Complete genome sequence of *Flavobacterium indicum* GPSTA100-9^T isolated from warm spring water

Paul Barbier, Jean-François Bernardet & Eric Duchaud

Unité de Virologie et Immunologie Moléculaires, INRA, Domaine de Vilvert, F-78350 Jouy-en-Josas, France.

Correspondence: pbarbier@jouy.inra.fr

Abstract

The number of described *Flavobacterium* species is constantly increasing and *Flavobacterium* strains occur in a wide range of ecological niches. *Flavobacterium indicum* is non-pathogenic and among the rare mesophilic bacterial species in the genus *Flavobacterium*. This environmental species, isolated from a warm spring water in Assam (India), is also of interest as a member of the *Flavobacteriaceae*, a bacterial family of high importance for the degradation/turnover of organic matter in terrestrial, freshwater and marine ecosystems [1,2]. The main objective of this study was to contribute to the scientific knowledge in the field of microbial ecology. We are using *Flavobacterium* diversity as a good context for comparative analysis of closely related organisms with different life-styles using genomic approaches. We determined the whole genome sequence of the type strain of *F. indicum* (GPSTA100-9^T = CIP 109464^T) [3] in order to perform comparative genomics studies between this environmental species and other *Flavobacterium* species whose genome sequences have been published: two fish-pathogenic species that severely impact aquaculture worldwide [1] (*F. psychrophilum* [4] and *F. branchiophilum* [5]) and another environmental species *F. johnsoniae* [6], a model organism for characterizing gliding motility [7] and biopolymer utilization in oligotrophic environments (*F. johnsoniae* strain A3 [8]).

The genome of the type strain of *F. indicum* was sequenced using a combination of Sanger (ABI3730, Applied Biosystem), performed on a genomic DNA library with an average fragment length of 10 kbp cloned in a low copies vector named pCNS; and 454 (GS-FLX, Roche) sequencing with a 2.6 fold and 17.4 fold coverage, respectively. The 454 reads were assembled in 145 contigs using Newbler. These contigs and the Sanger reads were assembled in 4 scaffolds (39 contigs) using Phrap. Scaffolds were ordered using an optical map (OpGen Technologies) [9] and gaps were closed using primer walking on gap-spanning clones or by PCR sequencing. A careful manual annotation of the genome was performed using the AGMIAL annotation platform [10].

The complete genome of *F. indicum* consists of a circular chromosome of 2,993,089 bp without plasmid with an overall G+C content of 31.8%. The genome is predicted to encode 2671 protein-coding genes, 55 tRNA genes and four rRNA operons. Genome comparisons allowed to define 23 large regions not found in previously sequenced *Flavobacterium* genomes. Twenty-two insertion sequences (IS) from four different families were identify and AlienHunter predicted 12 large regions horizontally acquired (Fig. 1), most of them carrying genes of unknown functions. Some phage scars were found in these regions. The whole genome of *F. indicum* will help in the definition of the *Flavobacterium* core-genome, which contains most of essentials genes and important metabolic pathways shared by all sequenced genomes in the genus.

