Supplementary Online Material

Comprehensive evaluation of statistical approaches for differential metaproteomics

Authors:

Tjorven Hinzke*#[1,2,3], Benoit J. Kunath*#[3,4,5], J. Alfredo Blakeley-Ruiz[3], Abigail Korenek[3], Simina Vintila[3], Paul Wilmes[4,6], Manuel Kleiner#[3]

*contributed equally

Corresponding authors: tjorven.hinzke@uni-greifswald.de, benoit.kunath@lih.lu, manuel_kleiner@ncsu.edu

1 University of Greifswald, partner of the Greifswald Mire Centre, Greifswald, Germany
2 Helmholtz Institute for One Health, Greifswald, Germany
3 Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, USA
4 Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
5 Multiomics Data Science, Department of Cancer Research, Luxembourg Institute of Health, Strassen, Luxembourg
6 Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Supplementary Material and Methods

Cultivation of microorganisms

Escherichia coli strain HS was grown on either LB complex medium (Bertani, 1951) or M9 minimal medium (Harwood and Cutting, 1990) with 1% glucose at 37 °C until the bacteria reached exponential phase; 3 h and 20 min for LB and 4 h and 30 min for M9. Some 1 L of cultures were grown in quadruplicates for each condition. Samples were pelleted by centrifugation at 8,000 g for 30 min. Supernatant was removed and pellets were resuspended in 100 mL of PBS. Suspended pellets were aliquoted into 64 tubes in 1 mL aliquots. Aliquots were centrifuged at 17,000 x g for 10-15 min. Supernatant was removed and frozen at -80 °C.
Bacterioides thetaiotaomicron was grown in degassed serum bottles in meat broth (Bacic and Smith, 2008) at 37 °C in quadruplicates overnight until turbid. 1.5 ml aliquots were removed and centrifuged at 13,000 x g for 5 min. Pellets were resuspended in 1 ml sterile PBS buffer and pelleted again at 13,000 x g for 5 min. Pellets were immediately stored at -80 °C.
Thermus thermophilus strain HB27 was grown in quadruplicates at either 70 °C (high temperature) or 55 °C (low temperature). Cultures were grown in 1 L TSB (Difco) supplemented with 0.4% yeast extract and 0.3% NaCl, for 24-48 h and harvested at 15,000 x g for 20 min at 4 °C. Pellets were resuspended in 100 ml sterile PBS buffer, aliquoted into 1.7 ml Eppendorf tubes and centrifuged again at 15,000 x g for 5 min at 4 °C. Pellets were immediately stored at -80 °C.
Chlamydomonas reinhardii CC-3403 was grown in TAP-medium (Gorman and Levine, 1965) supplemented with 267 mM arginine. Cultures were grown at room temperature under either 270 µE (high light) or 27 µE (low light) with a 12:12 light:dark cycle. Some 1 L of cultures were grown for 72-96 h in quadruplicates for each condition and harvested at 10,000 x g for 20 min at 4 °C. Pellets were resuspended in 110 ml sterile PBS, aliquoted into 1.7 ml Eppendorf tubes and centrifuged again at 10,000 x g for 5 min at 4 °C. Pellets were immediately stored at -80 °C.
Rhizobium leguminosarum strains RL3841 and VF39 were grown in 1L YEM-broth (Vincent, 1970). Each strain was grown in quadruplicates at 28 °C for 48 h. Cultures were harvested at 15,000 x g for 20 min at 4 °C. Pellets were resuspended in 110 ml sterile PBS, aliquoted into 1.7 ml Eppendorf tubes and centrifuged again at 15,000 x g for 5 min at 4 °C. Pellets were immediately stored at -80 °C.

Metaproteomics database generation

Databases for the different mixes were built from combining building blocks as follows: For mixes M01 to M06 based on the matrix made of standard mouse faecal pellets, the protein sequence database was constructed by combining the clustered microbial translated metagenome, the clustered mouse proteome, the clustered mouse diet proteome, and contaminants. The resulting matrix database was clustered at 99% using CD-HIT-2D (Li and Godzik, 2006) with the database of clustered microbial isolate genomes to remove sequences from the matrix database that were very similar to sequences from the pure cultures. The resulting matrix-specific protein sequences were concatenated with all microbial isolate protein sequences to generate the final database for the mixes M01 to M06.

For the mixes M07 to M11 based on the sterile corn root matrix, the matrix proteome database consisted of the clustered corn proteome and the contaminants. Like the mouse proteome, the resulting matrix database was clustered at 99% using CD-HIT-2D with the clustered microbial isolate database to remove sequences from the matrix database that were very similar to sequences from the pure culture isolates. The resulting matrix-specific protein sequences were concatenated with the microbial isolate proteomes to generate the final database for the mixes M07 to M11.

For the mixes M12 and M13 based on the gnotobiotic mouse matrix, the matrix protein sequence database consisted of the clustered proteins of the reduced microbial community, the clustered mouse proteome (see above), the clustered diet proteins and the contaminants (see above). The resulting matrix database was clustered at 99% using CD-HIT-2D with the clustered isolates to remove sequences from the matrix database that were very similar to sequences from the pure culture isolates. The resulting matrix-specific sequences were concatenated with the pure culture proteomes to generate the final database for the mixes M12 to M13.

Defined metaproteome protein sample generation

Pure culture peptide generation

Pellets of pure cultures were used for protein extraction and digested into peptides using the filter-aided sample preparation (FASP; Wiśniewski *et al.*, 2009; Blakeley-Ruiz *et al.*, 2025) with few modifications as follows. Briefly, for each pellet, ~300 µL of SDT lysis buffer (4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT) was added. Some samples underwent bead beating using lysing matrix E tubes (MP Biomedicals) for 5 cycles of 45 s at 6,45 m/s speed and 1 min pause between cycles (see Supplementary Table 1 for details). All samples were heated to 95 °C for 10 min and centrifuged at 21,000 x g for 5 min prior being transferred to several FASP filters (60 µL/filter). We mixed 60 µl of the lysate with 400 µL of UA solution (8 M urea in 0.1 M Tris/HCl pH 8.5) onto a modified PES 10 kDa 500 µL filter unit (VWR International) and centrifuged at 14,000 x g for 30 min. The filters were washed using 200 µL of UA solution and centrifuged at 14,000 g for 40 min followed by incubation with 100 µL IAA (0.05 M iodoacetamide in UA solution) for 20 min and centrifugation at 14,000 g for 20 min. Filters were then washed three times with 100 µL of UA buffer and 3 times using 100 µL of ABC (50 mM Ammonium Bicarbonate) buffer. The proteins were digested into peptides by adding 0.8 µg of MS grade trypsin (Thermo Scientific Pierce) solubilized in 40 µL of ABC buffer to the filters and incubating for 16 h in a wet chamber at 37°C. Peptides were eluted by centrifugation at 14,000 x g for 20 min followed by another elution using 50 µL of 0.5 M NaCl and centrifugation at 14,000 x g for 20 min. Extracted peptides from the same samples extracted on different filters were combined into ~800-1000 µL of resulting peptides, and 0.4 to 0.5 ml of 0.1% formic acid (FA) was added prior to desalting using C18 Sep-Pak Light cartridges (Waters) and syringes. Briefly, the Sep-Pak cartridge was washed and primed with 5 ml of wash solution (70% ACN, 0.1% FA) followed by 5 ml of 0.1% FA solution.

