Comprehensive analysis of the microbial consortium in the culture of flagellate *Monocercomonoides exilis*

Alejandro Jiménez-González[1,2], Sebastian Cristian Treitli[1], Priscila Peña-Diaz[1], Anna Janovská[3], Vladimír Beneš[4], Petr Žáček[3], Vladimír Hampl[1]

Affiliations:

[1] Charles University, Faculty of Science, Department of Parasitology, BIOCEV, Vestec 252 50, Czech Republic

[2] Uppsala Biomedicine Centre, Department of Cell and Molecular Biology, Molecular Evolution program, Uppsala University, Uppsala, Sweden

[3] Charles University, Faculty of Science, BIOCEV, Vestec 252 50, Czech Republic

[4] Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.

Corresponding author: Vladimir Hampl, vlada@natur.cuni.cz

Abstract

Monocercomonoides exilis is the only known amitochondriate eukaryote, making it an excellent model for studying the implications of mitochondrial reduction from a cellular and evolutionary point of view. Although *M. exilis* is an endobiotic heterotroph, it can grow *in vitro*, albeit with an uncharacterised and complex prokaryotic community. All attempts to grow *M. exilis* axenically have been unsuccessful. Here, we use metagenomic sequencing at different time points during culture growth to describe the composition and dynamics of this community. We assembled genomes of 24 from at least the 30 different bacterial species within. Based on DNA read abundances, *M. exilis* represent less than 1.5%, and the representation of dominant bacterial members changes over time. Genome-scale metabolic reconstruction, differential expression analysis and measurements of metabolites in the media showed that the community depends on organic carbon oxidation, fermentation, and hydrogen production without methanogenesis. This is consistent with the rapid decline of amino acids, nucleotides, glyceraldehyde, lactate, fatty acids, and alcohols in the media. The community recycles nitrogen compounds via nitrite ammonification, particularly through the nrfADH system, yet has a limited capacity to fix nitrogen gas. With the senescence of the culture, we observe changes in the expression of several metabolic pathways in *M. exilis*, particularly those adapting to starvation, and those potentially leading to encystation. We do not reveal any clear metabolic link to explain the dependence of *M. exilis* on prokaryotes.

Introduction

The gut microbiota refers to the community of microorganisms inhabiting the gastrointestinal system. Several studies have proved the influence of this community on human health [1–3], but these have mainly focused on the prokaryotic component since this represents the majority of cells and biomass. Yet, the animal gut microbiota also comprises viruses and eukaryotes, including unicellular protists. The effect of protists on the function and processes in the intestine, as well as their metabolic interactions with prokaryotes, are essentially unknown, except for some medically and veterinary important parasites, e.g., *Entamoeba histolytica, Giardia intestinalis, Cryptosporidium* spp., *Spironucleus salmonicida, Blastocystis* spp. [4, 5]. However, these represent only a fraction of the intestinal protist diversity and their presence often leads to pathological states that significantly depart from the conditions found in the healthy gut.

Many intestinal protists can be successfully maintained in xenic cultures using a medium such as TYSGM-9 [6, 7] and for many, this is the only feasible method of *in vitro* maintenance, given that the process of axenisation, i.e., removal of other unwanted organisms, has not been achieved. Xenic cultivation depends on establishing a stable community in culture, typically derived from the original sample, i.e., gut content or stool, which can be maintained for years by regular transfers. Whereas the community may be derived from the sampled environment, its eventual composition will most probably reduce as it adapts to the culture conditions (nutrient source, oxygen concentration, etc.) and thus will only distantly resemble the situation in the host gut. Still, the culture represents a simple and tractable model system from which information about these complex communities may be obtained. To our knowledge, such attempts are rare [8].

In the present study, we attempt to characterise the culture community derived from the stool of *Chinchilla lanigera*. It contains a single eukaryote, the commensal bacterivorous

flagellate *Monocercomonoides exilis*, for which a quality genome draft is available and functionally annotated [9–11]. This species belongs to Oxymonadida (Preaxostyla, Metamonada), known as inhabitants of animal intestines [12, 13]. The whole group, including *M. exilis*, has lost mitochondria, which is unique among eukaryotes and raises interest in their biochemistry and physiology [11, 14]. Light and electron microscopy has demonstrated that species of *Monocercomonoides* actively feed on prokaryotes, hence acting as predators within the intestinal microbiota [12]. The effect and scale of this grazing on the prokaryotic community and the level of prey selectivity are unknown. Specific roles and interactions with the prokaryotic community have been partially elucidated for other oxymonad species, such as *Streblomastix strix*, a protist that takes part in the complex community of the hindgut of termites and is involved in cellulose digestion [15].

We used a multi-omic approach to obtain a first insight into the processes of this multipartite culture community over five days of growth. Namely, we determined the species composition and changes in species abundances through time, predicted the metabolism pathways, followed gene expression changes of the most abundant members as the culture aged, and finally, measured the composition and changes of selected metabolites in the media. With all these data in hand, we attempted to reconstruct the major community biochemical processes and searched for potential interactions between its members.

Material and methods

Cell culturing

700 mL of TYSGM-9 medium in 2x 1 L glass bottles were bacterised with 500 µL of *Citrobacter freundii* culture. After inoculation, the flasks were closed and allowed to grow standing at 37°C. The next day, the two flasks were merged in a 2 L bottle and mixed thoroughly. 45 mL of culture were removed and used to isolate DNA, RNA and metabolomic

analyses of day 0. The rest of the medium was inoculated with approx. 125 mL of well-grown *M. exilis* culture ($2.5x10^5$ cells.mL^{-1}). After inoculation, the media was thoroughly mixed and divided into 50-mL tubes containing 45 mL of media, closed, and incubated at 37°C. Each day, three tubes were removed, one for each replicate. From each tube, 20 mL of culture were used for DNA isolation, 20 mL for RNA isolation and 5 mL for metabolomics. Before each isolation, *M. exilis* cells were counted using a Neubauer cytometer.

DNA isolation

All DNA isolations were performed using the Dneasy Blood and tissue Kit (Qiagen). Each culture was centrifuged at 1500 x*g* for 10 min at 4°C. The supernatant was separated into a fresh 50 mL tube and centrifuged again at 6000 x*g* for 10 min at 4°C. The pellet from the first centrifugation was resuspended in 200 µL of PBS and DNA was isolated using the cultured cells protocol of the Dneasy Blood and tissue Kit (Qiagen). The second pellet, obtained after centrifugation at 6000 x*g*, was resuspended in 200 µL of PBS and DNA was isolated using the gram-positive protocol of the Dneasy Blood and tissue Kit (Qiagen). For each isolation, the DNA was eluted in 100 µL of elution buffer. The DNA from the two pellets was merged into one DNA sample. The quality of the DNA was estimated using Nanodrop and the concentration was measured using the QuantiFluor ONE dsDNA System (Promega).

RNA isolation

Each culture was centrifuged at 1500 x*g* for 10 min at 4°C. The supernatant was separated into a fresh 50 mL tube and centrifuged again at 6000 x*g* for 10 min at 4°C. Both pellets were resuspended together in 1 mL Tri-Reagent (Sigma-Aldrich) and the total RNA was isolated according to the manufacturer's procedure. The obtained RNA pellet was resuspended in 60 µL of RNA-grade H_2O and heated at 37°C for 10 min to allow full resuspension.

The total RNA was further DNAse-treated. For this purpose, we added 6 µL of rDNAse buffer and 0.6 µL rDNAse (Macherey-Nagel) per sample and incubated at 37°C for 10 min.

Samples were subsequently re-isolated with Tri-Reagent (Sigma-Aldrich) as before and finally resuspended in 30 µL of RNA-grade H_2O. RNA concentration was measured using Nanodrop.

Identification of compounds within the growth medium

For metabolomics, cells were pelleted as described for DNA and RNA isolation (see above). The supernatant obtained in the second centrifugation was filter-sterilized and stored at -80ºC. Metabolites released to the media were determined using a non-targeted metabolomics approach. To capture both volatiles and non-volatiles, a combination of two techniques was employed – Liquid Chromatography analysis in connection with Orbitrap Fusion mass spectrometer (LC-MS/MS; Orbitrap Fusion, Q-OT-qIT, Thermo Scientific) and two-dimensional comprehensive gas chromatography in connection with mass spectrometer (GCxGC-MS; Pegasus 4D, Leco Corp.).

For LC-MS/MS, media samples harvested in triplicates were precipitated with acetonitrile in a ratio of 4:1 (ACN:media) and centrifugated for 20 min at 16000 xg at 5ºC. 100 µL of each supernatant was collected, fully evaporated and dissolved in 50 µL 10% (v/v) ACN. 15 µL were injected into a ProntoSIL column, 150 x 3.0 mm, 3 µm (Bischoff Chromatography). Compounds were eluted under the following conditions: 0 - 3 min 100% A, 3 - 18 min linear gradient to 100% B, 18 - 20 min 100% B using mobile phases. A: 2% (v/v) ACN, 10 mM $HCOONH_4$, 0.1% (v/v?) FA; B: 99% (v/v) ACN, 0.1% (v/v) FA with a flow of 0.4 $mL.min^{-1}$. The separated compounds were ionised in an electrospray ion source in positive polarity mode. Master scans of precursors from 80 to 1000 m/z were performed in Orbitrap at 120 K resolution with an intensity threshold filter 2.0 x 10^4. Tandem MS was performed by isolation in the quadrupole, HCD fragmentation with a stepped collision energy of 15, 30, and 45% and an isolation window of 1.6 m/z. The MS2 ions were detected in Orbitrap with a set resolution of 30k and a max injection time of 300 ms. The dynamic exclusion duration occurred every 20 s with a 10 ppm tolerance around the selected precursor. The data processing, including

chromatographic peaks alignment, was performed using Compound Discoverer 3.3 (Thermo Scientific). Quality control samples were employed for the data correction. The mzCloud libraries were used as a tool for a fragmentation spectra comparison and the compound structure assignment.