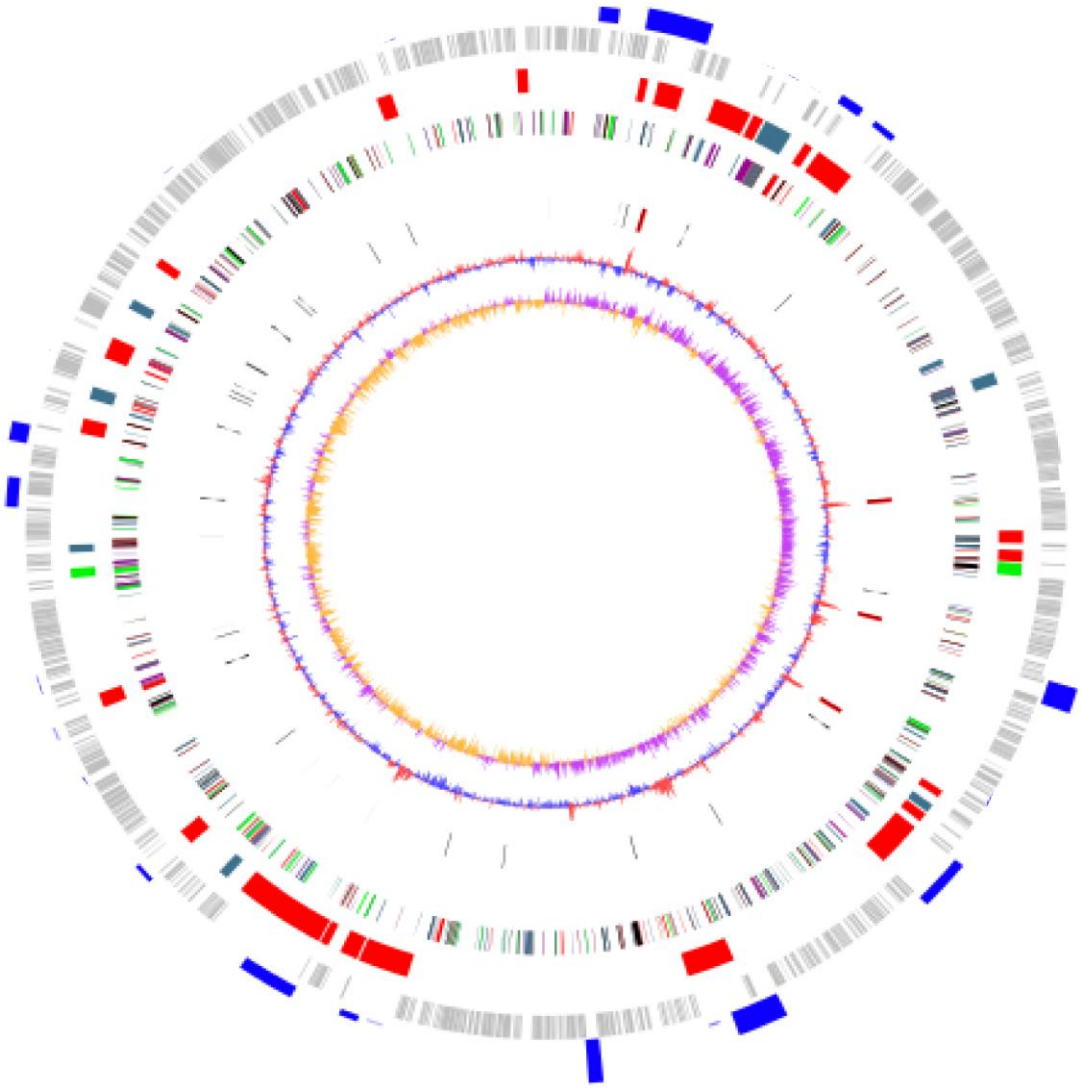

Figure 1. Circular representation of *F.indicum* whole genome. Circles represent the following (from the inside out): **1.** GC skew ; an inversion at the origin and the termini of replication is clearly visible; **2.** GC% local variations ; **3.** location of tRNA genes (black) and rRNA operons (red) ; **4.** Genes with orthologs in *F. psychrophilum* (green), *F. johnsoniae* (blue/grey) and *F. branchiophilum* (magenta) ; **5.** Regions with at least 10 genes not shared by all known *Flavobacterium* genomes. *F. indicum* specific regions (red), shared regions (at least 40% of genes) with *F. psychrophilum* (green) and *F. johnsoniae* (blue/grey) ; **6.** Core-genome genes in light grey ; **7.** Putative horizontal gene transfer regions (at least 10 genes) detected by Alien Hunter (blue).

Genome comparison with other *Flavobacterium* species confirms a loss of synteny at the genus level [5,6], likely due to the presence of many repeats (e.g. IS and rhs elements). Moreover, *F. indicum* genome, wich is two-fold smaller than the 6 Mbp-long genome of *F. johnsoniae* UW101^{T} [6], lacks the 2 Mbp region enriched in genes involved in polysaccharide utilization of the latter (Fig. 2). Among *Flavobacterium* environmental species that have been sequenced to date, *F. indicum* has the smallest genome. These findings corroborate the weak biopolymer-degrading ability of *F. indicum* and tends to suggest that *F. indicum* is adapted to a narrow ecological niche.

