



HAL
open science

A shotgun metagenomic method to characterise low abundant species and assign precisely taxonomy in complex microbial ecosystems

Anne-Laure Abraham, Mathieu Almeida, Nicolas Pons, Charlie Pauvert,
Sophie Schbath, Pierre Renault

► **To cite this version:**

Anne-Laure Abraham, Mathieu Almeida, Nicolas Pons, Charlie Pauvert, Sophie Schbath, et al.. A shotgun metagenomic method to characterise low abundant species and assign precisely taxonomy in complex microbial ecosystems. ECCB'08: 7. European Conference on Computational Biology, Sep 2014, Strasbourg, France. , 2014. hal-01204320

HAL Id: hal-01204320

<https://hal.science/hal-01204320>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A shotgun metagenomic method to characterise low abundant species and assign precisely taxonomy in complex microbial ecosystems

Anne-Laure ABRAHAM, Mathieu ALMEIDA, Nicolas PONS, Charlie PAUVERT, Sophie SCHBATH and Pierre RENAULT

A first step for a better understanding of complex microbial ecosystems, such as cheese or human gut microbiota, is the characterisation and quantification of their species composition. Once DNA is extracted from samples, two main techniques are usually used: the sequencing of evolutionary conserved genes, such as those coding for the 16S or 18S RNA or ITS, or whole genome shotgun sequencing. The first approach is widely used and provides a rapid view of the ecosystem, but often fails to provide taxonomy information more precise than the genus level. Shotgun metagenomic sequencing approaches may circumvent such issues. They are often based on the use of gene marker sets to discriminate species rapidly and without ambiguity, although the use of a small part of the genomes decreases sensitivity of this approach.

We present a new tool that gives encouraging results in the characterisation of low abundance species and may allow distinguishing strains under the subspecies taxonomic assignments. This approach is based on a mapping of metagenomic reads on the whole genomes of a dataset of reference genomes. We then identify species present in the sample based on the number and distribution of reads along each reference genome. Preliminary analyses indicate that we can determine the taxonomy up to the strain level, and identify low abundant species.

Our method is complementary to methods based on a set of marker genes. It is computationally more expensive, but can provide a more precise view of the ecosystem. Indeed, methods based on a set of marker genes usually fail to go up to the strain level in the taxonomy identification, and can miss low abundant species if the marker genes are not sequenced at a sufficient depth. We tested our tool on several datasets, including simulated reads and artificial metagenomic dataset in order to test the power and limits of this method. We will present examples on cheese ecosystems.

Keywords Metagenomics, next generation sequencing, microbial genomes.