Volatiles were analysed using GCxGC-MS and a headspace solid phase microextraction (HS-SPME) on fibre (DVB/CAR/PDMS_grey; Supelco, USA). A volume of 1 mL of media samples was placed in a 20 mL glass vial and 10 µL of an internal standard (2,2,2-trifluoro ethanol, Fluka, 0.11 $mg.mL^{-1}$) was added. Samples were incubated for 10 min at 40°C before extraction. The extraction was carried out for 5 minutes. The volatiles were analysed using a combination of polar and mid-polar separation columns for the separation (primary column: Stabilwax-DA (30 m x 0.25 mm, Restek, USA); secondary column BPX-50 (1.38 m x 0.1 mm, SGE, Australia)). Other parameters were set as follows: inlet temperature 220°C, split 5 modes, constant He flow 1 $mL.min^{-1}$, modulation time 3 s (hot pulse 0.7 s), modulation temperature offset with respect to the secondary oven 20°C. The temperature program applied on the primary oven was 30°C (hold 1.5 min), then increased to 110°C (8°C.min^{-1}), followed by an increase to 250°C (25°C.min^{-1}) to 250°C (hold 5 min). The temperature offset applied on the secondary column was +5°C. Transferline temperature was held at 280°C. The mass detector was equipped with an Electron Ionization ion source and Time-of-Flight analyser enabling a united mass resolution. The scanned mass range was 29 – 400 *m/z*. The ion source chamber was held at 250°C.

To extend the range of compounds analysable by gas chromatography to polar non-volatiles, the media samples underwent two consecutive derivatisation procedures – oximation (25 $mg.mL^{-1}$ of methoxyamine hydrochloride in anhydrous pyridine hydrochloride, Sigma-Aldrich) and silylation. Two different agents were used for silylation: N,O-Bis(trimethylsilyl)trifluoroacetamide with trimethylchlorosilane (BSTFA + 1% (v/v)TMSC)

and N-tert-Butyldimethylsilyl-N-methyl trifluoroacetamide (MTBSTFA), both from Sigma-Aldrich.

For silylation using BSTFA + 1% (v/v) TMSC, 50 μL were transferred to a 2 mL glass vial and 2 μL of internal standard were added (Adonitol, 424 $\mu L.mL^{-1}$). The liquid was dried in a vacuum concentrator and reconstituted back by adding 30 μL of anhydrous pyridine and 30 μL of an oximation agent. The samples were then incubated in a Thermomixer (Eppendorf) at 40ºC for 2 hours at constant shaking (1500 rpm), after which 40 μL of neat solution of BSTFA + 1% (v/v) TMSC mixture were added to the samples and were further incubated at 70ºC for 30 min at constant shaking (1500 rpm). Before analysis, 400 μL of hexane was added to the sample. For silylation using MTBSTFA, the derivatisation process is essentially the same as the derivatisation using BSTFA + TMSC except that 30 μL of media sample are used, plus 5 μL of nor-Valine (550 $\mu L.mL^{-1}$) as an internal standard and 20 μL of oximation agent.

Metabolites were analysed as oximated trimethylsilylated or tert-butyl silylated derivatives. A combination of nonpolar and mid-polar separation columns was used for the separation; a primary Rxi-5SIL MS column (30 m x 0.25 mm, Restek, USA) and a secondary BPX-50 column (1.39 m x 0.1 mm, SGE, Australia). Other parameters were set as follows: inlet temperature 290ºC, splitless mode, constant He flow 1 $mL.min^{-1}$, modulation time 3s (hot pulse 0.7s), modulation temperature offset with respect to the secondary oven 15ºC. The temperature program applied to the primary oven was 50ºC (hold 1 min), then increased ($8^{o}C.min^{-1}$) to 320ºC (hold 2 min). The temperature offset applied on the secondary column was +5ºC. Transferline temperature was held on 280ºC. The scanned mass range was 85 – 800 *m/z*. The ion source chamber was held at 280ºC. Data were processed in ChromaTOF v4.5 software. Detected analytes were quantified relatively after normalisation to the internal standards. Metabolites detected by all the methods used were identified by comparison of their mass spectra with those available in the main NIST mass library, Fiehn´s mass library of

silylated compounds, and in-house-built libraries. The retention index was determined using linear hydrocarbons. In the cases where the same compounds were detected in both derivatisations, the highest signal was taken to the final table.

Library preparation and sequencing

For the DNA samples, sequencing libraries were prepared from 1 µg of gDNA using the Illumina TruSeq DNA PCR free (Illumina) library preparation kit according to the manufacturer's protocol. The prepared libraries were sequenced on an Illumina HiSeq 4000 with 2x150 bp reads.

For RNA, 1 µg of total RNA per sample was used for library preparation. Ribosomal RNA depletion was performed using the NEBNext rRNA Depletion Kit for Bacteria (New England Biolabs), spiked with a custom oligo pool. The oligo pool was designed based on bacterial and eukaryotic ribosomal sequences that were identified from previous sequencing projects of *M. exilis* (Cite Karnkowska 2016 and Treitli 2021, the two genomic papers). After rRNA depletion, libraries were prepared using the NEBNext® Ultra™ Directional RNA Library Prep Kit (New England Biolabs) according to the manufacturer's protocol. The prepared libraries were sequenced using Illumina NextSeq 500 with 2x75bp reads.

Metagenome assembly and binning

The metagenomics raw reads were quality-checked with FASTQC v0.11.5 and trimmed using Trimmonatic v0.39 [16] (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10, LEADING:20, TRAILING:20, SLIDINGWINDOW:5:20 MINLEN:50). To reduce complexity during assembly, the eukaryotic reads were removed by mapping the reads against the *M. exilis* genome [9] using BBMap (minidentity=0.98, idfilter=0.98) included in BBTools v38.90 (https://sourceforge.net/projects/bbmap/) and taking only unmapped reads.

The filtered reads were assembled independently for each sample using metaSPAdes v3.14.0 [17] with k-mers 21,33,55,77,99,127. A reiterative process was used for binning

metagenome-assembled genomes (MAGs) within the samples. First, contigs above 2kbp were binned using MaxBin v2.2.7 [18] (minimum probability 0.9, marker set 107) and tetraESOM [19]. Then, bins were manually checked based on the markers identified by MaxBin using Blastp to avoid chimeric constructions. For each potential MAG, rRNA genes were identified using Prokka v1.14.6 [20] and used as controls. Lastly, reads corresponding to verified bins were removed from the data set by mapping the reads back to the identified bins using BBSplit (minidentity=0.98, idfilter=0.98) included in BBTools v38.90. This process was iterated and stopped when no new clean MAG was created. In total, 24 prokaryotic MAGs were assembled. The contribution of each prokaryotic MAG and *M. exilis* to the whole community per sample was calculated as the percentage of reads mapped to each genome assembly.

Identification of prokaryotic MAGs

16S rRNA genes identified by Prokka were used as blast queries against a 16S ribosomal reference database downloaded from NCBI (April 2021). For each query, 500 hits were kept. The resulting dataset was aligned using MAFFT v7.453 [21] (default parameters) and trimmed with BMGE v1.12 [22] (matrix: DNAPM1, maximum gap rate allowed per character 0.5). Maximum likelihood trees were computed using IQtree v1.6.12 [23] under an automatic model finder [24]. Branch supports were assessed using ultrafast bootstrap approximation (UFboot) with 1,000 bootstrap replicates [25].

Annotation and metabolic capacities of prokaryotic MAGs

The structural annotation of the 24 prokaryotic MAGs was performed using Prokka v1.14.6 automatic annotation pipeline. The functional annotation of each MAG was made using EggNOG-mapper v1.0.3-35 on emapper DB 2.0 [26, 27] (standard parameter except minimum % of query coverage: 50 and minimum % of subject coverage: 50). EggNOG-mapper predictions and Prokka annotations were combined using emapper2gbk, included in the Metage2Metabo pipeline [28], to create a compatible input for Pathway-Tools v24.5 [29].

A metabolic database was created for each MAG using Pathologic [30], included in Pathway-Tools (standard parameters except name matching that was turned off). All databases were manually curated using the Assign Probable Enzymes, Transport Inference Parser and Rescore Pathways tools included in Pathologic. Predicted pathways were manually checked to remove highly incomplete pathways.

