



Microbiome Profiling Report

Functional profiles

Biomcare ApS

30/12/2024

Customer	Tove Mariegaard Pedersen
Customer ID	DA00204-23
Project	Markens mikrobielle samfund, Aar 3
Sample Type	Soil
Number of samples	40 samples
Type of data	Shotgun metagenomic sequencing

Processing of the Shotgun metagenomic sequencing data through quality control and microbiome profiling across the 40 samples resulted in the detection of microorganisms (see report 2 on taxonomy) and profiling of the functional capacity of the microbes. The software HUMAnN3 was used to profile the functional capacity in terms of MetaCyc pathways and UniRef50 gene families. The latter was regrouped to Gene Ontology (GO) and KEGG Orthology (KO) terms. The functional profiling is described in detail in “**Report 1: Sequencing and Data Processing Report**” and the different databases used for the grouping of the functions is described below.

- **MetaCyc Metabolic Pathways:** Metacyc is a curated database of metabolic pathways across all domains of life. It only includes pathways that have been experimentally studied (in contrast, many other databases also contain computationally predicted pathways). MetaCyc contains 2937 pathways from 3295 different organisms. Website: <https://metacyc.org/> (<https://metacyc.org/>)
- **KO:** The KO database is a database of functions grouped by orthology (i.e. groups of genes with the same function across different organisms). The KO terms link pathways and functions and thereby, pathways can be identified from the KO terms. Website: <https://www.genome.jp/kegg/ko.html> (<https://www.genome.jp/kegg/ko.html>)
- **GO:** The GO database contains functional information of genes. As compared to KO, GO is not structured around pathways but instead, the so-called ontologies (a specific way of structuring functional information and functional relationships). Each gene is associated with a GO term describing the function of the gene. Website: <http://geneontology.org/docs/ontology-documentation/> (<http://geneontology.org/docs/ontology-documentation/>)

These databases all have different strengths and limitations and therefore, to be as exhaustive as possible, we use all three databases for the regrouping of the UniRef50 gene families.

Functional capacity versus functional activity

There are two types of data we can use for profiling of the functional capacity of the microbes; shotgun metagenomic sequencing and metatranscriptomics, where the former is sequencing of the DNA and the latter is sequencing of RNA. There are some important differences between the information we obtain with these two types of data. With sequencing of DNA we obtain sequences from the whole genome (both coding and non-coding sections) whereas with sequencing of RNA we only obtain sequences from the coding part of the genome. Furthermore, when we investigate the DNA we look at the functional capacity and we do not obtain information of the expression level of the genes. In contrast, with RNA sequencing we can analyse the functional activity, i.e. the expression level of the genes. For example, a gene can have a low relative abundance at DNA level but a high relative abundance at RNA level due to a high expression of the gene. These considerations are important to keep in mind when interpreting the data shown in this report. In this report, we look at the functional capacity as we have performed DNA sequencing.

Most abundant functional terms/pathways

This section lists the most abundant functional terms/pathways detected in your data for each of the described databases. This section is meant to give you a first impression of the format of the functional data, the types of terms and pathways detected, and how they are named. To evaluate further, you can try to google the terms or look them up in the individual databases using the links given above in order to get a better understanding of their interpretation and usage.

The tables show the most abundant functional terms/pathways. These will often be the household functions and are often not the most interesting for you to study and evaluate further. Therefore, based on your question of interest, we also highlight some functional terms/pathways that may be of special interest in your project. These can be seen in the tables in the section “**Functional terms/pathways of interest**”.

The values in all of the following tables range from 1-100, with the exception of some values that sum to more than 100 due to the structure of the databases (i.e. one gene is found in more than one term). Due to the high fraction of unmapped reads and the high number of functional terms/pathways, the numbers are quite small.

Most abundant MetaCyc pathways

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
UNMAPPED	5.40E+01	5.61E+01	5.66E+01	5.66E+01	5.70E+01	5.91E+01
UNINTEGRATED	3.80E+01	4.01E+01	4.06E+01	4.05E+01	4.09E+01	4.30E+01
VALSYN-PWY: L-valine biosynthesis	3.21E-02	3.33E-02	3.41E-02	3.42E-02	3.50E-02	3.74E-02
ILEUSYN-PWY: L-isoleucine biosynthesis I (from threonine)	3.07E-02	3.22E-02	3.28E-02	3.29E-02	3.36E-02	3.59E-02
PWY-5103: L-isoleucine biosynthesis III	2.91E-02	3.09E-02	3.15E-02	3.16E-02	3.23E-02	3.48E-02

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
BRANCHED-CHAIN-AA-SYN-PWY: superpathway of branched chain amino acid biosynthesis	2.86E-02	3.05E-02	3.08E-02	3.10E-02	3.15E-02	3.39E-02
PWY-7111: pyruvate fermentation to isobutanol (engineered)	2.28E-02	2.45E-02	2.50E-02	2.51E-02	2.56E-02	2.81E-02
PWY-7229: superpathway of adenosine nucleotides de novo biosynthesis I	2.14E-02	2.24E-02	2.31E-02	2.31E-02	2.35E-02	2.51E-02
PWY-6126: superpathway of adenosine nucleotides de novo biosynthesis II	2.04E-02	2.16E-02	2.22E-02	2.23E-02	2.28E-02	2.44E-02
PWY-7222: guanosine deoxyribonucleotides de novo biosynthesis II	1.85E-02	2.12E-02	2.17E-02	2.16E-02	2.20E-02	2.45E-02
PWY-7220: adenosine deoxyribonucleotides de novo biosynthesis II	1.85E-02	2.12E-02	2.17E-02	2.16E-02	2.20E-02	2.45E-02
PWY-7228: superpathway of guanosine nucleotides de novo biosynthesis I	1.86E-02	2.08E-02	2.12E-02	2.13E-02	2.18E-02	2.35E-02
PWY-6969: TCA cycle V (2-oxoglutarate synthase)	1.91E-02	2.04E-02	2.11E-02	2.11E-02	2.14E-02	2.32E-02
PWY-3001: superpathway of L-isoleucine biosynthesis I	1.96E-02	2.06E-02	2.10E-02	2.11E-02	2.14E-02	2.27E-02
PWY-6277: superpathway of 5-aminoimidazole ribonucleotide biosynthesis	1.80E-02	2.02E-02	2.08E-02	2.09E-02	2.15E-02	2.39E-02
PWY-6122: 5-aminoimidazole ribonucleotide biosynthesis II	1.80E-02	2.02E-02	2.08E-02	2.09E-02	2.15E-02	2.39E-02
NONOXIPENT-PWY: pentose phosphate pathway (non-oxidative branch) I	1.87E-02	1.99E-02	2.06E-02	2.08E-02	2.17E-02	2.40E-02
PWY-6125: superpathway of guanosine nucleotides de novo biosynthesis II	1.76E-02	1.98E-02	2.02E-02	2.03E-02	2.09E-02	2.26E-02