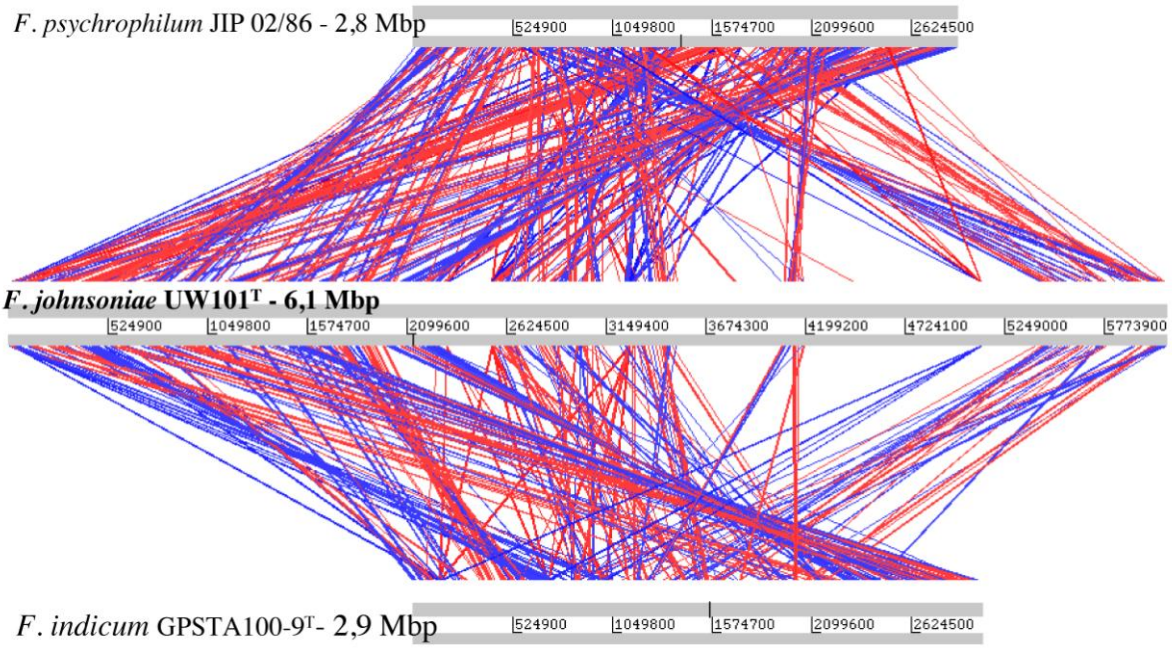

Figure 2. Nucleotidic pair-wise complete genomes comparison. The red and blue bands represent the forward and reverse matches, respectively.

The analysis of the *F. indicum* genome sequence revealed interesting features and genes in relation to its environmental life style. No CRISPR locus [11] was identified in *F. indicum*, which also lacks the toxin-encoding genes previously described in the three fish pathogenic species *F. psychrophilum* [4], *F. branchiophilum* [5] and *F. columnare* [12]. The genome is predicted to encode 38 adhesins, likely used for binding on different surfaces. Six glycoside hydrolase precursors, 50 endo- and exopeptidases and one polysaccharide utilization system (PUL) [13] were found likely in relation with its ability to degrade some macromolecules such as gelatin, casein, and starch [3]. We also identified genes encoding for secretion systems and for protein translocation across the outer membrane. We found 33 proteins with a conserved C-terminal sequence likely associated to the cell surface. In contrast with *F. psychrophilum* and *F. johnsoniae,* which contain flexirubin type pigments, the yellowish-orange color of *F. indicum* is only due to the presence of carotenoid pigments; genes for carotenoid biosynthesis were indeed found. Thirteen gliding motility machinery [14] encoding genes were also identified but the *gldA* gene, encoding the gliding motor, is frameshifted and the *gldE* gene, involved in gliding, is absent. Although not reported in the original description of *F. indicum* [3], weak gliding motility actually occurs [1].

