

# Applying LFQRatio Normalization in Quantitative Proteomic Analysis of Microbial Co-culture Systems

Mengxun Shi<sup>1</sup>, Caroline A. Evans<sup>1</sup>, Josie L. McQuillan<sup>1</sup>, Josselin Noirel<sup>2</sup> and Jagroop Pandhal<sup>1, \*</sup>

<sup>1</sup>School of Chemical, Materials and Biological Engineering, The University of Sheffield, Sheffield, UK

<sup>2</sup>GBCM Laboratory (EA7528), Conservatoire National des Arts et Métiers, HESAM Université, 2 rue Conté, Paris, France

\*For correspondence: [j.pandhal@sheffield.ac.uk](mailto:j.pandhal@sheffield.ac.uk)

## Abstract

Quantitative proteomic analysis plays a crucial role in understanding microbial co-culture systems. Traditional techniques, such as label-free quantification (LFQ) and label-based proteomics, provide valuable insights into the interactions and metabolic exchanges of microbial species. However, the complexity of microbial co-culture systems often leads to challenges in data normalization, especially when dealing with comparative LFQ data where ratios of different organisms can vary across experiments. This protocol describes the application of LFQRatio normalization, a novel normalization method designed to improve the reliability and accuracy of quantitative proteomics data obtained from microbial co-cultures. The method was developed following the analysis of factors that affect both the identification of proteins and the quantitative accuracy of co-culture proteomics. These include peptide physicochemical characteristics such as isoelectric point (pI), molecular weight (MW), hydrophobicity, dynamic range, and proteome size, as well as shared peptides between species. We then created a normalization method based on LFQ intensity values named LFQRatio normalization. This approach was demonstrated by analysis of a synthetic co-culture of two bacteria, *Synechococcus elongatus* cscB/SPS and *Azotobacter vinelandii* ΔnifL. Results showed enhanced accuracy of differentially expressed proteins, allowing for more reliable biological interpretation. This protocol provides a reliable and effective tool with wider application to analyze other co-culture systems to study microbial interactions.

## Key features

- Assessment of factors affecting the quantitative accuracy of co-culture proteomics.
- Provides a LFQRatio normalization method for label-free quantification of microbial co-cultures.
- Recommendations for co-culture proteomics for mixed microbial populations.

**Keywords:** Proteomics, LFQRatio normalization, Label-free quantification, Quantitative proteomics, Microbial co-culture

**This protocol is used in:** J Proteome Res (2024), DOI: 10.1021/acs.jproteome.3c00714; Commun Biol (2024), DOI: 10.1038/s42003-024-06098-5

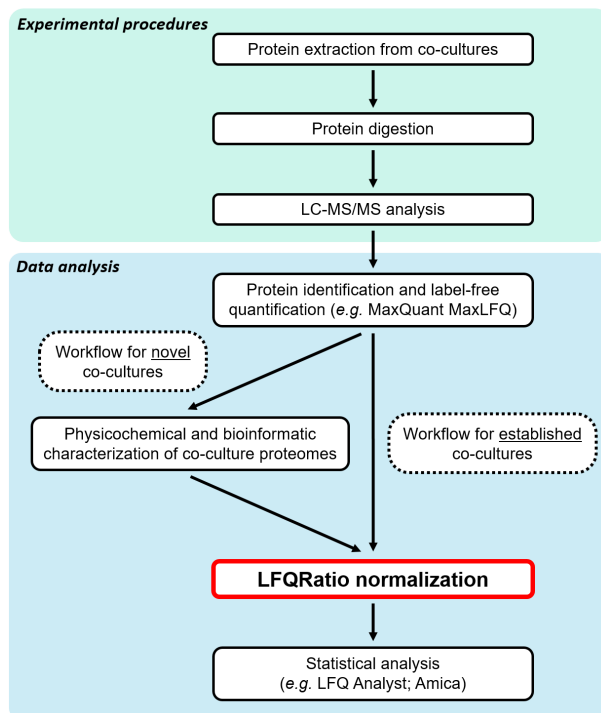
## Background

Microbial co-cultures are valuable models for studying interactions between different microbial species, revealing insights into symbiotic relationships, competition, and metabolic cooperation [1–3]. Quantitative proteomics provides a critical tool that allows for the comprehensive analysis of protein dynamics in various biological systems, including microbial co-cultures.

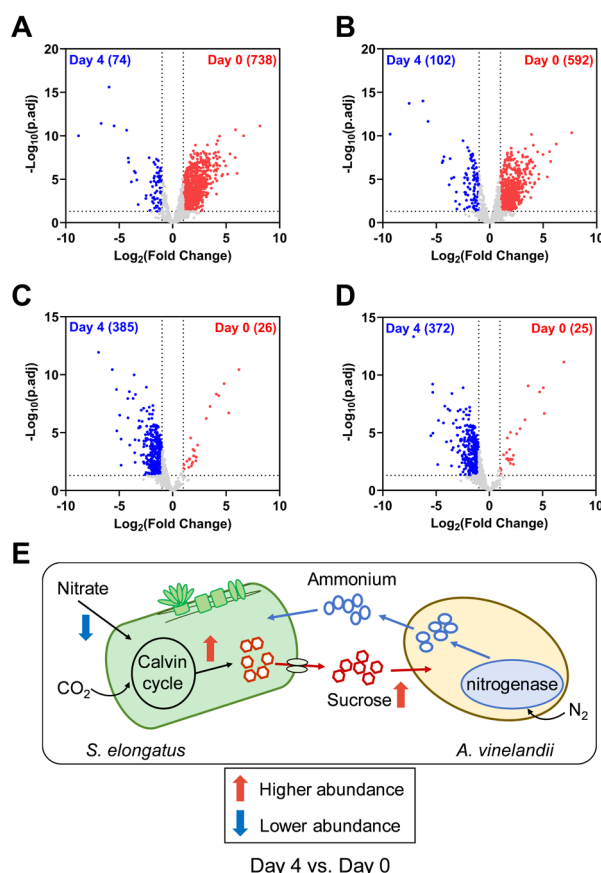
Quantitative analysis in proteomics has largely shifted to label-free methods due to the rapidly increasing sensitivity of liquid chromatography (LC) and mass spectrometry hardware and the accuracy of proteomics data analysis tools [4]. Label-free quantification (LFQ) has advantages such as being cost-saving, requiring less stringent chemicals for extraction buffers, and presenting no limitations on the number of samples compared to label-based methods [5]. However, sample complexity and variations in protein abundance can make data interpretation difficult.