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
CALVIN-PWY: Calvin-Benson-Bassham cycle	1.93E-02	1.99E-02	2.02E-02	2.04E-02	2.09E-02	2.20E-02
PWY-7221: guanosine ribonucleotides de novo biosynthesis	1.78E-02	1.97E-02	2.02E-02	2.02E-02	2.07E-02	2.26E-02
PWY-7977: L-methionine biosynthesis IV	1.82E-02	1.97E-02	2.01E-02	2.03E-02	2.07E-02	2.39E-02
PWY-7208: superpathway of pyrimidine nucleobases salvage	1.74E-02	1.95E-02	2.00E-02	2.01E-02	2.04E-02	2.29E-02

Table 1: Summary statistics for top 20 most abundant MetaCyc pathways. Summary statistics were computed for all pathways and the 20 most abundant were selected based on the median relative abundance across the samples.

The data set contains 641 different pathways, not considering the UNMAPPED and UNINTEGRATED.

- The UNMAPPED row contains information on the fraction of reads that did not map to a gene in the reference database
- The UNINTEGRATED row contains information on the fraction of reads that did map to a gene in the reference database, but where the gene is not contributing to a pathway in the used pathway database.

The UNMAPPED and UNINTEGRATED rows allow for evaluation of the fraction of data that does not contribute to the identified functions and therefore, is not used in further analysis. We do expect a fairly high number of UNINTEGRATED reads, as many known genes are not assigned to known pathways. Furthermore, for soil samples, a high percentage (>80%) of UNMAPPED reads is expected as the soil microbiome is very complex and not well-described in the reference databases.

Most abundant KO terms

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
UNMAPPED	5.40E+01	5.61E+01	5.66E+01	5.66E+01	5.70E+01	5.91E+01
UNGROUPED	3.75E+01	3.97E+01	4.01E+01	4.01E+01	4.06E+01	4.28E+01
K03088: RNA polymerase sigma-70 factor, ECF subfamily	2.52E-02	2.94E-02	3.22E-02	3.18E-02	3.44E-02	3.72E-02
K01990: ABC-2 type transport system ATP-binding protein	1.48E-02	1.62E-02	1.73E-02	1.72E-02	1.81E-02	2.00E-02

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
K02051: sulfonate/nitrate/taurine transport system substrate-binding protein	8.91E-03	1.03E-02	1.11E-02	1.12E-02	1.18E-02	1.42E-02
K12132: serine/threonine-protein kinase WNK1	8.35E-03	1.05E-02	1.10E-02	1.08E-02	1.15E-02	1.25E-02
K01992: ABC-2 type transport system permease protein	9.49E-03	1.05E-02	1.10E-02	1.12E-02	1.17E-02	1.37E-02
K02050: sulfonate/nitrate/taurine transport system permease protein	7.41E-03	8.65E-03	9.03E-03	9.11E-03	9.60E-03	1.21E-02
K01999: branched-chain amino acid transport system substrate-binding protein	7.24E-03	8.08E-03	8.65E-03	8.67E-03	8.96E-03	1.29E-02
K04078: chaperonin GroES	7.74E-03	8.18E-03	8.52E-03	8.50E-03	8.90E-03	9.22E-03
K02003: NO_NAME	6.89E-03	7.77E-03	8.20E-03	8.22E-03	8.70E-03	1.01E-02
K03704: cold shock protein (beta-ribbon, CspA family)	7.26E-03	7.64E-03	8.07E-03	8.08E-03	8.38E-03	9.26E-03
K01997: branched-chain amino acid transport system permease protein	6.55E-03	7.43E-03	7.92E-03	7.90E-03	8.33E-03	1.15E-02
K00059: 3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]	6.77E-03	7.61E-03	7.85E-03	7.85E-03	8.14E-03	8.60E-03
K01996: branched-chain amino acid transport system ATP-binding protein	6.21E-03	7.16E-03	7.42E-03	7.53E-03	7.83E-03	1.02E-02
K02518: translation initiation factor IF-1	6.41E-03	6.97E-03	7.24E-03	7.26E-03	7.60E-03	8.28E-03
K01358: ATP-dependent Clp protease, protease subunit [EC:3.4.21.92]	6.20E-03	6.96E-03	7.22E-03	7.32E-03	7.60E-03	8.64E-03
K00799: glutathione S-transferase [EC:2.5.1.18]	5.78E-03	6.96E-03	7.22E-03	7.26E-03	7.69E-03	9.14E-03
K02078: acyl carrier protein	6.31E-03	6.76E-03	7.09E-03	7.10E-03	7.37E-03	8.02E-03
K02004: NO_NAME	5.16E-03	6.04E-03	7.07E-03	7.05E-03	7.72E-03	9.40E-03

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
K02033: peptide/nickel transport system permease protein	5.86E-03	6.31E-03	6.68E-03	6.62E-03	6.87E-03	7.68E-03
K00331: NADH-quinone oxidoreductase subunit B [EC:1.6.5.3]	5.80E-03	6.42E-03	6.58E-03	6.69E-03	6.96E-03	8.11E-03

Table 2: Summary statistics for top 20 most abundant KO terms. Summary statistics were computed for all KO terms and the 20 most abundant were selected based on the median relative abundance across the samples (KO terms assigned as “subunit ribosomal protein” were not included in the table as these are highly abundant and not that informative).

The data set contains 7269 different KO terms, not considering the UNMAPPED and UNGROUPED.

- The UNMAPPED row contains information on the fraction of reads that did not map to a gene in the reference database
- The UNGROUPED row contains information on the fraction of reads that did map to a gene in the reference database, but where the gene is not contributing to a pathway in the used pathway database.

The UNMAPPED and UNGROUPED rows allow for evaluation of the fraction of data that does not contribute to the identified functions and therefore, is not used in further analysis. We do expect a fairly high number of UNGROUPED reads, as many known genes are not part of a known KO term. Furthermore, for soil samples, a high percentage (>80%) of UNMAPPED reads is expected as the soil microbiome is very complex and not well-described in the reference databases.

Most abundant GO terms

The GO terms are split into three main categories:

- Biological process [BP]
- Molecular function [MF]
- Cellular component [CC]

Of these three categories we find the biological process most informative and thus, we selected the top 20 most abundant BP terms from the GO data set. For more information about the individual GO terms, we refer to the QuickGo homepage: <https://www.ebi.ac.uk/QuickGO/> (<https://www.ebi.ac.uk/QuickGO/>).