Identification of metabolic capacities of *Monocercomonoides exilis*

The genome of *M. exilis* was downloaded from NCBI. The metabolic capacities of this species were predicted using Pathologic, included in Pathway-Tools, with standard parameters. The resulting database was manually curated using the Assign Probable Enzymes, Transport Inference Parser and Rescore Pathways tools included in Pathologic. Information from previous curations [9–11] and EggNOG-mapper was included at this step.

Carbohydrate-metabolising enzymes were identified by computing searches against CAZy databases using METABOLIC-G with default parameters [31].

Transcriptome processing and identification of differentially expressed genes

Meta-transcriptomic raw reads were quality-checked with FASTQC v0.11.5 and trimmed using Trimmonatic v0.39 (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:15, LEADING:15, TRAILING:15, SLIDINGWINDOW:4:15 MINLEN:50), resulting in 47 million read pairs per sample on average (except for sample D5R3 where the number of read pairs was 77 million). Reads were mapped and the expression of transcripts of prokaryotes and *M. exilis* was quantified using Salmon [32] with default parameters.

Due to the low mapping rate and a low number of reads, the differential expression was assessed only for *M. exilis* and the seven most abundant prokaryotic genomes: *Bacteroides fragilis, Bacteroides thetaiotaomicron, Bacteroides* sp., *Parabacteroides* sp., *Kerstersia gyiorum, Fusobacterium varium* and *Citrobacter freundii*. Raw reads were re-mapped onto the eight selected genomes using Bowtie2 [33]. Reads mapping to protein-coding sequences were

counted with the FeatureCounts program of the subread package v2.0.3 [34]. Read counts were normalised using the DESeq2 package v1.32.0 [35] in R v4.1.1 [36]. Normalisation was adapted from scripts provided in previous publications [37, 38].

Calculation of metabolic weights of processes in the community

The metabolic capacities of the prokaryotic community were also analysed using METABOLIC-C [31] with default parameters, except -tax genus, to understand their roles within the community's biochemical cycling processes. The analysis was run twice, once for DNA reads and a second time for RNA reads, to estimate the metabolic weight score of identified metabolic modules.

Identification of the metabolic relationships between *M. exilis* and the prokaryotic community

All identified reactions in *M. exilis* and the 24 prokaryotic MAGs metabolic database were exported as SBML format from Pathway-Tools. The Metacom tool implemented in the metage2metabo (m2m) pipeline was used to identify metabolic interactions between *M. exilis* and the rest of the community. In this analysis, the sampling days were compared pairwise, with *M. exilis* defined as the host. In each pairwise comparison, compounds that could be identified in the first sampling point were defined as seeds, meanwhile those whose concentration increased in the second sampling point, were defined as targets.

Results

Growth of the culture and sampling of DNA, RNA and media

The development of the culture was followed for seven days in an experiment performed in triplicates (Figure 1). An aliquot of the bacterised medium representing day 0 and aliquots of days 2, 3 and 5 representing the exponential, stationary and decline phases of *M. exilis* growth were chosen to sample metagenomic, metatranscriptomic and metabolomic data.

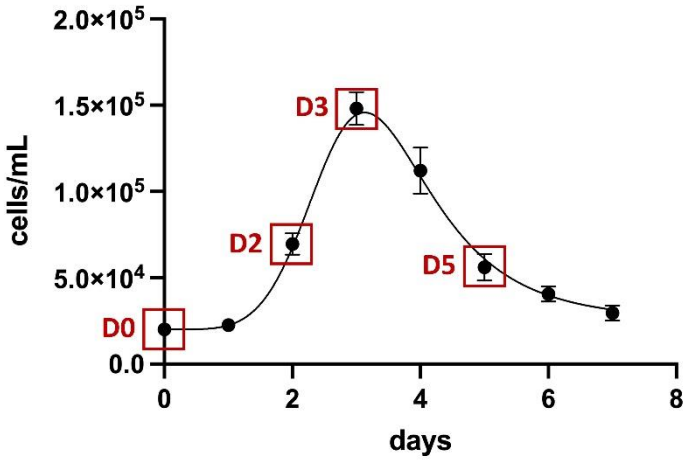


Figure 1: Growth curve of *Monocercomonoides exilis* and sampling points. Concentration of *M. exilis* (cells/mL^{-1}) grown in TYSGM-9 medium for seven days. Samples for thorough analysis were taken on days 0, 2, 3 and 5, as highlighted by red squares. Error bars provide standard deviations of the values from three measurements.

Community composition

Using metagenomic reads generated from triplicates of each sampling point, we assembled 24 metagenome-assembled genomes (MAGs) corresponding to the prokaryotic members of the culture community (Figure 2, Supplementary Table 1). In addition to the assembled MAGs, we identified six additional bacterial species based on the presence of 16S rRNA genes, but their read abundances were too low to assemble them into clean MAGs (Supplementary Table 1). This suggests the community consists of at least 30 bacteria, representing seven phyla, without any archaea or other eukaryote.

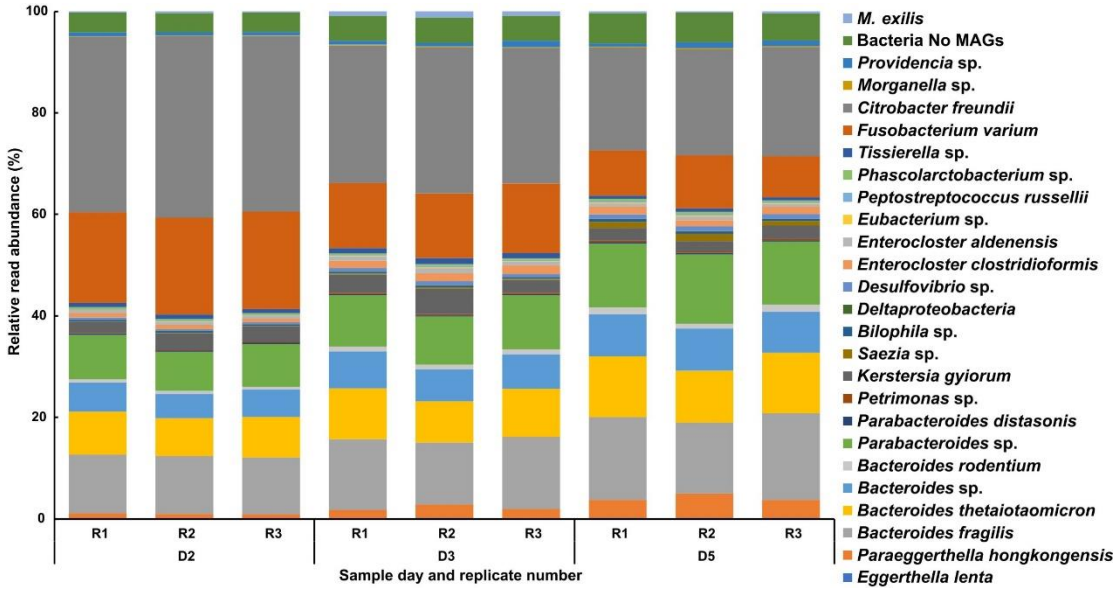


Figure 2: Relative read abundance of the members of the community. Each member of the community is represented as relative read abundance (%) corresponding to its genome on each day and replicate.

The quality of the 24 assembled MAGs varied, with the lowest completeness being around 83% and the vast majority having genome completeness greater than 95% (Supplementary Table 1). The contamination percentage shows levels lower than 5% overall, except for one MAG (Deltaproteobacteria sp.), whose contamination was estimated to be 11.8% (Supplementary Table 1). This MAG represents the less abundant MAG during the whole experiment. We calculated the relative read abundance of each bacterium and *M. exilis* for days 2, 3, and 5 of culture. While *M. exilis* represented only 0.25% to 1.23% of the reads, the seven most abundant bacterial species comprised 80 – 90% of the reads, with the remaining bacterial species never exceeding 1% of the reads per species (Figure 2 Supplementary Table 1). We observed a switch in the dominant groups throughout the culture growth. Not considering *Citrobacter freundii* (Gammaproteobacteria), which was inoculated into the medium before inoculation of the community ("bacterisation" of the medium), the most abundant bacterium on day two was *Fusobacterium varium* (Fusobacteriota). On day three and day five, the abundance of this bacterium decreased, and representatives of the phylum Bacteroidota dominated. *M. exilis* reached its peak of 1.23% on day 3.

***Monocercomonoides exilis* metabolic capacities**

Our analysis of *M. exilis* metabolism identified 933 enzymatic reactions, 31 transport and 148 enzymatic pathways (Supplementary Figure 1). As previously published [9, 11], *M. exilis* is equipped with a limited repertoire for energy metabolism relying on glycolysis to produce pyruvate, which is converted into acetyl-coenzyme A and further to acetate and ethanol. Our analysis showed that *M. exilis* can obtain glucose from the degradation of glycogen, maltose or starch but can also digest chitin and (1,3)- α-D-glucans present in fungal cell walls and peptidoglycan present in bacterial cell walls, which corroborates its bacterivorous feeding mode (Supplementary Table 5). The biosynthesis of amino acids and

nucleotides is limited and dependent on conversion and salvage pathways from compounds acquired from the diet or via transporters. However, *M. exilis* can synthesise dNTPs from NTPs using ribonucleoside-triphosphate reductase [10, 14].

