SUPPLEMENTARY MATERIAL: **Different lanthanide elements induce strong gene expression changes in a lanthanide-accumulating methylotroph**

Linda Gorniak[1], Julia Bechwar[1], Martin Westermann[2], Frank Steiniger[2], Carl-Eric Wegner[1,#]

[1]Institute of Biodiversity, Aquatic Geomicrobiology, Friedrich Schiller University, Dornburger Str. 159, 07743 Jena, Germany

[2]Electron Microscopy Center, Jena University Hospital, Ziegelmühlenweg 1, 07743 Jena, Germany

Corresponding author:

Carl-Eric Wegner

Email address: carl-eric.wegner@uni-jena.de

Running title: Lanthanides cause strong gene expression changes

Keywords: lanthanides, lanthanome, RNAseq, EDX, TEM, FFTEM

SUPPLEMENTARY MATERIAL

Supplementary Information. Complementary information about used materials and applied methods.

We also provide details about sequence data processing and differential gene expression analysis via the Open Science Framework (https://osf.io/) (https://osf.io/p2nf6/?view_only=b83c7bbd806b43bdac419ebc8117eaa0).

Quantitative PCR (qPCR)

DNA was diluted, and one to 20 ng of genomic DNA was used as a template for qPCR. The copy numbers of the *lanM* gene (coding for lanmodulin) were determined using a CFX96 instrument (Bio-Rad, Munich, Germany), Brilliant II SYBR® Green QPCR Master Mix (Agilent Technologies, Germany) to deduce cell numbers. *lanM* is a single-copy gene in Beijerinckiaceae bacterium RH AL1. We designed a primer set *lanM_f1/lanM_r1* (*lanM_f1:* 5'-GGATTTTCGCAAGACCCTTT-3', *lanM_r1:* 5'-CTTCTTGAACTTGGCCTTGG-3') and used it with the following cycling conditions: 15 min. denaturation at 94°C, followed by 30 cycles of 30 s at 94°C, 60 s at 55°C, and 45 s at 72°C. Plasmids containing a *lanM* fragment from strain AL1 were used for standard curves. Standard curves were linear from 5×10^8 to 5×10^2 copies with R^2 values and PCR efficiencies above 0.99 and 80%, respectively. Dilution series were used to assess potentially present PCR inhibitors. Melt curve analysis was done to verify PCR specificity. qPCR was done based on three biological replicates and three technical replicates per biological replicate.

RNAseq data pre-processing

Adaptor- and quality-trimming (settings: minlen= 75, qtrim = rl, ktrim = rl, k = 25, mink= 11, trimq = 20, qtrim = rl) were carried out with *bbduk* (v38.26) (1) using its included database of common sequence contaminants and adapters. Trimmed reads were filtered with *SortMeRNA* (v2.1) (2) and the SILVA (3) and Rfam (4) databases by removing rRNA-derived and non-coding RNA sequences. Read mapping onto the available reference genome of Beijerinckiaceae bacterium RH AL1 (EBI accession no. LR590083 [genome] and LR699074 [plasmid]) (5) was done with *bbmap* (v.38.26) (settings: slow, k = 11). The number of mapped reads per feature (e.g. coding genes) was determined from *.bam* files, that have been sorted and indexed with *samtools* (v1.3.1) (6), using *featureCounts,* which is part of the *Subread* package (v1.6.3) (7, 8). A snakemake workflow for sequence data processing is available on *github* (https://github.com/wegnerce/smk_rnaseq).

Differential gene expression analysis

Pseudo-counts ($\log_2$(counts+1)) (**Figure S1**) were generated for subsequent data exploration by means of MA (mean of the normalized counts versus the $\log_2$FC [fold change] for all genes tested) plots. Multidimensional scaling plots were generated using the *plotMDS* function of *limma* (v. 3.50.0) (9). Inter-dataset relationships were in addition assessed based on the hierarchical clustering of scaled read count data (**Figure S2**). Scatterplot matrices were plotted with the *plotSM* function of *bigPint* (v. 1.10.0) (10) to check how replicates of conditions behaved when compared to each other and to the replicates of other conditions. Genes with insufficient read count data were filtered with the *filterByExpr* function which is part of *edgeR* (v. 3.20.9)

(11). Libraries were normalized by calculating normalization factors and estimating dispersion. The biological coefficient of variation (BCV) (**Figure S3**) was used to identify the uncertainty regarding transcript abundance based on replicate groups. P-values derived from the exact test proposed by Robinson and Smyth (12) have been assessed for all carried-out comparisons (**Figure S4**).

