Identification of genomic regions for high-resolution taxonomic profiling using long-read sequencing technology

Jean Mainguy, Olivier Bouchez, Adrien Castinel, Sylvie Combes, Christine Gaspin, Denis Milan, Cecile Donnadieu, Carole Iampietro, Claire Hoede, Géraldine Pascal

Identification of genomic regions for high-resolution taxonomic profiling using long-read sequencing technology

Jean MAINGUY[1], Olivier BOUCHEZ[2], Adrien CASTINEL[2], Sylvie COMBES[3], Christine GASPIN[1], Denis MILAN[2], Cécile DONNADIEU[2], Carole IAMPIETRO[2], Claire HOEDE[1] and Géraldine PASCAL[3]

1 MIAT, PF Bioinfo GenoToul, Université de Toulouse, INRA, Chemin de Borde Rouge, 31320 Castanet-Tolosan, France

2 GeT-PlaGe, US 1426, Genotoul, INRA, 31320 Castanet-Tolosan, France

3 GenPhySE, Université de Toulouse, INRA, INPT, ENVT, Chemin de Borde Rouge, 31320 Castanet Tolosan, France

Contact :
geraldine.pascal@inra.fr
jean.mainguy@inra.fr

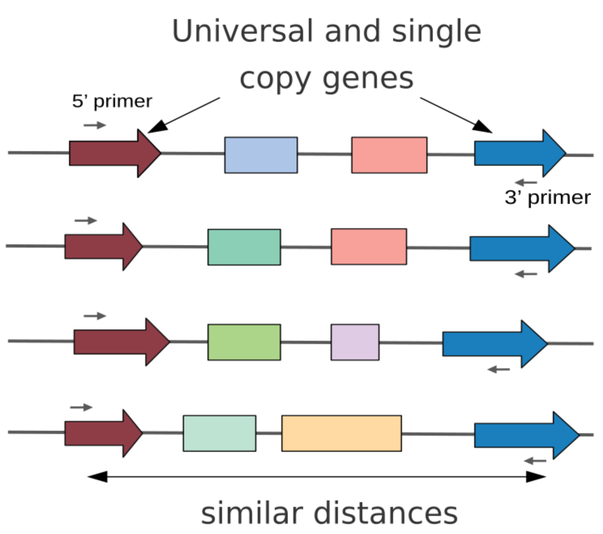
Taxonomic profiling: who is in the community?

The widespread MiSeq illumina amplicon sequencing approach:
Target a small part of the 16S RNA gene or alternative markers (rpoB, recA and gyrB) [1].

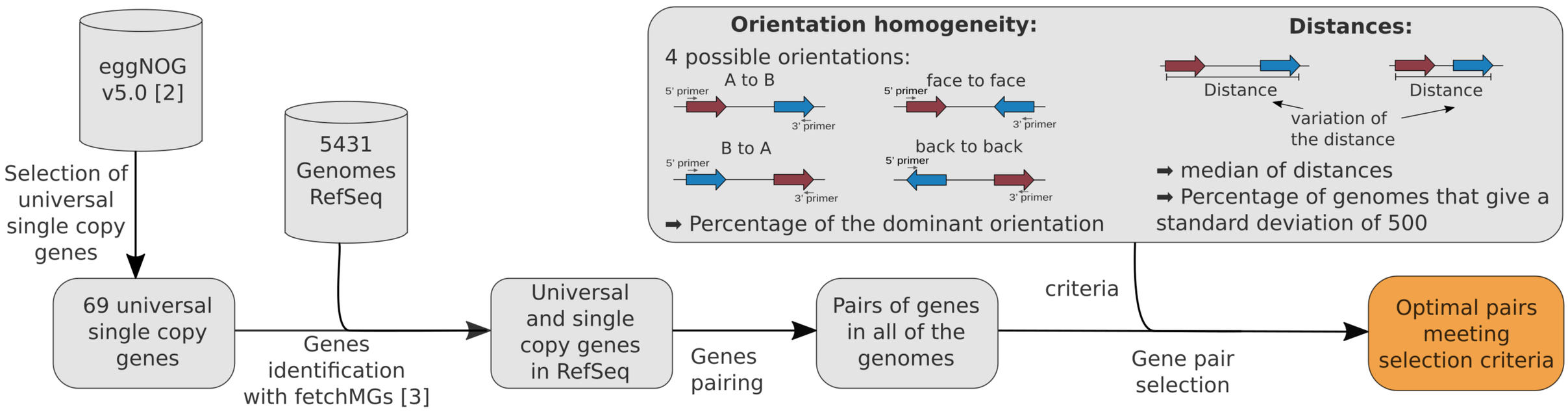
Limits of the approach:

- Amplicon size of around 450pb
- Poor resolution for closely related organisms
- Multicopy affects the abundance estimation

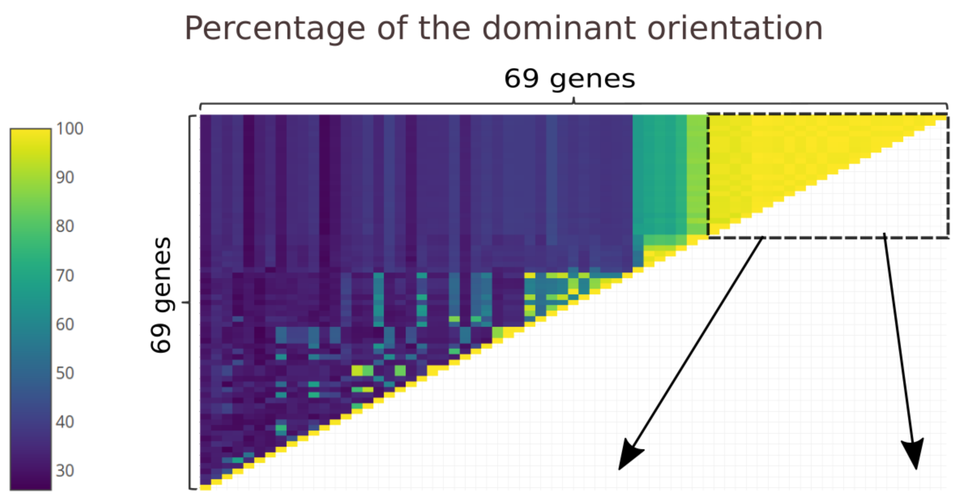
Our strategy using long read approach with Oxford Nanopore Technologies:
Identify genomic regions bounded by two universal and single copy genes.