Bacterial community metabolic capacities

To better understand the metabolic potential of the 24 assembled bacterial genomes, we reconstructed the metabolism of each MAG using Pathway-Tools. Our results showed a higher number of pathways and reactions compared to *M. exilis*, with a considerable diversity in the number of enzymatic reactions, transporters and pathways (from 443 in *Citrobacter freundii* to 218 in *Peptostreptococcus russellii*) (Supplementary Table 1). The metabolic maps of the seven most abundant MAGs are provided in Supplementary Figures 2-8. The number of reactions and pathways identified in closely related MAGs was very similar, even if the assembly quality differed between MAGs. This suggests that our metabolic analyses were not heavily affected by the level of fragmentation of the MAG assemblies.

We analysed the roles of community members at the genus level within the biogeochemical cycles of the community using METABOLIC-C. This allows assigning genes to major biochemical processes and reflects their DNA and RNA read abundances to calculate the metabolic weight score, i.e., their individual contribution to every monitored process. Metabolic weight scores of all monitored processes in all community members sum to 100. In the case of DNA reads, this value provides information about potential metabolic abilities, while in the case of RNA reads, it considers expression levels and informs about the processes taking place at a given time point. The results showed that the community likely depends on oxidation of organic carbon compounds (mainly amino acids and complex carbon compounds), fermentation and H_2 generation (Figure 3a and Supplementary Tables 2 and 3). The community is incapable of methanogenesis and the produced H_2 is probably further oxidised to water.

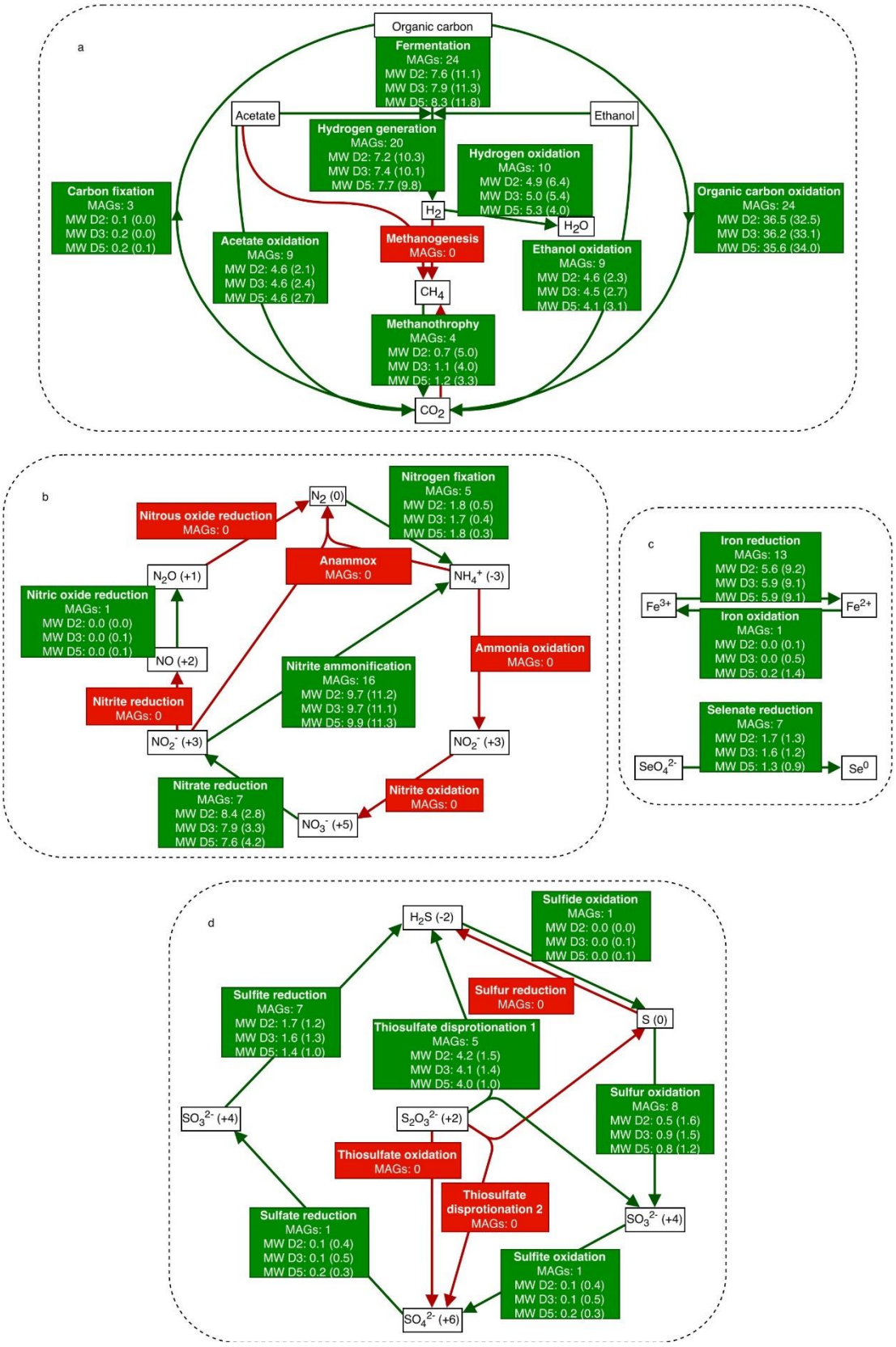


Figure 3: Contribution of the bacterial community to the biochemical cycling processes of carbon (a), nitrogen (b), iron (c) and sulphur (d): Labels represent the main steps. Present steps are identified in green. Absent steps are identified in red. MAGs: number of MAGs responsible for a step. MW: metabolic weight score. The first value represents the metabolic weight score from the DNA data, while the value in brackets represents the metabolic weight score from RNA data. Steps with identified responsible MAGs but a metabolic weight score of zero reflect that the original metabolic weight score was rounded to zero for being less than 0.05.

347

However, four MAGs are methanotrophic and oxidise methane to CO_2 – *Parabacteroides* sp, *Kerstersia gyiorum*, *Enterocloster clostridioformis* and Deltaproteobacteria spp. Though the source of this methane is unattained, it may theoretically originate from the degradation of methyl compounds (e.g., methylphosphonate degradation). A few community members, including *Fusobacterium varium*, can fix N_2 to ammonium. However, most of the ammonium is likely recycled via nitrite ammonification [39], as the transcriptome metabolic weight score of nitrite ammonification using either nrfADH or nirBD complexes is 20x higher than that for N_2 fixation (Figure 3b, Supplementary Table 3). The complete cycle of sulphur compound reductions and oxidations in the community is mainly operated by three members: *Desulphovibrio* sp, *Kerstersia gyiorum,* and *Fusobacterium varium*. The reduction of iron is another important observed activity, with 13 MAGs involved in the process and the total metabolic weight score of almost 6 for DNA and more than 9 for RNA (Figure 3c). In contrast, iron oxidation activity was identified only for *Saezia* sp. with a metabolic weight score of 1.3 and lower (Figure 3c and Supplementary Tables 2 and 3).

The consumption and production of compounds by the community

Triplicates of the cell-free medium were sampled at each time point and their composition was analysed by non-targeted metabolomics. LC-MS/MS and GCxGC-MS covered a broad range of structural classes of metabolites, detecting 270 distinctive compounds altogether, of which 171 take part in the biological reactions listed on the MetaCyc database [40], hence were the only ones used for the follow-up analyses (Supplementary Table 4).

On the day of inoculation (day 0), the medium was already affected by an overnight growth of *C. freundii*, yet it was relatively abundant in nucleotides and their derivates, glyceraldehyde, lactate, fatty acids, alcohols and amino acids, except asparagine, aspartate, cysteine, glutamine and serine, which were below the detection limit at this time point

(Supplementary Table 4). On day 2, a drastic evanesce of some compounds was registered, specifically nucleotides and their derivates, most amino acids, glyceraldehyde, glyceric acid, malate, citric acid, and citrulline, suggesting that the community had consumed them. On the contrary, a spectrum of fatty acids and aromatic compounds putatively derived from the degradation of amino acids, such as tyrosine, phenylalanine and tryptophan, was observed in the media on this day (Supplementary Table 4). Between days 2, 3 and 5, the measurements displayed little changes, suggesting that the composition of cell-free media stabilised.

We integrated the synthesis and consumption of compounds into an *in silico* metabolic capacity prediction, both for *M. exilis* and the bacterial community, using the Metage2Metabo (m2m) pipeline [28]. In these predictions, the culture days were compared pairwise, i.e., day 0 vs. day 2, day 2 vs. day 3 and day 3 vs. day 5. We then classified compounds either as seeds if detected on the first day in pairwise comparison or as targets if they were detected only or their concentration increased on the second sampling day of that pair. This was performed using measured differences in media composition, presence/absence and concentration changes of compounds between sampling days. These results are summarised in Figure 4. Between day 0 and day 2, 137 compounds were classified as seeds and 92 as targets, of which *M. exilis* can synthesise 63 by itself. However, the metabolic capacities of *M. exilis* are not limited to the target compounds, as it can synthesise other important compounds like pyruvate, amino acids, medium-chain fatty acids and vitamins. An additional 29 target compounds can be synthesised via metabolic interactions of *M. exilis* and the bacterial community (Figure 4). Among these compounds are several amino acids. In this pairwise comparison, m2m did not identify any bacteria as an essential member of the community. However, two Bacteroidota bacteria were classified as key members (species occurring in at least one of the minimal community models) (Figure 4). Between day 2 and day 3, we identified 141 compounds as seeds and 72 as targets. According to the m2m pipeline, *M. exilis* can synthesise 68 targets independently and 51

compounds through metabolic interaction with the community (Figure 4). In this comparison, no bacterial species are considered essential or key members of the community. The last comparison, day 3 vs. day 5, showed a similar result. *M. exilis* could synthesise 54 of the 55 compounds identified as targets using 140 compounds as seeds. The interaction between *M. exilis* and the bacterial community allows the synthesis of 49 more compounds and none of the bacteria were considered essential or key members (Figure 4).