Transmission elektron microscopy (TEM)

Biomass was harvested by centrifugation and fixed overnight at 4°C by use of glutaraldehyde (2.5 %, v/v) in cacodylate buffer (100 mM, pH 7.3). Cells were washed with cacodylate buffer three times and incubated 5, 10, and 15 minutes prior to pelleting the biomass. All centrifugation steps were performed at room temperature and 4000 × g for 10 minutes. The pellets were resuspended and kept in 500 µL cacodylate buffer at 4°C. Further processing was according to Wegner et al., 2021 (13). Fixed samples were dehydrated in an ethanol series and stained with 2% (w/v) uranyl acetate in 50% (v/v) ethanol. Araldite resin (Plano, Wetzlar, Germany) was used for embedding samples. Ultrathin sections (70 nm thickness) were cut using an ultramicrotome Ultracut E (Reichert-Jung, Vienna, Austria) and mounted on Formvar-carbon coated 100 mesh grids (Quantifoil, Großlöbichau, Germany). The ultrathin sections were stained with lead nitrate for 10 minutes, examined in a Zeiss CEM 902 A electron microscope (Carl Zeiss AG,Oberkochen, Germany), and imaged using a TVIPS 1k Fast-Scan CCD-Camera (TVIPS, Munich, Germany). Ultrathin sections and freeze-fracture replicas were examined in a digital Zeiss EM 900 electron microscope (Zeiss, Oberkochen, Germany; digital upgrade by Point Electronic, Halle, Germany) operated at 80 kV. Digitized images were taken

with a wide-angle dual-speed 2K CCD camera controlled by the Sharp:Eye base controller and operated by the Image SP software (TRS, Moorenweis, Germany).

Freeze fracture transmission electron microscopy (FFTEM)

Glutaraldehyde-fixed biomass was supplied with the cryoprotectant glycerol (final concentration 15%), and 2 µL samples were transferred into a pair of copper hat-type carriers. Copper sandwiches were incubated for 1 minute in a cooled propane-ethane mix to rapidly plunge-freeze the samples before transferring them into liquid nitrogen. A BAF400T freeze fracture unit (BAL-TEC, Liechtenstein) supplied with a double-replica stage was used to freeze-fracture the samples at -110°C and high vacuum conditions ($<10^{-4}$ Pa). Platinum/carbon replicates were created by use of an electron beam vaporizer (1350 V - 1950 V, 89 mA - 93 mA), by adding a 2 nm platin layer at a 35° angle onto the fractured surface, followed by vertical evaporation with coal (ca. 20 nm layer). Cell material was removed from replicas by use of sodium dodecyl sulfate. Replicas were washed four times in distilled water and transferred onto uncoated EM-grids for examination through TEM as outlined above.

Fig S1. Comparison of pseudocount distributions between RNAseq data sets. Pseudocounts were determined gene-wise by calculating Log_2([count of mapped reads]+1).

Fig S2. Hierarchical clustering of RNAseq data sets. The clustering was based on a distance matrix that was generated by subtracting the Spearman correlation between data sets from 1. The resulting distances ranged between 0 and 2. Spearman correlations were calculated from scaled CPM (counts per million).

Fig S3. Biological coefficient of variation (BCV) plot. The BCV indicates the variation coefficient with which the gene transcripts' true abundance varies between replicate RNAseq data sets. Log_2CPM = log_2 counts per million.

Fig S4. P-value distributions resulting from testing for differential gene expression between groups of RNAseq data sets. The p-values were derived from the exact test proposed by Robinson and Smyth (12) and implemented in *edgeR (11)*.

Fig S5. Soft-agar based motility assay. Fourteen days after inoculation, diameters of cell intrusion into the soft-agar medium were measured and respective areas were calculated. For each condition (supplementation with La, Nd, or lanthanide cocktail), increasing lanthanide concentration led to a decrease in motility.

Fig S6. Deconvolution of EDX spectra. (A) The deconvolution for La deposits was done based on measurements published previously (13) (left panel), and the right panel shows the deconvolution for an exemplary Nd deposit. (B) The deconvolution

for the deposit originating from cells grown with the Ln cocktail is shown for two different energy ranges. The panel on the right provides a more zoomed-in representation of the spectrum. We did three replicate measurements for each condition using different deposits and cells (**Table S15**).

Table S1. Members of the lanthanide series.

Table S2. Ln concentrations in MM2 supplemented with and without the prepared, equimolarly pooled Ln cocktail. No carbon source and no biomass were added. Concentrations were determined through ICP-MS. Values are based on three independent measurements. MM2 without Ln cocktail contained no detectable Ln. Conc. = concentrations, SD = standard deviation.

Table S3. Optical density readings of the incubations carried out for downstream RNAseq-based gene expression analysis. Two sets of incubations were done. The left table refers to Beijerinckiceae bacterium RH AL1 grown in MM2 medium with methanol as the carbon source (0.5%, v/v) and supplemented with either 50 nM or 1 µM La. The right table refers to strain RH AL1 grown in MM2 medium with methanol as the carbon source (0.5%, v/v) and supplemented with either 1 µM La, 1 µM Nd, or a cocktail of light and heavy lanthanides (Ce, Nd, Dy, Ho, Er, Yb; 0.9 µM Ln **Table S2**). Control refers to cultures incubated without added Ln. Incubations were done in triplicates. Time points used for calculating growth parameters are marked. STDEV = standard deviation.

