

METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks



C-DEBI Virtual Meeting Series
Feb 4, 2022

Goals for creating METABOLIC

- Enable metabolic and biogeochemical analyses for genomes and microbial communities
- Enable visualization of biogeochemical cycling potential and community-scale functional networks.

Software workflow



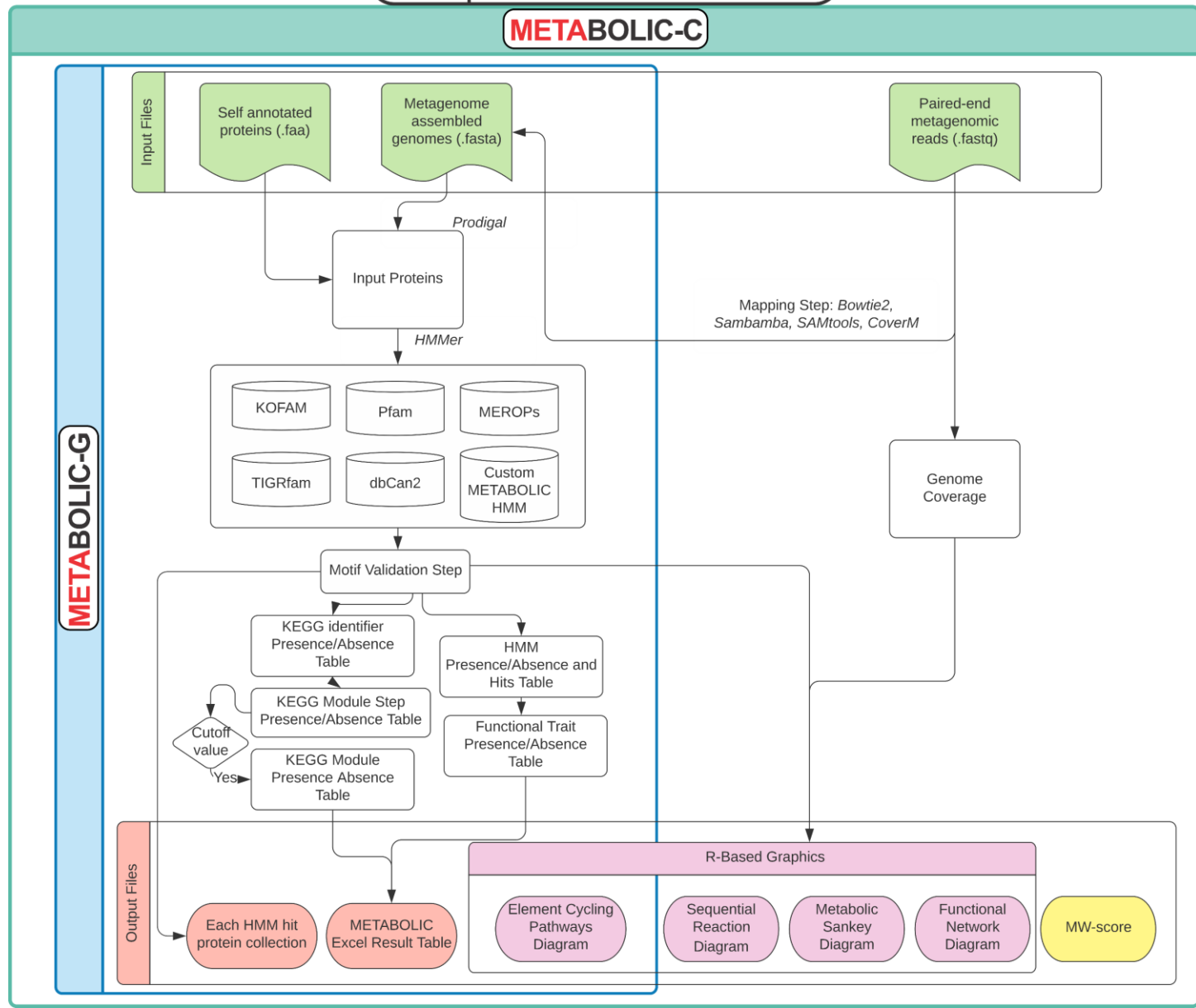
METABOLIC-G:

