Functional characterization of fermented milk products using quantitative metagenomic approach

Mathieu Almeida[1,2,3], Agnès Hébert[4], Anne-Laure Abraham[1], Simon Rasmussen[5], Christophe Monnet[4], Nicolas Pons[2], Céline Delbès[6], Valentin Loux[7], Jean-Michel Batto[2], Pierre Léonard[2], Sean Kennedy[2], S. Dusko Ehrlich[2], Mihai Pop[3], Marie-Christine Montel[6], Françoise Irlinger[4], Pierre Renault[1].

[1]INRA, MICALIS, Jouy-en-Josas, France ; [2]INRA, MGP, France ; [3]CBCB, University of Maryland, College Park, USA ; [4]AgroParisTech, Thiverval-Grignon, France.

[5]DTU, CBS, Denmark ; [6]INRA, Aurillac, France; [7]INRA, MIG, Jouy-en-Josas, France.

E-mail: mathieu.almeida@jouy.inra.fr ; pierre.renault@jouy.inra.fr
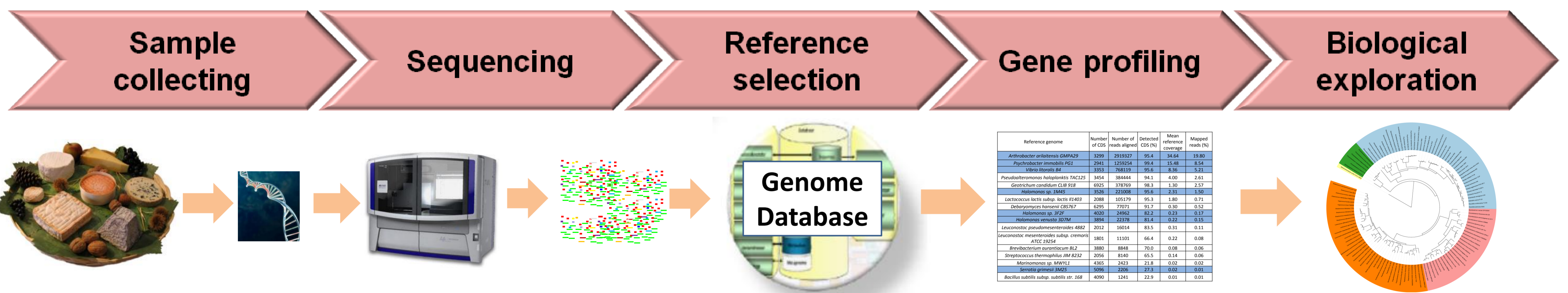
Introduction

Present dairy sample microbial analysis principally use marker genes to characterize microbial composition or presence of particular functions.

