Data-independent acquisition mass spectrometry in metaproteomics of gut microbiota – implementation and computational analysis

Juhani Aakko [1,#] Sami Pietilä[1,#], Tomi Suomi[1], Mehrad Mahmoudian[1,2], Raine Toivonen[3], Petri Kouvonen[1], Anne Rokka[1], Arno Hänninen[3,4], Laura L. Elo[1*]

[1] Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

[2] Department of Future Technologies, University of Turku, Finland

[3] Department of Medical Microbiology and Immunology, University of Turku, Finland

[4] TYKS Microbiology, Turku University Central Hospital, Turku, Finland

[#] The authors contributed equally to this work

* Email: laura.elo@utu.fi

Abstract

Metagenomic approaches focus on taxonomy or gene annotation but lack power in defining functionality of gut microbiota. Therefore, metaproteomics approaches have been introduced to overcome this limitation. However, the common metaproteomics approach uses data-dependent acquisition mass spectrometry, which is known to have limited reproducibility when analyzing samples with complex microbial composition. In this work, we provide a proof-of-concept for data-independent acquisition (DIA) metaproteomics. To this end, we analyze metaproteomes using DIA mass spectrometry and introduce an open-source data analysis software package *diatools*, which enables accurate and consistent quantification of DIA metaproteomics data. We demonstrate the feasibility of our approach in gut microbiota metaproteomics using laboratory assembled microbial mixtures as well as human fecal samples.

Keywords: proteomics, metaproteomics, mass spectrometry, data-independent acquisition, data analysis, bioinformatics, software, human gut microbiota, microbiota functionality

Introduction

Metaproteomics is an emerging research area to analyze an entire set of proteins from all microorganisms present in one ecosystem [1]. Its major benefit relies in its ability to reveal directly the functionality of gut microbiota, while the more widely used metagenomics provides insights only on the functional potential.

The common approach in metaproteomics is data-dependent acquisition (DDA) mass spectrometry [2,3]. However, the performance of DDA has been reported to decline when sample complexity increases [2], leading to limited reproducibility, bias toward high abundance peptides, and undersampling [4]. It has been suggested that one possible approach to overcome these limitations is data-independent acquisition (DIA) mass spectrometry [5], such as sequential window acquisition of all theoretical fragment-ion spectra (SWATH) [6]. However, DIA has not been applied to metaproteomics studies of gut microbiota before and currently there is a lack of DIA data analysis software that can reliably process the DIA metaproteomics data.

To this end, we introduce a DIA approach for metaproteomics and a software package *diatools* for analyzing the resulting DIA data. To demonstrate the feasibility of the approach, we chose here the commonly used DIA approach that first builds a spectral library from DDA data and then performs a targeted search for peptides that are present in the library. Although a prerequisite for peptide discovery is that it has been previously discovered through the DDA procedure of selecting the most intensive ions, the subsequent identification in DIA samples is more comprehensive, resulting in a more systematic identification of peptides across all samples [7]. Accordingly, the data processing in *diatools* is divided into two main steps: 1) generation of spectral library from DDA data, and 2) identification and quantification of peptides from the DIA data using the spectral library (**Supplementary Fig. 1**). The first step was implemented following the protocol by Schubert et al. [8],

while the second step builds on the OpenSWATH software [9]. To ensure easy and reproducible access to the software environment, *diatools* was implemented as a Docker image, which is freely available at Docker Hub (elolab/diatools). Further details on the installation and usage of our *diatools* package are provided in software documentation (https://github.com/elolab/diatools).

Experimental Section

The *diatools* software package

The *diatools* software package is designed for automatic analysis of DIA mass spectrometry data from raw data files to a matrix that contains the identified peptides as rows and their intensities in each sample as columns. In short, d*iatools* generates a spectral library from DDA data and uses it to identify and quantify peptides from the DIA data automatically. In essence, diatools has been developed to contain a selection of components chosen from multiple standard proteomics tools and workflows in such a way that they support metaproteomic data analysis of complex gut microbiota data, which otherwise is a challenging task.