The genome version

METABOLIC-C:

The community version

Include metagenomic reads for community analysis and visualization



Software workflow

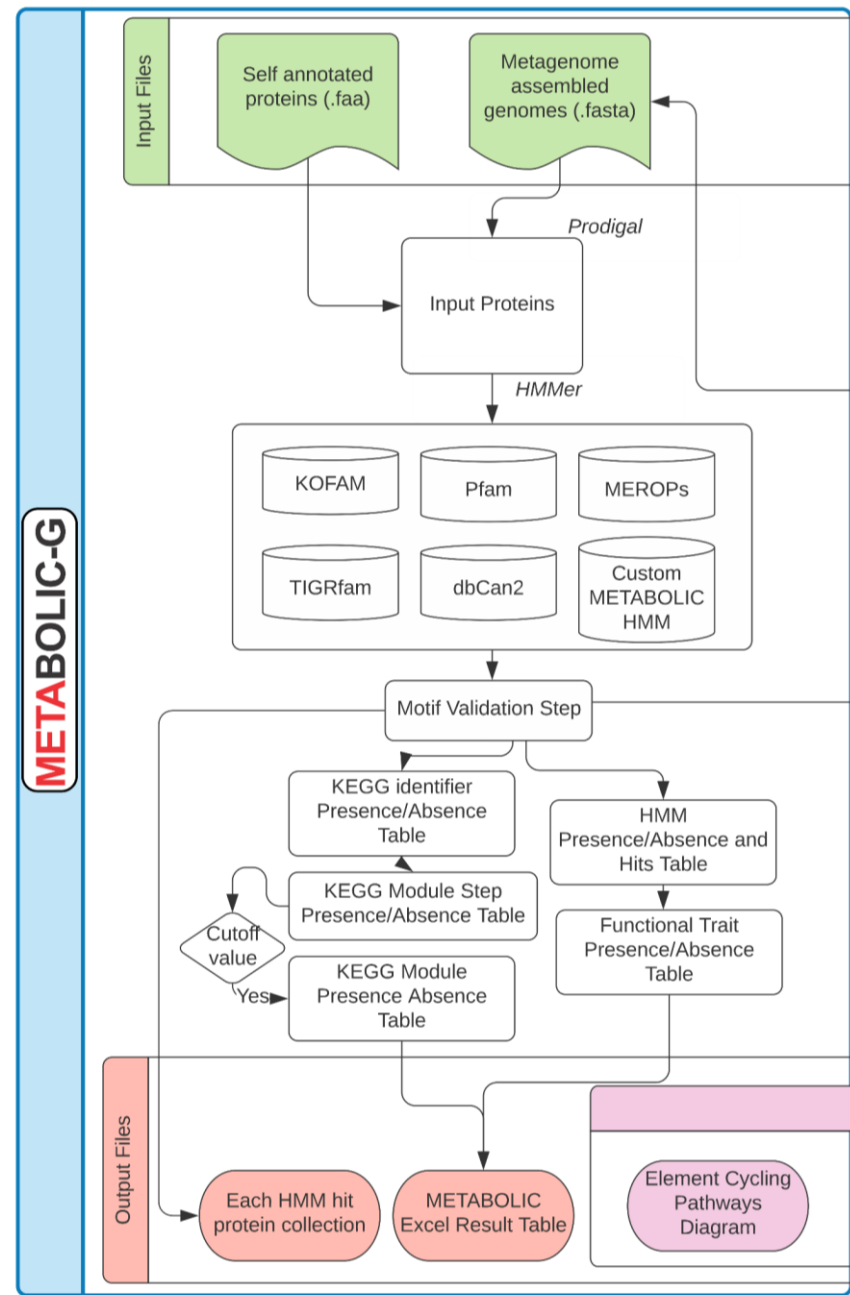
Intake microbial genomes (both *.faa and *.fasta) of: genomes of microbial isolates, MAGs, SAGs

