Chilean marine farms. We report here the draft genomes of the *T. dicentrarchi* and “*T. finnmarkense*” type strains. These genomes were compared with draft genomes from field isolates retrieved from Chile and Norway and with previously published *Tenacibaculum* genomes. We used Average Nucleotide Identity and core genome-based phylogeny as a proxy index for species boundary delineation. This work highlights evolution of closely related fish-pathogenic species and suggests that homologous recombination likely contributes to genome evolution. It also corrects the species affiliation of strain AYD7486TD claimed by Grothusen et al. (2016).

**Key words:** *Tenacibaculum*, tenacibaculosis, genomes, fish pathogens, virulence, evolution.

**Introduction**

The rapid development of intensive aquaculture has been associated with a dramatic increase in outbreaks of infectious diseases (FAO 2016; Bayliss et al. 2017). The rapid international spread of pathogens through the trade of fish and eggs or as a response to environmental changes has been documented (Brynilsrud et al. 2014; Rahmati-Holasoo et al. 2016). In this context, the success and sustainability of aquaculture largely depend on the control of pathogens. Among those, several species of the genus *Tenacibaculum* are responsible for diseases collectively designated tenacibaculosis (Avendaño-Herrera et al. 2006; Suzuki 2015). *Tenacibaculum dicentrarchi* and “*Tenacibaculum finnmarkense*” are two among those fish-associated, recently described species. The former was first isolated from European sea bass (*Dicentrarchus labrax*) in Spain (Piñeiro-Vidal et al. 2012) and recently also identified from Atlantic salmon (*Salmo salar*) in Chile (Avendaño-Herrera et al. 2016)
and Norway (Olsen et al. 2017) and from red conger eel (Genypterus chilensis) in Chile (Irgang et al. 2017). “Tenacibaculum finnmarkense” was isolated from Atlantic salmon with ulcerative disease in Norway (Småge et al. 2016).

Identification of the causative agent of tenacibaculosis was first based on the isolation of bacteria from tissues of diseased fish and their characterization by phenotypic, biochemical, and serological methods (Piñeiro-Vidal, Carballas et al. 2008; Piñeiro-Vidal, Riaza et al. 2008). The use of 16S rDNA sequencing improved the identification reliability (Cepeda et al. 2003; Fringuelli et al. 2012). However, these methods usually cannot differentiate closely related bacterial species. MLST was developed (Habib et al. 2014) and used to demonstrate the presence of T. dicentrarchi in Chile (Avendaño-Herrera et al. 2016) and to reveal the variety of Tenacibaculum spp. in a number of sea-farmed fish species in Norway (Olsen et al. 2017).

In this study, we present seven draft genomes of Tenacibaculum strains, including the T. dicentrarchi (USC 3509T) and “T. finnmarkense” (HFI) type strains, as well as five field isolates from Chile and Norway selected on the basis of available MLST data (Olsen et al. 2017). Comparison has been performed with available Tenacibaculum genomes from Genbank, including strain AYD7486TD originally described as T. dicentrarchi (Grothusen et al. 2016). We used Average Nucleotide Identity (ANI) to delineate species boundaries and core genome analysis to infer phylogenetic relationships between these strains. We also correct the species affiliation of strain AYD7486TD.

Materials and Methods

Bacterial Strains

The T. dicentrarchi type strain USC 3509T (Piñeiro-Vidal et al. 2012) was obtained from Dr Y. Santos (University of Santiago de Compostela, Spain) and the “T. finnmarkense” strain HFI (Småge et al. 2016) was obtained from Dr H. Duesund (Cermaq Group AS, Bergen, Norway). Strains TNO006, TNO010, and TNO020 were isolated from skin ulcers of Atlantic salmon in Norway whereas strain TNO021 was isolated from mouth ulcer of a corkwing-wrasse (Symphodus melops) also in Norway (Habib et al. 2014; Olsen et al. 2017). Strain TdChD05 was retrieved from external lesion of an Atlantic salmon in Chile (Avendaño-Herrera et al. 2016). All strains were routinely grown on marine agar 2216 (Difco) and in the corresponding broth at 170 rpm and 15°C (TNO006) or 22°C (all other strains).