The *diatools* software package was built on the Ubuntu operating system (version 17.04) and contains the following preinstalled software: OpenMS [10] (version 2.3), Trans-Proteomic Pipeline (TPP) [11] (version 5.0), msproteomicstools [12] (version 0.6.0), ProteoWizard [13] (version 3.0.11252), and R [14] (version 3.3.2) with the SWATH2stats [15] (version 1.8.1). The *diatools* package with the full software environment is distributed as a Docker image and can be downloaded from Docker Hub repository compbiomed/diatools. Step-by-step instructions to use the software are provided in the software documentation (https://github.com/elolab/diatools).

When developing the software package, we used a single OpenStack[16] virtual machine with 23 cores (Intel Haswell architecture). The available amount of RAM available in the server was 238 GB. In our datasets, the running time was 1 - 2 days per a dataset. Additionally, we tested that all our datasets can be successfully run with 128GB of RAM.

Microbial mixture and human fecal samples

In the present study, a laboratory assembled microbial mixture was used as sample material. The microbial mixture (12mix) contained twelve different strains isolated from fecal samples of three human donors grown on fastidious anaerobe agar (LAB 090; LAB M, UK) and annotated by sequencing their 16S-rDNA: *Bacteroides vulgatus*, *Parabacteroides distasonis*, *Enterorhabdus* sp., *Bifidobacterium pseudocatenulatum*, *Escherichia coli*, *Streptococcus agalactiae*, *Bacteroides fragilis*, *Alistipes onderdonkii*, *Collinsella aerofaciens*, *Clostridium sordellii*, *Eubacterium tenue,* and *Bifidobacterium bifidum*. Prior to mixing, the bacterial cell counts were equalized to 10×10^8 cells / ml using flow cytometry (Bacteria counting kit for FLO, Fisher Scientific) and 1×10^8 cells of each isolate were added to the final mixture.

In addition, six human fecal samples from anonymous individuals were analyzed under the permission of the Southwest Finland Hospital District.

Protein isolation

The protein isolation for the 12mix samples was performed using a Barocycler instrument NEP3229 (Pressure BioSciences Inc., South Easton, Easton, Massachusetts, USA), which uses pressure cycles to lyse the cells. From the human fecal samples, the proteins were isolated using the NoviPure

Microbial Protein Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Protein concentrations were determined using the Bradford method. Fifty µg of protein was used for trypsin digestion. The proteins were reduced with dithiothreitol (DTT) and alkylated with iodoacetamide. The trypsin digestion was performed conventionally in two steps: first trypsin was added in a 1:50 ratio and digested for 4h and then with a 1:30 ratio overnight at 37 °C. After digestion, the peptides were desalted using a SepPak C18 96-well plate (Waters Corporation, Milford, Massachusetts, USA).

LC-MS/MS setup for data-dependent and data-independent analyses

The LC-ESI-MS/MS analyses were performed on a nanoflow HPLC system (Easy-nLC1200, Thermo Fisher Scientific, Waltham, Massachusetts, USA) coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray ionization source. Five hundred ng of the digested protein samples were first loaded on a trapping column and subsequently separated inline on a 15 cm C18 column (75 µm × 15 cm, ReproSil-Pur 5 µm 200 Å C18-AQ, Dr. Maisch HPLC, Ammerbuch-Entringen, Germany). The mobile phase consisted of water with 0.1% formic acid (solvent A) or acetonitrile/water (80:20 volume/volume) with 0.1% formic acid (solvent B). A 90 min two-step gradient from 7% to 35% B, followed by wash with 100% B, was used to elute the peptides.

The mass spectrometry data was acquired automatically using Thermo Xcalibur 3.1 software (Thermo Fisher Scientific). The DDA method consisted of an Orbitrap MS survey scan of mass range 375-1500 m/z followed by higher energy collisional dissociation (HCD) fragmentation for 15 most intense peptide ions. The survey scan was done with 120 K resolution. AGC target was 3e6 and max injection time 50 ms. Monoisotopic masses were then selected for further fragmentation for ions with 2 to 5 charge within a dynamic exclusion range of 30 s and a minimum intensity threshold of 2e4 ions. Precursor ions were isolated using the quadrupole with an isolation window of 1.4 m/z,

NCE 27% was used, the AGC target was set at 1e5 and maximum injection time was 50 ms. For the DDA analysis, the samples of each type were pooled and spiked with indexed retention time peptides (HRM Calibration kit, Biognosys, Schlieren, Switzerland). The pooled 12mix samples were analyzed three times and the pooled human fecal samples six times in DDA mode.