Annotate using a comprehensive set of database: KEGG, TIGRfam, Pfam, custom HMM databases, dbCAN2 (for CAZymes), and MEROPS (for peptidases/inhibitors); manually curated cutoff scores for several HMMs to increase accuracy

Motif validation step by comparing protein motifs against a manually curated set of highly conserved residues in important proteins

Increase the annotation confidence on functionally important proteins with high sequence similarity

Provide user-friendly outputs in the form of tables and figures



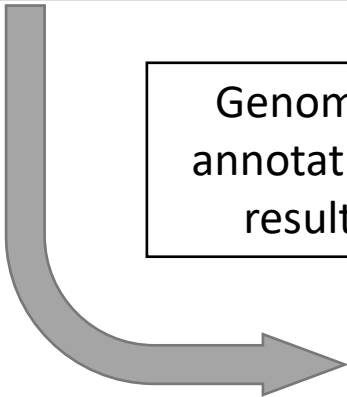
Software workflow

| Program Name | Program Description |
|----------------|--|
| METABOLIC-G.pl | Allows for classification of the metabolic capabilities of input genomes. |
| METABOLIC-C.pl | Allows for classification of the metabolic capabilities of input genomes, calculation of genome coverage, creation of biogeochemical cycling diagrams, and visualization of community metabolic interactions and contribution to biogeochemical processes by each microbial group. |

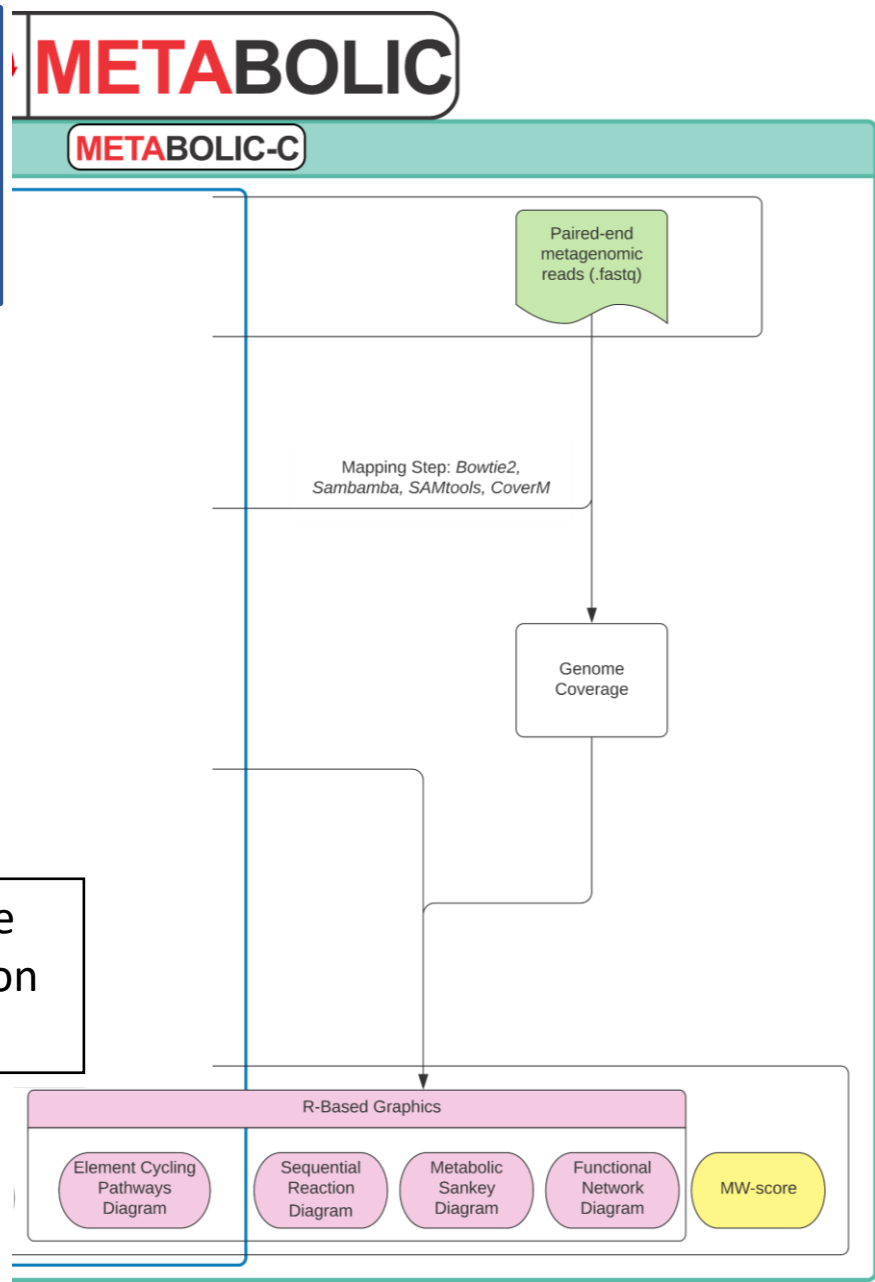
Metagenomic read mapping to individual genomes