Many attempts have been made to deal with systematic biases among samples and to improve the accuracy of LFQ proteomics, including the creation of new algorithms [6–8]. However, there are challenges in applying these common workflows to synthetic co-cultures, i.e., when several strain types are cultivated together. This is particularly important when comparing conditions with highly variable cell-type ratios, as this will cause large differences in protein abundance between samples [9]. Therefore, a method for interpreting the contribution of two distinct protein groups in a sample in quantitative proteomics analyses remains to be established.

To address these challenges, we started by analyzing the physicochemical properties and factors related to protein quantification between two co-cultured species. Then, we developed a new normalization method named LFQRatio, designed for the quantitative proteomic analysis of microbial co-culture systems (Figure 1). The LFQRatio normalization method is based on a more accurate LFQ intensity-based quantification approach [10, 11], which addresses the effects of changes in cell proportions between co-cultured microorganisms and facilitates comparisons of protein abundance under different conditions. This method enables more accurate identification of differentially expressed proteins, which is crucial for deciphering dynamics within microbial co-cultures (Figure 2).



**Figure 1. Label-free quantitative proteomics workflow with LFQRatio normalization for the analysis of co-cultures**



**Figure 2. Example results obtained when including LFQRatio normalization in our label-free quantitative proteomics analysis pipeline for an *S. elongatus*–*A. vinelandii* co-culture. (A–D) Volcano plots**

showing differentially enriched proteins (DEPs) between Day 0 and Day 4 of an *S. elongatus*–*A. vinelandii* co-culture, with and without LFQRatio normalization. (A) DEPs of *S. elongatus* between Day 0 and Day 4 without LFQRatio normalization. (B) DEPs of *S. elongatus* between Day 0 and Day 4 with LFQRatio normalization. (C) DEPs of *A. vinelandii* between Day 0 and Day 4 without LFQRatio normalization. (D) DEPs of *A. vinelandii* between Day 0 and Day 4 with LFQRatio normalization. Each data point represents an individual protein. Proteins significantly enriched on Day 0 and Day 4 are shown in red and blue, respectively. Differential expression analysis was performed using LFQ Analyst using the following parameters: p-value cutoff = 0.05;  $\text{Log}_2(\text{fold change})$  cutoff = 1; imputation = Perseus style. (E) Schematic showing the changes in cross-feeding dynamics between Day 4 and Day 0 in an *S. elongatus*–*A. vinelandii* co-culture uncovered using LFQRatio normalization. On Day 4, nitrate uptake transporters in *S. elongatus* were less abundant, while proteins involved in sucrose production in *S. elongatus* and sucrose import in *A. vinelandii* were enriched, compared to Day 0. These biological conclusions were able to be drawn by applying LFQRatio normalization to the LFQ intensity values acquired using MaxQuant MaxLFQ; for more details, see Shi et al. [11].

## Materials and reagents

### Biological materials

1. Bacterial monoculture 1: *Synechococcus elongatus* PCC 7942 cscB/SPS monoculture (from Abramson et al. [12]; kindly provided by Prof. Daniel Ducat). For growth monitoring information, see General notes
2. Bacterial monoculture 2: *Azotobacter vinelandii* DJ  $\Delta$ nifL monoculture (from Ortiz-Marquez et al. [13]; kindly provided by Prof. Leonardo Curatti)
3. Bacterial co-culture (combination of bacterial monocultures 1 and 2): *Synechococcus elongatus* cscB/SPS and *Azotobacter vinelandii*  $\Delta$ nifL co-culture

**Caution:** These strains are genetically modified and so need to be handled and disposed of according to the recommended safety guidelines of your institute.

### Reagents

1. Sodium dodecyl sulfate (SDS) (Sigma-Aldrich, catalog number: 71725)
2. Phosphate-buffered saline (PBS) tablets (Millipore, catalog number: 6500-OP)
3. 100× Halt protease inhibitor cocktail (Thermo Scientific, catalog number: 78446)
4. 425–600  $\mu\text{m}$  glass beads, acid-washed (Sigma-Aldrich, catalog number: G8772)
5. Urea, proteomics grade (Thermo Scientific, catalog number: 29700)
6. Trizma® base (Sigma-Aldrich, catalog number: T1503)
7. Buffers for adjusting pH (e.g., NaOH/HCl) (Sigma-Aldrich, catalog number: 06203/258148)
8. DL-Dithiothreitol (DTT) (Promega, catalog number: V3151)
9. Water, HPLC grade (Fisher, catalog number: W/0106/PB17)
10. Iodoacetamide (IAA) (Sigma-Aldrich, catalog number: 1149)
11. Calcium chloride ( $\text{CaCl}_2$ ) (Sigma-Aldrich, catalog number: C3306)
12. Trifluoroacetic acid, HPLC grade (TFA) (Sigma-Aldrich, catalog number: 15608360 or similar)
13. Sequencing-grade modified trypsin (Promega, catalog number: V5111)
14. Formic acid, LC-MS grade (Thermo Scientific, catalog number: 85178)