DIA quantification was performed with a resolution of 30 000. AGC target was set at 5e5 with automatic maximum injection time. Isolation window was 15 m/z, and the acquisition window covered a mass range from 400 to 1000 m/z through 40 consecutive isolation windows. For the DIA analysis, the samples were spiked with indexed retention time peptides (HRM Calibration kit, Biognosys, Schlieren, Switzerland) and each sample was injected once.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD008738.

Mass spectrometry data analysis

All data analysis steps from the raw DDA and DIA files to peptide intensity matrix were done using the *diatools* software package. The spectral libraries required for the DIA data analysis were generated separately for each sample type utilizing the data acquired in the DDA mode and two search algorithms: X!Tandem [17] (version 2016.01) and Comet [18] (version 2017.2.1.4). Parent ion mass tolerance was set to 10 ppm and fragment ion tolerance to 0.02 Da. The integrated gene catalogue [19] , containing 9.9 million protein sequences, was utilized as the sequence database when performing the peptide to spectrum matching of the DDA data. The false discovery rate (FDR) was set at 1% for peptide identifications from DDA data. For the TRIC feature alignment [20] of the DIA data, the target and maximum FDRs were 1% and 5%, respectively.

The peptides were functionally and taxonomically annotated using the annotations of the integrated reference catalog of the human gut microbiome [21]. For each peptide, annotations of all possible target protein sequences were retrieved. A functional annotation for the peptide was assigned only if the source proteins had exactly the same functional annotation. Otherwise, the peptide was annotated as ambiguous. A taxonomic annotation was assigned similarly.

Statistical analysis

The reproducibility of quantification between the technical replicates of the 12mix DIA samples was assessed by Pearson correlation.

Results and Discussion

To first demonstrate the technical feasibility of our DIA metaproteomics approach in a controlled setting, we performed liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of a microbial mixture containing twelve (12mix) different bacterial strains. Three technical replicates of the mixture were analyzed. The *diatools* data analysis identified a total of 14888 unique peptides (**Supplementary Tables 1-2**). The number was comparable to those reported earlier from DDA data with similar laboratory protocols [22], which is in line with the fact that the DIA method can identify only peptides that are present in the DDA data spectral library.

Investigation of the overlaps of the identified peptides between the technical replicates of the mixture and correlations between their intensities suggested that DIA had high reproducibility in terms of both identifications and quantifications. When comparing the overlaps of the identified peptides, more than 96% of the peptides were identified in all three technical replicates (**Fig. 1A**).

Moreover, pairwise correlations between the intensities across the replicates were high (r = 0.96, 0.96 and 0.97) (**Fig. 1B** and **Supplementary Fig. 2**).

Additionally, we compared the observed and known compositions of the mixture. At phylum level, 71% of the peptides could be assigned a unique annotation (**Supplementary Fig. 3A**). At genus level, 43% of the peptides could be annotated, respectively (**Fig. 1C**). Importantly, less than 0.1% of all the peptides were incorrectly annotated to phyla not present in the mix and less than 1% of genera were incorrectly annotated.

After confirming the technical feasibility of our DIA metaproteomics approach in the laboratory-grown bacterial mixture, we also included in the analysis fecal samples from human donors to show the applicability of our method in a more complex setting. For this, we assessed six human fecal samples and identified 12804 unique peptides using the *diatools* software package (**Supplementary Tables 1-2**). Again, the number of unique peptides was comparable to that reported earlier from DDA data with similar laboratory protocols [22].

In the more complex human fecal samples, overlaps of the identified peptides across individuals were relatively high with the DIA approach; the median overlap of the identified peptides between each sample pair was 73% (interquartile range IQR 67% – 80%) (**Fig. 2A**). As a reference, Kolmeder et al. [23] reported 30% (IQR 25% – 34%) overlap for DDA data in their study (**Fig. 2B**). Of the identified peptides, unique annotation could be assigned to 47% of the peptides at phylum level and to 28% of the peptides at genus level (**Supplementary Fig. 3B, Fig. 2C**). Although the majority of the peptides remained unannotated, the taxonomic composition of the metaproteome was similar to those reported earlier by others [3,23,24].