Obtain genome coverage information



Genome annotation result



METABOLIC outputs

METABOLIC output table

| Output File/Folder | File Description | Generated by METABOLIC-C | Generated by METABOLIC-G |
|---------------------------------------|---|--------------------------|--------------------------|
| All_gene_collections_mapped.depth.txt | The gene depth of all input genes | X | |
| Each_HMM_Amino_Acid_Sequence/ | The faa collection for each hmm file | X | X |
| intermediate_files/ | The hmmsearch, peptides (MEROPS), CAZymes (dbCAN2), and GTDB-Tk (only for METABOLIC-C) running intermediate files | X | X |
| KEGG_identifier_result/ | The hit and result of each genome by Kofam database | X | X |
| METABOLIC_Figures/ | All figures output from the running of METABOLIC | X | X |
| METABOLIC_Figures_Input/ | All input files for R-generated diagrams | | |
| METABOLIC_result_each_spreadsheet/ | TSV files representing each sheet of the created METABOLIC_result.xlsx file | | |
| MW-score_result/ | The resulted table for MW-score | | |
| METABOLIC_result.xlsx | The resulting excel file of METABOLIC | | |

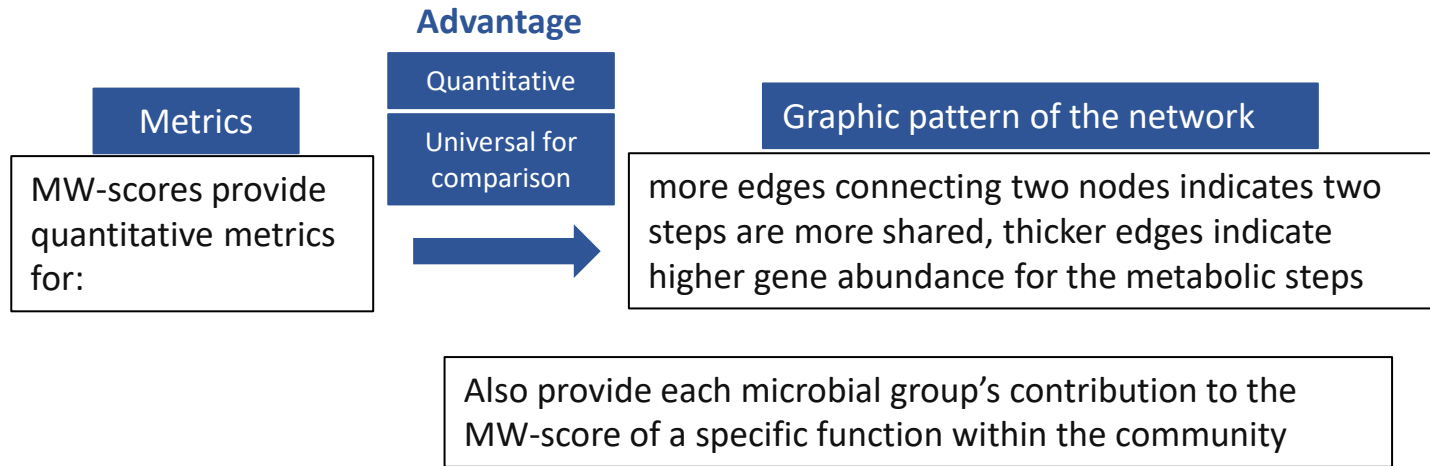
- 1) HMMHit: records the marker HMM hits – including presence/absence state, number of hits, and hit names
- 2) FunctionHit: records the combined HMM hit results to represent each function presence/absence
- 3) KEGG module Hit
- 4) KEGG module step hits to provide the details for each step
- 5) dbCAN2 result for CAZymes
- 6) MEROPS hit for peptidases/inhibitors