15. Acetonitrile (ACN), HPLC grade (Fisher, catalog number: A/0627/17)
16. Methanol (MeOH), HPLC grade (Fisher, catalog number: M/4056/17)

## Solutions

1. 1 M Tris-HCl (pH 8.5) (see Recipes)
2. 1 M DTT (see Recipes)
3. Lysis buffer (see Recipes)
4. Urea buffer (see Recipes)
5. 100 mM iodoacetamide (IAA) (see Recipes)
6. 1 mg/mL trypsin stock (see Recipes)
7. 100 mM CaCl<sub>2</sub> (see Recipes)
8. 50 mM Tris-HCl (pH 8.5)/10 mM CaCl<sub>2</sub> (see Recipes)
9. MS loading buffer (see Recipes)
10. 80% ACN (see Recipes)
11. MS mobile phase (see Recipes)

## Recipes

*Note: Storage conditions are room temperature unless otherwise stated.*

### 1. 1 M Tris-HCl (pH 8.5)

Weigh 12.114 g of Tris and dissolve in 80 mL of HPLC-grade water. Lower the pH to 8.5 with HCl and top up to 100 mL with HPLC-grade water.

### 2. 1 M DTT

Dissolve 1.542 g of DTT in 10 mL of HPLC-grade water. Aliquot and store at -20 °C or use immediately.

*Note: DTT oxidizes in air, so the solution must be prepared fresh or frozen aliquots must be used immediately and then discarded.*

### 3. Lysis buffer

Reagent	Final concentration	Quantity or Volume
SDS	2% (w/v)	400 mg
1 M Tris Tris-HCl (pH 8.5)	40 mM	0.8 mL
1 M DTT	60 mM	1.2 mL
HPLC-grade water		Up to 20 mL

### 4. Urea buffer

Reagent	Final concentration	Quantity or Volume
Urea	8 M	240 mg
1 M Tris-HCl (pH 8.5)	100 mM	50 µL
1 M DTT	5 mM	2.5 µL
HPLC-grade water		Up to 500 µL

*Note: We recommend using 1.5 mL LoBind Eppendorf tubes to prepare the solution. Always make up fresh urea solution. When in solution, urea degrades within a few hours into a compound that can alter the composition of your protein sample.*

#### 5. 100 mM iodoacetamide (IAA)

Dissolve 0.0185 g of iodoacetamide in 1 mL of HPLC-grade water.

*Note: Always prepare fresh as IAA is light-sensitive. Wrap in foil once prepared.*

#### 6. 1 mg/mL trypsin stock

Add 20  $\mu$ L of resuspension buffer (provided by the trypsin manufacturer) to one vial of lyophilized sequencing-grade modified trypsin (20  $\mu$ g) to generate the trypsin stock. Store at -80 °C for up to 1 month.

#### 7. 100 mM $\text{CaCl}_2$

Dissolve 1.47 g of  $\text{CaCl}_2$  in 100 mL of HPLC-grade water.

#### 8. 50 mM Tris-HCl (pH 8.5)/10 mM $\text{CaCl}_2$

Reagent	Final concentration	Quantity or Volume
1 M Tris-HCl (pH 8.5)	50 mM	1 mL
100 mM $\text{CaCl}_2$	10 mM	2 mL
HPLC-grade water		17 mL

#### 9. MS loading buffer

Reagent	Final concentration	Quantity or Volume
ACN	3% (v/v)	30 $\mu$ L
TFA	0.1% (v/v)	1 $\mu$ L
HPLC-grade water		969 $\mu$ L

**Caution:** Prepare in a fume cabinet.

#### 10. 80% ACN

Reagent	Final concentration	Quantity or Volume
ACN	80% (v/v)	800 mL
HPLC-grade water		200 mL

**Caution:** Prepare in a fume cabinet.

#### 11. MS mobile phase

##### Solvent A

Reagent	Final concentration	Quantity or Volume
Formic acid	0.1% (v/v)	1 mL
HPLC-grade water		999 mL

##### Solvent B

Reagent	Final concentration	Quantity or Volume
Formic acid	0.1% (v/v)	1 mL

80% ACN

999 mL

**Caution:** Prepare in a fume cabinet.

## Laboratory supplies

1. 1,000 µL pipette tips
2. 200 µL pipette tips
3. 10 µL pipette tips
4. 50 mL sterile conical tubes (Falcon, catalog number: 352070)
5. 1.5 mL protein LoBind tubes (Eppendorf, catalog number: 0030108116)
6. 2D Clean-Up kit (Cytiva, formerly GE Healthcare Life Sciences, catalog number: 80-6484-51)
7. Pierce™ C18 spin columns (Thermo Scientific, catalog number: 89870)
8. 0.3 µL PP snap ring HPLC vials (VWR, catalog number: 548-0120A)
9. ND11 open-top PE HPLC vial snap caps (VWR, catalog number: 548-3203A)