Genome Sequencing and Annotation

Genomic DNA was extracted with the Wizard Genomic DNA Purification Kit (PROMEGA). All strains were sequenced with Illumina (HiSeq 2x100 pair-end reads with 300 bp insert size). Sequencing reads were assembled using Velvet (Zerbino and Birney 2008) or SPAdes (Bankevich et al. 2012). Genome annotation, including manual curation, and genome comparison including core genome computation were performed using the MicroScope platform (Médiége et al. 2017 and references therein).

ANI and Phylogenetic Reconstruction

ANIs analyses were performed using the ANIm method described by Richter and Rosselló-Móra (2009) and implemented in the Python module Pyani (https://github.com/widdowquinn/pyani). Digital DNA–DNA hybridization was performed using the GGDC website (http://ggdc.dsmz.de/) and Formula 2 (Auch et al. 2010). Logistic regression was performed using the GGDNC website (http://ggdc.dsmz.de/). The resulting alignments were manually checked and concatenated for tree reconstruction using UGENE and PhyML with default parameters (Okonechnikov et al. 2012). Tree rendering was achieved using the Figtree software (http://tree.bio.ed.ac.uk/software/figtree/). The neighbor-net analysis in the Splits Tree 4 software (http://splitstree.org/; Huson and Bryant 2006) provided a phylogenetic network representing possible evolutionary relationships between the concatenated sequences of core genome genes. Minimal recombination breakpoints were identified using the four-gamete test (Hudson and Kaplan 1985). Putative recombination events were indicated as pairwise homoplasy index (PH; Bruen et al. 2006) calculated by Splits Tree 4.

Results and Discussion

General Genome Features

A summary of the genomes analyzed in this study is presented in supplementary table 1, Supplementary Material online. Strikingly, the sizes of the genomes reported here are the smallest among those available to date for the genus Tenacibaculum (e.g., T. maritimum: 3.4 Mb, T. soleae: 3.0 Mb, T. ovolyticum: 4.1 Mb). The average genome size is 2.7 Mb and 2.9 Mb for T. dicentrarchi and “T. finnmarkense”, respectively; at 2.4 Mb, strain TNO020 has the smallest genome. All strains studied are devoid of plasmid.
ANI Delineates *T. dicentrarchi* and “*T. finnmarkense*” and Allocates Strain AY7486TD to “*T. finnmarkense*”

ANI was reported to be an accurate and practical method for species delineation and a 95–96% identity was proposed as the threshold (Rodriguez-R and Konstantinidis 2014). ANI comparisons were computed using *T. ovolyticum* (Suzuki et al. 2001) as an outgroup and the results were plotted as a heatmap (supplementary fig. 1A and 1B, Supplementary Material online). Strains TdChD05 from Chile and TNO021 from Norway, both previously allocated to the species *T. dicentrarchi* (Piñeiro-Vidal et al. 2012; Olsen et al. 2017; Avendaño-Herrera et al. 2016), indeed form a highly cohesive group with *T. dicentrarchi* USC 3509\(^\text{1}\) (98% ANI and \(\geq88\%\) alignment coverage). In contrast, strain AY7486TD, also originally described as *T. dicentrarchi* (Grothusen et al. 2016), displays an ANI value of only 93% with the type strain of this species and consequently does not fall within the *T. dicentrarchi* cluster. Instead, strain AYD7486TD forms a cluster with strains TNO006, TNO010 and “*T. finnmarkense*” HFI\(^\text{1}\) (\(\geq96\%\) ANI and \(\geq85\%\) alignment coverage, above the species delineation threshold). ANI values delineate two subclusters, one grouping strains TNO010 and AYD7486TD (99% ANI) and the other grouping strain TNO006 and “*T. finnmarkense*” HFI\(^\text{1}\) (98% ANI). Supplementary figure 1, Supplementary Material online, also shows that although *T. dicentrarchi* and “*T. finnmarkense*” are distinct species they obviously display significant proximity in terms of sequences identity (93–94% ANI) and fraction of shared genomes (77–88%). Strain TNO020 does not belong to any of the previously defined clusters (88–89% ANI; 55–68% alignment coverage) and therefore likely belongs to a yet undescribed *Tenacibaculum* species as previously suggested (Habib et al. 2014; Olsen et al. 2017). The same conclusions (supplementary fig. 1C, Supplementary Material online) were drawn using genome-to-genome distance calculator (Auch et al. 2010). As expected, *T. ovolyticum* DSM 18103\(^\text{1}\) behaves as an outgroup displaying lower sequence identity (85%) and poor alignment coverage (17–34%) with all other strains studied. Using 895 core genome-encoding protein sequences, we constructed a phylogenetic tree from the concatenation of each individual alignment (fig. 1). Bootstrap values strongly support the division between *T. dicentrarchi* and “*T. finnmarkense*” and the core genome-based phylogenetic tree perfectly matches the ANI dendrogram. Furthermore, a correlation between the MLST clades defined by Olsen et al. (2017) and the clusters observed in figure 1 is obvious: “*T. finnmarkense*” HFI\(^\text{1}\) and strain TNO006 belong to clade III, strains TNO010 and AYD7486TD to clade I, all three *T. dicentrarchi* strains to clade II and strain TNO020 to clade IV. Using Splits Tree analysis, a reticulated network structure between the four clades, indicative of within and between species recombination events (fig. 2), is observed. The dense network joining clade I and clade III is in good agreement with the grouping of “*T. finnmarkense*” strains in two connected subclusters as previously observed in the MLST data set of Olsen et al. (2017). Whatever the group of strains considered (i.e., clades I and III strains, clades I, II, and III strains as well as clades I, II, III, and IV strains), the PHI test P-value is 0.0, indicating significant evidence of recombination. In addition, the high number of recombination breakpoints found in *T. dicentrarchi* and “*T. finnmarkense*” genomes (10 per kilobase in average) is another clue suggesting recombination events in these species.