Table S4. Quantitative PCR targeting the *lanM* gene. Time refers to the timepoints indicated in **Figure 2A+B**. Condition refers to the experimental condition and experiment refers to the two incubation setups: (1) low (50 nM) and high (1 µM) La, and (2) different Ln species (La, Nd, equimolarly pooled Ln cocktail; see the materials and methods section for details).

Table S5. Overview of gene expression data of Beijerinckiaceae bacterium RH AL1 grown with methanol (0.5%, v/v) as carbon source. And supplemented with either two different concentrations of La (50 nM or 1 µM) or different Ln elements (1 µM La, 1 µM Nd, equimolarly pooled Ln cocktail [Ce, Nd, Dy, Ho, Er, Yb; 0.9 µM Ln **Table S2**). Gene expression is given in $\log_2$ counts per million (CPM). La.1-3 = Strain RH AL1 grown with 1 µM La, Nd.1-3 = Strain RH AL1 grown with 1 µM Nd, Cocktail.1-3 = Strain RH AL1 grown with an equimolarly pooled cocktail of heavy and light Ln (0.9 µM), La_low.1-3 = Strain RH AL1 grown with 50 nM La, La_high.1-3 = Strain RH AL1 grown with 1 µM. The latter two groups of data sets originate from a different incubation than the first free groups of data sets.

Table S6. Overview of differentially expressed genes in Beijerinckiaceae bacterium RH AL1 when grown with either 50 nM or 1 µM La. logFC = $\log_2$Fold change, $\log_2$CPM = $\log_2$ Counts per million, FDR = false discovery rate.

Table S7. Overview of differentially expressed genes in Beijerinckiaceae bacterium RH AL1 when grown with either 1 µM La or 1 µM Nd. $\log_2$FC = $\log_2$ Fold change, $\log_2$CPM = $\log_2$ Counts per million, FDR = false discovery rate.

Table S8. Overview of differentially expressed genes in Beijerinckiaceae bacterium RH AL1 when grown with either 1 µM La or 0.9 µM of an equimolarly pooled lanthanide cocktail (Ce, Nd, Dy, Ho, Er, Yb). log_2FC = log_2 Fold change, log_2CPM = log_2 Counts per million, FDR = false discovery rate.

Table S9. Overview of differentially expressed genes in Beijerinckiaceae bacterium RH AL1 when grown with either 1 µM Nd or 0.9 µM of an equimolarly pooled lanthanide cocktail (Ce, Nd, Dy, Ho, Er, Yb). log_2FC = log_2 Foldchange, log_2CPM = log_2 Counts per million, FDR = false discovery rate.

Table S10. Overrepresentation and gene set enrichment analysis. Based on KEGG annotations and gene sets shared or unique for individual comparisons, we determined overrepresented/enriched functions with *clusterProfiler (14, 15)*. The gene sets considered are labelled A through F.

Table S11. Presence-absence matrix for differential gene expression in response to different Ln elements and La concentration. To summarize the results, only genes associated with selected aspects of metabolism were included, e.g. motility and chemotaxis, carbohydrate metabolism, PHA metabolism, and transport. The color code indicates if corresponding genes are differentially expressed for individual comparisons.

Table S12. Results from the carried out motility assay with diameters and calculated areas of cell intrusion into the MM2 soft-agar medium after 14 days of incubation. Data series (La, Nd, Ln cocktail) were normally distributed (p-value < 0.29,

Kolmogorov-Smirnov test) without outliers (p-values < 0.29, *Dean-Dixon test*). They showed a decreasing trend (p-value < 0.89, *Neumann trend test*) of motility with increasing lanthanide concentrations. Statistic tests were performed based on a 95% confidence interval.

Table S13. Overview of gene expression data for lanthanome(-related genes). Gene expression is given in log2 counts per million (CPM). Gene name refers to the gene name in *M. extorquens AM1*. Genes were grouped as described in the main text. Gene annotation data was taken from Wegner et al. 2020 (5).

Table S14. Lanthanome(-related) genes in *M. extorquens* AM1 and Beijerinckiaceae bacterium RH AL1. Gene homologs in strain AL1 have been identified by *blastp* (16) searches of AM1 amino acid sequences against a protein database containing all amino acid sequences encoded in the genome of strain AL1. The queried sequences comprised relevant gene products identified previously (17), especially gene products of the lut-cluster, as well as LC cluster gene products described by Zytnick and colleagues (18). pident = percentage identity, qstart = query start, qend = query end, sstart = subject start, send = subject end.

Table S15. Results from the deconvolution analysis of obtained EDX spectra. EDX-based elemental analysis was done for periplasmic deposits identified in cells grown with Nd (n=3), the lanthanide cocktail (n=4), and La (n=3, (13)). P:Ln ratios were calculated based on respective At % values. The Ln % values originating from the deconvolution analysis of Ln cocktail details have been summarized in a separate table.

