Genomic diversity of *Flavobacterium* species

Eric Duchaud[1], Paul Barbier[1], Jean-François Bernardet[1], Christian Michel[1], Céline Chantry[1], Christophe Habib[1&2], Pierre Nicolas[2], Valentin Loux[2], Eduardo Rocha[3] & Marie Touchon[3]

[1] *Unité de Virologie et Immunologie Moléculaires, INRA, Domaine de Vilvert, F-78350 Jouy-en-Josas, France*

[2] *Unité Mathématique, Informatique et Génomes, INRA, Domaine de Vilvert, F-78350 Jouy-en-Josas, France*

[3] *Groupe de Génomique Evolutive des Microbes, Institut Pasteur, F-75015 Paris, France*

Correspondence: educhaud@jouy.inra.fr

Abstract

The advances in sequencing and bioinformatics have revolutionized the study of bacterial species. Nowadays, high throughput sequencing enables large scale identification of the genetic differences across several genomes from related strains and species.

Flavobacterium species are members of the *Flavobacteriaceae*, a bacterial family of high importance for the degradation/turnover of organic matter in a wide range of ecological niches. Three species of the genus *Flavobacterium* are recognized fish pathogens, a small fraction of the 85 species currently described. This number keeps increasing rapidly owing to the abundance of *Flavobacterium* strains in diverse ecological niches including extreme habitats (e.g. glacier and polar ice, freshwater and marine sediment, warm spring, freshwater and hard-water, activated sludgs, earthworm gut and greenhouse soil).

In order to identify relevant features and genes in relation to their life styles, including virulence for the pathogenic species, we are carrying out a genomic program that rely on a collection of bacterial isolates that are characterized by two different sequencing approaches: systematic multi-locus sequence typing (MLST) and whole-genome sequencing of a subset of isolates.

An MLST scheme was developped to get information about the population structure within the species *F. psychrophilum*, an important pathogenic bacterium of salmonids. The published MLST scheme [1] was optimized by (i) selecting seven loci (trpB, gyrB, dnaK, fumC, murG, tuf, and atpA) instead of eleven; (ii) adding M13 forward and reverse extension to the primers for more convenient sequencing; and (iii) devising a touch down PCR. An international effort (encompassing laboratories in the USA, Japan, Chile and Europe) was done to analyze the diversity of *F. psychrophilum* at a worldwide scale. So far, genotyping of more than 1000 isolates revealed: (i) an overall low genetic diversity; (ii) a diversity mainly driven by homologous recombination; (iii) an epidemic population structure (including the prevalence of specific clones); (iv) an association between the genotype of the strain and the host fish species; and (v) some geographically contrasted bacterial populations.

Whole-genome sequences of *F. psychophilum* JIP02/86 [2], isolated from rainbow trout in France and THC02/90 [unpublished], isolated from a coho salmon in the USA were compared allowing the identification of strain-specific genes, most of them located in genomic islands.

Ten *F. psychophilum* draft genomes, chosen as diverse as possible using MLST data, were obtained by Illumina sequencing. Their analysis confirmed the presence of genomic islands and revealed specific traits (e.g. the CRISPR locus is highly diverse). These genomes were useful to identify core, accessory and pan-genome genes (Fig. 1).

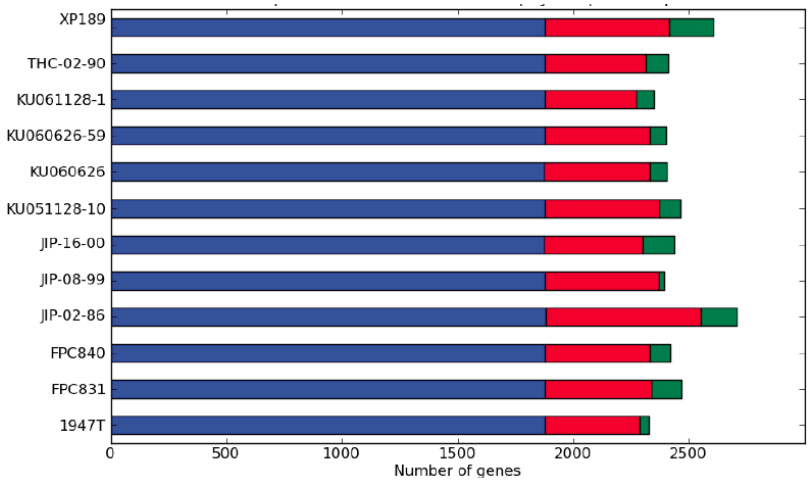

Figure 1. Distribution of *F. psychophilum* genes. In blue: core genome genes; in red: accessory genes identified in at least 2 different strains; in green accessory genes identified in only a single strain.

Complete genome sequences of *F. branchiophilum* FL15 [3], *F. indicum* GPSTA100-9^{T} [4], *F. frigidimaris* KUC1^{T}, *F. glaciei* 0499^{T} and *F. columnare* JIP14/00 were recently determined in our laboratory. Together with the complete genome of *F. johnsoniae* UW101^{T} [5] and *F. columnare* ATCC 49512 [6] they allow comparisons of their genomic content, including the diversity of toxin-encoding genes for the pathogenic species. Concateneted core genome proteins were used for phylogenetic tree reconstruction (Fig. 2). It reveals a very homogenous grouping of *F. psychophilum* strains and a split of the two *F. columnare* strains (that belong to two different genomovars).

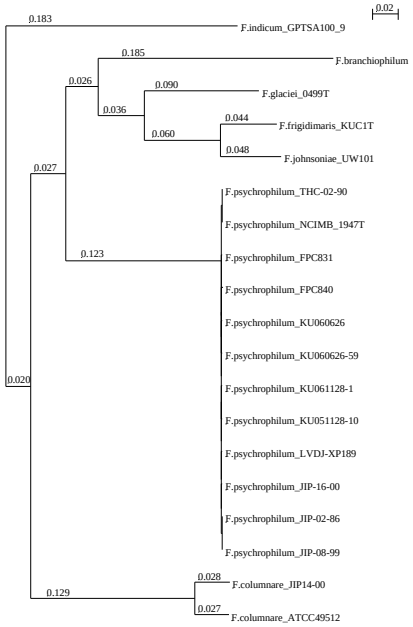

Figure 2. Phylogenetic tree reconstruction of *Flavobacterium* species using concateneted core genome proteins.

In conclusion, genomics is a valuable tool to analyse the diversity of *Flavobacterium* species.

References

[1] Nicolas P., Mondot S., Achaz G., Bouchenot C., Berdardet J.F., Duchaud E., 2008. Population structure of the fish-pathogenic bacterium *Flavobacterium psychrophilum*. Applied and Environmental Microbiology, 74:3702–3709

[2] Duchaud E., Boussaha M., Loux V., Berdardet 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

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

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

[5] 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

[6] 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