This study provides a new genome within the genus *Flavobacterium* [15] and therefore gives additional information about genome dynamics within the genus. The small genome size and a weak biopolymer degrading ability tend to suggest an adaptation to a likely restricted ecological niche. Many features reveal the environmental life style of the bacterium: no toxins were identified and the 38 genes encoding for adhesins suggest that *F. indicum* mostly adhere to surfaces in its habitat.

References

[1] Bernardet J.-F., Bowman J.P., 2011. Genus I. *Flavobacterium* Bergey et al. 1923. In W. Whitman (ed.), Bergey's Manual of Systematic Bacteriology, pp. 112-154, 2nd ed., Vol. 4, The Williams & Wilkins Co., Baltimore

[2] Brown M.V., Bowman J.P., 2001. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). FEMS Microbiology Ecology, 35:267–275

[3] Saha P., Chakrabarti T., 2006. *Flavobacterium indicum* sp. nov., isolated from warm spring water in Assam, India. International Journal of Systematic and Evolutionary Microbiology, 56:2617–2621

[4] Duchaud E., Boussaha M., Loux V., Bernardet J.-F., Michel C., Kerouault B., Mondot S., Nicolas P., Bossy R., Caron C., Bessières P., Gibrat J.-F., Claverol S., Dumetz F., Le Hénaff M., Benmansour A., 2007. Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum*. Nature Biotechnology, 25:763–769

[5] Touchon M., Barbier P., Bernardet J.-F., Loux V., Vacherie B., Barbe V., Rocha E.P., Duchaud E., 2011. Complete genome sequence of the fish pathogen *Flavobacterium branchiophilum.* Applied and Environmental Microbiology, 77:7656–7662

[6] McBride M.J., Xie G., Martens E.C., Lapidus A., Henrissat B., Rhodes R.G., Goltsman E., Wang W., Xu J., Hunnicutt D.W., Staroscik A.M., Hoover T.R., Cheng Y.Q., Stein J.L., 2009. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. Applied and Environmental Microbiology, 75:6864–6875

[7] McBride M.J., 2004. *Cytophaga-flavobacterium* gliding motility. Journal of Molecular Microbiology and Biotechnology, 7:63–71

[8] Sack E.L., van der Wielen P.W., van der Kooij D., 2011. *Flavobacterium johnsoniae* as a model organism for characterizing biopolymer utilization in oligotrophic freshwater environments. Applied and Environmental Microbiology, 77:6931–6938

[9] Latreille P., Norton S., Goldman B.S., Henkhaus J., Miller N., Barbazuk B., Bode H.B., Darby C., Du Z., Forst S., Gaudriault S., Goodner B., Goodrich-Blair H., Slater S., 2007. Optical mapping as a routine tool for bacterial genome sequence finishing. BMC Genomics, 8:321

[10] Bryson K., Loux V., Bossy R., Nicolas P., Chaillou S., van de Guchte M., Penaud S., Maguin E., Hoebeke M., Bessières P., Gibrat J.-F., 2006. AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system. Nucleic Acids Research, 34:3533–3545

[11] Karginov F.V., Hannon G.J., 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Molecular Cell, 37:7–19

[12] Tekedar H.C., Karsi A., Gillaspy A.F., Dyer D.W., Benton N.R., Zaitshik J., Vamenta S., Banes M.M., Gülsoy N., Aboko-Cole M., Waldbieser G.C., Lawrence M.L., 2012. Genome sequence of the fish pathogen *Flavobacterium columnare* ATCC 49512. Journal of Bacteriology, 194:2763–2764

[13] Reeves A.R., D'Elia J.N., Frias J., Salyers A.A., 1996. A *Bacteroides thetaiotamicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch. Journal of Bacteriology, 178:823–830

[14] Sato K., Naito M., Yukitake H., Hirakawa H., Shoji M., McBride M. J., Rhodes R.G., Nakayama K., 2010. A protein secretion system linked to *Bacteroidetes* gliding motility and pathogenesis. Proceedings of the National Academy of Sciences USA, 107:276–281

[15] Barbier P., Houel A., Loux V., Poulain J., Bernardet J.-F., Touchon M., Duchaud E., 2012. Complete genome sequence of *Flavobacterium indicum* GPSTA100-9^{T}, isolated from warm spring water. Journal of Bacteriology, 194:3024–3025