Table S16. Share of Ln left (in percent) in cell-free Beijerinckiaceae bacterium RH AL1 supernatant samples after cultivation in MM2 with methanol (0.5%, v/v) as the carbon source and the Ln cocktail (Ce, Nd, Dy, Ho, Er, Yb; 0.9 µM **Table S2**). The original Ln content was determined by ICP-MS analysis of MM2 medium supplemented with the Ln cocktail, the remaining Ln were determined by calculating Ln depletion based on ICP-MS data for cell-free supernatant samples taken during late-exponential phase. AVG = average (of three biological triplicates and three technical replicates per biological replicate), STDEV = standard deviation.

Note: All supplementary tables are provided in one combined spreadsheet.

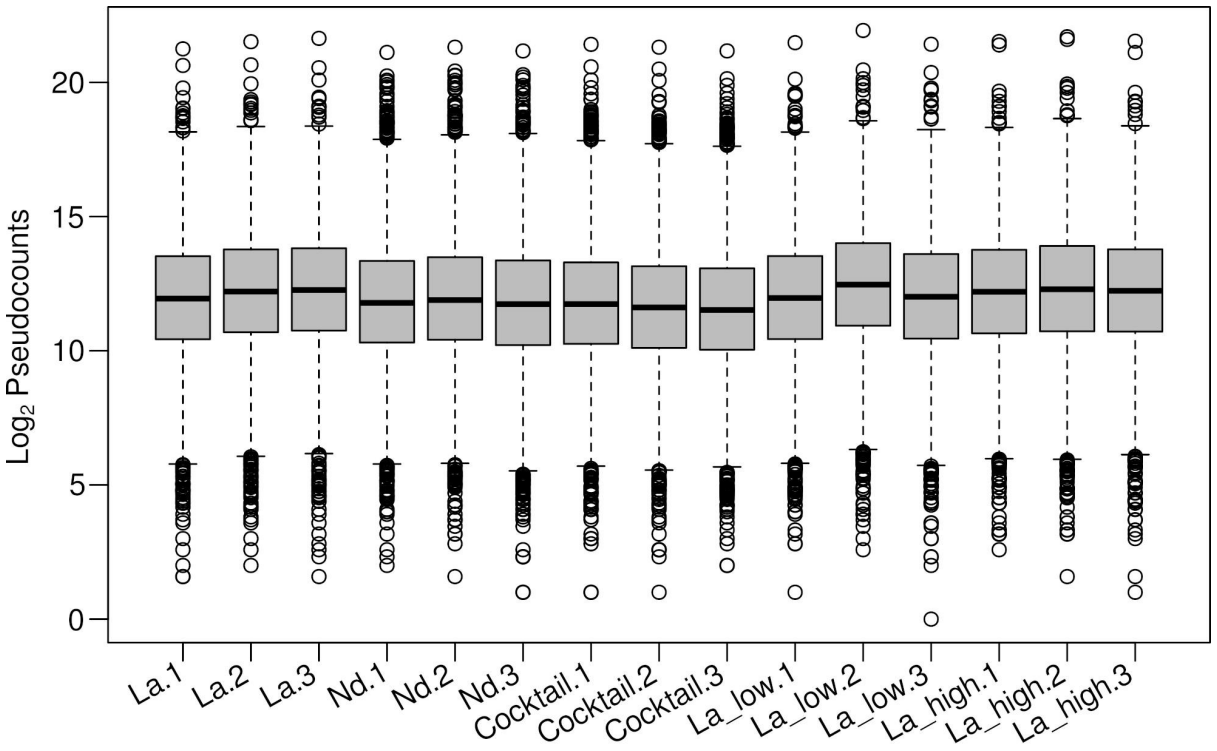
Log2 Pseudocounts
0
5
10
15
20
La.1
La.2
La.3
Nd.1
Nd.2
Nd.3
Cocktail.1
Cocktail.2
Cocktail.3
La_low.1
La_low.2
La_low.3
La_high.1
La_high.2
La_high.3