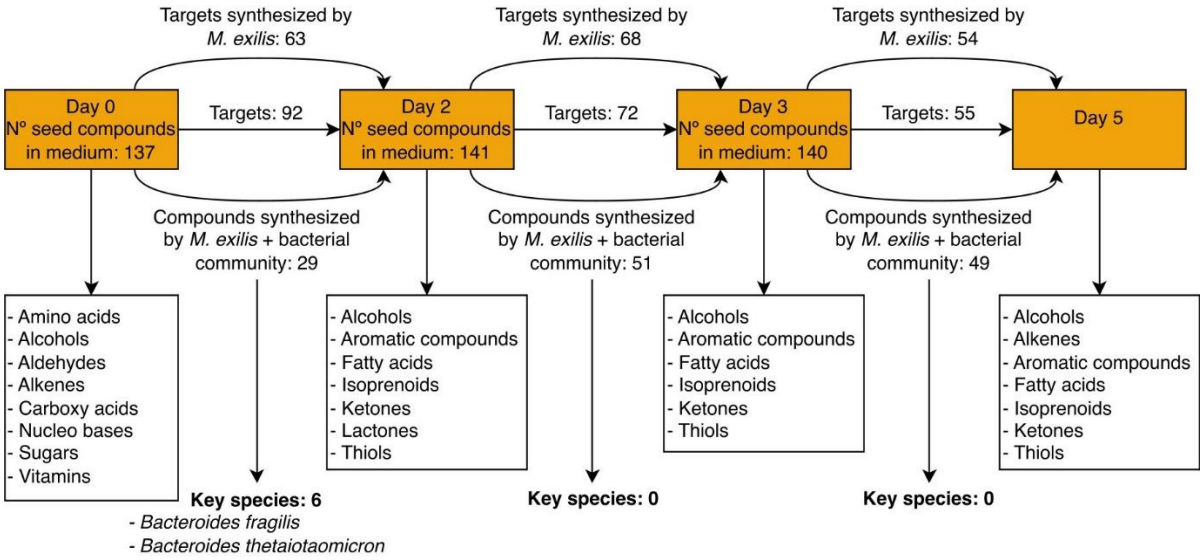
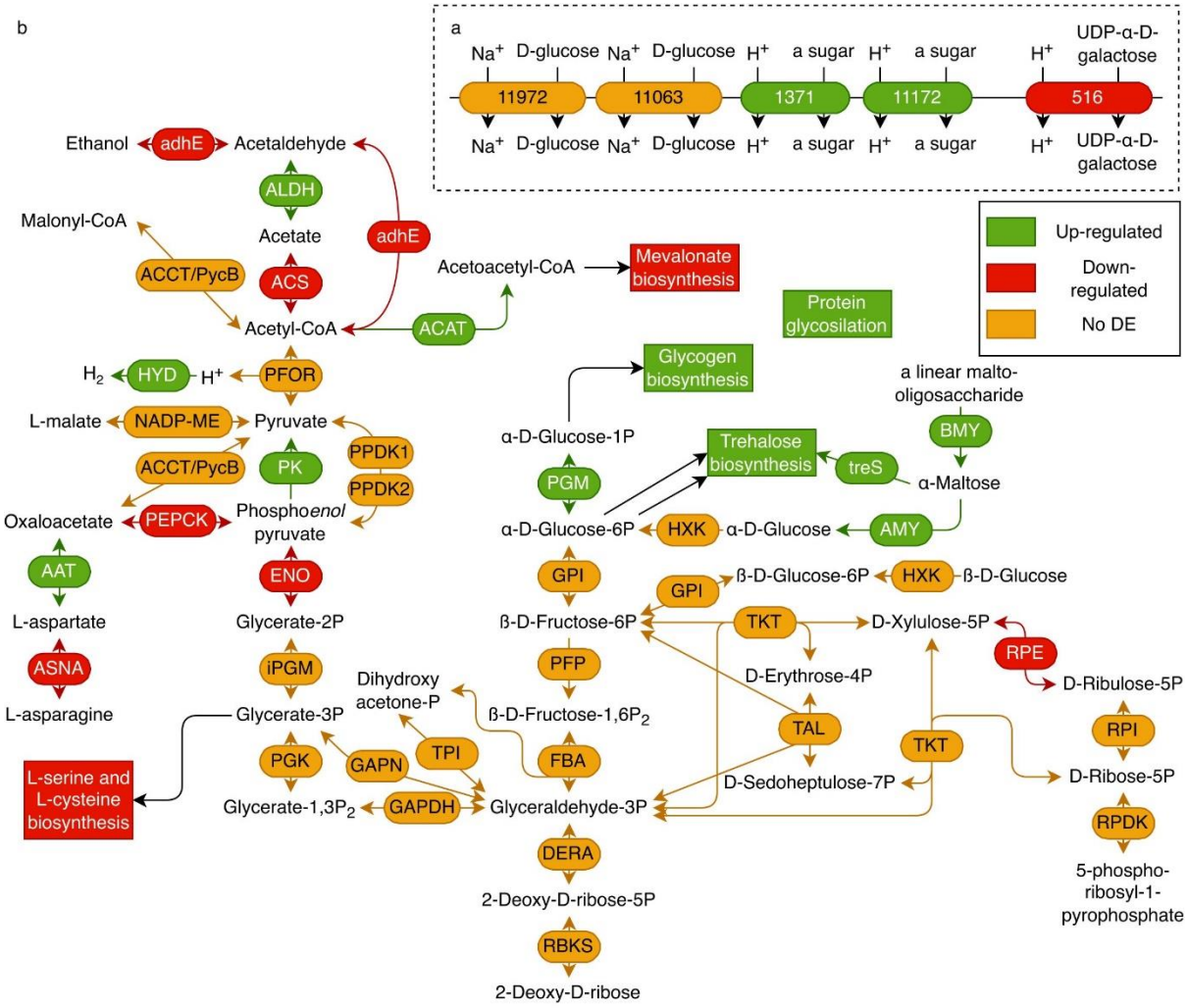


Figure 4: Schematic summary of changes in the culture medium. Identified compounds per day; number of compounds exhibiting an increase in concentration to the previous day (targets); number of targets that *M. exilis* can synthesise independently, and number of compounds that can be synthesised because of the interaction between *M. exilis* and the bacterial community are shown. Key bacterial species identified by the m2m pipeline are shown for each pairwise comparison. Chemical classification of significant compounds is shown for each day (orange arrow).

Monocercomonoides exilis differential expression analysis

To investigate how *M. exilis* metabolism changes during the growth curve, we also compared the expression of transcripts between days 2, 3 and 5, again pairwise. Most significant changes in gene expression were observed between day 2 and 5, while comparisons made between consecutive days revealed less significant changes (i.e., day 3 vs. day 2 and day 5 vs. day 3). 2,236 genes were up-regulated and 2,431 down-regulated between day 2 and 5 (Supplementary Figure 9), of which the majority, 1703 and 1,381 genes respectively, are annotated as hypothetical, hence their function could not be deciphered.

The changes in expression of all functionally annotated metabolic enzymes and pathways are shown in Supplementary Figure 1 and those discussed below are shown in Figures 5 and 6 in more detail. We observed an increase in the expression of importers of sugars such as α-D-xylose and *sn*-glycerol-3-phosphate towards day 5. Connected to this, there was an increment in the expression of enzymes responsible for the degradation of starch, (1->4)-α-glucan, chitin and other polysaccharides (Figure 5 and Supplementary Table 5). Furthermore, the interconversion between glucose-1-P and glucose-6-P was up-regulated, which may relate to the increased trehalose biosynthesis from glucose-6-P. Trehalose is often used as carbohydrate storage and as a stress protector. All three pathways for trehalose synthesis were up-regulated on day 5, including the synthesis directly from α-maltose, performed by both α- and β-amylases. The synthesis of glycogen, another storage polysaccharide, was also up-regulated. The expression of most genes involved in carbohydrate metabolism and glycolysis remained unchanged, however, some, such as enolase, exhibit down-regulation, suggesting that glycolysis may be tuned down in later days. The pentose phosphate pathway may also be down-regulated due to the repression of the ribulose-phosphate 3-epimerase (Figure 5). Interestingly, another key glycolytic enzyme, pyruvate kinase, was up-regulated. The expression of enzymes of extended glycolysis and ethanol fermentation, i.e., acetate synthase (ACS) producing acetate and aldehyde and alcohol dehydrogenases (ALDH, Adh) producing ethanol, decreases greatly on day 5 for all but one paralogue of ALDH (Supplementary Table S7). ACS1 was consistently the most abundant of these enzymes in all three time points and replicates. Hydrogenases, scavenging reduced cofactors (ferredoxins) were up-regulated on day 5. Overall, these expression data suggest that acetate and hydrogen are likely the preferred metabolic end products at all three time points.