Comparative Genomics Highlights Intricate Relationships between *T. dicentrarchi* and “*T. finnmarkense*”

The average number of predicted CDS for *T. dicentrarchi* and “*T. finnmarkense*” strains is 2,381 and 2,536, respectively, which is in good agreement with the observed genome sizes. The core genome is composed of 2,013 and 1,947 CDS for *T. dicentrarchi* and “*T. finnmarkense*,” respectively (supplementary fig. 2A and B, Supplementary Material online). These small gene sets, about half those of *Tenacibaculum agarivorans* HZ21 (Xu et al. 2017) and *Tenacibaculum jejuense* KCTC 22618\(^\text{1}\) (Ficko-Blean 2017), seem related to a deficient biopolymer-degrading ability of *T. dicentrarchi* and “*T. finnmarkense*.” Indeed, these genomes lack the pathways encoding for the degradation of marine carbohydrates (e.g., sulfatase, glycoside hydrolase, polysaccharide lyase) identified in the environmental species *T. agarivorans* and *T. jejuense*, in line with a restricted ecological niche (i.e., fish tissues) and an exclusive protein-based predicted regimen for these pathogenic species. Moreover, the presence of insertion sequences or their scars as well as genes remnants in *T. dicentrarchi*, “*T. finnmarkense*” and strain TNO020 argues for genome reduction trends in contrast to the horizontal transfer genes in *T. jejuense* and *T. agarivorans*. These findings support the expected small genome size of bacterial pathogens compared with their nonpathogenic relatives (Weinert and Welch 2017). The core genome of the seven strains belonging to both *T. dicentrarchi* and “*T. finnmarkense*” is composed of 1,818 CDS (supplementary fig. 2C, Supplementary Material online), a value close to those computed for each species. Each strain has ~180 specific genes essentially composed of prophages remnants, restriction/modification systems, toxin/antitoxin systems and transposases encoding genes or their scars as well as genes required for the biosynthesis of exopolysaccharides that likely account for the minor intraspecies genome size differences previously mentioned. These strain-specific genes, representing the accessory genome, do not seem linked to bacterial pathogenicity as no *bona fide* toxin or virulence factor-encoding genes have been identified in this gene pool. Therefore, virulence-encoding genes likely belong to the core genome common to both *T. dicentrarchi* and “*T. finnmarkense*.” Among those, peptidases containing a
The carboxy-terminal protein domain (TIGR04183), predicted to be required for T9SS-mediated secretion and cell surface exposure (Veith et al. 2013), were identified. These peptidases (i.e., TFINN_2500013, TFINN_140038, and TFINN_60057 from "T. finnmarkense" HFJT and their orthologs) are likely involved in the breakdown of proteinaceous compounds and the destruction of host tissues. The presence of a M9 family protease-encoding gene (TFINN_140038 and orthologs), similar to the 120 kDa collagenase of Clostridium perfringens (Matsushita et al. 1994) but different from the M43 family collagenase (encoded by MARIT_1085) identified in T. maritimum (Pérez-Pascual et al. 2017), suggests convergent evolution for some virulence-linked functions in fish-pathogenic Tenacibaculum species.