The sample was loaded in three steps on the cartridge, rinsed with 5 ml of wash solution and eluted slowly into two tubes using 1.5 ml of elution solution (90% ACN, 0.1% FA). The samples were concentrated and depleted of acetonitrile with a first vacuum centrifugation step to ~300 µL and solvent was exchanged by adding 600 µL of 0.1% FA solution followed by a second vacuum centrifugation step until a final volume of ~300µL was reached. Resulting peptide concentrations were measured in triplicate using the Pierce Micro BCA assay (Thermo Scientific Pierce) according to the manufacturer's instructions.

Matrix peptide generation

For the conventional mouse stool matrix, fecal pellets from a previous study (Blakeley-Ruiz *et al.*, 2025) were gathered from four different cages and homogenized together to form a representative sample of ~1.6 g fecal pellet, from which 500 mg were used for metagenomics analysis. Proteins and peptides were extracted and processed similarly to the pure cultures using the FASP protocol (see above). For obtaining corn roots, *Zea mays* seeds were sterilized using the protocol described in Parnell *et al.* (2024) by immersing the seeds in 70% ethanol for 3 min followed by immersion in 2% sodium hypochlorite for 3 min. Seeds were rinsed in de-ionized H_2O ten times before being placed individually in sterile WhirlPak bags (Nasco) containing 90 ml of sterile clay (Pro's Choice Rapid Dry, OIL-DRI) and watered with 90 ml sterile 0.5x Murashige-Skoog media. Bags were sealed with Aeraseal (Millipore Sigma) and plants were grown for 13 days at 25 °C with 16:8 hours light:dark cycle. Roots were harvested, chopped and frozen at -80 °C until further processing. For the corn roots and gnotobiotic mice samples (Mordant, Blakeley-Ruiz and Kleiner, 2025), the FASP protocol was not providing sufficient yield and the extractions were performed using the S-Trap mini MS sample prep kit (Protifi) instead. The original S-Trap manufacturer's protocol uses different reductant and alkylator as compared to the FASP protocol, which would have precluded mixing of the samples. Therefore, the S-Trap protocol was modified to use DTT and IAA and was applied as follows. For both matrices, 100 mg of material was mixed with 1 mL of SDS lysis buffer (5% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT), bead-beated using lysing matrix E tubes (MP Biomedicals; 5 cycles of 45 s at 6,45 m/s speed and 1 min pause between cycles), heated to 95 °C for 10 min, and centrifuged at 21,000 x g for 5 min. Samples were then aliquoted into 140 µL of sample. Reduction was performed by adding 6.4 µL of 500 mM DTT and incubating at 95 °C for 10 min. Subsequently, samples were alkylated by adding 12.7 µL of 500 mM IAA and incubating in the dark at room temperature for 30 min. Samples were then acidified using 16 µL of 12% phosphoric acid solution (to reach pH<1). Finally, 1050 µL of binding/wash buffer (100 mM TEAB in 90% methanol) was mixed with the samples to form protein colloids, loaded (two times 600 µL) onto S-Trap columns and centrifuged at 4,000 x g for 30 s. The column was washed three times with 400 µL of binding/wash buffer and centrifuged at 4,000 x g for 30 s, followed by a final centrifugation at 4,000 x g for 1 min to fully remove the buffer. Proteins were then digested into peptides by adding 0.8 µg of MS grade trypsin (Thermo Scientific Pierce) solubilized in 40 µL of digestion/elution buffer (50 mM TEAB in water) to the filters and incubating for 16 h in a wet chamber at 37 °C. Peptides were eluted by adding 80 µL of digestion/elution buffer, incubation 10 min at 37 °C and centrifugation at 4,000 x g for 1 min; followed by two more elutions by adding 80 µL of elution buffer 2 (0.2% FA) and centrifugation at at 4,000 x g for 1 min, and finally adding 80 µL of elution buffer 3 (50% ACN, 0.2% FA) and

centrifugation at at 4,000 x g for 1 min. The resulting 240 µL were concentrated and depleted of acetonitrile using vacuum centrifugation until a volume of ~60 µL was reached. Desalting of the peptides and their concentration measurement were performed as for the pure cultures (see above).

Defined metaproteome production

Following peptide quantification, the matrices and all quadruplicates of the pure cultures were diluted to 250 ng/µL and the dilutions were measured again in triplicate using the Pierce Micro BCA assay (Thermo Scientific Pierce) to ensure that the correct concentration was achieved. Aliquots were generated at all steps to subject samples to freeze-thaw cycles only once per preparation step. Peptides extracted from the different isolates and conditions were then mixed in different proportions with the respective matrix peptides to produce the various defined metaproteomes (see Main Text, Figure 1 and Supplementary Table 4).

Mouse metagenome sequencing and analysis

To generate generic mouse stool matrix protein data for the metaproteomics database, two times 250 mg of fecal pellets were used for DNA extraction with the DNeasy PowerSoil Pro Kit following manufacturer's instructions except that bead beating was used instead of vortexing (3.1 m/s for 3 cycles of 30 sec. with 1 min of cooling on ice in between each cycle) in 2 ml bead beating tubes (Lysing Matrix E, MP Biomedicals) using a Bead Ruptor Elite 24 (Omni International). The extracted DNA was eluted with 60 µL of solution C6, merged into one tube and DNA concentration of the combined extractions was assessed using a DS-11 FX+ Spectrophotometer (Denovix) with the Qubit™ dsDNA High Sensitivity Assay Kit (Invitrogen). Metagenomic DNA was submitted to the North Carolina State Genomic Sciences Laboratory (Raleigh, NC, USA) for Illumina library construction and sequencing to produce ~400 M 150 bp paired-end reads. Library construction was performed using an Illumina TruSeq Nano Library kit according to manufacturer's instructions. Libraries were sequenced on an Illumina NextSeq 500 sequencer.