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
UNMAPPED	5.40E+01	5.61E+01	5.66E+01	5.66E+01	5.70E+01	5.91E+01
UNGROUPED	1.54E+01	1.64E+01	1.68E+01	1.67E+01	1.70E+01	1.76E+01
GO:0055085 (GO:0055085): [BP] transmembrane transport	6.64E-01	7.24E-01	7.36E-01	7.35E-01	7.47E-01	7.83E-01

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0006355 (GO:0006355): [BP] regulation of transcription, DNA-templated	5.49E-01	5.76E-01	5.86E-01	5.84E-01	5.94E-01	6.18E-01
GO:0006412 (GO:0006412): [BP] translation	4.56E-01	4.80E-01	4.90E-01	4.92E-01	5.06E-01	5.29E-01
GO:0000160 (GO:0000160): [BP] phosphorelay signal transduction system	4.19E-01	4.55E-01	4.65E-01	4.62E-01	4.73E-01	4.87E-01
GO:0005975 (GO:0005975): [BP] carbohydrate metabolic process	3.80E-01	4.37E-01	4.49E-01	4.47E-01	4.58E-01	4.95E-01
GO:0006313 (GO:0006313): [BP] transposition, DNA- mediated	1.95E-01	2.65E-01	2.96E-01	2.95E-01	3.12E-01	4.66E-01
GO:0006310 (GO:0006310): [BP] DNA recombination	2.40E-01	2.53E-01	2.69E-01	2.69E-01	2.83E-01	3.06E-01
GO:0006281 (GO:0006281): [BP] DNA repair	2.26E-01	2.57E-01	2.63E-01	2.64E-01	2.70E-01	2.96E-01
GO:0015074 (GO:0015074): [BP] DNA integration	1.78E-01	2.28E-01	2.60E-01	2.56E-01	2.75E-01	3.83E-01
GO:0006352 (GO:0006352): [BP] DNA-templated transcription, initiation	2.17E-01	2.42E-01	2.48E-01	2.48E-01	2.56E-01	2.75E-01
GO:0009058 (GO:0009058): [BP] biosynthetic process	2.14E-01	2.34E-01	2.40E-01	2.40E-01	2.46E-01	2.57E-01
GO:0006260 (GO:0006260): [BP] DNA replication	1.45E-01	1.62E-01	1.64E-01	1.65E-01	1.70E-01	1.81E-01
GO:0051301 (GO:0051301): [BP] cell division	1.45E-01	1.55E-01	1.57E-01	1.59E-01	1.63E-01	1.81E-01
GO:0045454 (GO:0045454): [BP] cell redox homeostasis	1.45E-01	1.51E-01	1.54E-01	1.54E-01	1.57E-01	1.65E-01
GO:0055114 (GO:0055114): [BP] oxidation-reduction process	1.27E-01	1.32E-01	1.35E-01	1.35E-01	1.37E-01	1.48E-01
GO:0006099 (GO:0006099): [BP] tricarboxylic acid cycle	1.05E-01	1.19E-01	1.21E-01	1.21E-01	1.23E-01	1.33E-01

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0006457 (GO:0006457): [BP] protein folding	1.01E-01	1.05E-01	1.07E-01	1.07E-01	1.08E-01	1.11E-01
GO:0006807 (GO:0006807): [BP] nitrogen compound metabolic process	8.70E-02	9.78E-02	9.97E-02	9.95E-02	1.02E-01	1.12E-01
GO:0008360 (GO:0008360): [BP] regulation of cell shape	8.57E-02	9.45E-02	9.68E-02	9.78E-02	1.01E-01	1.12E-01
GO:0006633 (GO:0006633): [BP] fatty acid biosynthetic process	8.60E-02	9.32E-02	9.63E-02	9.63E-02	9.97E-02	1.06E-01

Table 3: Summary statistics for top 20 most abundant KO terms. Summary statistics were computed for all GO terms in the BP category and the 20 most abundant were selected based on the median relative abundance across the samples.

The data set contains 4535 different GO terms in the BP category, not considering the UNMAPPED and UNGROUPED.

- The UNMAPPED row contains information on the fraction of reads that did not map to a gene in the reference database
- The UNGROUPED row contains information on the fraction of reads that did map to a gene in the reference database, but where the gene is not contributing to a pathway in the used pathway database.

The UNMAPPED and UNGROUPED rows allow for evaluation of the fraction of data that does not contribute to the identified functions and therefore, is not used in further analysis. For soil samples, a high percentage (>80%) of UNMAPPED reads is expected as the soil microbiome is very complex and not well-described in the reference databases.

Functional terms of interest

To obtain an overview of functions of special interest in your project, we selected BP and GO terms related to the following functions: nitrogen metabolism, sulfur metabolism, methane metabolism, carbon fixation, symbiosis and photosynthesis. The selected functions are not exhaustive however, it provides a good overview of which aspects of the functional capacity that might be interesting to zoom in on.

Selected BP terms

To extract BP terms of special interest, we searched the BP terms using the following keywords: “nitrogen”, “nitrate”, “nitrite”, “nitrification”, “ammonium”, “ammonia”, “ammonification”, “urea”, “carbon”, “phosphorus”, “phosphate”, “methane”, “sulfate”, “sulfur”, “potassium”, “symbiotic”, “photosynthesis”, “respiration” and “starvation”. For more information about the individual GO terms, we refer to the QuickGo homepage: <https://www.ebi.ac.uk/QuickGO/> (<https://www.ebi.ac.uk/QuickGO/>).