Finally, we investigated the functional characteristics of the peptides in both the laboratory-grown bacterial mixture and the human fecal samples using annotations of the integrated reference catalog of the human gut microbiome [25]. For each peptide, we retrieved KEGG orthologous group (KOG) identifiers of all proteins from which it may have originated. For the majority of the peptides, the retrieved identifiers were identical and a single KOG could be assigned in the sample sets: 88% in the 12mix and 87% in the human fecal samples (**Supplementary Fig. 4**). This suggests that the peptide-level functional annotations can be assigned without the intermediate step of protein inference. Therefore, to avoid the potential ambiguities caused by protein inference, we argue that it may be more feasible to use peptide-based methods to detect functional differences in metaproteomics, similarly as suggested for single-organism proteomics [26]. The most common KOGs found in the human fecal samples (**Supplementary Table 3**) have been reported by others as well [23,27].

A main fallback of the DIA data analysis with a spectral library is that it might not represent the whole diversity of the samples. While the DIA approach aims to reduce the problem of undersampling in metaproteomics [4], the pooled DDA samples used to build the spectral libraries may be affected by peptides that are commonly detected in the sample set. However, the improved reproducibility of DIA is expected to be useful, for instance, when assessing metaproteome differences between groups of individuals, in which having a set of reproducibly detectable peptides across samples is crucial.

Conclusions

Our results demonstrate for the first time the utility of DIA in metaproteomics analysis of complex microbial samples. We provide an open-source software package *diatools* that enables easy

application of the method and also future refinements and improvements of the method, which we present in this work mainly as a proof-of-concept.

Acknowledgements

LLE reports grants from the European Research Council ERC (677943), European Union's Horizon 2020 research and innovation programme (675395), Academy of Finland (296801, 304995, 310561 and 313343), Juvenile Diabetes Research Foundation JDRF (2-2013-32), Tekes – the Finnish Funding Agency for Innovation (1877/31/2016) and Sigrid Juselius Foundation, during the conduct of the study. MM received funding from Doctoral Program in Mathematics and Computer Sciences of the University of Turku Graduate School. AH was supported by Päivikki and Sakari Sohlberg Foundation and Turku University Central Hospital.

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

A preprint of this manuscript has been deposited to bioRxiv (doi: https://doi.org/10.1101/413021).

Author contributions

JA, SP, MM, PK, AR, and LLE designed the study. JA, SP and TS wrote the first draft of the manuscript. RT and AH provided the clinical material and processed it together with PK and AR. JA, SP and TS wrote the software package. JA, SP, TS and MM performed the data analysis. LLE conceived and supervised the study and participated in writing the manuscript. All authors contributed to interpreting the results as well as editing the manuscript. All authors have read the final version of the manuscript and approved of its content.

Competing Financial Interests

The authors declare no competing financial interests.

SUPPORTING INFORMATION: The following files are available free of charge at ACS website: http://pubs.acs.org:

Supplementary Code: *Diatools* workflow commands and parameters.

Supplementary Figure 1. The workflow of our data-independent acquisition (DIA) approach for metaproteomics.

Supplementary Figure 2. Scatter plot illustrating the correlations between all pairs of technical replicates.

Supplementary Figure 3. Proportions of the identified peptides taxonomically annotated on phylum level.

Supplementary Figure 4. Number of KEGG orthologous groups assigned to each peptide.

Supplementary Table 1. Statistics of the spectral libraries used for analyzing the DIA data.

Supplementary Table 2. Functional and taxonomical annotation yields per sample type.

Supplementary Table 3. The most abundant KOGs (>1%) in the human fecal dataset.

Supplementary Table 4. Raw spectrum files of 12mix dataset.

Supplementary Table 5. Raw spectrum files of human fecal dataset.

References

(1) Wilmes, P.; Bond, P. L. Metaproteomics: Studying Functional Gene Expression in Microbial Ecosystems. *Trends Microbiol.* **2006**, *14* (2), 92–97.

(2) Tanca, A.; Palomba, A.; Pisanu, S.; Deligios, M.; Fraumene, C.; Manghina, V.; Pagnozzi, D.; Addis, M. F.; Uzzau, S. A Straightforward and Efficient Analytical Pipeline for Metaproteome Characterization. *Microbiome* **2014**, *2* (1), 49.