Removal of the Illumina Truseq3 adapters and quality trimming of the raw reads was performed using Trimmomatic v0.39 (Bolger, Lohse and Usadel, 2014), removing all bases on the 3′-end with a Phred score lower than 20 (if present) and excluding all reads shorter than 40 bp. The trimmed reads were assembled using MEGAHIT v1.2.9 (Li *et al.*, 2015) with default parameters, minimum and maximum k-mer sizes of 25 and 99 respectively and a k-mer increment of 4. The trimmed reads were mapped against the resulting assembly contigs using the Burrows-Wheeler Aligner with maximal exact matches (BWA-MEM; Li and Durbin, 2009) algorithm requiring 100% identity. The coverage was used to bin the assembly contigs with several binners, namely, MetaBAT v2.2.15 (Kang *et al.*, 2019), Maxbin 2.2.7 (Wu, Simmons and Singer, 2016), SemiBin v1.0.3 (Pan *et al.*, 2022), Binny v2 (Hickl *et al.*, 2022) and MetaDecoder v1.0.13 (Liu *et al.*, 2022) with default parameters. The bins from the different tools were then refined with DAS Tool v1.1.4 (Sieber *et al.*, 2018) using a score threshold of 0.7 and keeping both the contigs belonging to refined bins and the unbinned fraction. Open reading frames (ORFs) were then predicted on both

fractions using Prodigal v2.6.3 (Hyatt *et al.*, 2010) and the resulting ORFs were used for database generation.

Statistical analyses: normalization and transformation

For normalization, which is done to correct for differences in overall quantification (sum of AUC or PSMs) between mass spectrometry runs, as well as unequal abundances of the species to be compared between samples in the case of species-level normalization (Kleiner, 2017), we used total sum scaling (TSS), cumulative sum scaling (CSS), and (organism-level) normalized spectral abundance factors ((org)NSAF; Zybailov *et al.*, 2006; Mueller *et al.*, 2010).
Additionally, we used the trimmed mean of M-values (TMM) normalization (Robinson and Oshlack, 2010), and its new variant TMM with singleton pairing (TMMwsp; Chen *et al.*, 2025), which was developed for more robust analysis of low-abundant variables, as well as the relative log expression (RLE; Anders and Huber, 2010). The voom normalization for limma (Law *et al.*, 2014) uses the mean-variance trend for precision weight estimation. These weights are then used in subsequent analysis. DeqMS (Zhu *et al.*, 2020), another limma normalization method, which was explicitly developed for proteomics data, models the prior variance per group depending on the number of PSMs or peptides. Additionally, we employed variance stabilization normalization (vsn; (Huber *et al.*, 2002), which aims to make the standard deviation independent of the mean.
We used different transformations to make the data, or the residuals of a linear modelling approach, normally distributed. We used log2, logit, sqrt and arcsine square root (ast) transformation. The sqrt of TSS is also called Hellinger transformation (Legendre and Gallagher, 2001). Additionally, we used the centered log ratio (clr) transformation, which was proposed to remedy the issue of compositionality (Aitchison, 1982). More recently, the chiPower (chiP) transformation was proposed as an alternative to clr, which does not require zero replacement (Greenacre, 2024). For chiP transformation, protein group abundances are first raised to power lambda, then devided by the sample total, chi-square standardized (divided by the square root of the mean of that protein across samples), BoxCox-transformed and centered. We tested different values of lambda. See Table 3 of the Main Text for combinations of normalizations/transformations and tests.

Software implementation

For versions of R 4.3.1 (R Core Team, 2023) packages used, please see Supplementary Table 3.

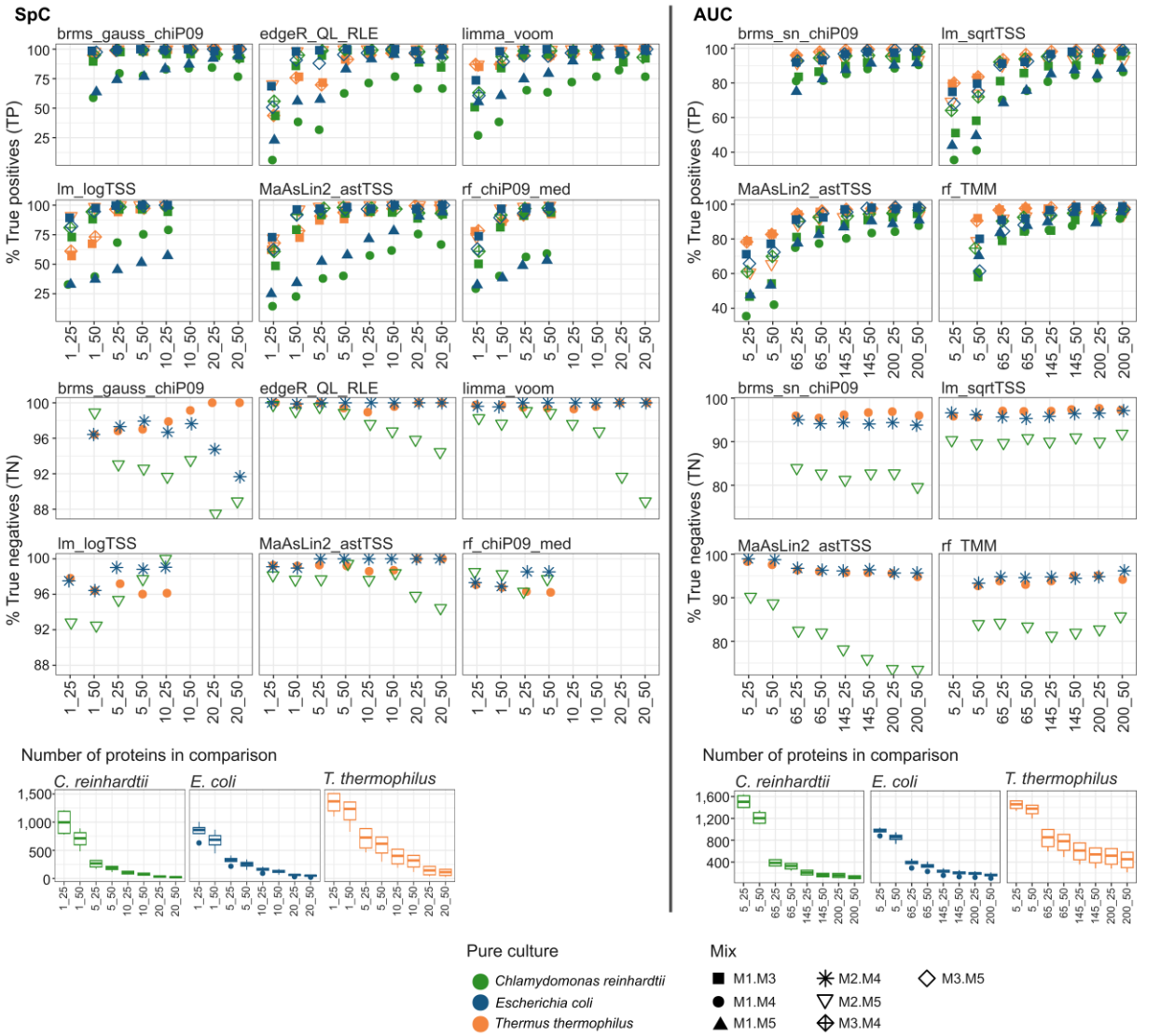
Supplementary Table 3: R Packages and versions used in this study.