Nitrogen

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0006807 (GO:0006807): [BP] nitrogen compound metabolic process	8.70E-02	9.78E-02	9.97E-02	9.95E-02	1.02E-01	1.12E-01
GO:0042128 (GO:0042128): [BP] nitrate assimilation	3.24E-02	3.74E-02	3.85E-02	3.81E-02	3.92E-02	4.08E-02
GO:0006808 (GO:0006808): [BP] regulation of nitrogen utilization	2.04E-02	2.13E-02	2.18E-02	2.19E-02	2.25E-02	2.35E-02
GO:0071705 (GO:0071705): [BP] nitrogen compound transport	1.29E-02	1.52E-02	1.61E-02	1.61E-02	1.66E-02	1.86E-02
GO:0009399 (GO:0009399): [BP] nitrogen fixation	1.18E-02	1.29E-02	1.37E-02	1.38E-02	1.42E-02	2.19E-02
GO:0019740 (GO:0019740): [BP] nitrogen utilization	7.57E-04	1.01E-03	1.15E-03	1.14E-03	1.25E-03	1.48E-03
GO:0042126 (GO:0042126): [BP] nitrate metabolic process	5.69E-04	8.01E-04	9.03E-04	9.47E-04	1.04E-03	1.45E-03
GO:0019333 (GO:0019333): [BP] denitrification pathway	3.74E-04	6.05E-04	6.60E-04	6.93E-04	8.05E-04	1.02E-03
GO:0019676 (GO:0019676): [BP] ammonia assimilation cycle	4.09E-04	4.64E-04	4.87E-04	4.87E-04	5.15E-04	5.48E-04
GO:1901565 (GO:1901565): [BP] organonitrogen compound catabolic process	0.00E+00	1.87E-05	2.78E-05	2.89E-05	3.52E-05	8.82E-05
GO:0006995 (GO:0006995): [BP] cellular response to nitrogen starvation	0.00E+00	7.96E-06	1.52E-05	1.61E-05	2.36E-05	4.74E-05
GO:1901698 (GO:1901698): [BP] response to nitrogen compound	0.00E+00	0.00E+00	6.75E-06	8.14E-06	1.42E-05	2.86E-05
GO:1902025 (GO:1902025): [BP] nitrate import	0.00E+00	0.00E+00	0.00E+00	6.38E-07	0.00E+00	2.55E-05
GO:0090294 (GO:0090294): [BP] nitrogen catabolite activation of transcription	0.00E+00	0.00E+00	0.00E+00	4.05E-06	0.00E+00	4.17E-05

GO:0090293 (GO:0090293): [BP] nitrogen catabolite regulation of transcription	0.00E+00	0.00E+00	0.00E+00	2.63E-06	0.00E+00	2.15E-05
GO:0071417 (GO:0071417): [BP] cellular response to organonitrogen compound	0.00E+00	0.00E+00	0.00E+00	8.98E-07	1.66E-06	6.58E-06
GO:0051410 (GO:0051410): [BP] detoxification of nitrogen compound	0.00E+00	0.00E+00	0.00E+00	2.19E-07	0.00E+00	8.74E-06
GO:0043562 (GO:0043562): [BP] cellular response to nitrogen levels	0.00E+00	0.00E+00	0.00E+00	6.28E-06	1.10E-05	4.17E-05
GO:0015707 (GO:0015707): [BP] nitrite transport	0.00E+00	0.00E+00	0.00E+00	7.14E-06	1.37E-05	3.69E-05
GO:0015706 (GO:0015706): [BP] nitrate transport	0.00E+00	0.00E+00	0.00E+00	1.86E-07	0.00E+00	7.45E-06
GO:0015696 (GO:0015696): [BP] ammonium transport	0.00E+00	0.00E+00	0.00E+00	1.14E-06	0.00E+00	2.85E-05
GO:0010243 (GO:0010243): [BP] response to organonitrogen compound	0.00E+00	0.00E+00	0.00E+00	3.61E-06	7.06E-06	1.47E-05
GO:0010167 (GO:0010167): [BP] response to nitrate	0.00E+00	0.00E+00	0.00E+00	1.86E-07	0.00E+00	7.45E-06

Urea						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0043419 (GO:0043419): [BP] urea catabolic process	6.94E-03	8.24E-03	8.58E-03	8.49E-03	8.79E-03	9.56E-03
GO:0019627 (GO:0019627): [BP] urea metabolic process	2.05E-03	2.28E-03	2.38E-03	2.38E-03	2.48E-03	3.09E-03
GO:0000050 (GO:0000050): [BP] urea cycle	7.26E-04	8.80E-04	9.06E-04	9.18E-04	9.60E-04	1.11E-03
GO:0071918 (GO:0071918): [BP] urea transmembrane transport	0.00E+00	0.00E+00	0.00E+00	2.05E-07	0.00E+00	8.22E-06
GO:0015840 (GO:0015840): [BP] urea transport	0.00E+00	0.00E+00	0.00E+00	2.05E-07	0.00E+00	8.22E-06

Carbon						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0006730 (GO:0006730): [BP] one-carbon metabolic process	2.26E-02	2.33E-02	2.41E-02	2.42E-02	2.48E-02	2.72E-02
GO:0015977 (GO:0015977): [BP] carbon fixation	7.14E-03	8.00E-03	8.41E-03	8.46E-03	8.91E-03	1.07E-02
GO:0015976 (GO:0015976): [BP] carbon utilization	4.65E-03	5.93E-03	6.27E-03	6.22E-03	6.52E-03	7.33E-03
GO:0043427 (GO:0043427): [BP] carbon fixation by 3-hydroxypropionate cycle	3.11E-04	3.96E-04	4.29E-04	4.28E-04	4.50E-04	5.95E-04
GO:0042206 (GO:0042206): [BP] halogenated hydrocarbon catabolic process	8.23E-05	1.32E-04	2.12E-04	2.20E-04	2.95E-04	4.52E-04
GO:0015701 (GO:0015701): [BP] bicarbonate transport	0.00E+00	1.08E-04	1.24E-04	1.24E-04	1.34E-04	1.98E-04
GO:0045013 (GO:0045013): [BP] carbon catabolite repression of transcription	1.42E-05	7.34E-05	1.00E-04	9.84E-05	1.19E-04	2.07E-04
GO:0034465 (GO:0034465): [BP] response to carbon monoxide	0.00E+00	0.00E+00	0.00E+00	1.80E-07	0.00E+00	7.19E-06
GO:0030634 (GO:0030634): [BP] carbon fixation by acetyl-CoA pathway	0.00E+00	0.00E+00	0.00E+00	8.45E-07	0.00E+00	3.38E-05
GO:0019386 (GO:0019386): [BP] methanogenesis, from carbon dioxide	0.00E+00	0.00E+00	0.00E+00	6.34E-06	1.23E-05	3.91E-05
GO:0009854 (GO:0009854): [BP] oxidative photosynthetic carbon pathway	0.00E+00	0.00E+00	0.00E+00	4.49E-06	8.28E-06	3.71E-05
Phosphorus						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0006817 (GO:0006817): [BP] phosphate ion transport	2.52E-02	2.63E-02	2.68E-02	2.69E-02	2.74E-02	2.94E-02