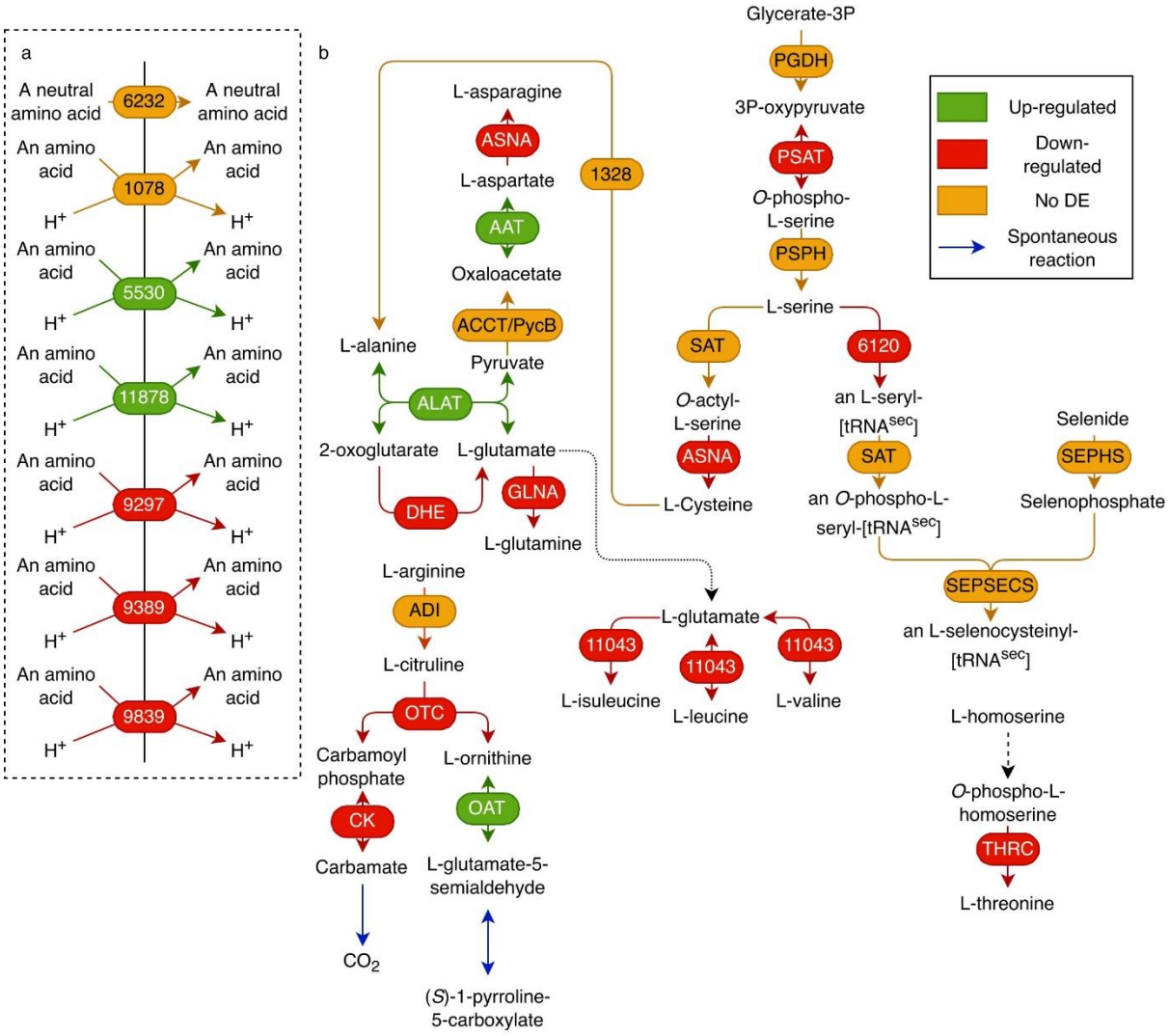
439

Figure 5 Differential expression of genes involved in carbohydrate metabolism and transport on *Monocercomonoides exilis* between day 5 and day 2. a) Identified sugar transporters. b) Glycolysis, pentose phosphate pathways and related reactions (adapted from Karnkowska et al. 2019). Up-regulated genes and processes are shown in green. Down-regulated genes and processes are shown in red. Non-differentially expressed genes and processes are shown in orange. adhE: Bifunctional acetaldehyde-CoA/alcohol dehydrogenase; ALDH: Aldehyde dehydrogenase; ACS: Acetyl-CoA synthetase (ADP-forming); ACCT/PycB: putative malonyl-CoA:pyruvate transcarboxylase; ACAT: Acetyl-CoA acetyltransferase; PFOR: Pyruvate-ferrodoxin oxidoreductase; HYD: [FeFe]-hydrogenase; NADP-ME: NADP-dependent malic enzyme; PK: Pyruvate kinase; PPDK: Pyruvate phosphate dikinase; PEPCK: Phosphoenolpyruvate carboxykinase; AAT: Aspartate aminotransferase; ASNA: Asparagine synthetase; ENO: Enolase; iPGM: 2,3-bisphosphoglycerate independent phosphoglycerate mutase; PGK: Phosphoglycerate kinase; GAPDH: Glyceraldehyde-3-phosphate dehydrogenase; GAPN: NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase; TPI: Triose phosphate isomerase; DERA: Deoxyribose-phosphate aldolase; RBKS: Ribokinase; FBA: Fructose-bisphosphate aldolase; PFP: Phosphofructokinase (pyrophosphate-based); GPI: Glucose-6-phosphate isomerase; PGM: Phosphoglucomutase; BMY: β-amylase; AMY: α-amylase; treS: Maltose alpha-D-glucosyltransferase/ alpha-amylase; HXK: Hexokinase; RPDK; ribose-phosphate diphosphokinase; RPI: Ribose-5-phosphate isomerase; RPE: Ribulose-phosphate 3-epimerase; TKT: Transketolase; TAL: Transaldolase.

440

441

442



443

Figure 6 Differential expression of genes involved in amino acid metabolism and transport on *Monocercomonoides exilis* between D5 and D2. a) Identified amino acid importers. b) Identified amino acid biosynthesis and arginine catabolism. Up-regulated genes and processes are shown in green. Down-regulated genes and processes are shown in red. Non-differentially expressed genes and processes are shown in orange. Spontaneous reactions are shown in blue. PGDH: Phosphoglycerate dehydrogenase; PSAT: Phosphoserine transaminase; PSPH: putative phosphoserine phosphatase; SAT: Serine O-acetyltransferase; SEPHS: Selenophosphate synthetase; SEPSECS: O-phospho-L-seryl-tRNA[Sec]:L-selenocysteinyl-tRNA synthase; THRC: Threonine synthase: ASNA: Asparagine synthetase; ALAT: Alanine aminotransferase; ACCT/PycB: putative malonyl-CoA:pyruvate transcarboxylase; AAT: Aspartate aminotransferase; ASNA: Asparagine synthetase; DHE: Glutamate dehydrogenase; GLNA: Glutamine synthetase: ADI: Arginine deiminase: OTC: Ornithine transcarbamylase: CK: Carbamate kinase; OAT: Lysine/Ornithine aminotransferase.

444

445 Genes responsible for the phosphorylation of (d)NDP, DNA and RNA polymerisation
446 and protein ubiquitination were up-regulated on day 5 (Supplementary Figure 1). In contrast,
447 nucleoside import and the salvage pathway of adenosine were down-regulated. Similarly, the
448 mevalonate pathway for the synthesis of isoprenoids from acetoacetyl-CoA was down-

regulated (Figure 4 and Supplementary Figure 1). Most amino acid importers displayed lower expression or showed no significant change in expression (Figure 6). This relates to the fact that amino acid content in the media significantly decreased from day 2. Asparagine, serine, and cysteine biosynthesis were down-regulated, as was the ATP-producing arginine deaminase pathway (Figure 6).

Another system that showed significant changes during cultivation was the endomembrane system. Among the proteins involved in the regulation of endomembrane transport, some show significant changes in expression (Supplementary Table 6). The proteins involved in the early secretory pathway show little change, but several proteins involved in the late steps [41–43] were significantly up-regulated on day 5 (Supplementary Table 6). These include proteins involved in trans-Golgi-network (TGN) to plasma membrane trafficking (DSCR3A and exocyst complex components Exo84, Sec5 and Sec15, DSCR), multivesicular body formation and sorting (Vsp2, 4a-c, 23, 24a and 60b, charged multivesicular body protein 7), plasma membrane uptake (AP-2a, TSET complex protein TPLATE subunit B), endocytic/digestive pathway (Vps34bc, Vps41b) and endosomes to TGN transport (Vps51b, TRAPP component Bet3 paralogue A, B, syntaxin 6D and 16). Alternatively, four Rab7 paralogues were consistently down-regulated, implying a decrease in lysosome biogenesis on day 5 [44].