Package	Version	Reference
General data reformatting		
dplyr	1.1.4	(Wickham, François, *et al.*, 2023)
plyr	1.8.9	(Wickham, 2011)
tidyr	1.3.1	(Wickham, Vaughan and Girlich, 2024)
tidyverse	2.0.0	(Wickham *et al.*, 2019)
Statistics and helper functions for statistics		
brms	2.22.0	(Bürkner, 2018)
cmdstanr	0.8.1	(Gabry *et al.*, 2024)
compositions	2.0-6	(van den Boogaart, Tolosana-Delgado and Bren, 2023)
corncob	0.3.1	(Martin, Witten and Willis, 2022)
DEqMS	1.2.0	(Zhu *et al.*, 2020)
DESeq2	1.40.2	(Love, Huber and Anders, 2014)
edgeR	4.0.16	(Chen *et al.*, 2025)
irr	0.84.1	(Gamer *et al.*, 2019)
limma	3.58.1	(Ritchie *et al.*, 2015)
lme4	1.1-36	(Bates *et al.*, 2015)
lmerTest	3.1-3	(Kuznetsova, Brockhoff and Christensen, 2017)
Maaslin2	1.16.0	(Mallick *et al.*, 2021)
matrixTests	0.2.3	(Koncevičius, 2023)
ranger	0.16.0	(Wright and Ziegler, 2017)
rstatix	0.7.2	(Kassambara, 2023)
TAF	4.2.0	(Magnusson and Millar, 2023)
tidymodels	1.2.0	(Kuhn and Wickham, 2020)
zCompositions	1.5.0-4	(Palarea-Albaladejo and Martín-Fernández, 2015)
Parallel job execution		
doParallel	1.0.17	(Microsoft Corporation and Weston, 2022)
foreach	1.5.2	(Microsoft and Weston, 2022)
Figure generation		
circlize	0.4.16	(Gu *et al.*, 2014)
ComplexHeatmap	2.18.0	(Gu, 2022)
egg	0.4.5	(Auguie, 2019)
gghalves	0.1.4	(Tiedemann, 2022)
ggplot2	3.5.1	(Wickham, 2016)
patchwork	1.1.3	(Pedersen, 2023)
RColorBrewer	1.1-3	(Neuwirth, 2022)
scico	1.5.0	(Pedersen and Crameri, 2023); color palette: batlow (Crameri, Shephard and Heron, 2020)
svglite	2.1.1	(Wickham, Henry, *et al.*, 2023)

Supplementary Results and Discussion

Filtering increases true positives and impacts true negatives

We tested the effect of filtering by evaluating within-condition abundance changes, i.e., evaluations which are independent of comparing with the pure-culture proteomes, as a pure culture proteome ground truth would also be dependent on the filtering (and therefore would need to be evaluated for each filtering level). For SpC, we varied filter criteria between at least 1 PSM in at least 25% of samples (lowest stringency) up to at least 20 PSM in at least 50% of samples (highest stringency). For AUC, we varied between an intensity of at least 5.000.000 in at least 25% of samples (lowest stringency) to at least 200.000.000 in at least 50% of samples (highest stringency). Increasingly stringent filtering by the number of samples which need to contain a certain protein group abundance value and the protein group abundance value itself generally increased the percentage of true positives, but did not impact or decreased the percentage of true negatives (Supplementary Figure 1). Obviously, also the absolute number of protein groups in the comparison decreased with more stringent filtering. For random forests using SpC data, too-harsh filtering criteria caused the p-value calculation to fail, as no negative importances were found. In the case of AUC data, too-lenient filtering criteria led to brms and the random forest being prohibitively slow.