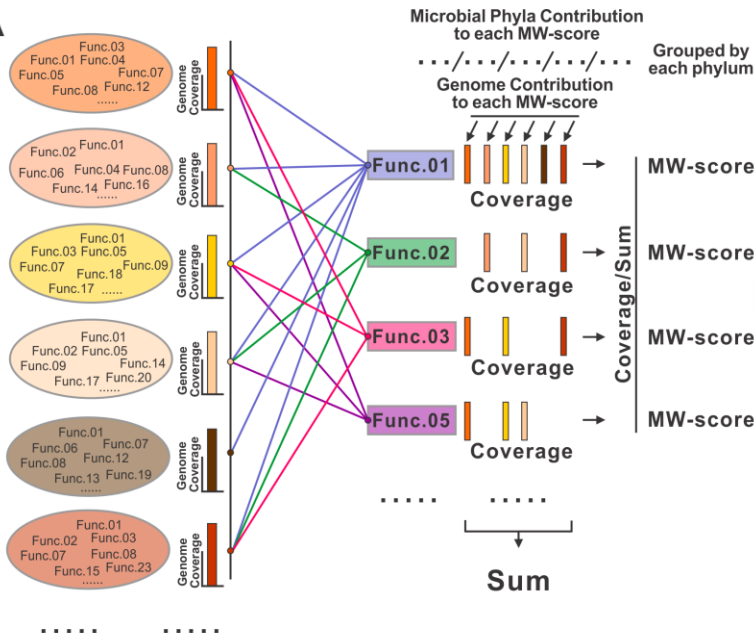
METABOLIC outputs

Purpose: To address the lack of quantitative and reproducible measures to represent potential metabolic interactions in microbial communities, we developed a new metric that we termed **MW-score (metabolic weight scores)**

MW-scores (Metabolic-Weight scores) resolves metabolic capacity and abundance in the co-sharing functional networks as studied and visualized in the above section



