Striking the current metaproteomics dogma for deeper characterization of microbiota

Olivier Pible, Karen Culotta, Duarte Gouveia, Karim Hayoun, Virginie Jouffret, Guylaine Miotello, Gérard Steinmetz, Béatrice Alpha-Bazin, Lucia Grenga, Jean Armengaud

SMAP 2019

16-19 Septembre
Strasbourg

Striking the current metaproteomics dogma for deeper characterization of microbiota

Authors: Pible Olivier, Culotta Karen, Gouveia Duarte, Hayoun Karim, Jouffret Virginie, Miotello Guylaine, Steinmetz Gérard, Alpha-Bazin Béatrice, Grenga Lucia, **Armengaud Jean**

Introduction: Metaproteomics is the analysis of complex samples to describe how they function. The data may give novel taxonomical information based on peptide information. The current dogma in metaproteomics is i) to rely only on taxon-specific peptides for extracting the taxonomical information, and ii) to interpret MS/MS data with metagenomics data acquired on the same sample. We propose a new metaproteomics pipeline that allows a quick identification of any microorganism present in the sample based on the whole dataset, without need of additional costly metagenomics information.

Methods: We developed in python an in-house pipeline to assign taxonomical information to each detected peptide against a generalist database and for the deconvolution of this complex signal. In parallel, we developed a procedure to regularly update and curate the database to avoid taxonomic misassignment. We also optimized the sample preparation for extracting proteins of any organisms and the tandem mass spectrometry acquisition to maximize the results.

Results: We discovered the principle of a mathematical signature describing the number of peptide sequences shared with all other organisms calculated by modeling and phylogenetic relationships. This principle allows deciphering the precise content of a sample by the linear combination of such signatures applied on any experimental metaproteomic dataset. A sample can thus be described by its peptide-specified taxa and their respective relative ratios defined from the global peptide information. Its efficiency is exemplified with artificial mixtures. We developed the informatic pipeline to interpret quickly MS/MS datasets in terms of taxonomy, obtain response linearity regarding the label-free quantitation of biomass contributions, and establish the most appropriate protein sequence database for each sample. Several examples will be commented such as a deep characterization of human feces and a comparative analysis of sentinel animal intestines.

Conclusions: This methodology paves the way to accurate label-free quantitative metaproteomics without the need of metagenomic information. It has been proved robust with artificial mixtures of microorganisms and has been applied to a large panel of samples of interest in terms of clinics, biotechnology, and environment.

Novel Aspect: A new procedure for interpreting MS/MS metaproteomic dataset allows characterizing any sample in terms of taxonomy and improve the functional characterization.

References:

- Pible & Armengaud (2015) Improving the quality of genome, protein sequence, and taxonomy databases: a prerequisite for microbiome meta-omics 2.0. Proteomics 15(20):3418-23.

- Grenga et al. (2019). Pathogen proteotyping: a rapidly developing application of mass spectrometry to address clinical concerns. Clinical Mass Spectrometry, *In press*.
- Pible et al (2019) Improving relative quantitation of the biomass contributions of microorganisms present in a mixture by tandem mass spectrometry and the phylopeptidomics concept. Submitted.
- Gouveia et al (2019) A new metaproteomics approach reveals the composition and protein function of the intestinal microbiota in a crustacean sentinel species. Submitted.
- Hayoun et al. (2019) Evaluation of sample preparation methods for fast proteotyping of microorganisms by tandem mass spectrometry. Submitted.

Keywords: Metaproteomics, bioinformatics, Microbiota.

For information please contact: smap2019-abstracts@sciencesconf.org