Bacterial community differential expression analysis

Due to the low abundance of most community members, our data only allowed differential expression analysis for the seven most abundant bacteria: *Bacteroides fragilis*, *B. thetaiotaomicron*, *Bacteroides* sp., *Parabacteroides* sp., *Kerstersia gyiorum*, *Fusobacterium varium* and *Citrobacter freundii*. These species represent 80-90% of the community DNA reads, a proxy measure of their abundance. These bacteria showed a variable number of up- or

down-regulated genes (Supplementary Table 1); for example, only 9% of *Bacteroides* sp. transcripts changed in expression between the D2 vs. D5, while in *Parabacteroides* sp., it was 51%. Expression changes of metabolic enzymes and pathways of these bacteria are indicated in Supplementary Figures 2-8.

We focused specifically on pathways absent in *C. freundii* but present in other members of the community, as these may explain why *M. exilis* cannot grow with only *C. freundii* as food. First, the community can ferment lysine to acetate and butanoate, but in *C. freundii*, this pathway is incomplete. Furthermore, the community encodes genes for the synthesis of the essential polyamine spermidine and the phospholipid phosphatidylcholine. While phosphatidylcholine is one of the major components of cell membranes [45], spermidine is involved in an array of biological processes, such as maintenance of membrane potential, control of intracellular pH, lipid metabolism and regulation of cell growth, proliferation, and death [46]. In both cases, our metabolic analysis fails to identify a pathway that can synthesise these compounds *de novo* in *M. exilis*, suggesting that the eukaryote relies on the uptake of these compounds from the bacterial community.

Iron uptake is another relevant function for which no pathway is known in *M. exilis*. Several bacteria encode for transporters of both forms of iron, Fe^{2+} and Fe^{3+}, as well as hemin and heme b. We do not observe any differential expression amongst these transporters between day 2 and day 5, except an up-regulation in the Fe^{3+} transporter of *F. varium* and *K. gyiorum* (Supplementary Figure 2 – 8). Of importance is another type of molecule used for iron chelation in bacteria known as siderophores, which are secreted into the environment. Iron-bound siderophores are subsequently selectively imported [47]. Our data showed that *C. freundii* and *Saezia* sp. synthesise enterobactin, while *F. varium* synthesises staphylopine, and *K. gyiorum* a periplasmic iron-binding protein. Interestingly, all bacteria, but not *M. exilis*, encode for transporters of siderophores, suggesting that the bacterial community depends on these few

species able to synthesise siderophores and contain "cheaters" that uptake siderophores produced by others [48, 49]. The siderophores' producers are, from this point of view, the key players for the prokaryotic community but do not facilitate the uptake of iron into *M. exilis.* Once iron is imported into the cell, it cannot be kept free in the cytoplasm. One solution is its coordination within different kinds of porphyrins [50–52]. We observed two main porphyrins, siroheme and heme *b*. Heme *b* can be synthesised aerobically or in an oxygen-independent manner, while siroheme is synthesised only independently of oxygen [50, 52]. Our data showed that *C. freundii* can synthesise both compounds, but under anaerobic conditions, the synthesis of heme *b* is down-regulated, while that of siroheme is up-regulated on day 5 (Supplementary Figure 2). The opposite situation was observed on *Bacteroides fragilis*, where oxygen-independent synthesis of heme *b* was up-regulated, yet siroheme synthesis underwent down-regulation (Supplementary Figure 4). Some bacteria in the community can synthesise heme *b* from siroheme, however, due to their low abundance, we lacked transcriptome data for their assessment.

Discussion

The first and only study of chinchilla gut microbiota [53] reports a composition of Betaproteobacteria as the dominant phylum, and *Parabacteroides* and *Barnesiella* as the dominant genera. Our study of an *M. exilis* culture derived from the gut of chinchilla identified 30 bacterial species, of which *Fusobacterium varium* was the most abundant at the beginning of the experiment, steadily replaced by the Bacteroidetes group, represented by *Bacteroides fragilis*, *Parabacteroides* sp. and *Bacteroides* sp. on day 5. *Parabacteroides* sp. was the second most abundant bacteria on day 5, yet we failed to detect any *Barnesiella* species. Only two Betaproteobacteria species were identified in our study, but they are not among the most abundant species. At the level of bacterial phyla, Bacteroidota and Bacillota are the most

abundant in our culture which on a broad scale resembles the composition of the gut microbiota of Guinea pigs [54]. These comparisons indicate similar compositions of the prokaryotic community between the gut of caviomorph rodents and our *M. exilis* culture; yet the community has changed as a consequence of long-term cultivation.

Given that *M. exilis* represents around 1% of the biomass, the biochemical processes in the community are driven by prokaryotes. Our analyses infer that the community dominantly oxidises organic carbon substrates, performs fermentation and generates hydrogen, which is further oxidised to water (Figure 3 and Supplementary Tables 2 and 3). Since the ability to fix carbon from CO_2 is minimal (<0.1 metabolic weight in the transcriptome) and present only in the two least-abundant species of *Enterocloster*, the community depends on the supply of organic carbon. This is consistent with our observation that organic carbon sources (glyceraldehyde, lactate, fatty acids, alcohols and most amino acids) present on day 0 are quickly consumed.

On the contrary, the nitrogen metabolism of the community seems to be in equilibrium. These processes are dominated by nitrite ammonification (metabolic weight score 10), which recycles nitrites and nitrates generated by oxidative decay (not drawn in Figure 3) of nitrogen compounds (e.g., amino acids) to ammonia, exploitable for the building of organic mass. Two systems for nitrite ammonification are present there: the respiratory membrane-bound nitrate reductase nrfADH, which uses nitrates as electron acceptor and allows building membrane potential, and the assimilatory cytoplasmic nitrite reductase NirBD [39]. The former has a 10x higher metabolic weight score (Supplementary Table 3). The bacterial community can also fix a small amount of nitrogen gas (N_2), mainly thanks to *F. varium*. These results suggest that the community can recycle nitrogen compounds and, unlike carbon, is not dependent on their external source. The same, in principle, applies to sulphur metabolism, where reactions converting all basic inorganic sulphur compounds are present (Figure 3).

The metabolomic analyses of the soluble media clearly showed that on day 2, when the population of *M. exilis* only began to grow, essentially all soluble carbon sources, nucleotides and most amino acids had been already consumed (Supplementary Table 4). The fact that *M. exilis* maintains active growth after day 2 strongly suggests that the protist uses bacterial prey as the major source of carbon, nitrogen, and energy to build its biomass. A putative exception could be arginine, which was detected in later days and may serve as a potential substrate for energy metabolism [55, 56]. Grazing on bacteria has been previously documented by electron microscopy [12] and the fact that the protist encodes lysozymes amongst a broad array of glycosyl hydrolases (Supplementary Table S5) also provides molecular tools for feeding on bacteria. There is no consistent pattern that would indicate up- or down-regulation of lysozymes as the culture ages. The protist, however, displays a significant up-regulation of several genes involved in the late pathways of the endomembrane system on day 5. At the same time expression of many genes encoding for proteins regulating these processes remained unchanged and, notably, four paralogues of Rab7 small GTPase, known for its role in lysosome activation [44], were significantly down-regulated on day 5 (Supplementary Table 6). We interpret this pattern as an indication of up-regulation of exocytosis and secretion rather than bacterial feeding, which is also consistent with the decrease in *M. exilis* cell counts at this stage of the culture.

Most *M. exilis* genes that exhibit differential expression completely lack functional annotation, providing us with a rather large transcriptomic dark matter consisting of approx. three thousand genes that may convey some functional and adaptive value to the protist. Among the known metabolic genes, we observed an up-regulation of the importers of sugars, α-D-xylose and *sn*-glycerol-3-phosphate, which is likely the reaction of the cell to the absence of these substrates. Another functional category of genes up-regulated on day 5 is the enzymes responsible for starch degradation and synthesis of trehalose and glycogen, indicating the onset

of starvation. The degradation of nucleosides and nucleotides and the biosynthesis of 5-phospho-α-D-ribose 1-diphosphate (PRPP) are up-regulated, suggesting an increment in the storage of nucleotides and phosphate donors. On the other hand, glycolysis, extended glycolysis, fermentation, pentose phosphate pathway, amino acids biosynthesis, arginine deaminase pathway and mevalonate pathway involved in isoprenoid biosynthesis are generally down-regulated on day 5, with few exceptions. The same holds for the synthesis of very long and ultra-long fatty acids and the aminoacyl-tRNA synthetases, indicating the decline of proteosynthesis.

Although we hoped to find some metabolic dependencies between *M. exilis* and prokaryotes, potentially explaining the inability to grow the flagellate axenically using nutrients dissolved in the medium, we did not reveal any clear link. This obviously does not mean that some have remained unobserved or unnoticed, as the amount and complexity of the data is significant. The *in silico* metabolic capacity prediction by Metage2Metabo (Figure 4) has shown that the compounds detected in the media at the time points can be synthesised by the reactions and pathways found in the community. Importantly, none of the community members has been found to be essential at any time point, though some may provide key compounds for others, e.g. siderophores. This suggests that the community is based on broad redundant foundations, which certainly contributes to its long-term stability.