New NGS technologies may allow global exploration of food samples, including genes involved in microbial metabolism.
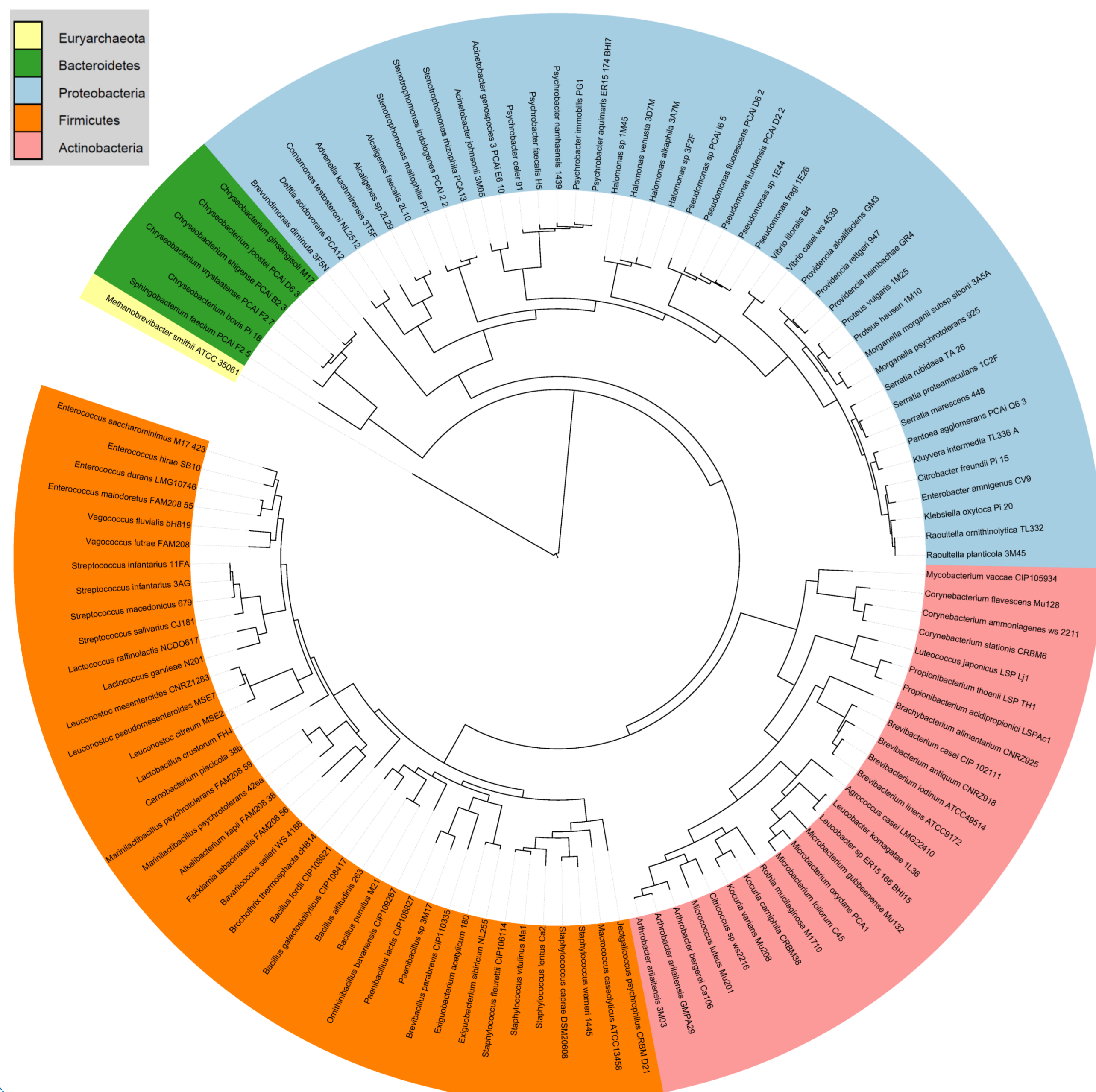
However, NGS analysis requires reference genome database adapted to the studied ecosystem to assign accurately NGS reads.
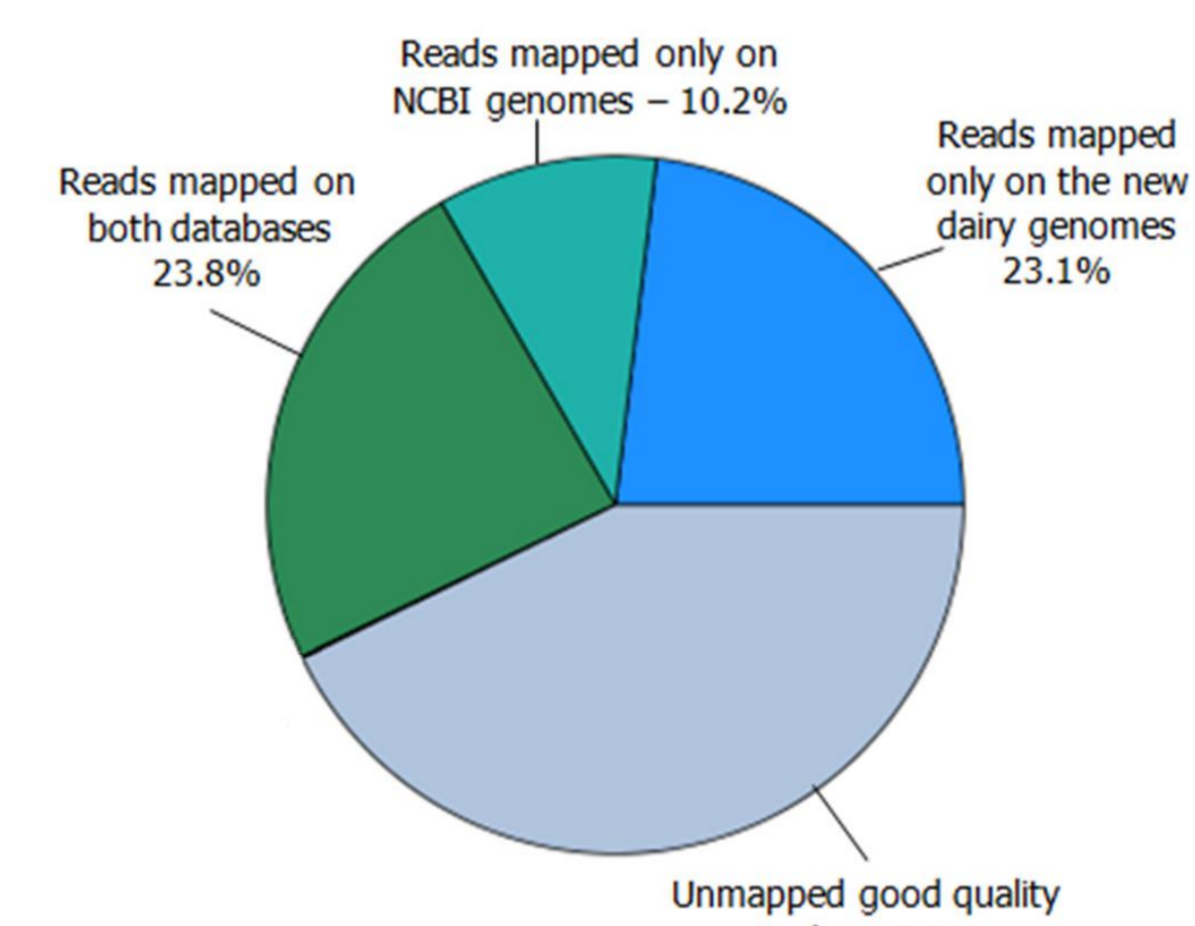
Method