## Equipment

1. Pipettes (p1000, p200, p10, p2.5) (Eppendorf)
2. Spectrophotometer (Jenway, model: 6305)
3. Temperature-controlled centrifuge (e.g., Thermo Scientific, model: SL 1ER)
4. Vortex with attachment for multiple microfuge tubes (Scientific Industries, model: Vortex-Genie 2)
5. Heated water bath (e.g., Grant, model: JB Nova)
6. Ultrasonic water bath (Fisherbrand, model: FB15051)
7. Nanodrop spectrophotometer (Thermo Scientific, model: Nanodrop 2000)
8. Incubator capable of maintaining 30–37 °C (e.g., Thermo Scientific, model: MaxQ 6000)
9. Vacuum concentrator (Eppendorf, model: Concentrator Plus)
10. Nano-flow liquid chromatography (Thermo Fisher Scientific, model: U3000 RSLCnano)
11. Hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific, model: Q Exactive HF)
12. Easy-Spray C18 column, 75 µm × 50 cm, Pepmap RSLC, 2 µm, 100 Å (Thermo Scientific, catalog number: E5903)

## Software and datasets

1. Uniprot (<https://www.uniprot.org/>, 02/03/2020)
2. MaxQuant (2.0.3.0)
3. RStudio (4.2.0)
4. LFQ-Analyst (<https://analyst-suite.monash-proteomics.cloud.edu.au/apps/lfq-analyst/>, 08/09/2023)
5. All code has been deposited to GitHub: <https://github.com/Mengxun-Shi/bio-protocol> (01/13/2025)

## Procedure

### A. Protein extraction

1. Grow cell cultures to a growth phase appropriate for your experiment and measure the optical density at the wavelength suitable for your culture. Centrifuge 20 mL of cells at  $4,000 \times g$  for 10 min at  $4^\circ\text{C}$  in 50 mL centrifuge tubes.

*Note: We grew our co-cultures in 500 mL flasks over an 8-day period, sampling on Days 0, 4, and 8 [11] (see General note 1). However, we encourage the reader to select their harvest points based on their own biological questions and comparisons.*

2. Wash the pellets by resuspending in 10 mL of PBS buffer and repeating the centrifugation. Pellets can be stored at  $-20^\circ\text{C}$  for up to 1 month before processing.

3. Resuspend the pellet in 500  $\mu\text{L}$  of lysis buffer per 1 OD unit (e.g., if  $\text{OD} = 2$ , add 1 mL of lysis buffer) and add the appropriate amount of  $100\times$  Halt protease inhibitor cocktail (e.g., if the total volume of the pellet + lysis buffer is  $\sim 700\ \mu\text{L}$ , add 7  $\mu\text{L}$  of protease inhibitor).

4. Freeze the cells at  $-80^\circ\text{C}$  for  $>16$  h and quickly thaw them in a water bath at  $37^\circ\text{C}$  for 3–5 min to allow partial cell breakage. Immediately transfer to ice after thawing.

*Note: The frozen pellets can remain at  $-80^\circ\text{C}$  in lysis buffer for  $\sim 1$  month.*

5. Label and pre-cool 1.5 mL LoBind microcentrifuge tubes on ice. Transfer the lysed cell sample to the LoBind tubes. For the remainder of the procedure, keep the samples on ice.

6. Add approximately 500  $\mu\text{L}$  of 425–600  $\mu\text{m}$  acid-washed glass beads to the samples, leaving 2–3 mm of cell suspension above the level of settled beads.

7. Break the cells by vigorous vortex mixing of the tube 20 times in cycles of vortexing for 30 s and cooling on ice for 30 s using a vortex with a multi-tube attachment.

8. Incubate samples at  $95^\circ\text{C}$  for 5 min using a dry heat block.

9. Collect the supernatant in a fresh 1.5 mL LoBind tube following centrifugation at  $15,000 \times g$  for 10 min at  $4^\circ\text{C}$ . The crude protein sample can be stored at  $-20^\circ\text{C}$  for  $\sim 1$  month.

*Note: At this point, it would be a good idea to quantify your crude proteins. Our crude protein concentration was typically 2–3 mg/mL. This can be done using, for example, a detergent-compatible Bradford reagent (Pierce, catalog number: 23246). It is also advisable to run the protein samples on an SDS-PAGE gel.*

10. Purify the crude samples using a 2D Clean-Up kit following the manufacturer's instructions to remove interfering substances and, if required, to concentrate the proteins. The resuspension buffer will be the urea buffer (see Section B).

**Caution:** Note the hazard statements in the 2D Clean-Up kit for the precipitant and wash buffer, which include “causes severe burns and eye damage” and “suspected to cause cancer.”

*Note: At the wash step, the protein pellet can be stored in wash buffer and wash reagent for up to 1 week at  $-20^\circ\text{C}$ . However, once the pellet has been dried for  $<5$  min, it is advisable to immediately proceed to the digestion steps and resuspend the pellet in urea buffer.*

### B. Protein digestion

1. Dissolve protein pellet in 50  $\mu\text{L}$  of urea buffer. Sonicate the sample in an ultrasonic water bath to solubilize the proteins; the suspension should be transparent.



*Note: If the sonication step takes longer than 10 min, cool the samples on ice for 5 min and continue the sonication until the samples are transparent.*

2. Take 1  $\mu\text{L}$  of resuspended protein pellet and dilute 1:10 by adding 9  $\mu\text{L}$  of  $\text{H}_2\text{O}$ . Estimate protein concentration using the Nanodrop: Use 1  $\mu\text{L}$  of 1:10-diluted urea buffer as a blank. Add 1–2  $\mu\text{L}$  diluted protein sample to the Nanodrop to estimate protein concentration at 280 nm.

