Association of
Biomolecular Resource Facilities

Research • Technology • Communication • Education

iPRG-2020 Proteome Informatics Research Group Study on Metaproteomics

--- Phase 2 ---

Dear iPRG-2020 Study Participant,

Thank you for participating or expressing your interest in participating in this year's Proteome Informatics Research Group (iPRG) Study. This letter provides the instructions needed to participate in Phase 2 of the Study, including how to access the mass spectrometry data and protein sequence files, complete your analysis, and submit your results. Phase 2 submissions are due **Monday November 2, 2020** to be included in the study report that will be presented at the 2021 ABRF Meeting (March 7-10th). **Note that everyone is welcome to submit in Phase 2. Participation in Phase 1 is not a prerequisite for participation in Phase 2!**

Phase 2

As in Phase 1, participants are free to choose whichever software or workflow they want, or to combine several approaches. Results can be reported in any format, including plain text. Jupyter notebooks or R Markdown are also welcome. In Phase 1, we did not reveal the specific taxonomical composition of the sample to the participants. We have now assembled a protein sequence database that should cover all species detectable in the sample. Preliminary data suggests some taxa are abundant and **some may be below the limit of detection in LC-MS/MS**. The species present in the sample may include *Bacillus subtilis*, *Citrobacter freundii*, *Clostridium butyricum*, *Enterobacteria phage T4*, *Escherichia coli*, *Salmonella bongori* and *Salmonella typhimurium*. **In Phase 2, we ask you to use the provided FASTA protein sequence database, which also includes common contaminants.**

Deliverables Phase 2 (due **Monday November 2, 2020**)

2.1 Provide a detailed description of how you performed the analysis after being provided with the FASTA sequence databases of the species present in the sample.
2.2 Provide list of the species or taxa identified in the sample, along with metrics of their relative abundances including number of PSMs, distinct peptides or proteins, and/or their confidences.
2.3 Describe any biologically interesting phenomena (proteins, pathways, etc.) you can observe.

Bonus questions

2.4 Did you find chimeric spectra in the dataset? If so, did you report the peptide/spectrum match and how did you decide between the multiple options?
2.5 There are public datasets for some of the species which are likely to be found in this sample. Comparing the proteins identified in this sample to other public resources, what can you suggest about the physiology/state of the organisms in this sample?
2.6 For peptides corresponding to important taxonomy or functions in a metaproteomics study, how would you validate them?

Raw LC-MS/MS files and the FASTA file are available on https://cpm.lumc.nl/export/iPRG2020/. See Phase 1 instructions for information how the data was acquired. Note the four datasets represent four biological replicates, and that no biological differences are expected. Your Phase 2 results should be submitted via e-mail to the study anonymizer, Sue Weintraub (weintraub@uthscsa.edu). The anonymizer will not share participant identities with anyone, including other members of the iPRG.

Important note to vendors and commercial laboratories: ABRF imposes strict guidelines on the use of study data for marketing purposes. These guidelines are described here:

https://abrf.org/sites/default/files/temp/Resources/research_group_study_participation_guidelines_2010.pdf

We thank you for our support of the ABRF and look forward to receiving your results for the study!

Sincerely,

The ABRF Proteome Informatics Research Group (iPRG)
https://abrf.org/research-group/proteome-informatics-research-group-iprg