doi 10.18699/bgrs2024-2.1-12

# Challenges of metabolic pathways proteogenomic mapping

Chudinov I.K.[1, 2]*, Butenko I.O.[1], Vvedenskii A.V.[1], Korneenko E.V.[1, 3], Bondar S.V.[1], Speranskaya A.S.[1]

[1] *Scientific Research Institute for Systems Biology and Medicine, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology (National Research University), Moscow, Russia*

[3] *Saint Petersburg Pasteur Institute, Federal Service on Consumer Rights Protection and Human Well-Being Surveillance, St. Petersburg, Russia*

* *ikchudinov@gmail.com*

**Key words:** genomics; proteomics; proteogenomics; metabolic pathways

*Motivation and Aim*: Dried multispecies biotechnological bacterial preparations, known as starter cultures, are used in the production of fermented dairy products. According to the European Food Safety Authority (EFSA) approach, each bacterial species has additional specifications for safe use based on existing knowledge. Therefore, appropriate methods of investigation are required for bacterial typing and detection of potential hazards. Well-established bacterial typing approaches, including 23S RNA sequencing or multilocus typing, may be insufficient for resolution of different close strains and lack insight into metabolic capabilities of these preparations. In this work, we tested the applicability of a proteogenomic approach for metabolic pathway mapping in *Lactococcus lactis*.

*Methods and Algorithms*: Bacterial isolates were obtained from lactic acid starter cultures of six different manufacturers. Selected isolates were grown on Petri dishes and collected for DNA isolation and bulk growth.

DNA sequencing was performed using PE100 on the MGI platform according to the manufacturer's recommendations. Genome assembly and annotation was performed using pipeline bacass [1].

Five biological replicates for each of six isolates were collected for proteomic analysis. Samples were lysed with sonication, total protein was precipitated, resolubilized, disulfide bonds were reduced, free tiol groups were alkilated and total protein fraction was digested with trypsin, dried and resolubilized. Each peptide sample was subjected to panoramic label-free data-dependent LC-MS analysis.

MSFragger [2] with Percolator [3] and Sage [4] with Mokapot [5] were used independently for spectral data evaluation with sequencing-based annotations and the reference proteome UP000326269 [6]. Pathway reconstruction was performed using QuickGO [7]. Subsequent data processing was performed with in-house Python scripts.

*Results*: Six draft genomes were obtained and annotated from the sequencing data. All isolates were typed as *Lactococcus lactis* subsp. *lactis bv. diacetylactis* strain SD96. Each of the samples is characterized with a shotgun proteomic data in multiple biological replicates. Peptides from more than 1000 proteins have been detected within all spectral identification approaches of the samples. Among the detected proteins, most of them are related to translation, peptidoglycan biosynthesis process and regulation of DNA-templated transcription processes. It was not possible to conclusively distinguish samples by characteristic protein signatures or metabolic pathways, but it is still possible to infer possible pathways in the culture of interest.

*Conclusion*: Existing approaches allow to determine the presence of genes and proteins that characterize the main metabolic pathways, but are not always sufficient to separate bacterial strains or to infer the possibility of producing potentially dangerous metabolites. To the best of our knowledge, there are no "out-of-the-box" solutions that address the ongoing challenges of computational omics research.

*Funding*: The study is supported by the state project "Development of algorithms to identify novel, unique DNA or RNA sequences in metagenomes and their phenotypic characterization in vitro" (No. 12203090069-4).

*References*

1. Peltzer A., Straub D. et al. nf-core/bacass: v2.1.0 nf-core/bacass: "Navy Steel Swordfish" (2.1.0). *Zenodo.* 2023. doi 10.5281/zenodo.10036158
2. Kong A., Leprevost F., Avtonomov D. et al. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nat Methods.* 2017;14:513-520. doi 10.1038/nmeth.4256
3. The M., MacCoss M.J., Noble W.S., Käll L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J Am Soc Mass Spectrom*. 2016;27(11):1719-1727. doi 10.1007/s13361-016-1460-7
4. Lazear M.R. Sage: An Open-Source Tool for Fast Proteomics Searching and Quantification at Scale. *J Proteome Res*. 2023;22(11):3652-3659. doi 10.1021/acs.jproteome.3c00486
5. Fondrie W.E., Noble W.S. mokapot: Fast and Flexible Semisupervised Learning for Peptide Detection. *J Proteome Res*. 2021;20(4):1966-1971. doi 10.1021/acs.jproteome.0c01010
6. Dorau R., Chen J., Jensen P.R., Solem C. Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* bv. diacetylactis SD96. *Microbiol Resour Announc*. 2020;9(3):e01140-19. doi 10.1128/MRA.01140-19.
7. Binns D., Dimmer E., Huntley R., Barrell D., O'Donovan C., Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*. 2009;25(22):3045-3046. doi 10.1093/bioinformatics/btp536