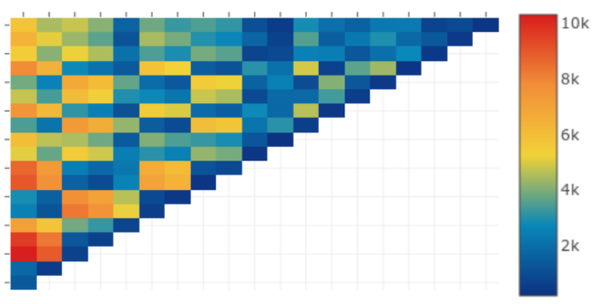


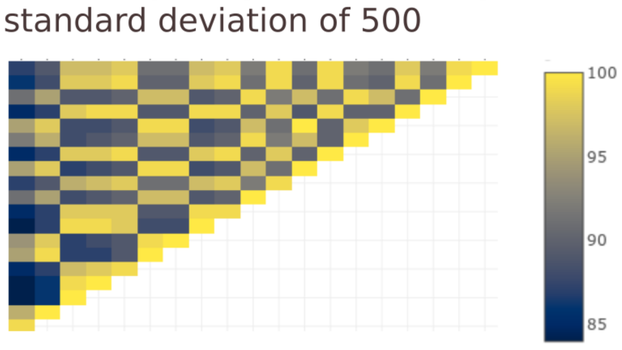
Genomic regions identification workflow

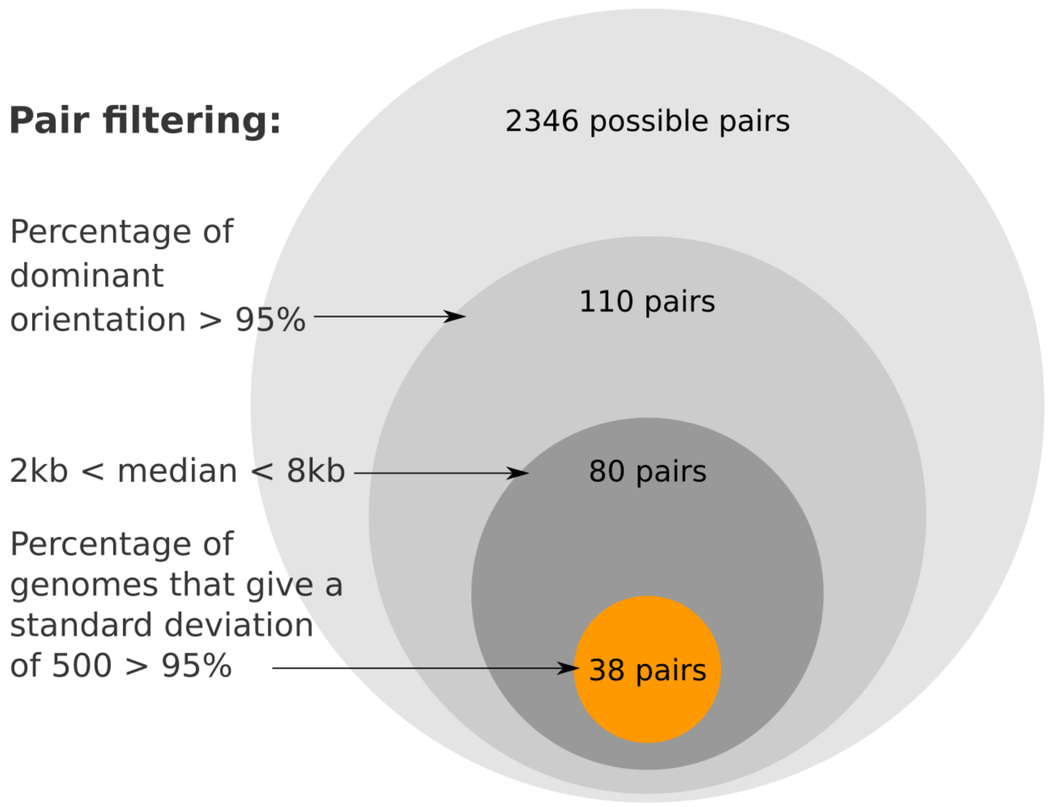


Optimal pair selection









Conclusion and perspectives

Our approach enables the identification of several pairs of genes found in **single copy** and displaying **consistent distances and orientations** across bacterial genomes. These pairs are very promising targets to enhance taxonomic profiling with long read sequencing technologies.

The next steps of the project are:

- ➡ Investigate how well these regions can discriminate closely related organisms.
- ➡ Design primers in order to amplify the chosen regions.
- ➡ Test out the chosen regions in a wet lab experiment on a real microbial community.

[1] Poirier, Simon, et al. "Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using gyrB amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing." PloS one 13.9 (2018): e0204629.
[2] Huerta-Cepas, Jaime, et al. "eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses." Nucleic acids research 47.D1 (2018): D309-D314.
[3] Milanese, Alessio, et al. "Microbial abundance, activity and population genomic profiling with mOTUs2." Nature communications 10.1 (2019): 1014.