Fig. S1

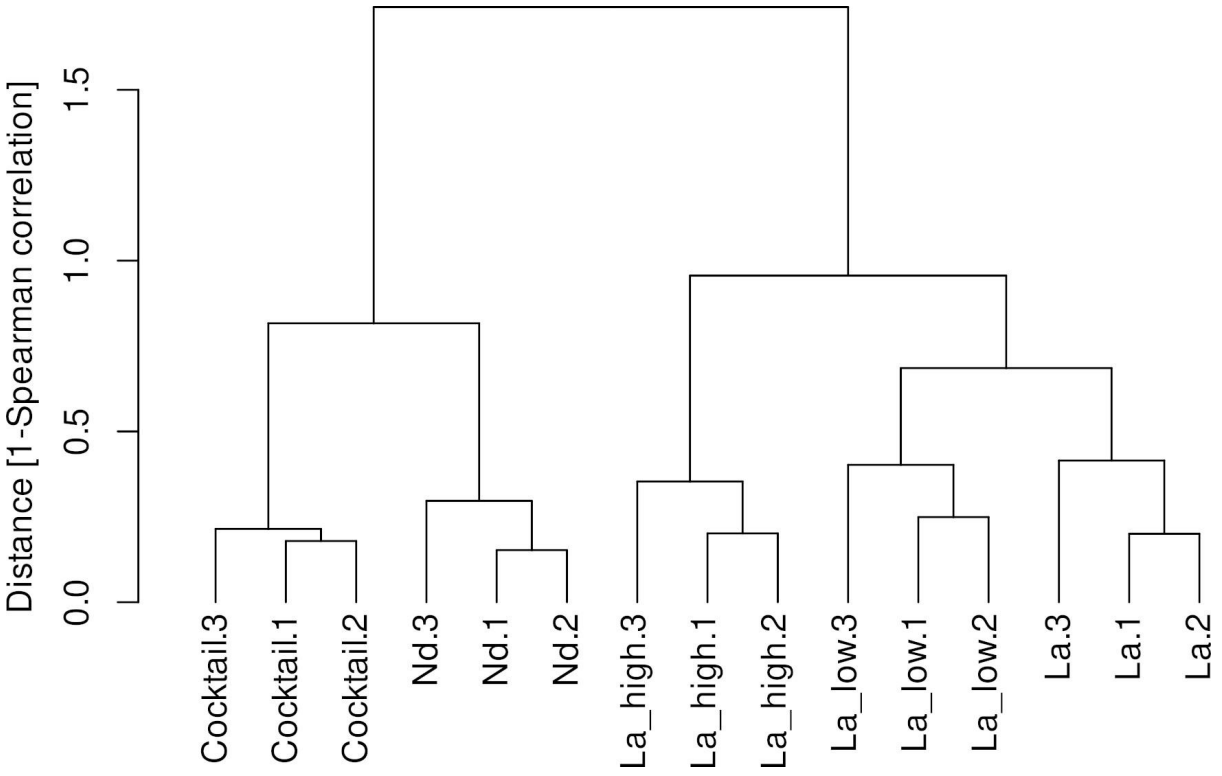
Distance [1-Spearman correlation]
0.0
0.5
1.0
1.5
Cocktail.3
Cocktail.1
Cocktail.2
Nd.3
Nd.1
Nd.2
La_high.3
La_high.1
La_high.2
La_low.3
La_low.1
La_low.2
La.3
La.1
La.2

Fig. S2

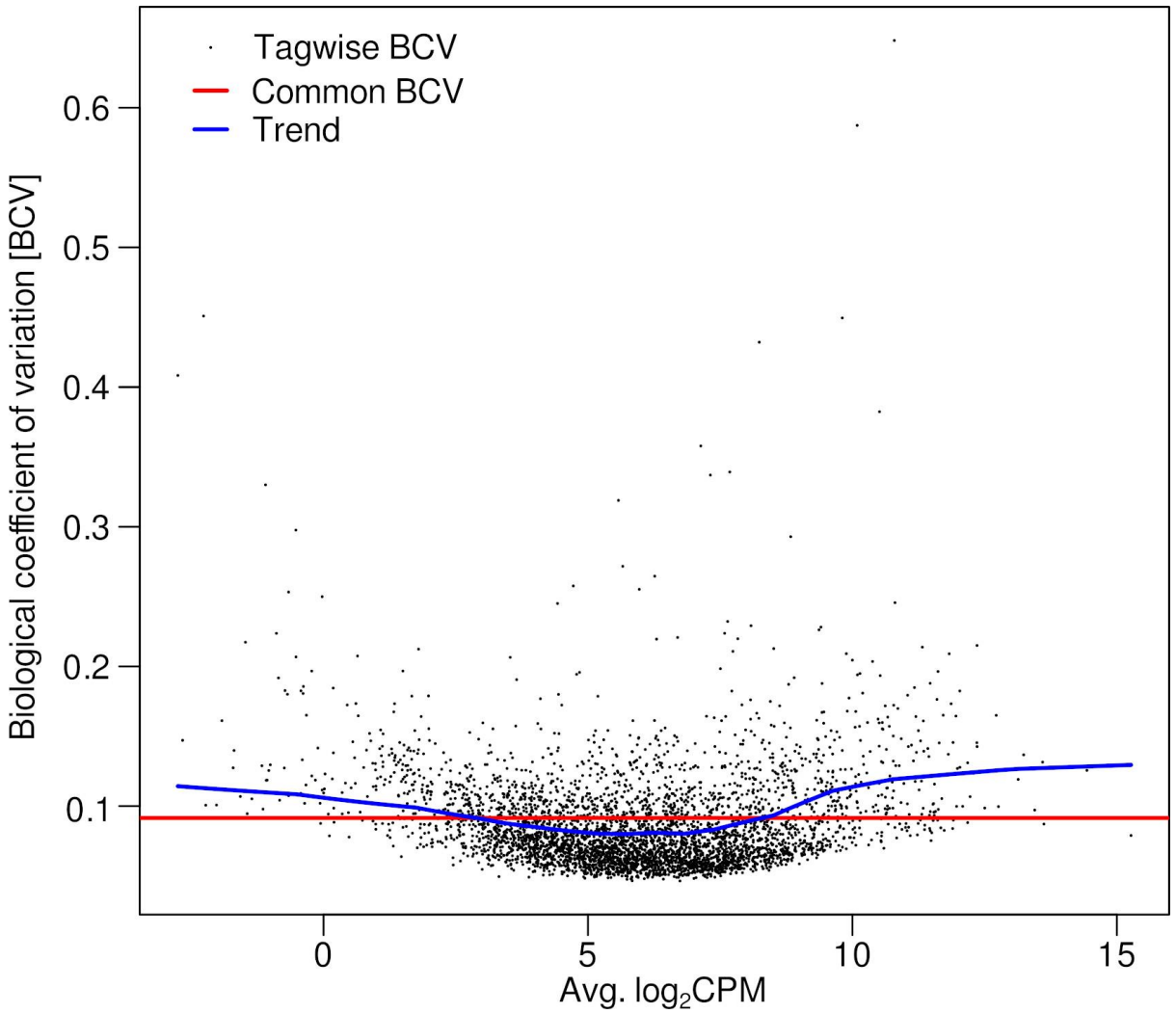
Tagwise BCV
Common BCV
Trend
Biological coefficient of variation [BCV]
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
Avg. log2CPM