(3) Kolmeder, C. A.; de Been, M.; Nikkilä, J.; Ritamo, I.; Mättö, J.; Valmu, L.; Salojärvi, J.; Palva, A.; Salonen, A.; de Vos, W. M. Comparative Metaproteomics and Diversity Analysis of Human Intestinal Microbiota Testifies for Its Temporal Stability and Expression of Core Functions. *PLoS One* **2012**, *7* (1), e29913.

(4) Michalski, A.; Cox, J.; Mann, M. More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority Is Inaccessible to Data-Dependent LC-MS/MS. *J. Proteome Res.* **2011**, *10* (4), 1785–1793.

(5) Muth, T.; Renard, B. Y.; Martens, L. Metaproteomic Data Analysis at a Glance: Advances in Computational Microbial Community Proteomics. *Expert Rev. Proteomics* **2016**, *13* (8), 757–769.

(6) Huang, Q.; Yang, L.; Luo, J.; Guo, L.; Wang, Z.; Yang, X.; Jin, W.; Fang, Y.; Ye, J.; Shan, B.; et al. SWATH Enables Precise Label-Free Quantification on Proteome Scale. *Proteomics* **2015**, *15* (7), 1215–1223.

(7) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Miladinović, S. M.; Cheng, L.-Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; et al. Extending the Limits of Quantitative Proteome Profiling

with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Molecular & Cellular Proteomics*. 2015, pp 1400–1410. https://doi.org/10.1074/mcp.m114.044305.

(8) Schubert, O. T.; Gillet, L. C.; Collins, B. C.; Navarro, P.; Rosenberger, G.; Wolski, W. E.; Lam, H.; Amodei, D.; Mallick, P.; MacLean, B.; et al. Building High-Quality Assay Libraries for Targeted Analysis of SWATH MS Data. *Nat. Protoc.* **2015**, *10* (3), 426–441.

(9) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; et al. OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nat. Biotechnol.* **2014**, *32* (3), 219–223.

(10) Sturm, M.; Bertsch, A.; Gröpl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; et al. OpenMS - an Open-Source Software Framework for Mass Spectrometry. *BMC Bioinformatics* **2008**, *9*, 163.

(11) Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Farrah, T.; Lam, H.; Tasman, N.; Sun, Z.; Nilsson, E.; Pratt, B.; Prazen, B.; et al. A Guided Tour of the Trans-Proteomic Pipeline. *Proteomics* **2010**, *10* (6), 1150–1159.

(12) msproteomicstools. msproteomicstools/msproteomicstools https://github.com/msproteomicstools/msproteomicstools (accessed Nov 20, 2018).

(13) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–920.

(14) Website R: The R Project for Statistical Computing [Internet]. [cited 20 Nov 2018]. Available: https://www.r-project.org/ (accessed Jan 25, 2019).

(15) Blattmann, P.; Heusel, M.; Aebersold, R. SWATH2stats: An R/Bioconductor Package to Process and Convert Quantitative SWATH-MS Proteomics Data for Downstream Analysis Tools. *PLoS One* **2016**, *11* (4), e0153160.

(16) Build the future of Open Infrastructure https://www.openstack.org/ (accessed May 28, 2019).

(17) Craig, R.; Beavis, R. C. TANDEM: Matching Proteins with Tandem Mass Spectra. *Bioinformatics* **2004**, *20* (9), 1466–1467.

(18) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An Open-Source MS/MS Sequence Database Search Tool. *Proteomics* **2013**, *13* (1), 22–24.

(19) Li, J.; Jia, H.; Cai, X.; Zhong, H.; Feng, Q.; Sunagawa, S.; Arumugam, M.; Kultima, J. R.; Prifti, E.; Nielsen, T.; et al. An Integrated Catalog of Reference Genes in the Human Gut Microbiome. *Nat. Biotechnol.* **2014**, *32* (8), 834–841.

(20) Röst, H. L.; Liu, Y.; D'Agostino, G.; Zanella, M.; Navarro, P.; Rosenberger, G.; Collins, B. C.; Gillet, L.; Testa, G.; Malmström, L.; et al. TRIC: An Automated Alignment Strategy for Reproducible Protein Quantification in Targeted Proteomics. *Nat. Methods* **2016**, *13* (9), 777–783.