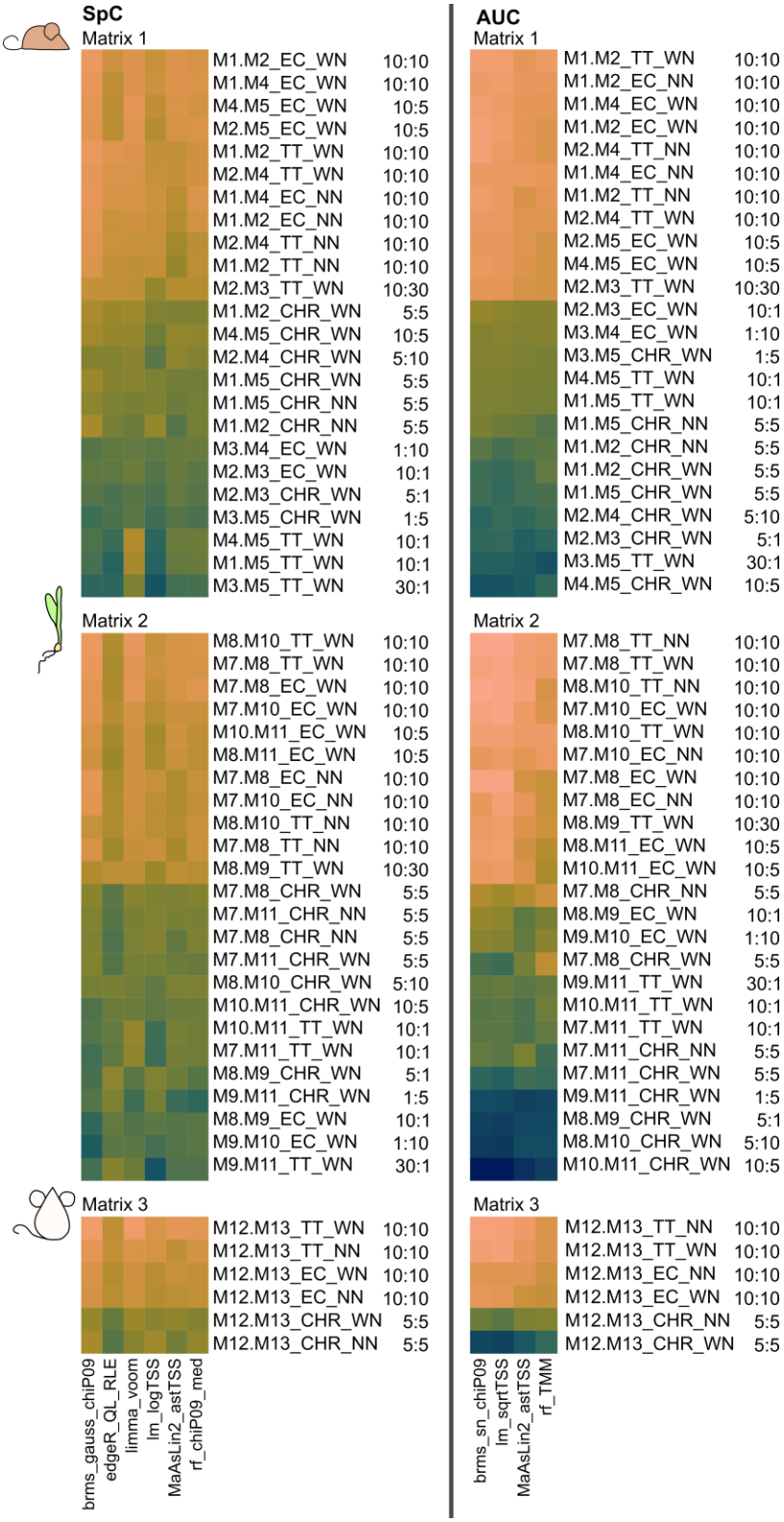


Supplementary Figure 1: Effect of filtering on the percentage ot true positives and true negatives found for the respective comparison. For spectral count (SpC) data, numbers on the x axis indicate the minimum number of spectra required in a certain share of samples, e.g., 1_25 means that at least 1 PSM is required in at least 25% of the samples. For peptide intensity data (AUC), numbers on the x axis indicate the minimum intensity (times 10^6) required in a certain share of samples, e.g., 5_25 means an intensity of 5.000.000 required in at least 25% of samples. Missing values indicate that the calculation was not possible, either because too many values were missing (rf for SpC data) or because the model evaluation got prohibitively slow due to a too-large dataset (brms and rf for AUC).

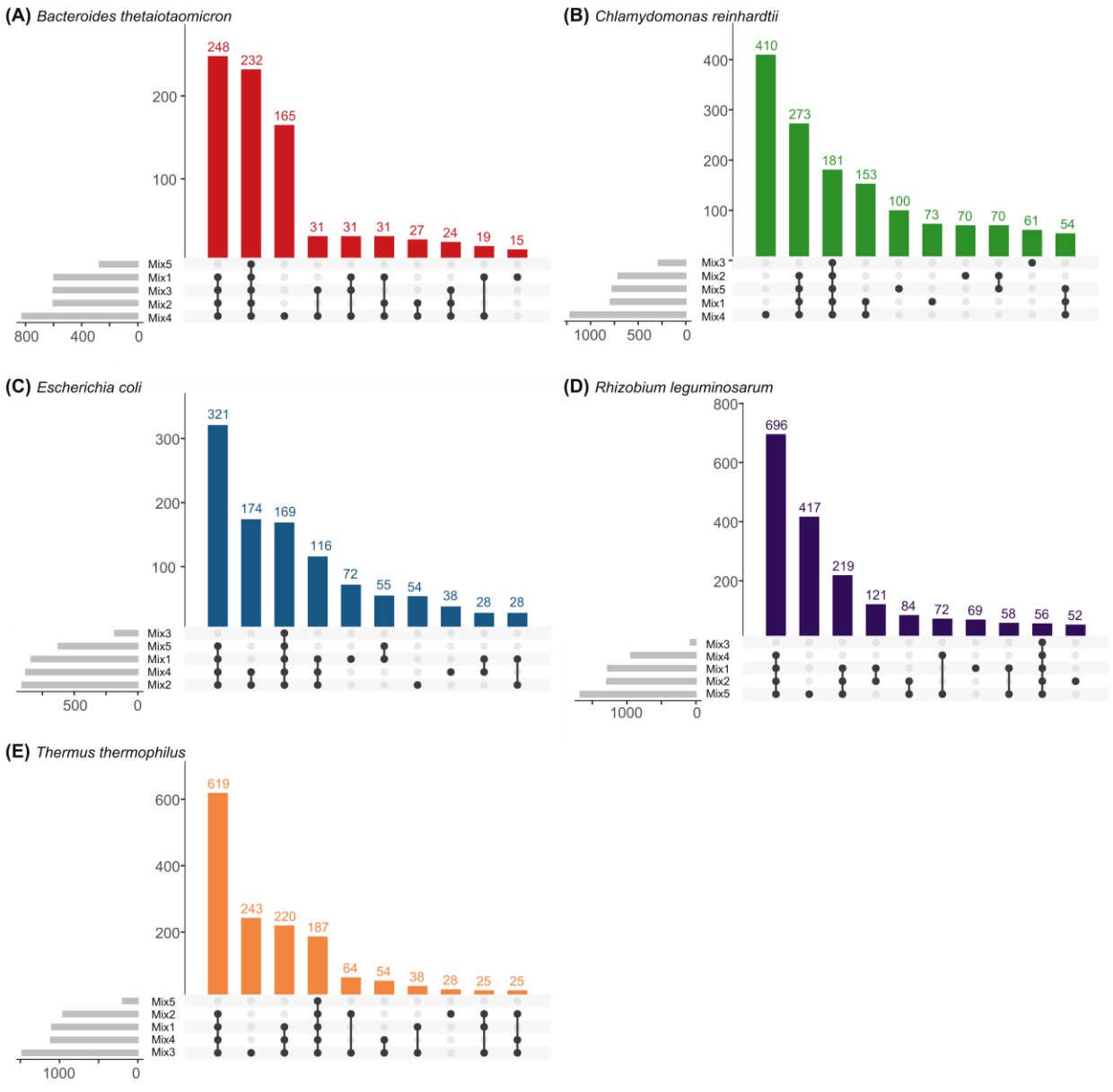
Matrix complexity and abundance changes impact AUC data evaluation performance more than SpC data evaluation performance

Marked differences existed in P4 scores depending on abundances of organisms compared, but also depending on the organisms themselves, with *C. reinhardtii* producing consistently lower results (Supplementary Figures 1 and 5, Supplementary Tables 5e, 5h, 5i, 6e, 6h, 6i). While AUC comparisons performed better for the higher-abundant *T. thermophilus* and *E. coli* comparisons, their performance loss was relatively and absolutely higher than that of SpC-based comparisons

