Streptococci in the human digestive tract
A metagenomic study

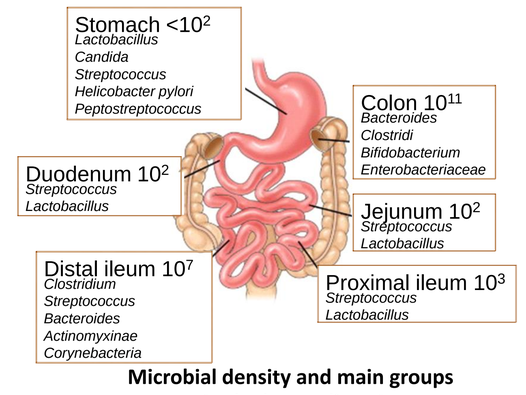
Mathieu Almeida, E. Achouri, B. Moumen, E. Le Chatelier, N. Pons, J.M. Batto, P. Léonard, S. Layec, F. Boumezbeur, S. Kennedy, C. Delorme, E. Guédon, S.D. Ehrlich, P. Renault.

UMR MICALIS 1319: Microbiologiste de l'Alimentation au service de la Santé. « Pôle Ecosystème », INRA, 78352 Jouy-en-Josas cedex, France.
E-mail: mathieu.almeida@jouy.inra.fr ; pierre.renault@jouy.inra.fr

Several species of Streptococci are responsible for numerous pathologies in humans and animals, including throat, pneumonia, meningitis, endocarditis… However, many of them occur as natural commensals in many flora such as the skin, the oral cavity, the gastrointestinal or urogenital tract.

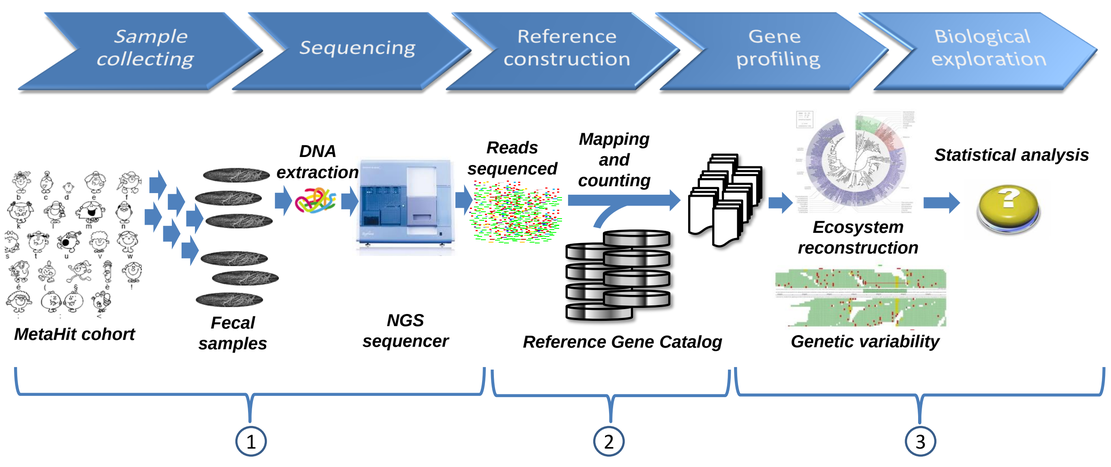
Occurrence of Streptococci in the lower part of the digestive tract is occasionally reported. Therefore we lack a global view on occurrence and level of streptococcal populations in this complex and extremely rich microbial niche.

We have developed a method based on NGS (Next Generation Sequencing) data analysis that enable to characterize and quantify accurately dominant and sub-dominant species in metagenomic samples. We applied this method to study 149 samples of human feces to provide an insight of major commensal Streptococci population of the human intestinal flora.



Microbial density and main groups present in the lower digestive tract (adapted from Startor, 2008).

Method for Gene Detection in Metagenomic Samples

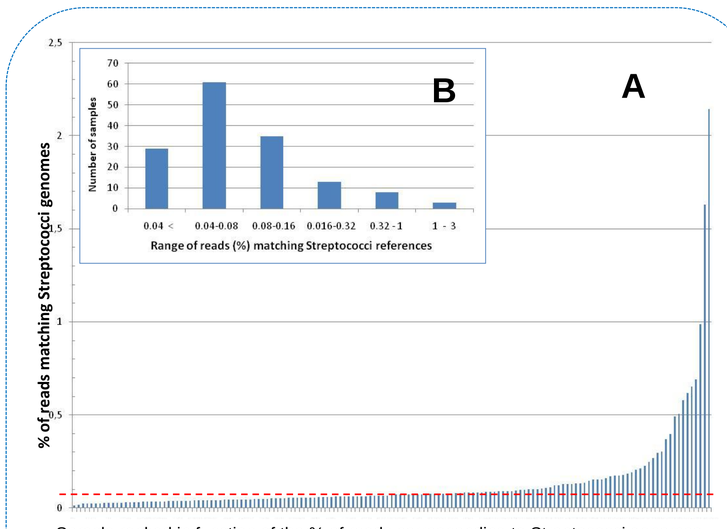


① **DNA Extraction and Illumina sequencing**. 149 human fecal samples from MetaHIT Project were isolated and sequenced. ~50 million of short sequences (<90 b) were generated for each metagenomic samples.

② **Reference Gene Catalog Establishment and Mapping Procedure**. 31 streptococcals species and 5 related outgroup species were used to create a reference catalog gene. Genes corresponding to phage, mobile genetic elements and highly conserved core genes were discarded, yielding an average of 450 markers for each species. All reads were aligned on the reference catalog gene with **BOWTIE** mapper (Langmead et al., 2009).