Phosphorus						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0035435 (GO:0035435): [BP] phosphate ion transmembrane transport	1.49E-02	1.82E-02	1.88E-02	1.88E-02	1.95E-02	2.21E-02
GO:0042823 (GO:0042823): [BP] pyridoxal phosphate biosynthetic process	1.03E-02	1.20E-02	1.24E-02	1.27E-02	1.34E-02	1.70E-02
GO:0045936 (GO:0045936): [BP] negative regulation of phosphate metabolic process	1.11E-02	1.17E-02	1.21E-02	1.21E-02	1.24E-02	1.40E-02
GO:0030643 (GO:0030643): [BP] cellular phosphate ion homeostasis	1.11E-02	1.17E-02	1.21E-02	1.21E-02	1.24E-02	1.40E-02
GO:0046855 (GO:0046855): [BP] inositol phosphate dephosphorylation	4.42E-03	4.78E-03	5.25E-03	5.24E-03	5.59E-03	6.28E-03
GO:0046386 (GO:0046386): [BP] deoxyribose phosphate catabolic process	1.48E-03	1.88E-03	2.02E-03	2.04E-03	2.19E-03	2.46E-03
GO:0035975 (GO:0035975): [BP] carbamoyl phosphate catabolic process	4.15E-04	8.17E-04	9.67E-04	1.02E-03	1.23E-03	1.64E-03
GO:0006793 (GO:0006793): [BP] phosphorus metabolic process	5.51E-04	6.89E-04	7.43E-04	7.56E-04	8.13E-04	9.94E-04
GO:0070409 (GO:0070409): [BP] carbamoyl phosphate biosynthetic process	4.14E-04	5.33E-04	5.66E-04	5.69E-04	6.04E-04	7.56E-04
GO:0016036 (GO:0016036): [BP] cellular response to phosphate starvation	1.76E-04	2.30E-04	2.67E-04	2.83E-04	3.22E-04	4.78E-04
GO:2000186 (GO:2000186): [BP] negative regulation of phosphate transmembrane transport	1.05E-04	1.62E-04	1.86E-04	1.84E-04	2.04E-04	2.45E-04

Phosphorus						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0080040 (GO:0080040): [BP] positive regulation of cellular response to phosphate starvation	2.94E-05	5.06E-05	5.88E-05	6.02E-05	6.78E-05	1.02E-04
GO:0019693 (GO:0019693): [BP] ribose phosphate metabolic process	0.00E+00	0.00E+00	2.25E-05	2.19E-05	3.28E-05	8.01E-05
GO:0006753 (GO:0006753): [BP] nucleoside phosphate metabolic process	0.00E+00	0.00E+00	2.25E-05	2.19E-05	3.28E-05	8.01E-05
GO:0036108 (GO:0036108): [BP] 4-amino-4-deoxy-alpha- L-arabinopyranosyl undecaprenyl phosphate biosynthetic process	0.00E+00	8.05E-06	1.34E-05	1.38E-05	1.88E-05	5.06E-05
GO:0061624 (GO:0061624): [BP] fructose catabolic process to hydroxyacetone phosphate and glyceraldehyde-3-phosphate	0.00E+00	0.00E+00	0.00E+00	6.52E-07	0.00E+00	1.56E-05
GO:0043647 (GO:0043647): [BP] inositol phosphate metabolic process	0.00E+00	0.00E+00	0.00E+00	1.99E-07	0.00E+00	6.56E-06
GO:0000971 (GO:0000971): [BP] tRNA exon ligation utilizing 2',3' cyclic phosphate of 5'-exon as source of linkage phosphate	0.00E+00	0.00E+00	0.00E+00	4.54E-06	0.00E+00	3.24E-05

Potassium						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0006813 (GO:0006813): [BP] potassium ion transport	2.01E-02	2.39E-02	2.49E-02	2.48E-02	2.59E-02	2.95E-02
GO:0071805 (GO:0071805): [BP] potassium ion transmembrane transport	3.60E-04	5.57E-04	6.45E-04	6.53E-04	7.60E-04	9.00E-04

Potassium						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:1901381 (GO:1901381): [BP] positive regulation of potassium ion transmembrane transport	0.00E+00	0.00E+00	4.40E-06	1.32E-05	2.65E-05	5.06E-05
GO:1903818 (GO:1903818): [BP] positive regulation of voltage-gated potassium channel activity	0.00E+00	0.00E+00	0.00E+00	1.64E-07	0.00E+00	6.56E-06
GO:1901379 (GO:1901379): [BP] regulation of potassium ion transmembrane transport	0.00E+00	0.00E+00	0.00E+00	3.89E-07	0.00E+00	1.56E-05
GO:0055075 (GO:0055075): [BP] potassium ion homeostasis	0.00E+00	0.00E+00	0.00E+00	1.52E-07	0.00E+00	6.06E-06
GO:0035865 (GO:0035865): [BP] cellular response to potassium ion	0.00E+00	0.00E+00	0.00E+00	1.80E-07	0.00E+00	7.19E-06
Sulfate						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0016226 (GO:0016226): [BP] iron-sulfur cluster assembly	3.42E-02	3.71E-02	3.82E-02	3.82E-02	3.89E-02	4.14E-02
GO:0097428 (GO:0097428): [BP] protein maturation by iron-sulfur cluster transfer	1.20E-02	1.35E-02	1.39E-02	1.40E-02	1.45E-02	1.55E-02
GO:0000103 (GO:0000103): [BP] sulfate assimilation	8.71E-03	9.78E-03	1.00E-02	1.01E-02	1.03E-02	1.10E-02
GO:0019379 (GO:0019379): [BP] sulfate assimilation, phosphoadenylyl sulfate reduction by phosphoadenylyl-sulfate reductase (thioredoxin)	5.14E-03	6.09E-03	6.33E-03	6.40E-03	6.72E-03	7.40E-03
GO:0019419 (GO:0019419): [BP] sulfate reduction	3.24E-03	3.54E-03	3.75E-03	3.73E-03	3.90E-03	4.44E-03

Sulfate	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0006790 (GO:0006790): [BP] sulfur compound metabolic process	2.56E-03	2.72E-03	3.03E-03	2.98E-03	3.16E-03	3.64E-03
GO:0019417 (GO:0019417): [BP] sulfur oxidation	8.67E-04	1.22E-03	1.36E-03	1.34E-03	1.46E-03	2.11E-03
GO:0019346 (GO:0019346): [BP] transsulfuration	4.47E-04	5.60E-04	5.86E-04	5.87E-04	6.16E-04	7.35E-04
GO:0010438 (GO:0010438): [BP] cellular response to sulfur starvation	2.36E-04	3.84E-04	4.48E-04	4.39E-04	5.02E-04	6.72E-04
GO:0015012 (GO:0015012): [BP] heparan sulfate proteoglycan biosynthetic process	1.09E-04	2.42E-04	3.47E-04	3.87E-04	5.16E-04	9.23E-04
GO:0044273 (GO:0044273): [BP] sulfur compound catabolic process	1.09E-04	1.91E-04	2.17E-04	2.19E-04	2.57E-04	3.34E-04
GO:0008272 (GO:0008272): [BP] sulfate transport	1.48E-04	1.87E-04	2.09E-04	2.07E-04	2.30E-04	2.59E-04
GO:0000096 (GO:0000096): [BP] sulfur amino acid metabolic process	1.73E-05	1.63E-04	1.90E-04	1.85E-04	2.04E-04	2.50E-04
GO:0018909 (GO:0018909): [BP] dodecyl sulfate metabolic process	5.16E-05	9.96E-05	1.49E-04	1.51E-04	2.00E-04	2.58E-04
GO:0010134 (GO:0010134): [BP] sulfate assimilation via adenylyl sulfate reduction	1.73E-05	4.79E-05	5.91E-05	6.53E-05	8.98E-05	1.19E-04
GO:0030206 (GO:0030206): [BP] chondroitin sulfate biosynthetic process	0.00E+00	2.45E-05	3.44E-05	3.58E-05	4.11E-05	1.04E-04
GO:0000098 (GO:0000098): [BP] sulfur amino acid catabolic process	0.00E+00	1.76E-05	2.70E-05	2.55E-05	3.43E-05	5.77E-05
GO:0009970 (GO:0009970): [BP] cellular response to sulfate starvation	0.00E+00	7.92E-06	1.39E-05	1.33E-05	1.72E-05	5.55E-05