Fig. S3

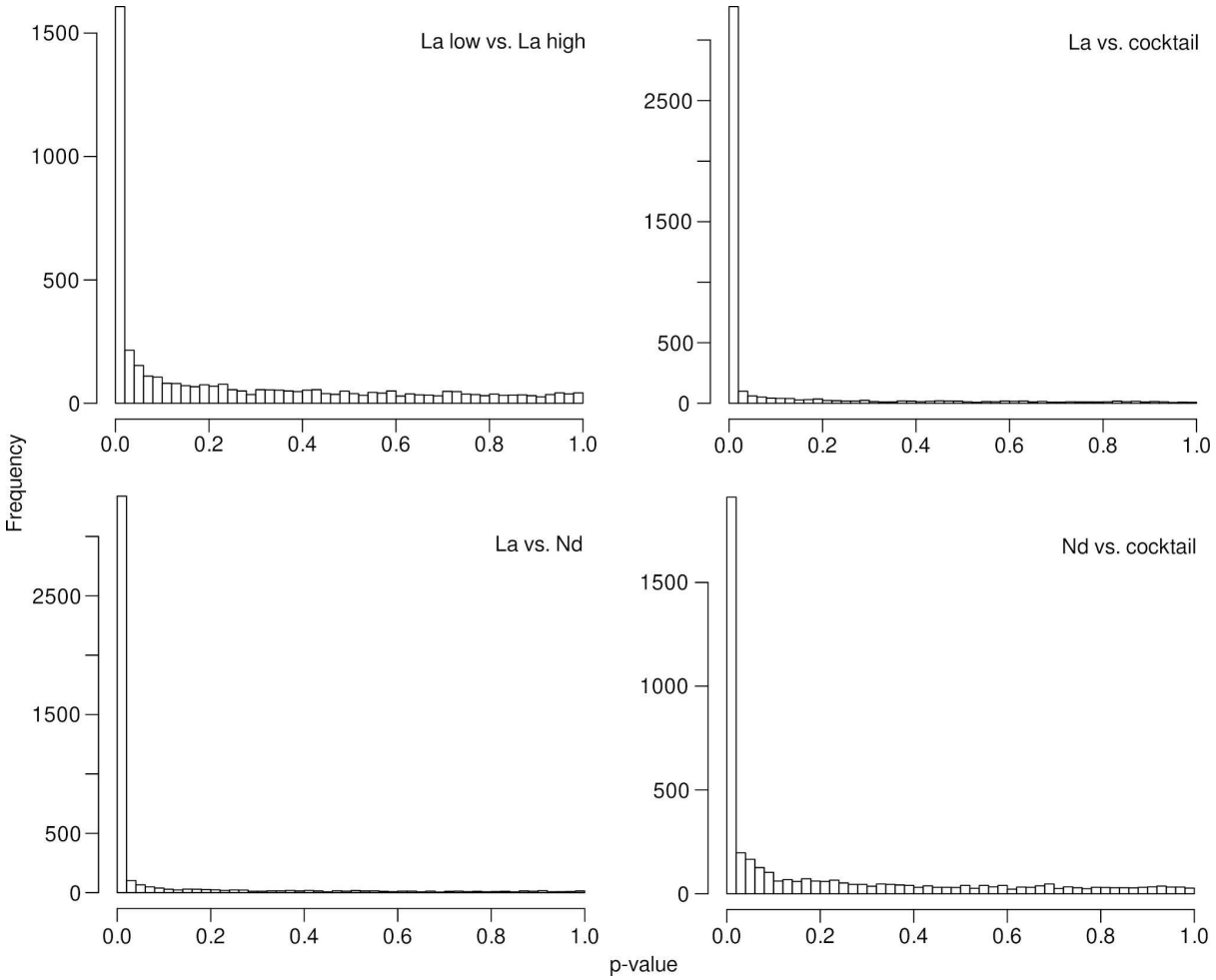
La low vs. La high
La vs. cocktail
La vs. Nd
Nd vs. cocktail
Frequency
p-value
0
500
1000
1500
2500
0.0
0.2
0.4
0.6
0.8
1.0