A



B

| Function | MW-score for each function | Microbial genome phyla | | | | | | | | | | | | | | | |
|--|----------------------------|------------------------|----------------|---------------------|---------------|---------------|-------------|------------------|----------------|----------------|--------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | | Proteobacteria | Actinobacteria | Alphaproteobacteria | Bacteroidetes | Cyanobacteria | Chloroflexi | Gemmatimonadetes | Planctomycetes | Thaumarchaeota | Thermotomarchaeota | Thermoplasmales | Thermoplasmales | Thermoplasmales | Thermoplasmales | Thermoplasmales | Thermoplasmales |
| C-S-01:Organic carbon oxidation - CO oxidation | 2.2 | 0.5 | 1.9 | 2.4 | 0.5 | 1.8 | 0.5 | 2.3 | 1.3 | 1.3 | 0.9 | 4.5 | 3.4 | 2 | 0.5 | 0.5 | 0.5 |
| C-S-01:Organic carbon oxidation - amino acid utilization | 8.5 | 0.2 | 5 | 1.6 | 1.6 | 0.5 | 1.3 | 0.4 | 0.5 | 0.3 | 2.9 | 1.1 | 0.8 | 3.4 | 6.7 | 5 | 0.5 |
| C-S-01:Organic carbon oxidation - aromatics degradation | 4.2 | 0.4 | 0 | 1 | 0.3 | 1 | 0.4 | 7.7 | 0 | 0.5 | 0 | 0.2 | 1.6 | 7 | 0 | 9.9 | 0.6 |
| C-S-01:Organic carbon oxidation - complex carbon degradation | 7.3 | 0.2 | 5.8 | 1.4 | 1.9 | 0.1 | 0 | 89 | 0.6 | 0.4 | 2 | 3.4 | 1.1 | 0.9 | 2.4 | 10 | 0.6 |
| C-S-01:Organic carbon oxidation - fatty acid degradation | 3.9 | 0.4 | 1.1 | 1.6 | 1 | 0 | 0 | 76 | 0.3 | 0 | 0 | 6.3 | 0 | 1.7 | 6.3 | 0 | 4.5 |
| C-S-01:Organic carbon oxidation - formaldehyde oxidation | 4.1 | 10 | 2.1 | 0.3 | 0.6 | 0 | 0 | 7.7 | 0 | 0.3 | 0 | 6 | 0 | 1.1 | 1 | 0 | 1.1 |
| C-S-01:Organic carbon oxidation - formate oxidation | 4.4 | 0.4 | 0 | 2.4 | 0 | 0 | 0 | 70 | 0.4 | 0.4 | 0 | 0 | 0 | 17 | 6.3 | 0 | 0.3 |
| C-S-01:Organic carbon oxidation - methanol oxidation | 3.3 | 0.5 | 0 | 1.9 | 0 | 0 | 0 | 26 | 1 | 0 | 0 | 7.5 | 0 | 0 | 1.2 | 0 | 1.3 |
| C-S-01:Organic carbon oxidation - methyl amine oxidation | 0.9 | 0 | 0 | 0 | 0 | 0.5 | 0 | 67 | 1 | 0 | 0 | 28 | 0 | 0 | 0 | 6 | 4.9 |
| C-S-02:Carbon fixation - 3HP/4HB | 0.1 | 0 | 0 | 5.3 | 0 | 0 | 0 | 31 | 0.7 | 0 | 0 | 0 | 0 | 0 | 2.6 | 0 | 0 |
| C-S-02:Carbon fixation - CBB cycle (Rubisco) | 2.1 | 0 | 0 | 0 | 0 | 0 | 13 | 59 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| C-S-02:Carbon fixation - Reverse TCA cycle | 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C-S-03:Ethanol oxidation | 2.5 | 0 | 0.5 | 0.6 | 0 | 0 | 0 | 95 | 0 | 0 | 4.3 | 0 | 0 | 0 | 0 | 0 | 0 |