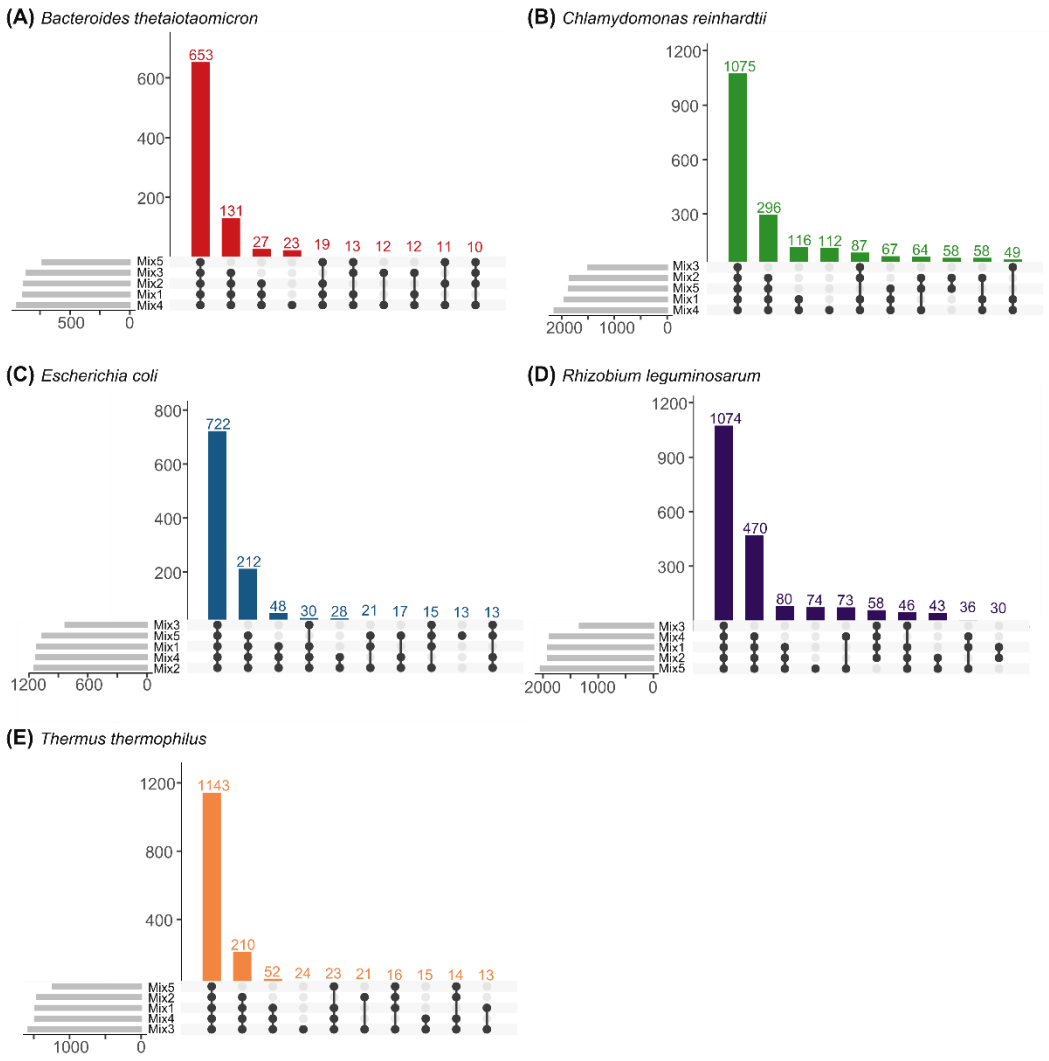
for the other tests (see Main Text). The differential performance of SpC vs. AUC-based statistics in general might partially be caused by a match-between-runs transfer of peptide identifications and AUC quantifications, while SpC-based data does not use match-between-runs, concomitant with less consistent protein identifications especially for *C. reinhardtii* (Supplementary Figures 3, 4, and Main Text).



Supplementary Figure 2: P4 scores for the best-performing tests of SpC and AUC comparisons per comparison. WN: with species-level normalization, NN: without species-level normalization.



Supplementary Figure 3: Numbers of protein groups identified per pure culture organism and mix, or combination of mixes, based on spectral count (SpC) data in the mouse faecal pellet matrix.

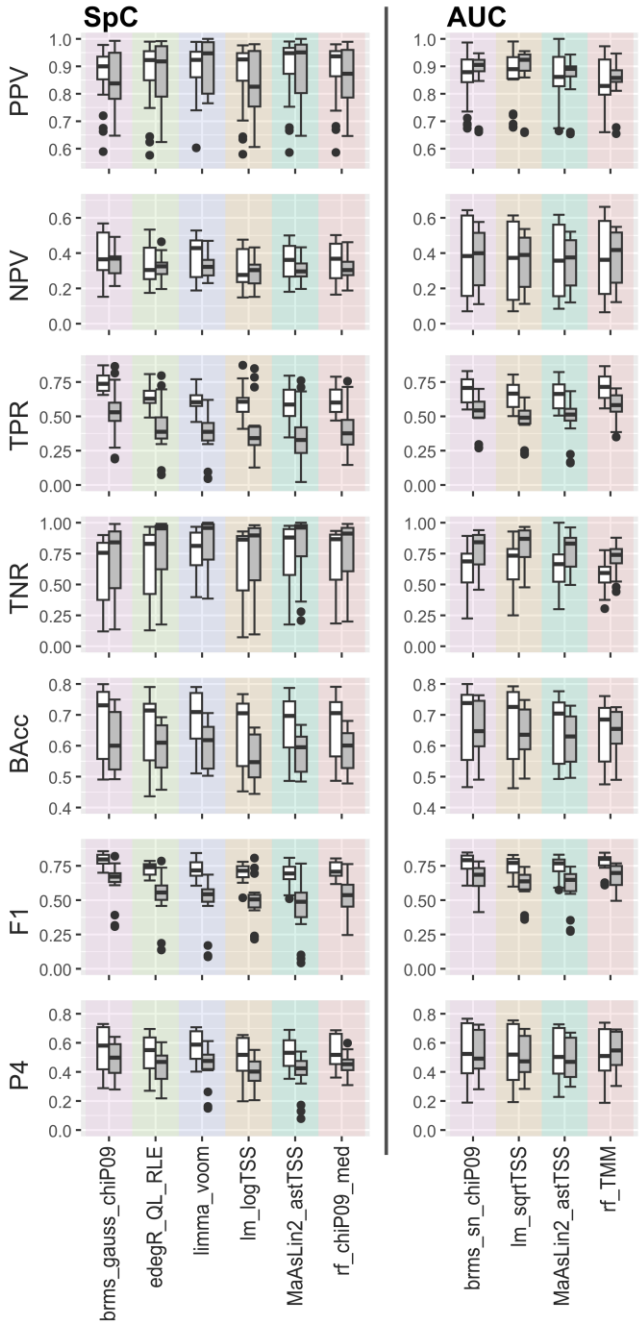


Supplementary Figure 4: Numbers of protein groups identified per pure culture organism and mix, or combination of mixes, based on peptide intensity (AUC) data in the mouse faecal pellet matrix.

