

Bacterial strains identification using Oxford Nanopore sequencing

Grégoire SIEKANIEC^{1,2}, Emeline ROUX^{1,2,3}, Eric GUEDON² and Jacques NICOLAS¹

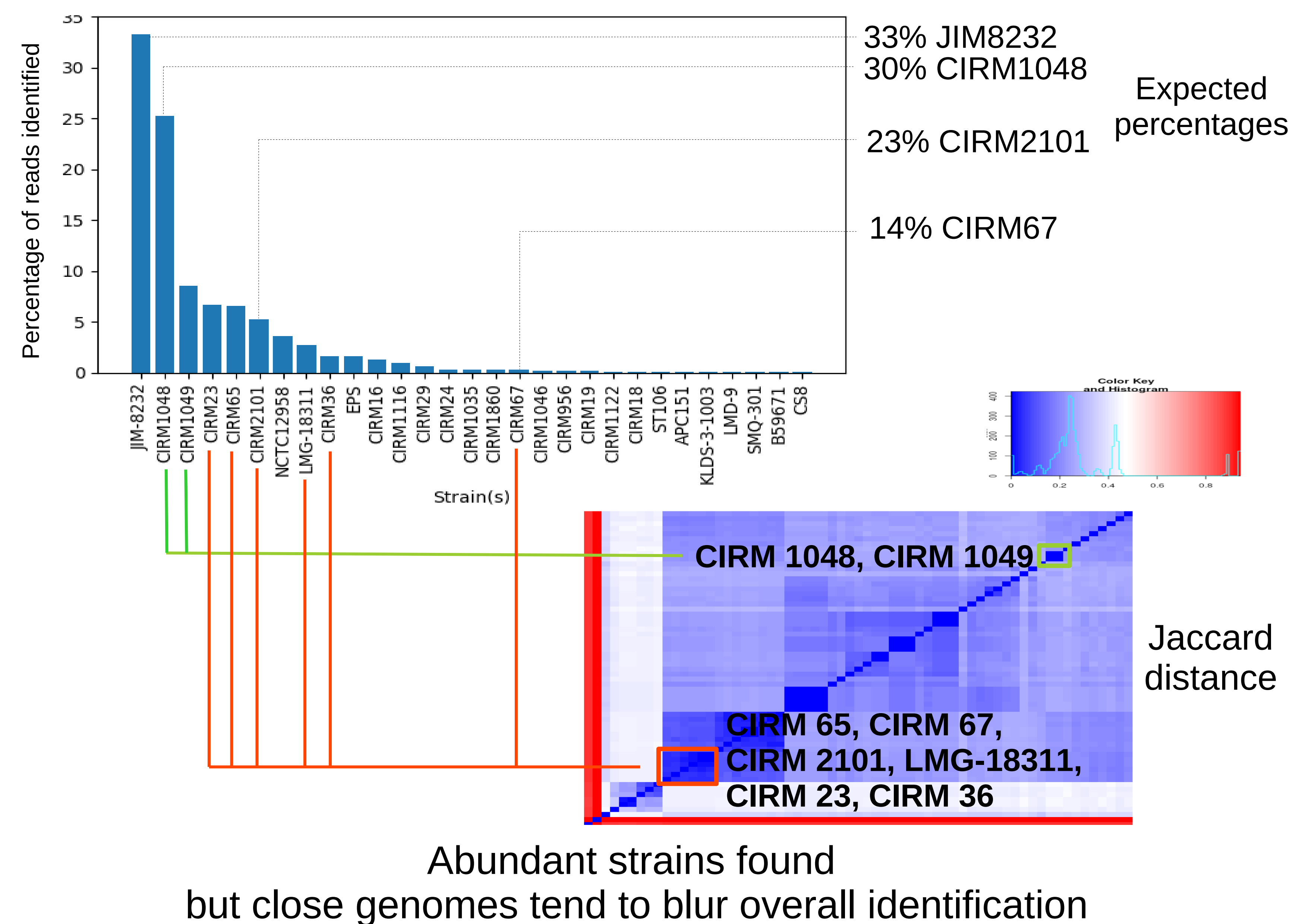
Context

- Strain level identification of bacteria is a complex problem.
- Design a new taxonomic assignment method using Nanopore long reads (MinION device).
- The challenge is to tackle with a relatively high error level (about 7 % on raw uncorrected reads) and show that, together with reference genome data, they allow to precisely distinguish one species from another and even to go down to the strain level.

