Reporting checklist for metaproteomics experiments

Version 0.1 - *manuscript in revision*

This checklist, prepared by the Metaproteomics Initiative (www.metaprotemics.org), summarizes key elements that we highly recommend be reported in any metaproteomics study. When using this checklist, please cite *Van Den Bossche, et al. 2026. Preprint on ChemRxiv (https://doi.org/10.26434/chemrxiv-2025-tm1ms).*

Item to report	Recommended reporting	Reported?
Sample Information		
Sample source	Describe the origin of the sample (e.g., soil, human gut, marine), the surrounding environment, and relevant parameters.	
Collection method	Report how the sample was collected, including the method used, sample volume or weight, and any pre-storage handling or processing.	
Storage and preservation	Specify storage conditions (e.g., temperature, duration), preservation method, time between collection (removal of sample from environment) and preservation, and time kept in preservation until processing.	
Experimental design		
Replicates	Report the number and type of replicates included in the study, clearly distinguishing between biological and technical replicates. Indicate how replicates were processed.	
Statistical design	Describe the overall study design and statistical considerations, including randomization strategy, use of controls, blocking, and batch correction. If applicable, provide power calculations or a justification for the chosen number of replicates.	
Sample (pre-)processing		
Pre-processing	Specify sample pre-processing steps, such as fractionation e.g., by centrifugation or filtering, and concentration e.g., with filters or lyophilization.	
Cell lysis and homogenization	Specify the lysis method (e.g., mechanical, enzymatic, chemical), including any cell separation or inactivation steps. Report whether protease inhibitors were used.	
Protein extraction	Describe the method used to extract proteins from lysed cells, including buffers or chemicals applied and any precipitation or cleanup steps.	
Protein fractionation	Describe protein fractionation method (e.g., 1D-PAGE or chromatography)	

Protein digestion	Detail the digestion protocol, including protein quantity, buffer composition, reduction and alkylation conditions, enzyme(s) used, incubation time and temperature, enzyme-to-protein ratio, and peptide cleanup method.	
Peptide fractionation	Describe peptide-level prefractionation strategies applied prior to LC-MS/MS analysis (e.g., high-pH reversed-phase liquid chromatography, strong cation exchange).	
Labelling	If chemical, metabolic or isotopic labeling was used (e.g., TMT, iTRAQ, Protein-SIP), describe the labeling protocol and any post-labeling processing steps.	
Contamination control measures	Specify relevant measures used to prevent, monitor, or account for contamination, such as blank controls, digestion controls, or carryover checks, where applicable.	
LC-(IM)-MS/MS		
LC separation conditions	Report LC parameters, including instrument, trap-and-elute vs. direct inject, column type, length, particle size, solvent composition, gradient types and elution conditions, flow rate, and peptide load.	
Ion mobility separation conditions	If ion mobility was used, specify the technique (e.g., TIMS, TWIMS, FAIMS), drift gas type and pressure, temperature, electric field settings, and calibration method. If collision cross section (CCS) values were reported, include the calibration procedure and associated uncertainty.	
MS acquisition parameters	Specify the mass spectrometer model and configuration. Report acquisition mode (e.g., DDA or DIA), scan range, resolution, dynamic exclusion, and fragmentation settings. For DDA, include MS2 resolution and top N selection strategy; for DIA, specify isolation window size and scan strategy.	
Bioinformatics analysis		
Protein database construction	Report the number of sequences in the search database and the download date if sourced from a public repository. If a custom database was used, describe in detail how it was generated, including the origin of the sequences, how taxa were selected or excluded, how redundancy was handled, and how both functional and taxonomic annotations were assigned. Indicate how contaminant sequences were incorporated.	

Search engine and parameters	State which search engine has been used and which version number. Mention parameters that have been used in the search, including but not limited to precursor- and fragment ion tolerances, fixed (static) and variable (dynamic) modifications, number of allowed missed cleavages, and enzymatic specificity (e.g., tryptic, semi-tryptic)	
FDR control	Define the false discovery rate (FDR) approach used and the thresholds applied at the peptide-spectrum match (PSM), and, where applicable, peptide, and protein or protein (sub)group levels. Specify whether global or local FDR was used.	
Protein identification criteria	Specify the criteria used to report proteins or protein (sub)groups, including the minimum number of peptides required for identification and whether peptides must be unique or may include shared peptides.	
Protein inference	Describe the algorithm or method used to group peptides into proteins and protein (sub)groups, including version and parameters where applicable.	
Identification metrics	Report the number of PSMs, peptides, and proteins or protein (sub)groups identified at the defined FDR threshold.	
Protein quantification	Indicate the quantification strategy applied (label-free or labeling), the normalization method, the proportion of missing values, and whether imputation was performed (including the method, if applicable).	
Taxonomic profiling	Report the method or software used for taxonomic annotation, including version and relevant parameters. Provide the total number of taxonomic units identified (e.g., genera, species) and the number of peptides or proteins assigned to each. If taxonomic units were quantified, describe the quantification method. Include any filtering or binning steps (e.g., grouping low-abundance taxa) and justify these choices.	
Functional annotation	Report the method or software used for functional annotation, including version and relevant parameters. Provide the number of functional groups detected and the number of peptides or proteins assigned to each. If functional groups were quantified, describe the quantification method in sufficient detail to ensure reproducibility.	
Post-processing	List any tools or scripts used for downstream processing, such as normalization, visualization, statistical analysis, or	

	rescoring. Include software versions and key parameters. Make any custom scripts critical to result reproduction publicly available.	
Data, metadata and code availability		
Data availability	All data, including but not limited to raw files, protein database (FASTA file), DIA library, primary search results, processed tables, and figure source data, should be deposited in a ProteomeXchange repository, such as PRIDE. We recommend a complete submission that includes processed identification results in a PSI format (e.g., mzIdentML) to enable linking peptides to spectra and ensure integration, visualization, and reanalysis. The submission to the repository must include a table that allows understanding the linkage between samples and data files (e.g., SDRF-proteomics format).	
Metadata availability	All metadata should be submitted in SDRF-Proteomics format. While the Metaproteomics Initiative is currently developing dedicated templates for different microbial environments in collaboration with HUPO-PSI, we recommend using the default SDRF-Proteomics template.	
Code availability	Make analysis scripts and pipelines publicly available through a code-sharing platform, such as GitHub. For long-term access and citation, archive the code in a repository that provides a DOI, such as Zenodo. Clearly indicate the exact version used, for example by referencing a GitHub tag or commit hash. Include enough detail to reproduce results, figures, and tables.	