A computational genomics pipeline for prokaryotic sequencing projects: application on marine bacteria producing sulfated exopolysaccharides

Soumaya Messaoudi [1], Tristan Barbeyron [2], Laure Quintric [3], Sylvia Colliec-Jouault[1], Mirjiam Czjzek [2], Gurvan Michel [2], Christine Delbarre-Ladrat [1]

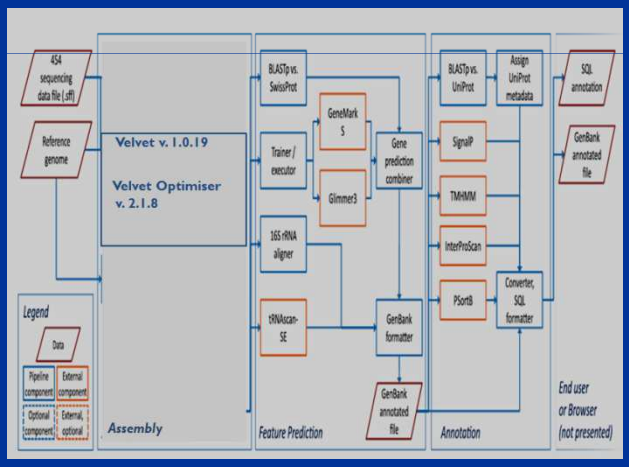
1) Laboratoire de Biotechnologie et Molécules Marines-BMM, IFREMER Nantes, France
2) UMR 7139 CNRS Paris VI, Station Biologique de Roscoff.
3) Service Ressources Informatiques et Communications de l'Unité Informatique et Données Marines, IFREMER Brest

Abstract

Although sulfated carbohydrates are biological components that exhibit many interesting properties with medical and cosmetical applications, the discovery of innovative molecules is hindered by the difficulty to manage the sulfation reactions at high scale and in a specific manner. Sulfotransferases are the bio-catalysts that realize the sulfation reactions but, a very few of them have been described as bioconversion tools. To find molecules naturally bearing sulfate at a high rate but also enzymes able to (over-)sulfate molecules in a specific and controlled reaction compatible with the environment remains a challenge. On the other hand, marine microbial biodiversity has proved to be an interesting source of new compounds, in particular sulfated glycopolymers. Its investigation for the discovery of carbohydrate active enzymes especially carbohydrate sulfotransferases usable in industrial processes must be developed.
For this purpose, the genomes of two marine bacteria able to produce sulfated polysaccharides, *Alteromonas infernus* and *Pseudoalteromonas sp. HYD721* were sequenced and analyzed by GC-pipeline.
Concerning *Alteromonas infernus* GY785, its genome (4.3Mb) was assembled by GATC. It consists of 86 contigs (3 scaffolds and 6 contigs unlocated). About *Pseudoalteromonas* Hyd 721, its genome (4.7Mb), was assembled with Velvet to 365 contigs.

Bioinformatics tools set up

Installation of a computational genomics pipeline for prokaryotic sequencing projects (Kislyuk et al 2010)



- A pipeline for prokaryotic genome analysis has been set up on Ifremer server (CAPARMOR: CAlcul PARallèle Mutualisé pour l'Océanographie et la recherche). This analytical pipeline consists of three integrated subsystems: genome assembly, feature prediction and functional annotation. Each subsystem consists of a top-level execution script managing the input, output, format conversion and combination of results for a number of distinct software components.

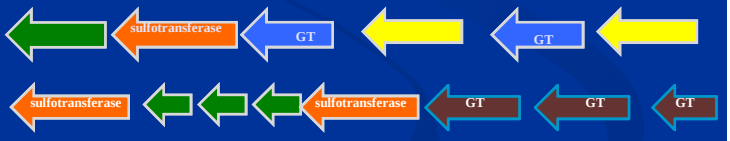
Results

Sulfotransferases and CAZymes identified

Bacteria / enzymes	*A.infernus* GY785	*Pseudoalteromonas* Hyd721
Sulfotransferases	4	5
Glycosyl Transferases (GT)	54	36
Glycoside Hydrolases (GH)	43	42
Carbohydrate Esterases (CE)	12	12
Polysaccharide Lyases	1	2

Clusters potentially involved in the biosynthesis of polysaccharides

Clusters identified on the genome of *A.infernus* GY785



Clusters identified on genome of *Pseudoalteromonas* Hyd721

Conclusion

-This pipeline has been used for assembly, gene prediction and genome annotation of the genomes of two marine bacterial strains.
- Expert annotation to identify sulfotransferase and carbohydrate active enzyme genes of both genomes shows the presence of several genes encoding sulfotransferases from different families. We were also able to identify clusters potentially involved in the biosynthesis of polysaccharides in both bacteria

Perspectives

- Cloning and Characterization of sulfotransferases
- Cloning the clusters of exopolysaccharide biosynthesis and determining the role of each glycosyl transferase in this biosynthesis