(21) Li, J.; Jia, H.; Cai, X.; Zhong, H.; Feng, Q.; Sunagawa, S.; Arumugam, M.; Kultima, J. R.; Prifti, E.; Nielsen, T.; et al. An Integrated Catalog of Reference Genes in the Human Gut Microbiome. *Nat. Biotechnol.* **2014**, *32* (8), 834–841.

(22) Zhang, X.; Li, L.; Mayne, J.; Ning, Z.; Stintzi, A.; Figeys, D. Assessing the Impact of Protein Extraction Methods for Human Gut Metaproteomics. *J. Proteomics* **2018**, *180*, 120–127.

(23) Kolmeder, C. A.; Salojärvi, J.; Ritari, J.; de Been, M.; Raes, J.; Falony, G.; Vieira-Silva, S.; Kekkonen, R. A.; Corthals, G. L.; Palva, A.; et al. Faecal Metaproteomic Analysis Reveals a Personalized and Stable Functional Microbiome and Limited Effects of a Probiotic Intervention in Adults. *PLoS One* **2016**, *11* (4), e0153294.

(24) Tanca, A.; Abbondio, M.; Palomba, A.; Fraumene, C.; Manghina, V.; Cucca, F.; Fiorillo, E.; Uzzau, S. Potential and Active Functions in the Gut Microbiota of a Healthy Human Cohort. *Microbiome* **2017**, *5* (1), 79.

(25) Li, J.; Jia, H.; Cai, X.; Zhong, H.; Feng, Q.; Sunagawa, S.; Arumugam, M.; Kultima, J. R.; Prifti, E.; Nielsen, T.; et al. An Integrated Catalog of Reference Genes in the Human Gut Microbiome. *Nat. Biotechnol.* **2014**, *32* (8), 834–841.

(26) Suomi, T.; Corthals, G. L.; Nevalainen, O. S.; Elo, L. L. Using Peptide-Level Proteomics Data for Detecting

Differentially Expressed Proteins. *J. Proteome Res.* **2015**, *14* (11), 4564–4570.
(27) Tanca, A.; Abbondio, M.; Palomba, A.; Fraumene, C.; Manghina, V.; Cucca, F.; Fiorillo, E.; Uzzau, S. Potential and Active Functions in the Gut Microbiota of a Healthy Human Cohort. *Microbiome* **2017**, *5* (1), 79.

Figures and tables

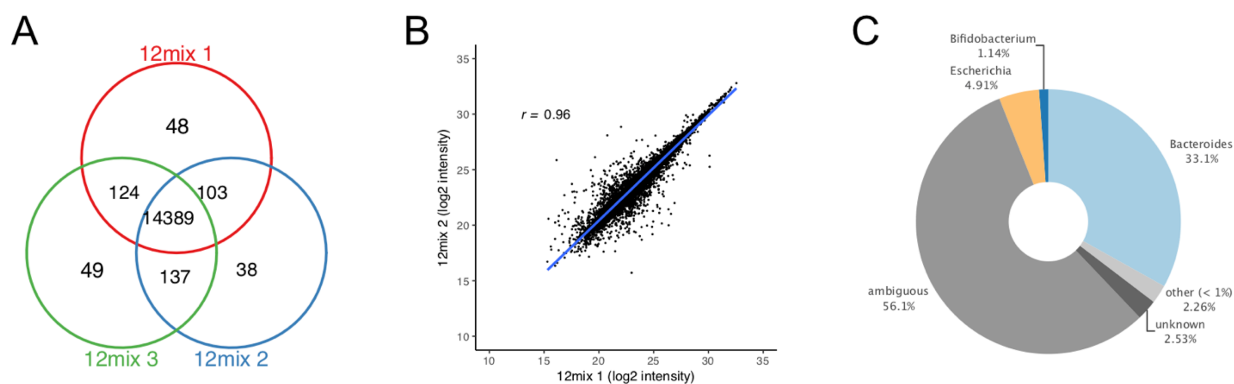


Figure 1. (**A**) Overlaps of the identified peptides between the three technical replicates of the 12mix samples. (**B**) Representative example of correlations of peptide quantifications between two technical replicates of the 12mix samples. (**C**) Proportions of the identified peptides taxonomically annotated on genus level in the 12mix samples.