| C-S-04:Acetate oxidation | 6.2 | 0.3 | 6.3 | 1.7 | 1.9 | 0.6 | 1.7 | 70 | 0.4 | 0.4 | 1.2 | 3.9 | 1.5 | 1.1 | 4.7 | 0.6 | 3.1 |
| C-S-06:Fermentation | 7.4 | 0.2 | 5.3 | 1.4 | 1.6 | 0.5 | 1.4 | 81 | 0.3 | 0.4 | 1.2 | 3.3 | 1.2 | 0.9 | 4 | 10 | 5.7 |
| C-S-06:Methanotrophy | 2.2 | 0 | 0 | 0.7 | 0 | 0 | 0 | 38 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.4 | 0 |
| C-S-09:Hydrogen oxidation | 0.5 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| N-S-03:Nitrite oxidation | 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| N-S-04:Nitrate reduction - napAB | 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| N-S-04:Nitrate reduction - narGH | 1.3 | 0 | 0 | 1.3 | 0 | 0 | 0 | 37 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.1 |
| N-S-05:Nitrite reduction - nirKS | 2 | 0 | 6.3 | 0 | 3.6 | 0 | 0 | 41 | 0 | 0 | 0 | 2.7 | 4 | 3.3 | 0 | 35 | 4.5 |
| N-S-06:Nitric oxide reduction | 0.6 | 0 | 0 | 0 | 0 | 0 | 0 | 39 | 3 | 0 | 25 | 0 | 3.5 | 11 | 12 | 0 | 7.3 |
| N-S-07:Nitrous oxide reduction | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| N-S-08:Nitrite ammonification - nirBD | 2.8 | 0 | 0 | 1.1 | 0 | 0 | 0 | 39 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| N-S-08:Nitrite ammonification - nirAB | 0.5 | 3.3 | 0 | 0 | 0 | 0 | 0 | 34 | 2.1 | 1.3 | 0 | 3.1 | 0 | 16 | 26 | 0 | 2.9 |
| O-S-01:Metal reduction | 0.3 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| O-S-02:Arsenate reduction | 6.1 | 0.3 | 2.3 | 1.9 | 1.6 | 0 | 0 | 74 | 0.6 | 0.1 | 2.4 | 4 | 1.2 | 1.1 | 4.8 | 0 | 0.8 |
| O-S-03:Sulfate reduction | 0.3 | 0 | 0 | 0 | 0 | 0 | 0 | 7.7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| O-S-04:Selenate reduction | 0.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S-S-01:Sulfide oxidation - sgr | 0.3 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S-S-03:Sulfur oxidation - dsrAB | 2.2 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S-S-03:Sulfur oxidation - sds | 5.7 | 0.3 | 7.4 | 1.8 | 1.4 | 0.1 | 0.3 | 64 | 0.2 | 0.2 | 0 | 4.3 | 1.4 | 1.2 | 1.6 | 13 | 2.6 |
| S-S-04:Sulfite oxidation | 3.9 | 0 | 0 | 1.9 | 0 | 0 | 0 | 65 | 0.2 | 1.2 | 6.3 | 0 | 1.7 | 3.8 | 18 | 0 | 1.1 |
| S-S-05:Sulfate reduction | 3.9 | 0 | 0 | 1.9 | 0 | 0 | 0 | 65 | 0.2 | 1.2 | 6.3 | 0 | 1.7 | 3.8 | 18 | 0 | 1.1 |
| S-S-07:Thiosulfate oxidation | 2.3 | 0.8 | 0 | 0.6 | 0 | 0 | 0 | 39 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S-S-08:Thiosulfate disproportionation (to sulfate + sulfur) | 6.1 | 0.3 | 0 | 1 | 0 | 0 | 0 | 78 | 0 | 0 | 2.4 | 0.9 | 0.3 | 1.1 | 2.4 | 12 | 0 |