Sulfate						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0018283 (GO:0018283): [BP] iron incorporation into metallo-sulfur cluster	0.00E+00	0.00E+00	7.45E-06	1.80E-05	3.17E-05	9.29E-05
GO:0042340 (GO:0042340): [BP] keratan sulfate catabolic process	0.00E+00	0.00E+00	0.00E+00	2.90E-07	0.00E+00	1.16E-05
GO:0030200 (GO:0030200): [BP] heparan sulfate proteoglycan catabolic process	0.00E+00	0.00E+00	0.00E+00	4.54E-07	0.00E+00	1.82E-05
GO:0019420 (GO:0019420): [BP] dissimilatory sulfate reduction	0.00E+00	0.00E+00	0.00E+00	8.68E-07	0.00E+00	3.47E-05
GO:0018146 (GO:0018146): [BP] keratan sulfate biosynthetic process	0.00E+00	0.00E+00	0.00E+00	2.90E-07	0.00E+00	1.16E-05
GO:0015709 (GO:0015709): [BP] thiosulfate transport	0.00E+00	0.00E+00	0.00E+00	5.15E-07	0.00E+00	8.24E-06
GO:0006791 (GO:0006791): [BP] sulfur utilization	0.00E+00	0.00E+00	0.00E+00	6.88E-07	0.00E+00	1.74E-05
GO:0000101 (GO:0000101): [BP] sulfur amino acid transport	0.00E+00	0.00E+00	0.00E+00	6.88E-07	0.00E+00	1.74E-05
Starvation						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0009267 (GO:0009267): [BP] cellular response to starvation	5.29E-03	5.82E-03	5.97E-03	5.98E-03	6.19E-03	6.65E-03
GO:0010438 (GO:0010438): [BP] cellular response to sulfur starvation	2.36E-04	3.84E-04	4.48E-04	4.39E-04	5.02E-04	6.72E-04
GO:0016036 (GO:0016036): [BP] cellular response to phosphate starvation	1.76E-04	2.30E-04	2.67E-04	2.83E-04	3.22E-04	4.78E-04

Starvation						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0042594 (GO:0042594): [BP] response to starvation	1.68E-04	1.84E-04	1.97E-04	2.05E-04	2.18E-04	2.74E-04
GO:0010106 (GO:0010106): [BP] cellular response to iron ion starvation	3.18E-05	5.77E-05	6.74E-05	6.81E-05	7.78E-05	1.06E-04
GO:0080040 (GO:0080040): [BP] positive regulation of cellular response to phosphate starvation	2.94E-05	5.06E-05	5.88E-05	6.02E-05	6.78E-05	1.02E-04
GO:0006995 (GO:0006995): [BP] cellular response to nitrogen starvation	0.00E+00	7.96E-06	1.52E-05	1.61E-05	2.36E-05	4.74E-05
GO:0009970 (GO:0009970): [BP] cellular response to sulfate starvation	0.00E+00	7.92E-06	1.39E-05	1.33E-05	1.72E-05	5.55E-05
GO:0034198 (GO:0034198): [BP] cellular response to amino acid starvation	0.00E+00	1.73E-06	9.13E-06	1.00E-05	1.79E-05	3.02E-05
GO:1990928 (GO:1990928): [BP] response to amino acid starvation	0.00E+00	0.00E+00	0.00E+00	2.10E-07	0.00E+00	6.13E-06
GO:1990641 (GO:1990641): [BP] response to iron ion starvation	0.00E+00	0.00E+00	0.00E+00	1.07E-06	0.00E+00	2.18E-05
GO:0042149 (GO:0042149): [BP] cellular response to glucose starvation	0.00E+00	0.00E+00	0.00E+00	7.12E-07	0.00E+00	1.78E-05
Symbiosis						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0052143 (GO:0052143): [BP] chemotaxis on or near host involved in symbiotic interaction	1.28E-05	5.59E-05	7.34E-05	8.05E-05	1.02E-04	1.74E-04

Symbiosis						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0052501 (GO:0052501): [BP] positive regulation by organism of apoptotic process in other organism involved in symbiotic interaction	0.00E+00	0.00E+00	1.19E-05	1.03E-05	1.41E-05	3.02E-05
GO:0044657 (GO:0044657): [BP] pore formation in membrane of other organism during symbiotic interaction	0.00E+00	0.00E+00	0.00E+00	1.55E-07	0.00E+00	6.22E-06
Photosynthesis						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GO:0019684 (GO:0019684): [BP] photosynthesis, light reaction	3.96E-03	5.03E-03	5.38E-03	5.36E-03	5.75E-03	6.19E-03
GO:0015979 (GO:0015979): [BP] photosynthesis	1.85E-03	2.37E-03	2.55E-03	2.55E-03	2.85E-03	3.29E-03
GO:0009765 (GO:0009765): [BP] photosynthesis, light harvesting	0.00E+00	8.59E-05	1.09E-04	1.13E-04	1.52E-04	1.79E-04
GO:0019685 (GO:0019685): [BP] photosynthesis, dark reaction	2.59E-05	6.36E-05	9.83E-05	1.13E-04	1.48E-04	3.00E-04
GO:0010109 (GO:0010109): [BP] regulation of photosynthesis	0.00E+00	1.65E-05	2.19E-05	2.19E-05	2.90E-05	4.05E-05
GO:0009768 (GO:0009768): [BP] photosynthesis, light harvesting in photosystem I	0.00E+00	1.15E-05	2.16E-05	2.22E-05	2.66E-05	6.84E-05
GO:0009769 (GO:0009769): [BP] photosynthesis, light harvesting in photosystem II	0.00E+00	3.42E-06	1.01E-05	9.62E-06	1.40E-05	2.63E-05

Table 4: Summary statistics for GO terms of interest. Summary statistics were computed for the selected GO terms. The selection of GO terms was performed using the keywords found above the table.