Sample collecting → Sequencing → Reference selection → Gene profiling → Biological exploration

Genome Database

Result

Selection and sequencing of 117 dairy bacterial genomes to improve present genome databases

Legend:
- Euryarchaeota
- Bacteroidetes
- Proteobacteria
- Firmicutes
- Actinobacteria

Characterization of a cheese surface using 6000 NCBI genome database and the 117 dairy genomes

Reads mapped only on NCBI genomes – 10.2%
Reads mapped only on the new dairy genomes 23.1%
Reads mapped on both databases 23.8%
Unmapped good quality reads - 42.8%

Reference genome	Number of CDS	Number of reads aligned	Detected CDS (%)	Mean reference coverage	Mapped reads (%)
Arthrobacter arilaitensis GMPA29	3299	2919327	95.4	34.64	19.80
Psychrobacter immobilis PG1	2941	1259254	99.4	15.48	8.54
Vibrio litoralis B4	3353	768119	95.6	8.36	5.21
Pseudoalteromonas haloplanktis TAC125	3454	384444	94.1	4.00	2.61
Geotrichum candidum CLIB 918	6925	378769	98.3	1.30	2.57
Halomonas sp. 1M45	3526	221008	95.6	2.31	1.50
Lactococcus lactis subsp. lactis Il1403	2088	105179	95.3	1.80	0.71
Debaryomyces hansenii CBS767	6295	77071	91.7	0.30	0.52
Halomonas sp. 3F2F	4020	24962	82.2	0.23	0.17
Halomonas venusta 3D7M	3894	22378	81.4	0.22	0.15
Leuconostoc pseudomesenteroides 4882	2012	16014	83.5	0.31	0.11
Leuconostoc mesenteroides subsp. cremoris ATCC 19254	1801	11101	66.4	0.22	0.08
Brevibacterium aurantiacum BL2	3880	8848	70.0	0.08	0.06
Streptococcus thermophilus JIM 8232	2056	8140	65.5	0.14	0.06
Marinomonas sp. MWYL1	4365	2423	21.8	0.02	0.02
Serratia grimesii 3M25	5096	2206	27.3	0.02	0.01
Bacillus subtilis subsp. subtilis str. 168	4090	1241	22.9	0.01	0.01

Conclusion

The 117 dairy bacterial genomes improve significantly quantitative metagenomic analysis.

The quantitative metagenomic analysis of the cheese surface reveal occurrence of poorly characterized dominant or sub-dominant bacteria.

The presence of still significant percentage of unassigned reads suggests that additional genomes should be determined.

Acknowledgment