Confounding effects decrease significant results

Comparisons of species (EC, TT, CRH) from M1 to M5 with the same species in M6 (Figure 1), where these species were mixed in from two culturing conditions (i.e., where these two conditions are confounded), generally decreased test performance, with the exception of NPV and TNR, which can increase if fewer proteins are detected as significantly differentially abundant overall. Additionally, AUC-based tests fared better overall on TPR as compared to SpC-based tests (Supplementary Figure 5). This result highlights that in many real-life datasets, e.g., biofilms with several phenotypes of the same organism being present (Stewart & Franklin, 2008), decreases in performance of the statistics have to be expected – meaning that only relatively stronger changes of expressed phenotypes can be accessed, if not accounting explicitly for spatial variation in the experimental design. There is no straightforward statistical approach to disentangle small

changes in protein abundances and the presence of subpopulations with differing changes in protein abundances. Theoretically, changes in the protein abundances in different subpopulations can cancel each other out. Therefore, enrichment of sub-populations (e.g.,(Babin *et al.*, 2017; Hinzke *et al.*, 2021) or activity-based labeling such as stable isotope probing (Sachsenberg *et al.*, 2015; Kleiner *et al.*, 2023) should be considered in order to draw meaningful conclusions about samples where sub-populations are to be expected.



Supplementary Figure 5: Comparison of test statistics for samples in which each species is only present from one culturing condition (white bars, M1 to M5 compared against each other for three species TT, EC, CRH) vs. a species present in a mix of two culture conditions (grey bars, M1 to M5 compared with EC, TT, CRH in M6). SpC: Spectral count quantification data; AUC: Area under the curve quantification data.

References

Aitchison, J. (1982). The Statistical Analysis of Compositional Data, Journal of the Royal Statistical Society Series B: Statistical Methodology, 44(2), pp. 139–160. doi: 10.1111/j.2517-6161.1982.tb01195.x.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data, Genome Biology, 11(10), p. R106. doi: 10.1186/gb-2010-11-10-r106.

Auguie, B. (2019). egg: Extensions for “ggplot2”: Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size. R package version 0.4.5. CRAN.R-project: package=egg.

Babin, B.M. et al. (2017). Selective Proteomic Analysis of Antibiotic-Tolerant Cellular Subpopulations in *Pseudomonas aeruginosa* Biofilms, mBio. Edited by J.P. Boyle, 8(5), pp. e01593-17. doi: 10.1128/mBio.01593-17.

Bacic, M.K. and Smith, C.J. (2008). Laboratory Maintenance and Cultivation of *Bacteroides* Species, Current Protocols in Microbiology, 9(1). doi: 10.1002/9780471729259.mc13c01s9.

Bates, D. et al. (2015). Fitting Linear Mixed-Effects Models Using lme4, Journal of Statistical Software, 67(1). doi: 10.18637/jss.v067.i01.

Bertani, G. (1951). Studies on Lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*, Journal of Bacteriology, 62(3). doi: 10.1128/jb.62.3.293-300.1951.

Blakeley-Ruiz, J.A. et al. (2025). Dietary protein source alters gut microbiota composition and function, The ISME Journal, 19(1), p. wraf048. doi: 10.1093/ismejo/wraf048.

Bolger, A.M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

van den Boogaart, K., Tolosana-Delgado, R. and Bren, M. (2023). compositions: Compositional Data Analysis. R package version 2.0-6. CRAN.R-project: package=compositions.

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms, The R Journal, 10(1), p. 395. doi: 10.32614/RJ-2018-017.

Chen, Y. et al. (2025). edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets, Nucleic Acids Research, 53(2), p. gkaf018. doi: 10.1093/nar/gkaf018.

Gabry, J. et al. (2024). cmdstanr: R Interface to “CmdStan”. R package version 0.8.1. mc-stan: cmdstanr/.

Gamer, M. et al. (2019). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. CRAN.R-project: package=irr.

Gorman, D.S. and Levine, R.P. (1965). Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardi*, Proceedings of the National Academy of Sciences, 54(6), pp. 1665–1669. doi: 10.1073/pnas.54.6.1665.

Greenacre, M. (2024). The chiPower transformation: a valid alternative to logratio transformations in compositional data analysis, Advances in Data Analysis and Classification, 18(3), pp. 769–796. doi: 10.1007/s11634-024-00600-x.

Gu, Z. et al. (2014). circlize implements and enhances circular visualization in R, Bioinformatics, 30(19), pp. 2811–2812. doi: 10.1093/bioinformatics/btu393.

Gu, Z. (2022). Complex heatmap visualization, iMeta, 1(3), p. e43. doi: 10.1002/imt2.43.

Harwood, C.R. and Cutting, S.M. (1990). Chemically defined growth media and supplements, in Molecular Biological Methods for *Bacillus*, p 548.

Hickl, O. et al. (2022). binny : an automated binning algorithm to recover high-quality genomes from complex metagenomic datasets, Briefings in Bioinformatics, 23(6), p. bbac431. doi: 10.1093/bib/bbac431.

Hinzke, T. et al. (2021). Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis, eLife, 10, p. e58371. doi: 10.7554/eLife.58371.

Huber, W. et al. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression, Bioinformatics, 18(suppl_1), pp. S96–S104. doi: 10.1093/bioinformatics/18.suppl_1.S96.

Hyatt, D. et al. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification, BMC Bioinformatics, 11(1), p. 119. doi: 10.1186/1471-2105-11-119.

Kang, D.D. et al. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies, PeerJ, 7, p. e7359. doi: 10.7717/peerj.7359.

Kassambara, A. (2023). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.2. CRAN.R-project: package=rstatix.

Kleiner, M. (2017). Normalization of metatranscriptomic and metaproteomic data for differential gene expression analyses: The importance of accounting for organism abundance. doi: 10.7287/peerj.preprints.2846v1.