3. Reduction step: Transfer 50  $\mu\text{g}$  of protein to a fresh 1.5 mL LoBind tube. Dilute to 10  $\mu\text{L}$  with urea buffer. Incubate at 37 °C in a water bath for 30 min.

*Note: DTT, the reducing agent, is included in the urea buffer.*

4. Alkylation step: Add 1.5  $\mu\text{L}$  of 100 mM IAA to the protein solution and incubate in the dark at room temperature for 30 min.

5. Digestion step: Add 1  $\mu\text{L}$  of 1 mg/mL sequencing-grade modified trypsin stock to the protein sample and then add 67.5  $\mu\text{L}$  of 50 mM Tris-HCl (pH 8.5)/10 mM  $\text{CaCl}_2$  to a total volume of 80  $\mu\text{L}$ . Digest overnight at 37 °C.

*Note: The protease:protein ratio recommended by the trypsin manufacturer is 1:100 to 1:20 (w/w); here, we used 1:50.*

6. Terminate the digestion: Add 0.8  $\mu\text{L}$  of formic acid to a final concentration of 1% (recommend 0.1–1% v/v). Once added, it can be stored at -20 °C for up to 48 h.

7. Use C18 Spin columns to purify the digested peptides and remove contaminants following the manufacturer's instructions. Purified samples should be dried using a vacuum concentrator and stored at -20 °C (days/weeks) or -80 °C (months) until mass spectrometry analysis.

*Note: The binding capacity of the C18 Spin columns is 30  $\mu\text{g}$ . By overloading the column with 50  $\mu\text{g}$  of peptides, 30  $\mu\text{g}$  will reliably be captured and eluted without peptide loss.*

## C. Mass spectrometry

1. Dissolve dried peptide pellets in 30  $\mu\text{L}$  of loading buffer for a peptide concentration of 1  $\mu\text{g}/\mu\text{L}$ . Sonicate samples in a water bath for 3 min at room temperature to suspend the samples.

2. Centrifuge the sample at  $15,000 \times g$  for 2 min. Take 5  $\mu\text{L}$  from the top of the centrifuged sample and transfer to an HPLC vial. Top up with 15  $\mu\text{L}$  of loading buffer (20  $\mu\text{L}$  total volume; 250 ng/ $\mu\text{L}$  peptide concentration), ensuring not to introduce bubbles to the vial.

3. Inject 500 ng of purified peptides (2  $\mu\text{L}$ ) into the LC-MS system with Easy-Spray C18 column. Perform LC-MS/MS by nanoflow liquid chromatography coupled to a hybrid quadrupole-orbitrap mass spectrometer or equivalent LC-MS/MS equipment. Here, peptides were separated by reverse-phase high-performance liquid chromatography using an Easy-Spray C18 column (75  $\mu\text{m} \times 50 \text{ cm}$ ) at a flow rate of 300 nL/min and column temperature of 40 °C. Two mobile phases were used: mobile phase A was composed of water with 0.1% formic acid v/v, and mobile phase B was composed of 80% acetonitrile in water, 0.1% v/v formic acid. Both were mixed over time for gradient elution. A 2-step gradient was employed: from 3% B to 10% B over 5 min, then from 10% B to 50% B (0.1% formic acid in 80% acetonitrile) to 75 min.

4. The mass spectrometer was programmed for data-dependent acquisition with 10 product ion scans (resolution 30,000, automatic gain control  $1 \times 10^5$ , maximum injection time 60 ms, isolation window 2.0 m/z, fixed first mass 100 m/z, normalized collision energy 28, and intensity threshold  $3.3 \times 10^4$ ) per full MS scan (resolution 120,000, automatic gain control  $1 \times 10^6$ , maximum injection time 60 ms, scan range 375–1,500 m/z) with a 35 s dynamic exclusion time. Make sure to include both biological ( $\geq 3$ ) and technical ( $> 2$ ) replicates and ensure the appropriate control samples are included for meaningful sample comparison(s) to ensure the biological question is addressed.

## Data analysis

### A. Protein identification and quantification

1. Create a reference database (FASTA format) using all protein sequences of both strains; in this case, *S. elongatus* PCC 7942 and *A. vinelandii* DJ appended with the amino acids sequences of sucrose permease (CscB) from *Escherichia coli* and sucrose phosphate synthase (SPS) from *Synechocystis* sp. PCC6803 from Uniprot (Dataset S3).

*Note: This can be done with a simple notepad program, saving the file name with the appendage .fasta.*

2. Load the raw MS data files and database to MaxQuant (2.0.3.0) [7].
3. Select modifications: *Oxidation (M)* and *Acetyl (Protein N-term)* for variable modification, and *Carbamidomethyl (C)* for fixed modification.
4. Select *Trypsin/P* for enzymatic digestion.
5. Filter peptide-spectrum matches and protein identifications using a target-decoy approach at a false discovery rate (FDR) of 1%.
6. Select *LFQ* and intensity-based absolute quantification (*iBAQ*) options for quantitative analysis. Other parameters are shown in the supplementary files (Table S1).
7. If your synthetic co-culture is a novel combination of two bacteria, and this is the first proteomics experiment conducted for this consortium, continue to Section B. Otherwise, proceed to Section C.