Fig. S4

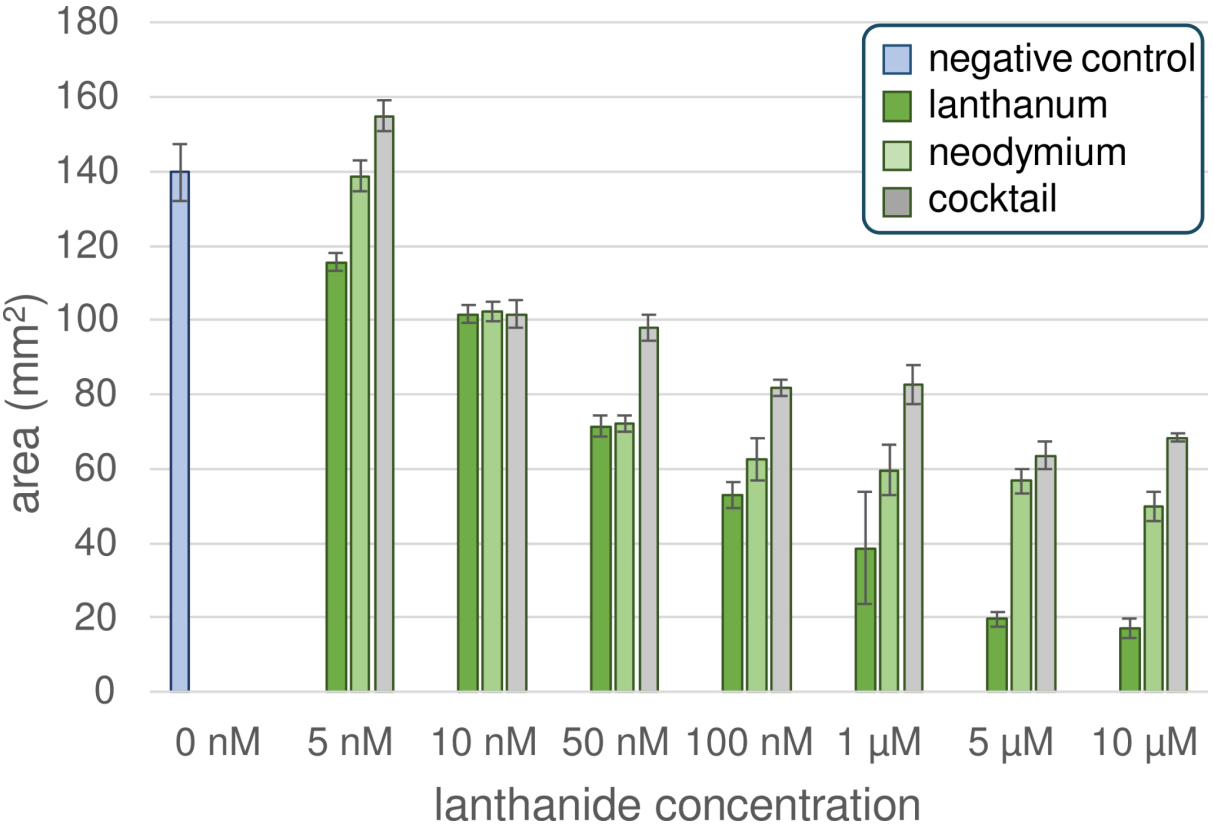
180
160
140
120
100
80
60
40
20
0
area (mm²)
negative control
lanthanum
neodymium
cocktail
0 nM
5 nM
10 nM
50 nM
100 nM
1 µM
5 µM
10 µM
lanthanide concentration

