Taking a bioinformatics hike off the beaten track: metaproteomics and proteomics of non-model organisms

Lennart Martens

Computational Omics and Systems Biology (CompOmics) Group, Department of Medical Protein Research, VIB, Ghent, Belgium, and Department of Biochemistry, Ghent University, Ghent, Belgium

Mass spectrometry based proteomics has traditionally relied heavily on the availability of sequence databases to search the acquired (fragmentation) mass spectra against, and match these to known peptides and proteins. However, not all researchers can optimally make use of this standard strategy. A first example is provided by investigators working in metaproteomics. They encounter the problem of search engine resolution, where a search against too many (related) proteomes leads to high (yet very often strongly underestimated) false discovery rates. Furthermore, even confidently assigned peptides carry substantial ambiguity in them, since they can map to multiple proteins, and multiple species. A second example of researchers unable to optimally use database searching, is found in the analysis of non-model organisms, where (well-annotated) genome or proteome sequences are usually missing altogether. Since there is no database in these cases, search engines become much less useful tools.

We here present tools and analyses that are aimed at alleviating these two issues: on the one hand the broad scope and ambiguous results of a metaproteomics search, and on the other hand the absence of accurate sequence databases when working with non-model organisms.