### B. Assessment of factors relating to protein identification for first-time proteomic analyses of new synthetic co-cultures

The physicochemical characteristics of proteins, such as the range of pI [14], MW [15], and hydrophobicity [16], as well as the dynamic range distribution of protein abundances within the proteome [17], can affect factors such as protein extraction efficiency and ionization efficiency in the mass spec, and hence how many peptides are detected. Similarly, a much larger proteome or highly abundant protein could mean a bias toward identifying peptides from a specific organism [18]. Finally, shared peptides between the strains will conflict with quantitative data and so should be taken into account [19]. We therefore advise researchers to identify variations in these factors between co-culture species, which could affect quantitative accuracy and therefore the biological interpretation of LFQ proteomics data.

1. Preliminary physicochemical characteristics analysis
  - a. Assess the theoretical pI ranges of the proteins in *S. elongatus* and *A. vinelandii* using R script (File S1).
  - b. Assess the MW of the proteins in *S. elongatus* and *A. vinelandii* using R script (File S2).
  - c. Assess the hydrophobicity of proteins in the *S. elongatus* and *A. vinelandii* databases via the grand average of hydropathy (GRAVY) scores using R script (File S3).
  - d. Dynamic range: Quantify the abundance of detected *S. elongatus* and *A. vinelandii* proteins by absolute quantification (*iBAQ*). Use the values from MaxQuant outputs-ProteinGroups.txt file, calculating log<sub>10</sub> *iBAQ* intensity to show the dynamic range of both strains. Plot log<sub>10</sub> *iBAQ* intensity against protein number.

*Note: The dynamic range is the range of MS1 peak intensities over which peptides can be detected. A wide dynamic range or a protein that is extremely abundant in one strain may affect the quantification of the co-culture proteome.*

Therefore, it is critical to examine the dynamic range of proteins for each strain in the microbial community before further analysis [11].

## 2. Proteome size analysis

a. Run MaxQuant multiple times against *S. elongatus* and *A. vinelandii* individual and merged databases. For example, we run MaxQuant four times: i) *S. elongatus* monoculture against the *S. elongatus* database (Dataset S1); ii) *S. elongatus* monoculture against the merged database (Dataset S3); iii) *A. vinelandii* monoculture against the *A. vinelandii* database (Dataset S2); and iv) *A. vinelandii* monoculture against the merged database (Dataset S3).

b. Compare the identified protein numbers of *S. elongatus* and *A. vinelandii* taken from proteinGroups.txt files of each run. Filter for proteins with two or more unique peptides.

*Note: We expect the same strain to have the same/similar number of proteins identified in both individual and combined database runs, which means that database size does not affect protein identification in this case. However, this should be considered in co-culture proteomics workflows, particularly where proteome sizes of co-culture members vary more widely.*

## 3. Shared peptides analysis

a. Calculate theoretical shared peptides using R script (File S4) of tryptic peptides of the specified size (8–25 amino acids).

b. Analyze measured shared peptides by comparison of the resulting peptide sequences of each strain from MaxQuant outputs–Peptides.txt file.

## C. LFQRatio normalization and proteomic analysis

1. Open the ProteinGroup.txt file output from MaxQuant and select identified proteins with two or more peptides.  
2. Normalize LFQ intensity data using the LFQRatio normalization method as shown in the following equation, i.e., for each individual protein detected, its LFQ intensity is divided by the sum of all protein LFQ intensities for its respective strain:

$$\text{Normalized LFQ intensity} = \frac{\text{LFQ intensity of one protein}}{\text{total LFQ intensities of the strain}}$$

3. Use the normalized proteomics data to analyze differentially expressed proteins using an LFQ proteomics analysis platform such as LFQ-Analyst [20]. The following steps will be for using LFQ-Analyst; however, other platforms can be used (see General notes).

4. Load two input files: the LFQRatio-normalized proteinGroups.txt file and an experimental design file (an exemplary experimental design file can be downloaded from the LFQ-Analyst website).

5. In the *Advanced Options*, the adjusted p-value cutoff and Log2 fold change cutoff can be selected; we applied the default options of 0.05 and 1, respectively.

6. The default imputation type (Perseus type) can be applied; however, this is not necessarily recommended (see General notes and troubleshooting)

7. The type of FDR correction can be selected; we applied Benjamini–Hochberg for our experiment.

8. A list of differentially expressed proteins for each pair-wise comparison will be generated, as well as quality control plots, volcano plots, and expression heatmaps. Example volcano plots showing differential expression data generated with and without application of LFQRatio are shown in Figure 2.

9. Other parameters are shown in Table 1.

**Table 1. Parameter settings of LFQ-Analyst used for differentially expressed proteins analysis**

Parameters	Settings
Adjusted p-value cutoff	0.05
Log2 fold change cutoff	1
Imputation type	Perseus type* (see General notes)

## Validation of protocol

This protocol or parts of it has been used and validated in the following research article(s):

- Shi et al. [11] LFQRatio: A normalization method to decipher quantitative proteome changes in microbial coculture systems. *Journal of Proteome Research* (Figure 9).
- Kratzl et al. [21] *Pseudomonas putida* as saviour for troubled *Synechococcus elongatus* in a synthetic co-culture–interaction studies based on a multi-OMICs approach. *Communications Biology* (Figure 4).

## General notes and troubleshooting

1. Growth details for *S. elongatus* and *A. vinelandii*: In our model system, we performed a time-course experiment to examine the growth and physiology of our *S. elongatus* and *A. vinelandii* co-culture over 8 days, harvesting samples for proteomics analysis on Days 0, 4, and 8. We monitored the growth both by optical density (OD) at 750 nm, which did not distinguish between strains but showed overall growth of the system, and by i) flow cytometry for *S. elongatus* and ii) calculating colony forming units for *A. vinelandii*. For more information on media, starting cell ratios, growth monitoring, and maintenance of these strains, we direct the reader to our paper [11].