Conclusion
Since the pioneering work of Wakabayashi and col. on T. maritimum in the eighties (Wakabayashi et al. 1986), many other Tenacibaculum species have been described. Some of them are important fish pathogens and an unexpected diversity at different levels (e.g., genetic, fish host, geographical) has been reported (Habib et al. 2014; Olsen et al. 2017). Tenacibaculum dicentrarchi strains were previously identified...
from several farmed fish species in Spain, Norway and Chile, whereas “T. finnmarkense” strains were exclusively isolated in Norway so far (Olsen et al. 2017). Thanks to genomes comparison, we were able to correct the affiliation of strain AYD7486TD which actually belongs to the species “T. finnmarkense” rather than to T. dicentrarchi as previously claimed (Grothusen et al. 2016). Importantly, this result demonstrates that “T. finnmarkense” is also present in Chilean fish farms. Our data set suggests that T. dicentrarchi strains form a cohesive group whereas “T. finnmarkense” strains are split into two subclusters. Similar subclusters, referred to as genomovars, were reported in Flavobacterium columnare, another fish pathogen of the family Flavobacteriaceae with a broad host range. Correlations between genomovars, fish hosts and virulence have been suggested (Evenhuis and LaFrentz 2016; Olivares-Fuster et al. 2007). Hence, the same type of genomic heterogeneity observed in “T. finnmarkense” may account for specific traits such as host specificity or level of virulence. Strikingly, the virulence factors described in the closely related species T. maritimum (i.e., a sphingomyelinase, a ceramidase, a chondroitin AC lyase, a slialidase and a M43 family collagenase; Pérez-Pascual et al. 2017) have not been identified in any of the genomes described in the present study. In full agreement with this observation, a parallel evolution of pathogenicity in the species encompassed in the genus Tenacibaculum has been proposed (Habib et al. 2014). However, the grouping of T. dicentrarchi and “T. finnmarkense” as well as T. soleae and T. ovolycicum (Habib et al. 2014; Olsen et al. 2017; Småge et al. 2016) in a single clade exclusively encompassing fish-associated bacteria suggests evolution of these four species from a pathogenic ancestor. In addition, our data support recombination as an important force shaping genome evolution of these fish pathogens as previously observed in other members of the family Flavobacteriaceae (Nicolas et al. 2008; Vos and Didelot 2009). The genome sequences reported here should therefore facilitate future epidemiological studies and provide new insights into pathogenicity and niche adaptation of emerging fish pathogens.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments

This study was funded by the EU EMIDA ERA-NET project “Control Flavobacteriaceae Infections in European Fish farms” and by the Agence Nationale pour la Recherche (contract ANR-14-CE19-0020). This work has benefited from the facilities and expertise of the high-throughput sequencing platform of I2BC (Centre de Recherche de Gif; http://www.i2bc.paris-saclay.fr/spip.php?article3998&lang=en). We also wish to thank the following structures for providing computational resources: The INRA MIGALE bioinformatics platform (http://migale.jouy.inra.fr), the LABGeM, and the National Infrastructure “France Génomique” funded as part of the Investissement d’avenir program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). A.-H. acknowledges Grant FONDECYT 1150695 and the CONICYT/FONDAP/15110027 from the Comisión Nacional de Investigación Científica y Tecnológica (CONICYT, Chile). The authors are very grateful to Sophie Pasek and Mathilde Carpenter for fruitful discussion.

Literature Cited


Associate editor: Howard Ochman