Fig. S5

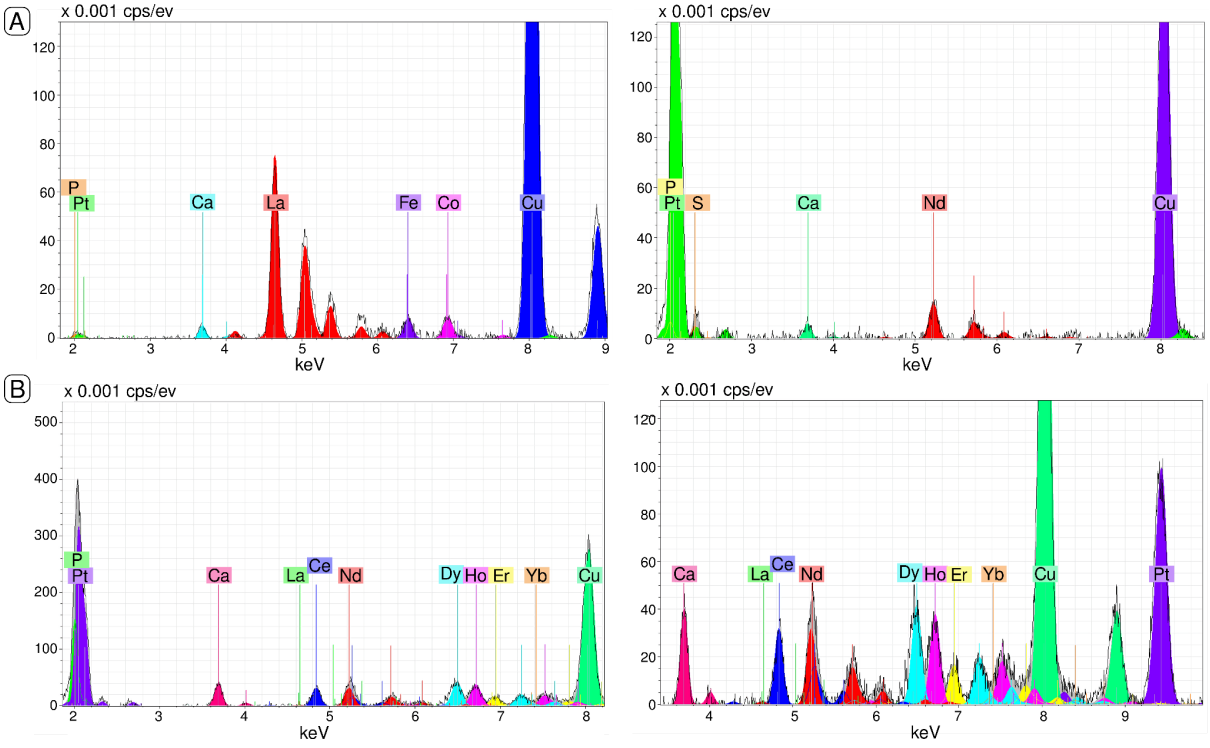
A
x 0.001 cps/ev
P
Pt
Ca
La
Fe
Co
Cu
keV
x 0.001 cps/ev
P
Pt
S
Ca
Nd
Cu
keV
B
x 0.001 cps/ev
P
Pt
Ca
La
Ce
Nd
Dy
Ho
Er
Yb
Cu
keV
x 0.001 cps/ev
Ca
La
Ce
Nd
Dy
Ho
Er
Yb
Cu
Pt
keV

Fig. S6

References

1. Bushnell B. 2016. BBMap short read aligner. https://www.sourceforge.net/projects/bbmap/.

2. Kopylova E, Noé L, Touzet H, Noe L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.

3. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–6.

4. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. 2013. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res 41:D226–32.

5. Wegner C-E, Gorniak L, Riedel S, Westermann M, Küsel K. 2020. Lanthanide-Dependent Methylotrophs of the Family Beijerinckiaceae: Physiological and Genomic Insights. Appl Environ Microbiol 86:e01830–19.

6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

7. Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41:e108.

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

9. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47.

10. Rutter L, Cook D. 2020. bigPint: A Bioconductor visualization package that makes big data pint-sized. PLoS Comput Biol 16:e1007912.

11. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

12. Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 9:321–332.

13. Wegner C-E, Westermann M, Steiniger F, Gorniak L, Budhraja R, Adrian L, Küsel K. 2021. Extracellular and Intracellular Lanthanide Accumulation in the Methylotroph Beijerinckiaceae Bacterium RH AL1. Appl Environ Microbiol 87:e03144–20.

14. Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16:284–287.

15. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) 2:100141.

16. Altschul SF, Madden TL, Schäffer A a., Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

17. Roszczenko-Jasińska P, Vu HN, Subuyuj GA, Crisostomo RV, Cai J, Lien NF, Clippard EJ, Ayala EM, Ngo RT, Yarza F, Wingett JP, Raghuraman C, Hoeber CA, Martinez-Gomez NC, Skovran E. 2020. Gene products and processes contributing to lanthanide homeostasis and methanol metabolism in Methylorubrum extorquens AM1. Sci Rep 10:12663.

18. Zytnick AM, Good NM, Barber CC, Phi MT, Gutenthaler SM, Zhang W, Daumann LJ, Cecilia

Martinez-Gomez N. 2022. Identification of a biosynthetic gene cluster encoding a novel lanthanide chelator in Methylorubrum extorquens AM1. bioRxiv.