Microbial communities show the greatest organisms diversity on earth. Culture-independent molecular techniques have revealed this extraordinary diversity which remains difficulty accessible with current metagenomic approaches. It appears therefore necessary to reduce the complexity of metagenomics samples to explore microbial diversity. Barcoding efficiently reduces this biological complexity through the PCR amplification of phylogenetic biomarkers (e.g. 16S rRNA genes). It enables an easy and rapid access to taxonomic microbial diversity, but only targets a reduced portion of biomarkers. Variability observed on such small genomic regions is therefore not sufficient to have a good resolution in sequence affiliation and a reliable assignation in lower taxonomic ranks (family, genus, species). Other approaches providing information on complete phylogenetic biomarkers and spanning regions are consequently crucial to overcome those limitations and enhance taxonomic affiliation.

Based on a solution hybrid selection method combined with next-generation sequencing, we developed an innovative gene capture approach associated with a specific bioinformatic treatment to enable the access to the phylogenetic information of the complete targeted biomarker and its flanking regions.

To benefit from capture advantages, a suitable bioinformatics treatment of generated data is necessary. Capture sequencing data analysis may be achieved according to two complementary workflows, sharing key sequence treatment steps.

Analyses realized on samples from different environments show that gene capture reveals a broader taxonomic diversity than observed with barcoding approach and that this diversity leads to a more reliable representation of the studied ecosystems. Moreover, gene capture enables to access the phylogenetic information of the complete targeted biomarker and its informative flanking regions, such as 23S rRNA and ITS when 16S rRNA is targeted, providing a reliable and resolvent affiliation until lowest taxonomic ranks like species and even strain. This innovative complexity reduction method combined with an original bioinformatics treatment thus enables a better exploration of microbial ecosystems.