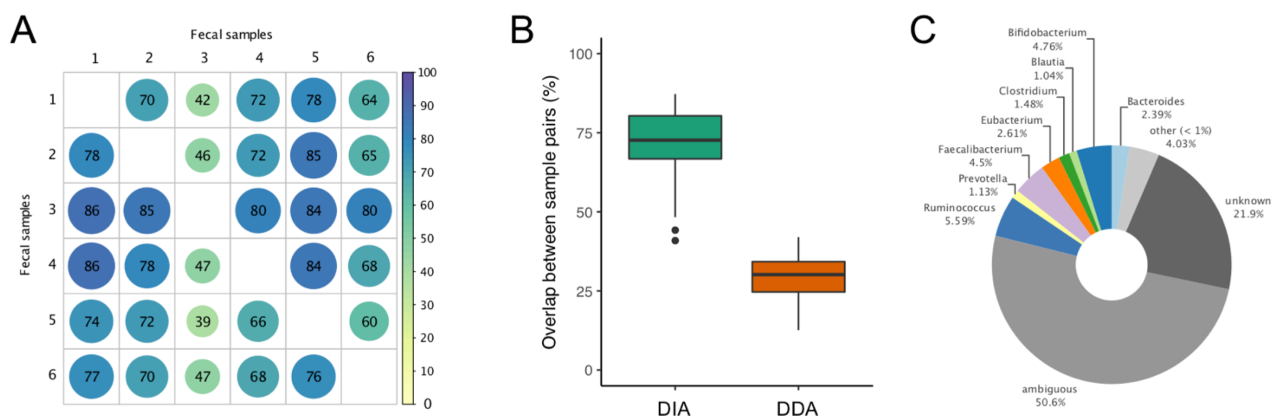
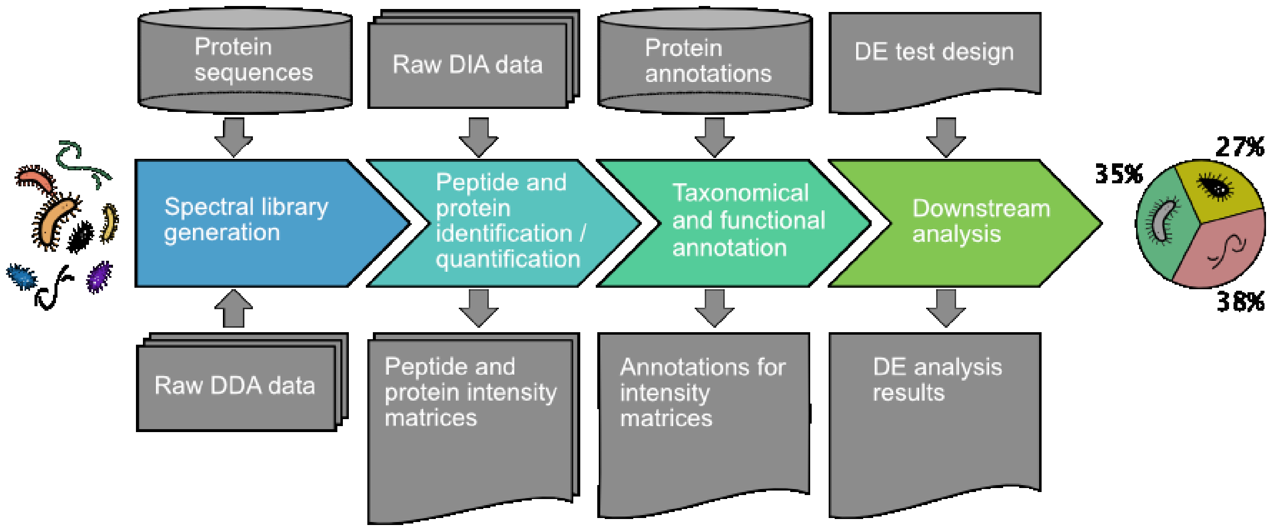


Figure 2. (**A**) Overlaps of the peptide identifications between each possible pair of human fecal samples. The proportions were calculated by dividing the number of common peptides by the total number of peptides in either one of the samples of a sample pair, which results in an asymmetric matrix of percentages. (**B**) Overlaps of peptide identifications across sample pairs in the current DIA study and a recent DDA study [23]. (**C**) Proportions of the identified peptides taxonomically annotated on genus level in the human fecal samples.



For TOC Only