

Rapid Microorganism Identification Database (www.RMIDb.org)

ASMS
2006

Nathan J. Edwards¹, Fernando Pineda²

¹Center for Bioinformatics & Computational Biology, University of Maryland, College Park;
²Molecular Microbiology & Immunology, Johns Hopkins Bloomberg School of Public Health.

Overview

www.RMIDb.org

- Comprehensive bioinformatics platform for microorganism identification by mass spectrometry of proteins and peptides.
- Supports intact protein and trypsin or acid digest product protocols.
- Facilitates exploratory protocol development.
- Private guest and registered user workspaces.
- Sound statistical significance estimation.

Introduction

Statistically significant rapid microorganism identification using mass spectrometry and protein sequence databases has been demonstrated for various protocols, such as intact proteins from bacterial cells (1) and tryptic (2) and acid (3) digest products from bacterial spores and viruses (4).

Protocol development for rapid microorganism identification requires the determination of preferentially observed proteins or peptides ("identification model") and an appropriate method for estimating the *statistical significance* of mass matches to the model (5).

The Rapid Microorganism Identification Database (www.RMIDb.org) provides each of these informatics needs in an easy to use web interface.

Public Web Interface

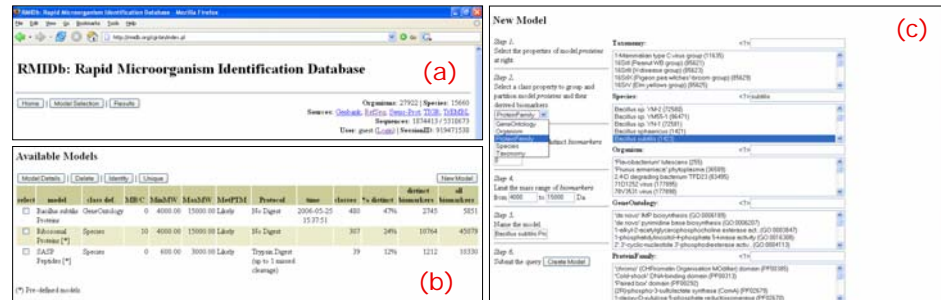


Figure 1: (a) Comprehensive bacteria & virus protein database; (b) System and user defined identification models; (c) New model interface.

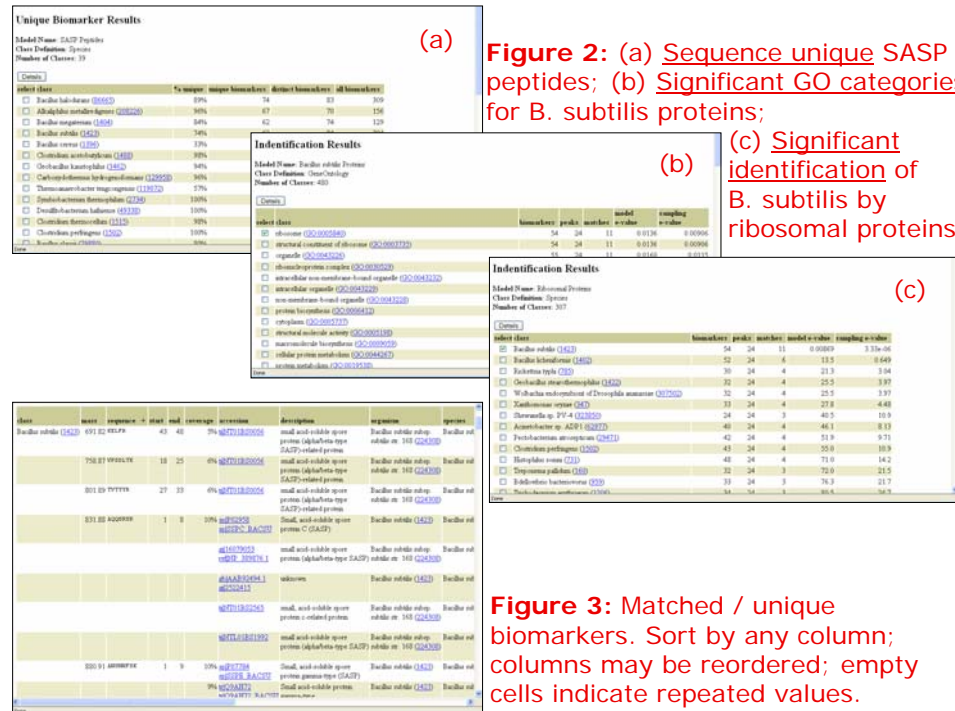


Figure 2: (a) Sequence unique SASP peptides; (b) Significant GO categories for B. subtilis proteins;

(c) Significant identification of B. subtilis by ribosomal proteins.

Statistical Significance

The statistical significance of microorganism identifications are commonly estimated using the method of (5). This model based approach permits an explicit *p*-value computation, but requires problematic assumptions:

- Protein masses are uniformly distributed,
- Input masses match independently.

We instead randomly sample appropriately sized sets of (sequence distinct) biomarkers without regard to the class labels, and count the number of samples that match at least as many peaks as the class of interest. We draw samples until the *p*-value estimate is within 10% of its true value, with 99% confidence. Finally, we report *e*-values to correct for multiple-testing.

Conclusions

The Rapid Microorganism Identification Database (www.RMIDb.org) provides a flexible, easy-to-use bioinformatics platform for microorganism identification. The RMIDb facilitates protocol development by supporting user-defined identification models. A new statistical significance model eliminates problematic assumptions of previous methods.

References

1. Pineda, Antoine, Demirev, Feldman, Longenecker, Lin. Anal. Chem. 2003.
2. English, Warscheid, Fenselau, Cotter. Anal. Chem. 2003.
3. Yao, Demirev, Fenselau. Anal. Chem. 2002.
4. Swatkoski, Russell, Edwards, Fenselau. Anal. Chem. 2006.
5. Pineda, Lin, Fenselau, Demirev. Anal. Chem. 2000.