2. Alternative in-browser programs for analyzing differentially expressed proteins: LFQ-Analyst was suggested here as an accessible, in-browser option for performing differential expression analysis of LFQ proteomics data [20]. Since performing our initial analysis [11], a developer version of LFQ-Analyst has been released (<https://analyst-suites.org/apps/lfq-analyst-dev/>), which allows for more control over parameter settings for differential expression analysis, as well as presence/absence analysis. Another powerful in-browser tool that we recommend is amica, which provides the user with an in-depth quality control analysis and parameter selection [22].

3. Dealing with missing values: Missing values in LFQ proteomics datasets are to be expected [23]. They can be broadly categorized as either “missing not at random” (MNAR), where the protein is either not expressed or has an abundance lower than the mass spectrometer’s limit of detection (i.e., biological reasons), or “missing at random” (MAR), where the protein is not identified for reasons related to the protocol and instrumentation (for more detail, see Jin et al. [24]). So as not to lose valuable data, we can apply various imputation algorithms to fill in these missing values.

We observed that missing values were particularly prevalent in mixed co-culture samples compared to their mono-culture counterparts, and the extent of this will likely differ between co-culture types. In our example application of LFQRatio, we applied the default imputation settings (Perseus type) for our DEP analysis [11]; we also applied this same imputation method in another study [21] in which we achieved a good correlation of the proteomics data with transcriptomics data, suggesting that this imputation method was acceptable. However, we advise researchers to consider which imputation method is most appropriate for their own dataset; for more information and guidance, we refer the reader to this paper: <https://www.nature.com/articles/s41598-021-81279-4>.

4. LFQ-Analyst details: LFQ-Analyst analyzes differentially expressed proteins according to the following steps: Filter out contaminant proteins, reverse sequences, and proteins identified “only by site;” remove proteins that have been only identified by a single peptide and proteins not identified/quantified consistently in the same condition; transform the LFQ data to log2 scale; group samples by conditions and missing values imputed using the selected imputation type [here, we used Perseus-type imputation, a MNAR method that takes random draws from a left-shifted Gaussian distribution of 1.8 StDev (standard deviation) apart with a width of 0.3]; use protein-wise linear models combined with empirical Bayes statistics for the differential expression analyses; and use the limma package from R Bioconductor to generate a list of differentially expressed proteins for each pair-wise comparison [20].

## Supplementary information

The following supporting information can be downloaded [here](#):

1. Table S1. Parameter settings of MaxQuant software used for protein identification and quantification.
2. File S1. R script for analyzing the pI of *S. elongatus* or *A. vinelandii* proteome.
3. File S2. R script for analyzing the MW of *S. elongatus* or *A. vinelandii* proteome.
4. File S3. R script for analyzing the hydrophobicity of *S. elongatus* or *A. vinelandii* proteome.
5. File S4. R script for analyzing shared peptides between *S. elongatus* and *A. vinelandii*.

*Note: These scripts can be used for other organisms by changing the “sequence.fasta” file to the corresponding FASTA files.*

6. Dataset S1. Protein sequences of *S. elongatus* cscB/SPS.
7. Dataset S2. Protein sequences of *A. vinelandii*  $\Delta$ nifL.
8. Dataset S3. Protein sequences of *S. elongatus* cscB/SPS and *A. vinelandii*  $\Delta$ nifL co-culture (merged database).

## Acknowledgments

The authors acknowledge the Chinese Scholarship Council for funding a PhD studentship (201808370169) and EPSRC funding Engineering Microbial Consortia for Industry (EP/S020705/1). The QExactive HF orbitrap mass spectrometer was funded by BBSRC UK (award no. BB/M012166/1). This protocol was used in [11,21].

## Competing interests

The authors declare no conflicts of interest.

Received: January 13, 2025; Accepted: March 23, 2025; Available online: April 15, 2025; Published: May 05, 2025