Acknowledgement

We thank to Professor Joel Dacks for his advice on the analysis of the differentially expressed protein related with the endomembrane system. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 771592 to VH) and the Centre for research of pathogenicity and virulence of parasites (registration no.

CZ.02.1.01/0.0/0.0/16_019/0000759). The metagenome assemblies were computed using the resources provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Competing Interest

The authors declare that they have no conflict of interest.

Data Availability Statement

DNA reads, RNA reads and the complete 24 MAGs are available under a BioProject with accession number PRJEB70623. MAGs are available with accession numbers from ERS17699982 to ERS17699999.

References

1. Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasikala M, Reddy DN. Role of the normal gut microbiota. *World J Gastroenterol* 2015; **21**: 8836–8847.
2. Thursby E, Juge N. Introduction to the human gut microbiota. *Biochem J* 2017; **474**: 1823–1836.
3. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* 2021; **19**: 55–71.
4. Partida-Rodriguez O, Nieves-Ramirez M, Laforest-Lapointe I, Brown EM, Parfrey L, Valadez-Salazar A, et al. Exposure to Parasitic Protists and Helminths Changes the Intestinal Community Structure of Bacterial Communities in a Cohort of Mother-Child Binomials from a Semirural Setting in Mexico. *mSphere* 2021; **6**.

5. von Huth S, Thingholm LB, Kofoed P-E, Bang C, Rühlemann MC, Franke A, et al. Intestinal protozoan infections shape fecal bacterial microbiota in children from Guinea-Bissau. *PLoS Negl Trop Dis* 2021; **15**: e0009232.

6. Diamond LS. A new liquid medium for xenic cultivation of *Entamoeba histolytica* and other lumen-dwelling protozoa. *J Parasitol* 1982; **68**: 958–9.

7. Hamann E, Gruber-Vodicka H, Kleiner M, Tegetmeyer HE, Riedel D, Littmann S, et al. Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. *Nature* 2016; **534**: 254–258.

8. Hamann E, Tegetmeyer HE, Riedel Di, Littmann S, Ahmerkamp S, Chen J, et al. Syntrophic linkage between predatory *Carpediemonas* and specific prokaryotic populations. *ISME J* 2017; **11**: 1205–1217.

9. Treitli SC, Peña-Diaz P, Hałakuc P, Karnkowska A, Hampl V. High quality genome assembly of the amitochondriate eukaryote *Monocercomonoides exilis*. *Microb Genomics* 2021; **7**.

10. Karnkowska A, Treitli SC, Brzoň O, Novák L, Vacek V, Soukal P, et al. The Oxymonad Genome Displays Canonical Eukaryotic Complexity in the Absence of a Mitochondrion. *Mol Biol Evol* 2019; **36**: 2292–2312.

11. Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, et al. A eukaryote without a mitochondrial organelle. *Curr Biol* 2016; **26**: 1274–1284.

12. Treitli SC, Kotyk M, Yubuki N, Jirounková E, Vlasáková J, Smejkalová P, et al. Molecular and Morphological Diversity of the Oxymonad Genera *Monocercomonoides* and *Blattamonas* gen. nov. *Protist* 2018; **169**: 744–783.

13. Hampl V. Preaxostyla. In: Archibald JM, Simpson AGB, Slamovits CH (eds). *Handbook of the Protists*. 2017. Springer International Publishing, Cham, pp 1139–1174.

14. Novák LVF, Treitli SC, Pyrih J, Hałakuc P, Pipaliya S V., Vacek V, et al. Genomics of Preaxostyla Flagellates Illuminates the Path Towards the Loss of Mitochondria. *PLoS Genet* 2023; **19**: 1–30.

15. Treitli SC, Kolisko M, Husník F, Keeling PJ, Hampl V. Revealing the metabolic capacity of *Streblomastix strix* and its bacterial symbionts using single-cell metagenomics. *Proc Natl Acad Sci* 2019; **116**: 19675–19684.

16. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

17. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res* 2017; **27**: 824–834.

18. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016; **32**: 605–607.

19. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, et al. Community-wide analysis of microbial genome sequence signatures. *Genome Biol* 2009; **10**.

20. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–2069.

21. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002; **30**: 3059–3066.

22. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 2010; **10**: 210.

23. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQTree: A fast and effective

stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol* 2015; **32**: 268–274.

24. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017; **14**: 587.

25. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018; **35**: 518–522.

26. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019; **47**: D309–D314.

27. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* 2021; **38**: 5825–5829.

28. Belcour A, Frioux C, Aite M, Bretaudeau A, Hildebrand F, Siegel A. Metage2metabo, microbiota-scale metabolic complementarity for the identi1cation of key species. *Elife* 2020; **9**: 1–38.

29. Karp PD, Paley SM, Midford PE, Krummenacker M, Billington R, Kothari A, et al. Pathway Tools version 24.0: Integrated Software for Pathway/Genome Informatics and Systems Biology. *arXiv e-prints* 2020.

30. Karp PD, Latendresse M, Caspi R. The pathway tools pathway prediction algorithm. *Stand Genomic Sci* 2011; **5**: 424–429.

31. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* 2022; **10**.

32. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017; **14**: 417–419.

33. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**: 357–359.

34. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; **30**: 923–930.

35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550.

36. R Core Team. R: A language and environment for statistical computing. 2021.

37. Klingenberg H, Meinicke P. How to normalize metatranscriptomic count data for differential expression analysis. *PeerJ* 2017; **2017**.

38. Christel S, Herold M, Bellenberg S, Buetti-Dinh A, Hajjami M El, Pivkin I V., et al. Weak iron oxidation by Sulfobacillus thermosulfidooxidans maintains a favorable redox potential for chalcopyrite bioleaching. *Front Microbiol* 2018; **9**: 1–12.

39. Cabello P, Luque-Almagro VM, Roldán MD, Moreno-Vivián C. Nitrogen cycle. *Encycl Microbiol* 2019; 301–310.

40. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 2014; **42**: 459–471.

41. Law KC, Chung KK, Zhuang X. An Update on Coat Protein Complexes for Vesicle Formation in Plant Post-Golgi Trafficking. *Front Plant Sci* 2022; **13**: 1–9.

42. Shimizu Y, Uemura T. The sorting of cargo proteins in the plant trans-Golgi network. *Front Plant Sci* 2022; **13**: 1–10.

43. Tanaka T, Goto K, Iino M. Diverse Functions and Signal Transduction of the Exocyst Complex in Tumor Cells. *J Cell Physiol* . 2017. , **232**: 939–957

44. Bucci C, Thomsen P, Nicoziani P, McCarthy J, Van Deurs B. Rab7: A key to lysosome biogenesis. *Mol Biol Cell* 2000; **11**: 467–480.

45. Zhukov A, Popov V. Eukaryotic Cell Membranes: Structure, Composition, Research Methods and Computational Modelling. *Int J Mol Sci* 2023; **24**.

46. Michael AJ. Polyamines in eukaryotes, bacteria, and archaea. *J Biol Chem* 2016; **291**: 14896–14903.

47. Kramer J, Özkaya Ö, Kümmerli R. Bacterial siderophores in community and host interactions. *Nat Rev Microbiol* 2020; **18**: 152–163.

48. Leventhal GE, Ackermann M, Schiessl KT. Why microbes secrete molecules to modify their environment: The case of iron-chelating siderophores. *J R Soc Interface* 2019; **16**.

49. Butaite E, Baumgartner M, Wyder S, Kümmerli R. Siderophore cheating and cheating resistance shape competition for iron in soil and freshwater Pseudomonas communities. *Nat Commun* 2017; **8**.

50. Dailey HA, Dailey TA, Gerdes S, Jahn D, Jahn M, O'Brian MR, et al. Prokaryotic Heme Biosynthesis: Multiple Pathways to a Common Essential Product. *Microbiol Mol Biol Rev* 2017; **81**: 1–62.

51. Bryant DA, Hunter CN, Warren MJ. Biosynthesis of the modified tetrapyrroles—the pigments of life. *J Biol Chem* 2020; **295**: 6888–6925.

52. Layer G. Heme biosynthesis in prokaryotes. *Biochim Biophys Acta - Mol Cell Res* 2021; **1868**: 118861.

53. O' Donnell MM, Harris HMB, Ross RP, O'Toole PW. Core fecal microbiota of domesticated herbivorous ruminant, hindgut fermenters, and monogastric animals. *Microbiologyopen* 2017; **6**.

54. Crowley EJ, King JM, Wilkinson T, Worgan HJ, Huson KM, Rose MT, et al.

Comparison of the microbial population in rabbits and guinea pigs by next generation sequencing. *PLoS One* 2017; **12**.

55. Yarlett N, Martinez MP, Moharrami MA, Tachezy J. The contribution of the arginine dihydrolase pathway to energy metabolism by *Trichomonas vaginalis*. *Mol Biochem Parasitol* 1996; **78**: 117–125.

56. Schofield PJ, Edwards MR, Matthews J, Wilson JR. The pathway of arginine catabolism in *Giardia intestinalis*. *Mol Biochem Parasitol* 1992; **51**: 29–36.