③ A metagenomic pipeline called **METEOR (METagenome ExploratOR)** and a specific module, **GeDI (Genome Detector from Illumina reads)** were developed to analysis the result of mapping. The pipeline allows to (i) quantify species in each samples, (ii) detect specific genes marker for species or strains (iii) characterize SNP variants in each marker genes.

Streptococci in the intestinal microbiota of 149 individuals

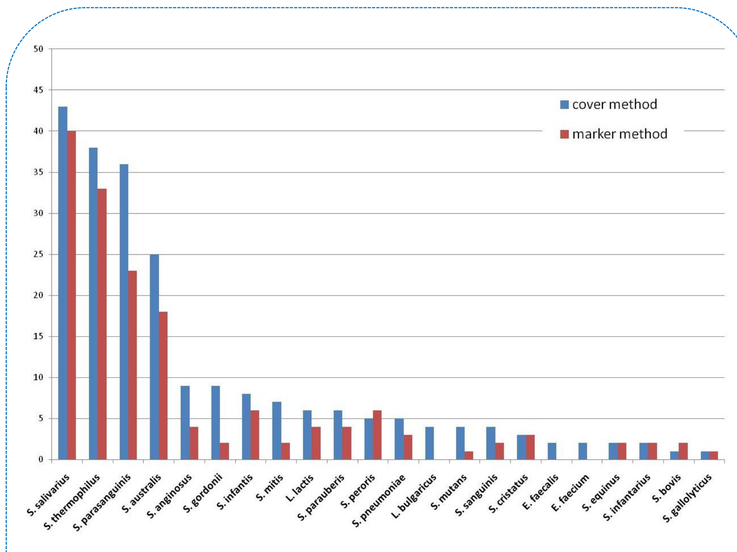


Sample ranked in function of the % of reads corresponding to Streptococci genomes.

Level of Streptococci reads in human feces samples.
a) Number of reads matching streptococcal genome in samples. Mapping parameters : 3 mismatches, 35 b paired-end reads. A level of 0.04%(>20,000 reads) was considered as positive signal (red doted line).
b) Number of samples classified in function of their level of read.

Streptococci are likely present in 110 samples out of 149.

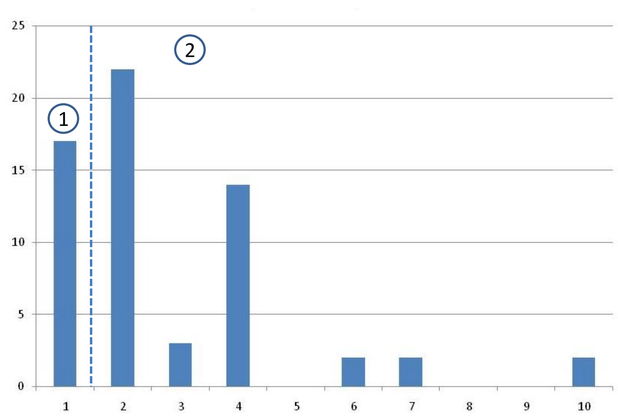
- For 3 individuals, over 1% of the read samples matched streptococcal genomes. This level would classify Streptococci as **part of the dominant flora** in their microbiota.
- The majority of individuals contains an average of 0.07±0.02% of the total reads. This level corresponds to that of the subdominant species.
- More sensitive methods suggest that species such as *S. salivarius* are almost ubiquitous, although present at low frequency in the last 1/3 individuals.



Streptococci species and their prevalence in human feces samples.
The histogram presents the number of individuals containing the streptococcal species. Two methods were used, one (in blue) based on the % of coverage of marker genes, or the % of marker in which more than 2 reads are matched (red).

The most frequently found streptococci species are the oral commensal *S.salivarius, S. parasanguinis,* and *S. australis* and the food species *S. thermophilus.*

- They are present in more than 15% of the human feces samples, but other species, often from the oral flora, are found in fewer samples.
- None of the two major pathogens *S. agalactiae, S. pyogenes*, as well as animal carried species such as *S. criceti, S. suis, S. downei…* have been detected.
- Interestingly, *Streptococci* are more frequently found than *Enterococci* and their population level is higher.



Occurence number of streptococcal species in each individuals.

Streptococci species are often associated together.

① Species found alone are the two commensals, *S. salivarius* (n=7) and *S. infantarius* (n=1), and the food strain, *S. thermophilus* (n=10).

② The other species, such as *S. parasanguinins*, were always found associated with *S. salivarius* or *S. infantarius* as dominant strains.

Conclusion

17 non food streptococcal species were characterized, the majority come from oral flora. Interestingly, many individuals carry more than one commensal species, indicating that certain individual backgrounds are permissive for efficient colonization by Streptococci.

Streptococci are part of the subdominant intestinal flora. They are generally present at subdominant level ($>10^7$ cell/g of feces in about 2/3 of the individuals) and occasionally at dominant level ($>10^9$ cell/g of feces in 3 individuals).

The high abundance of oral streptococci in intestinal flora may indicates their ability to broadly colonize the gastrointestinal tract, from ileum to colon. In contrast, the clear presence of the food species *S. thermophilus* and *L. lactis* should come from dairy products of the diet (yogurt or fresh cheese which can bring 10^8-10^9 bacteria per gram).

A better insight in species detection. In a first Metahit analysis, by using the metagenomic gene catalog (Qin et al. 2010), only 3 species, *S. salivarius*, *S. parasanguinis* and *S. thermophilus* were detected (Almeida et al, unpublished results). This fact was due to the limitation in the assembly method for the low frequency species. This new method uses genome reference and specific markers to overpass this limitation.

This technique has been successfully used to characterize streptococci community in human microbiomes. Further analysis can now be conducted to explore the ecology of human intestinal flora.