Selected KO terms involved in Nitrogen metabolism

The GO terms in the above table are quite broad terms and thus, in order to investigate specific genes involved in nitrogen metabolism, we selected KOs in the KEEG modules found in “Nitrogen metabolism” at <https://www.genome.jp/brite/ko00002> (<https://www.genome.jp/brite/ko00002>). For some of the broad modules, we selected specific parts of the module and only included the relevant KOs. Thus, not all of the module names seen in the below boxplot correspond to a KEEG module name (instead, they are part of a larger KEEG module).

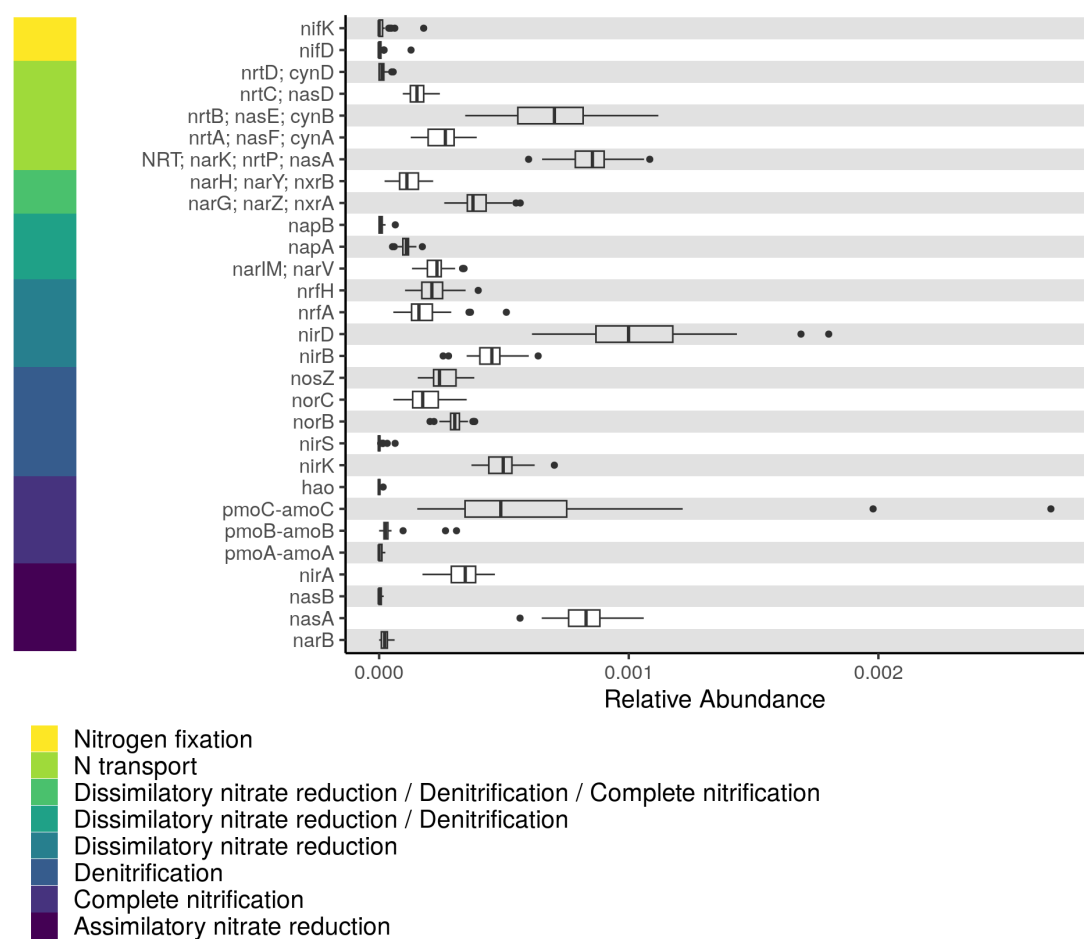
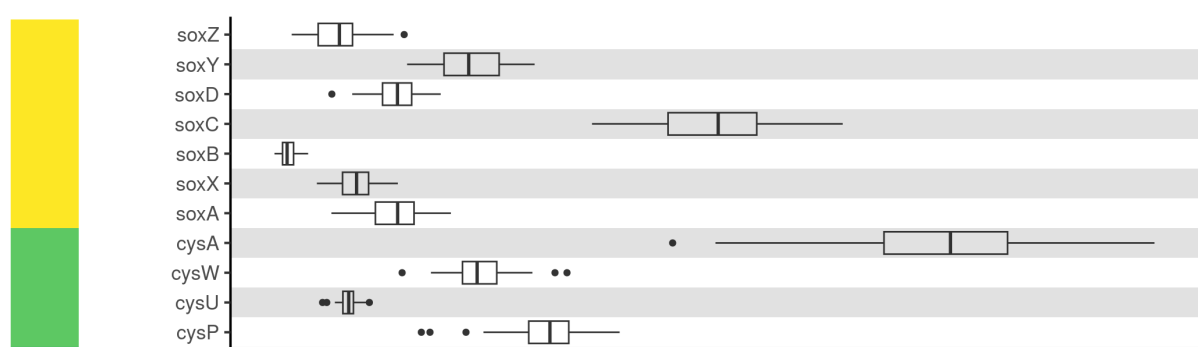


Figure 1: Boxplot of selected KO terms found in Nitrogen metabolism. The colored bar indicates which KEEG module each KO belongs to. Each row in the boxplot is one KO group (some groups have multiple names as seen in the figure). Furthermore, some KOs are found in several modules as indicated in the legend of the colored bar.

Selected KO terms involved in Sulfur metabolism

The GO terms are quite broad terms and thus, in order to investigate specific genes involved in Sulfur metabolism, we selected KOs in the KEEG modules found in “Sulfur metabolism” at <https://www.genome.jp/brite/ko00002> (<https://www.genome.jp/brite/ko00002>). For some of the broad modules, we selected specific parts of the module and only included the relevant KOs. Thus, not all of the module names seen in the below boxplot correspond to a KEEG module name (instead, they are part of a larger KEEG module).