Kleiner, M. et al. (2023). Ultra-sensitive isotope probing to quantify activity and substrate assimilation in microbiomes, Microbiome, 11(1), p. 24. doi: 10.1186/s40168-022-01454-1.

Koncevičius, K. (2023). matrixTests: Fast Statistical Hypothesis Tests on Rows and Columns of Matrices. R package version 0.2.3. CRAN.R-project: package=matrixTests.

Kuhn, M. and Wickham, H. (2020) Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. www.tidymodels.org.

Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models, Journal of Statistical Software, 82(13). doi: 10.18637/jss.v082.i13.

Law, C.W. et al. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts, Genome Biology, 15(2), p. R29. doi: 10.1186/gb-2014-15-2-r29.

Legendre, P. and Gallagher, E.D. (2001). Ecologically meaningful transformations for ordination of species data, Oecologia, 129(2), pp. 271–280. doi: 10.1007/s004420100716.

Li, D. et al. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, Bioinformatics, 31(10), pp. 1674–1676. doi: 10.1093/bioinformatics/btv033.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics, 22(13), pp. 1658–1659. doi: 10.1093/bioinformatics/btl158.

Liu, C.-C. et al. (2022). MetaDecoder: a novel method for clustering metagenomic contigs, Microbiome, 10(1), p. 46. doi: 10.1186/s40168-022-01237-8.

Love, M.I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biology, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Magnusson, A. and Millar, C. (2023). TAF: Transparent Assessment Framework for Reproducible Research. R package version 4.2.0. CRAN.R-project: package=TAF.

Mallick, H. et al. (2021). Multivariable association discovery in population-scale meta-omics studies, PLOS Computational Biology. Edited by L.P. Coelho, 17(11), p. e1009442. doi: 10.1371/journal.pcbi.1009442.

Martin, B., Witten, D. and Willis, A. (2022). corncob: Count Regression for Correlated Observations with the Beta-Binomial. R package version 0.3.1.

Microsoft Corporation and Weston, S. (2022). doParallel: Foreach Parallel Adaptor for the “parallel” Package. R package version 1.0.17. CRAN.R-project: package=doParallel.

Microsoft and Weston, S. (2022). foreach: Provides Foreach Looping Construct. R package version 1.5.2. CRAN.R-project: package=foreach.

Mordant, A., Blakeley-Ruiz, J.A. and Kleiner, M. (2025). Metaproteomics-based stable isotope fingerprinting links intestinal bacteria to their carbon source and captures diet-induced substrate switching, The ISME Journal, 19(1), p. wraf127. doi: 10.1093/ismejo/wraf127.

Mueller, R.S. et al. (2010). Ecological distribution and population physiology defined by proteomics in a natural microbial community, Molecular Systems Biology, 6(1), p. 374. doi: 10.1038/msb.2010.30.

Neuwirth, E. (2022). RColorBrewer: ColorBrewer Palettes. R package version 1.1-3. CRAN.R-project: package=RColorBrewer.

Palarea-Albaladejo, J. and Martín-Fernández, J.A. (2015). zCompositions — R package for multivariate imputation of left-censored data under a compositional approach, Chemometrics and Intelligent Laboratory Systems, 143, pp. 85–96. doi: 10.1016/j.chemolab.2015.02.019.

Pan, S. et al. (2022). A deep siamese neural network improves metagenome-assembled genomes in microbiome datasets across different environments, Nature Communications, 13(1), p. 2326. doi: 10.1038/s41467-022-29843-y.

Parnell, J.J. et al. (2024). Effective Seed Sterilization Methods Require Optimization Across Maize Genotypes, Phytobiomes Journal, 8(4), pp. 418–424. doi: 10.1094/PBIOMES-12-23-0137-R.

Pedersen, T. (2023). patchwork: The Composer of Plots. R package version 1.1.3. CRAN.R-project: package=patchwork.

Pedersen, T. and Crameri, F. (2023). scico: Colour Palettes Based on the Scientific Colour-Maps. R package version 1.5.0. CRAN.R-project: package=scico.

R Core Team (2023). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria [Preprint].

Ritchie, M.E. et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Research, 43(7), pp. e47–e47. doi: 10.1093/nar/gkv007.

Robinson, M.D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data, Genome Biology, 11(3), p. R25. doi: 10.1186/gb-2010-11-3-r25.

Sachsenberg, T. et al. (2015). MetaProSIP: Automated Inference of Stable Isotope Incorporation Rates in Proteins for Functional Metaproteomics, Journal of Proteome Research, 14(2), pp. 619–627. doi: 10.1021/pr500245w.

Sieber, C.M.K. et al. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy, Nature Microbiology, 3(7), pp. 836–843. doi: 10.1038/s41564-018-0171-1.

Tiedemann, F. (2022). gghalves: Compose Half-Half Plots Using Your Favourite Geoms. R package version 0.1.4. CRAN.R-project: package=gghalves.

Vincent, J.M. (1970) A manual for the practical study of the root-nodule bacteria. Blackwell Scientific Publications, Oxford.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis, Journal of Statistical Software, 40(1). doi: 10.18637/jss.v040.i01.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. 2nd ed. 2016. Cham: Springer International Publishing : Imprint: Springer (Use R!).

Wickham, H. et al. (2019). Welcome to the Tidyverse, Journal of Open Source Software, 4(43), p. 1686. doi: 10.21105/joss.01686.

Wickham, H., François, R., et al. (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.4. CRAN.R-project: package=dplyr.

Wickham, H., Henry, L., et al. (2023). svglite: An “SVG” Graphics Device. R package version 2.1.3. CRAN.R-project: package=svglite.

Wickham, H., Vaughan, D. and Girlich, M. (2024). tidyr: Tidy Messy Data. R package version 1.3.1.

Wiśniewski, J.R. et al. (2009). Universal sample preparation method for proteome analysis, Nature Methods, 6(5), pp. 359–362. doi: 10.1038/nmeth.1322.

Wright, M.N. and Ziegler, A. (2017). ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, Journal of Statistical Software, 77(1). doi: 10.18637/jss.v077.i01.

Wu, Y.-W., Simmons, B.A. and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets, Bioinformatics, 32(4), pp. 605–607. doi: 10.1093/bioinformatics/btv638.

Zhu, Y. et al. (2020). DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis, Molecular & Cellular Proteomics, 19(6), pp. 1047–1057. doi: 10.1074/mcp.TIR119.001646.

Zybailov, B. et al. (2006). Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*, Journal of Proteome Research, 5(9), pp. 2339–2347. doi: 10.1021/pr060161n.