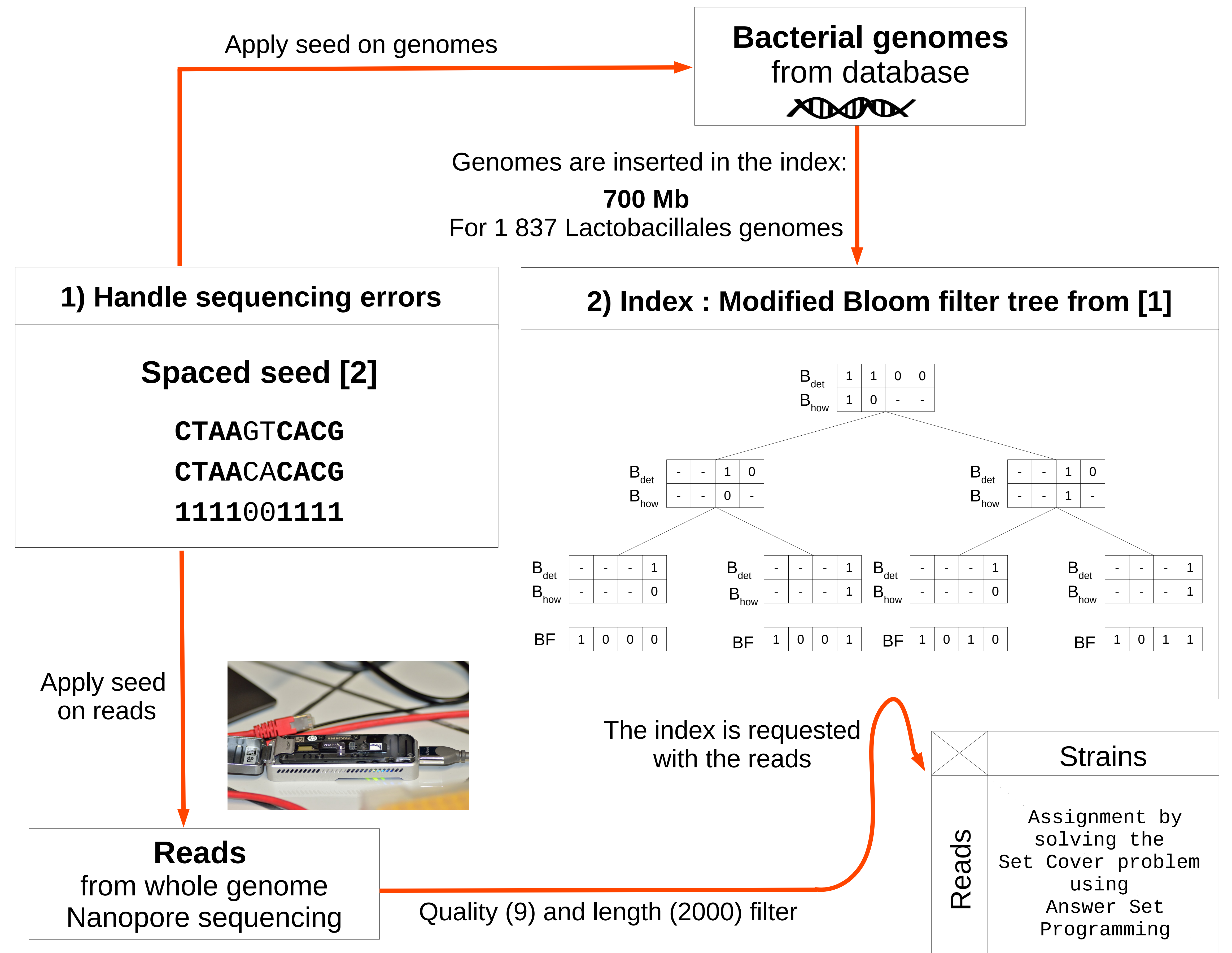
First results

- Single strain identification :**
4000 reads from *Streptococcus thermophilus* JIM 8232,
Index: 31 *S. thermophilus*, 1 *S. macedonicus*,
1 *L. delbrueckii subsp. bulgaricus* strains genomes
Almost a perfect identification (99% reads assigned to JIM8232)

Mixture of 4 *Streptococcus thermophilus* strains: 1226 reads



Methods



[1] Robert S Harris and Paul Medvedev, Improved representation of sequence bloom trees, Bioinformatics, btz662

[2] Laurent Noé, Best hits of 11110110111: model-free selection and parameter-free sensitivity calculation of spaced seeds, Algorithms for Molecular Biology, volume 12, issue 1, 2017.

Conclusion

We succeed in achieving identification of species and of distant strains within a same species. We are working on a clustering preprocessing step and a two precision levels method to tackle with the existence of many very close strains.

¹ UNIV RENNES, INRIA, CNRS, IRISA, RENNES F-35000, FRANCE

² STLO, INRAE, AGROCAMPUS OUEST, RENNES, FRANCE

³ CALBINOTOX, UNIVERSITÉ DE LORRAINE, F-54000, NANCY, FRANCE