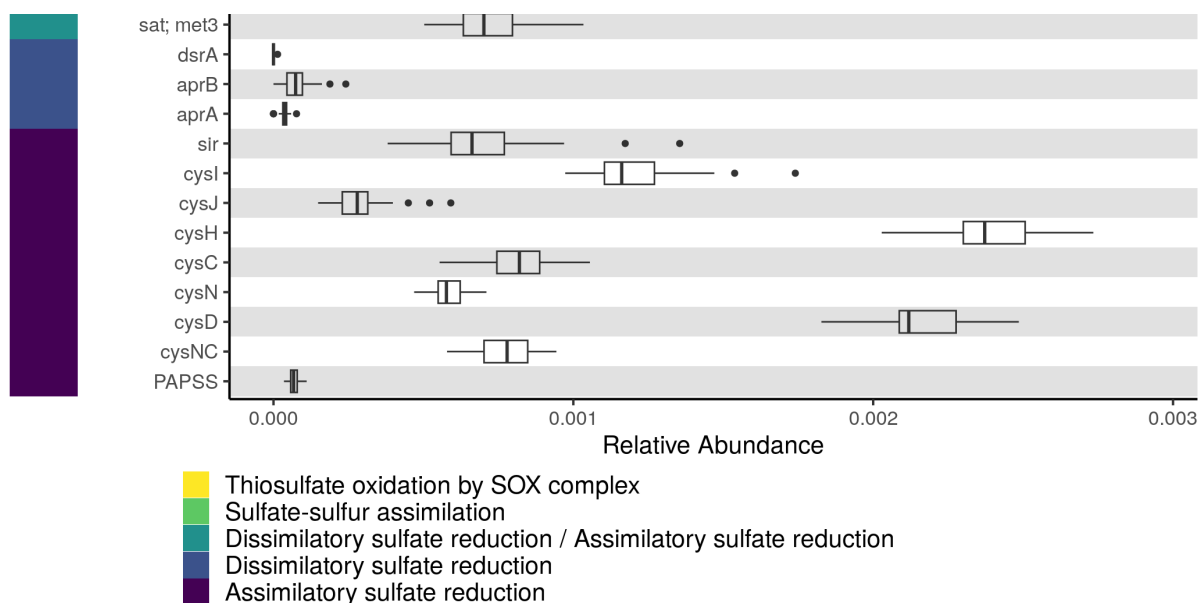
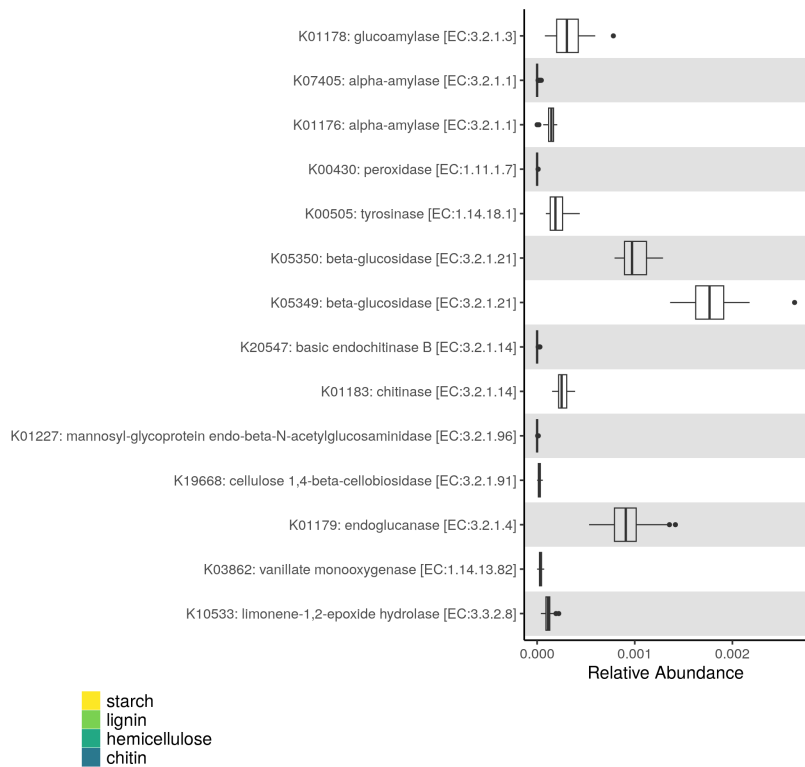


Figure 2: Boxplot of selected KO terms found in Sulfur metabolism. The colored bar indicates which KEGG module each KO belongs to. Each row in the boxplot is one KO group (some groups have multiple names as seen in the figure). Furthermore, some KOs are found in several modules as indicated in the legend of the colored bar.

Selected KO terms involved in carbon metabolism

For the carbon metabolism, the KEGG modules are very broad and we therefore find it more interesting to zoom in on specific aspects of the carbon metabolism. Therefore, we selected KO terms based on a recent article (Kelly et al., 2021). The KOs were selected based on their Enzyme Commission Number (EC Number) and the included EC numbers were grouped by the type of compound they metabolize as seen in the article and in **figure 3**.

Reference: Charlene N. Kelly, Geoffrey W. Schwaner, Jonathan R. Cumming, Timothy P. Driscoll, *Metagenomic reconstruction of nitrogen and carbon cycling pathways in forest soil: Influence of different hardwood tree species, Soil Biology and Biochemistry, Volume 156, 2021.*



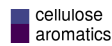


Figure 3: Boxplot of selected KO terms found in Carbon metabolism. The colored bar indicates which compound group the enzymes act on. Each row in the boxplot is one KO group.

Package	Version	Package	Version
OS	Ubuntu 20.04.4 LTS	cachem	1.1.0
R	4.3.3	yaml	2.3.9
sass	0.4.9	tools	4.3.3
utf8	1.2.4	tzdb	0.4.0
generics	0.1.3	ggsignif	0.6.4
rstatix	0.7.2	colorspace	2.1-0
xml2	1.3.6	broom	1.0.6
stringi	1.8.4	vctrs	0.6.5
hms	1.1.3	R6	2.5.1
digest	0.6.36	lifecycle	1.0.4
evaluate	0.24.0	car	3.1-2
grid	4.3.3	pkgconfig	2.0.3
timechange	0.3.0	pillar	1.9.0
fastmap	1.2.0	bslib	0.7.0
jsonlite	1.8.8	gtable	0.3.5
backports	1.5.0	glue	1.7.0
gridExtra	2.3	systemfonts	1.1.0
fansi	1.0.6	highr	0.11
scales	1.3.0	xfun	0.46
codetools	0.2-20	tidyselect	1.2.1
jquerylib	0.1.4	rstudioapi	0.16.0
abind	1.4-5	knitr	1.48
cli	3.6.3	farver	2.1.2
crayon	1.5.3	htmltools	0.5.8.1
rlang	1.1.4	labeling	0.4.3
cowplot	1.1.3	carData	3.0-5
munsell	0.5.1	rmarkdown	2.27
withr	3.0.0	svglite	2.1.3

Table 5: Software and package versions.