## References

1. Minty, J. J., Singer, M. E., Scholz, S. A., Bae, C. H., Ahn, J. H., Foster, C. E., Liao, J. C. and Lin, X. N. (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proc Natl Acad Sci USA*. 110(36): 14592–14597. <https://doi.org/10.1073/pnas.1218447110>
2. Pandhal, J. and Noirel, J. (2014). Synthetic microbial ecosystems for biotechnology. *Biotechnol Lett*. 36(6): 1141–1151. <https://doi.org/10.1007/s10529-014-1480-y>
3. Christie-Oleza, J. A., Sousoni, D., Lloyd, M., Armengaud, J. and Scanlan, D. J. (2017). Nutrient recycling facilitates long-term stability of marine microbial phototroph–heterotroph interactions. *Nat Microbiol*. 2(9): e100. <https://doi.org/10.1038/nmicrobiol.2017.100>
4. Ma, Q., Zhou, J., Zhang, W., Meng, X., Sun, J. and Yuan, Y. J. (2011). Integrated Proteomic and Metabolomic Analysis of an Artificial Microbial Community for Two-Step Production of Vitamin C. *PLoS One*. 6(10): e26108. <https://doi.org/10.1371/journal.pone.0026108>
5. Chignell, J. F., Park, S., Lacerda, C. M. R., De Long, S. K. and Reardon, K. F. (2018). Label-Free Proteomics of a Defined, Binary Co-culture Reveals Diversity of Competitive Responses Between Members of a Model Soil Microbial System. *Microb Ecol*. 75(3): 701–719. <https://doi.org/10.1007/s00248-017-1072-1>
6. Cappadona, S., Baker, P. R., Cutillas, P. R., Heck, A. J. R. and van Breukelen, B. (2012). Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino Acids*. 43(3): 1087–1108. <https://doi.org/10.1007/s00726-012-1289-8>
7. Cox, J., Hein, M. Y., Lubner, C. A., Paron, I., Nagaraj, N. and Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol Cell Proteomics*. 13(9): 2513–2526. <https://doi.org/10.1074/mcp.m113.031591>
8. Ammar, C., Schessner, J. P., Willems, S., Michaelis, A. C. and Mann, M. (2023). Accurate Label-Free Quantification by directLFQ to Compare Unlimited Numbers of Proteomes. *Mol Cell Proteomics*. 22(7): 100581. <https://doi.org/10.1016/j.mcpro.2023.100581>
9. Kleiner, M. (2019). Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *mSystems*. 4(3): e00115–19. <https://doi.org/10.1128/msystems.00115-19>
10. Goeminne, L. J., Gevaert, K. and Clement, L. (2018). Experimental design and data-analysis in label-free quantitative LC/MS proteomics: A tutorial with MSqRob. *J Proteomics*. 171: 23–36. <https://doi.org/10.1016/j.jprot.2017.04.004>
11. Shi, M., Evans, C. A., McQuillan, J. L., Noirel, J. and Pandhal, J. (2024). LFQRatio: A Normalization Method to Decipher Quantitative Proteome Changes in Microbial Coculture Systems. *J Proteome Res*. 23(3): 999–1013. <https://doi.org/10.1021/acs.jproteome.3c00714>



12. Abramson, B. W., Kachel, B., Kramer, D. M. and Ducat, D. C. (2016). Increased Photochemical Efficiency in Cyanobacteria via an Engineered Sucrose Sink. *Plant Cell Physiol.* 57(12): 2451–2460. <https://doi.org/10.1093/pcp/pcw169>
13. Ortiz-Marquez, J. C. F., Do Nascimento, M., Dublan, M. d. I. A. and Curatti, L. (2012). Association with an Ammonium-Excreting Bacterium Allows Diazotrophic Culture of Oil-Rich Eukaryotic Microalgae. *Appl Environ Microbiol.* 78(7): 2345–2352. <https://doi.org/10.1128/aem.06260-11>
14. Kozłowski, L. P. (2016). Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res.* 45: D1112–D1116. <https://doi.org/10.1093/nar/gkw978>
15. Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W. and Smith, R. D. (2012). Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem Soc Rev.* 41(10): 3912–3928. <https://doi.org/10.1039/c2cs15331a>
16. Warwood, S., Byron, A., Humphries, M. J. and Knight, D. (2013). The effect of peptide adsorption on signal linearity and a simple approach to improve reliability of quantification. *J Proteomics.* 85: 160–164. <https://doi.org/10.1016/j.jprot.2013.04.034>
17. Zubarev, R. A. (2013). The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics.* 13(5): 723–726. <https://doi.org/10.1002/pmic.201200451>
18. Kumar, D., Yadav, A. K. and Dash, D. (2017). Choosing an Optimal Database for Protein Identification from Tandem Mass Spectrometry Data. In: Keerthikumar, S. and Mathivanan, S. (eds). *Proteome Bioinformatics.* 17–29. [https://doi.org/10.1007/978-1-4939-6740-7\\_3](https://doi.org/10.1007/978-1-4939-6740-7_3)
19. Helliwell, K. E., Pandhal, J., Cooper, M. B., Longworth, J., Kudahl, U. J., Russo, D. A., Tomsett, E. V., Bunbury, F., Salmon, D. L., Smirnoff, N., et al. (2017). Quantitative proteomics of a B<sub>12</sub>-dependent alga grown in coculture with bacteria reveals metabolic tradeoffs required for mutualism. *New Phytol.* 217(2): 599–612. <https://doi.org/10.1111/nph.14832>
20. Shah, A. D., Goode, R. J. A., Huang, C., Powell, D. R. and Schittenhelm, R. B. (2019). LFQ-Analyst: An Easy-To-Use Interactive Web Platform To Analyze and Visualize Label-Free Proteomics Data Preprocessed with MaxQuant. *J Proteome Res.* 19(1): 204–211. <https://doi.org/10.1021/acs.jproteome.9b00496>
21. Kratzl, F., Urban, M., Pandhal, J., Shi, M., Meng, C., Kleigrew, K., Kremling, A. and Pflüger-Grau, K. (2024). *Pseudomonas putida* as saviour for troubled *Synechococcus elongatus* in a synthetic co-culture – interaction studies based on a multi-OMICs approach. *Commun Biol.* 7(1): 452. <https://doi.org/10.1038/s42003-024-06098-5>
22. Didusch, S., Madern, M., Hartl, M. and Baccarini, M. (2022). amica: an interactive and user-friendly web-platform for the analysis of proteomics data. *BMC Genomics.* 23: 817. <https://doi.org/10.1101/2021.11.23.466958>
23. Liu, M. and Dongre, A. (2020). Proper imputation of missing values in proteomics datasets for differential expression analysis. *Briefings Bioinf.* 22(3): 1–17. <https://doi.org/10.1093/bib/bbaa112>
24. Jin, L., Bi, Y., Hu, C., Qu, J., Shen, S., Wang, X. and Tian, Y. (2021). A comparative study of evaluating missing value imputation methods in label-free proteomics. *Sci Rep.* 11(1): 1760. <https://doi.org/10.1038/s41598-021-81279-4>