METABOLIC performance demonstration

Test the prediction of 3HP and 3HP/4HB cycles in three groups of microorganisms

- METABOLIC result has the same annotation compared to that in KEGG genome pathway

| | | | | METABOLIC result | | KEGG genome pathways | |
|--|----------------------------------|---------------------------|----------------|---------------------------------|---|---------------------------------|---|
| | | | | Carbon fixation | | Carbon fixation | |
| Accession ID | Organism | KEGG Organism Code | Group | 3 Hydroxypropionate cycle (3HP) | 3-hydroxypropionate/4-hydroxybutyrate cycle (3HP/4HB) | 3 Hydroxypropionate cycle (3HP) | 3-hydroxypropionate/4-hydroxybutyrate cycle (3HP/4HB) |
| GCA_000011905.1 | Dehalococcoides mccartyi 195 | det | Chloroflexi | Absent | Absent | Absent | Absent |
| GCA_000017805.1 | Roseiflexus castenholzii DSM 11 | rca | Chloroflexi | Present | Absent | Present | Absent |
| GCA_000018865.1 | Chloroflexus aurantiacus J-10-fl | cau | Chloroflexi | Present | Absent | Present | Absent |
| GCA_000021685.1 | Thermomicrobium roseum DSM | tro | Chloroflexi | Absent | Absent | Absent | Absent |
| GCA_000021945.1 | Chloroflexus aerogans DSM 94 | cag | Chloroflexi | Present | Absent | Present | Absent |
| GCA_000299395.1 | Nitrosopumilus sediminis AR2 | nir | Thaumarchaeota | Absent | Present | Absent | Present |
| GCA_000698785.1 | Nitrososphaera viennensis EN76 | rwn | Thaumarchaeota | Absent | Present | Absent | Present |
| GCA_000875775.1 | Nitrosopumilus piranensis D3C | nid | Thaumarchaeota | Absent | Present | Absent | Present |
| GCA_000812185.1 | Nitrosopelagicus brevis CN25 | nbv | Thaumarchaeota | Absent | Present | Absent | Present |
| GCA_900696045.1 | Nitrosocosmicus franklandus NF | nfn | Thaumarchaeota | Absent | Present | Absent | Present |
| GCA_000015145.1 | Hyperthermus butylicus DSM 54 | nbu | Crenarchaeota | Absent | Absent | Absent | Absent |
| GCA_000017945.1 | Caldisphaera lagunensis DSM 11 | clg | Crenarchaeota | Absent | Present | Absent | Present |
| GCA_000148385.1 | Vulcanisaeta distributa DSM 144 | vdn | Crenarchaeota | Absent | Absent | Absent | Absent |
| GCA_000193375.1 | Thermoproteus uzoniensis 768-2 | tuz | Crenarchaeota | Absent | Present | Absent | Present |
| GCA_003431325.1 | Acidilobus sp. 7A | acia | Crenarchaeota | Absent | Absent | Absent | Absent |
| Remarks: Only Chloroflexii contain 3HP. 3HP/4HB could only be detected in Crenarchaeota and Thaumarchaeota | | | | | | | |
| The reference of biochemical evidence of the existence of corresponding carbon fixation pathways | | | | | | | |
| | 3HP | 3HP/4HB | | | | | |
| Chloroflexi | 10.1073/pnas.1710798114 | | | | | | |
| Crenarchaeota | | 10.1126/science.1149976 | | | | | |
| Thaumarchaeota | | 10.1016/j.mib.2011.04.007 | | | | | |

- METABOLIC result annotation is also consistent to the reference of biochemical evidence

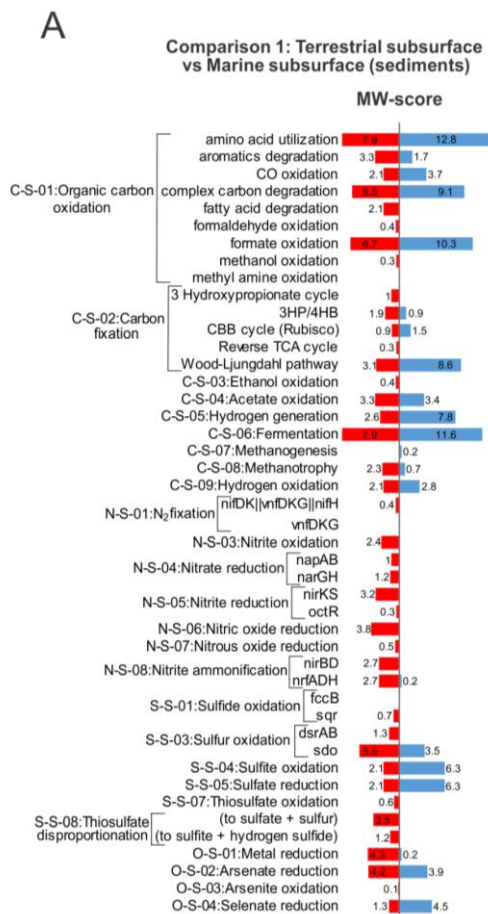
METABOLIC performance demonstration

- Wide application and good performance of METABOLIC in eight different samples:

marine subsurface,
terrestrial subsurface,
deep-sea hydrothermal vent,
freshwater lake

gut microbiome from patients with colorectal cancer,
gut microbiome from healthy control,
meadow soil,
wastewater treatment facility

- Community metabolism comparison based on MW-scores



Terrestrial subsurface contains more abundant metabolic functions related to nitrogen cycling compared to the marine subsurface

Consistency:

Consistent with the previous characterization of these two environments ([Anantharaman K et al., Nat Commun. 2016;7:13219](#); [Glass JB et al., Environ Microbiol. 2021 Aug;23\(8\):4646-4660](#))

Visualizations:

Four types of diagrams

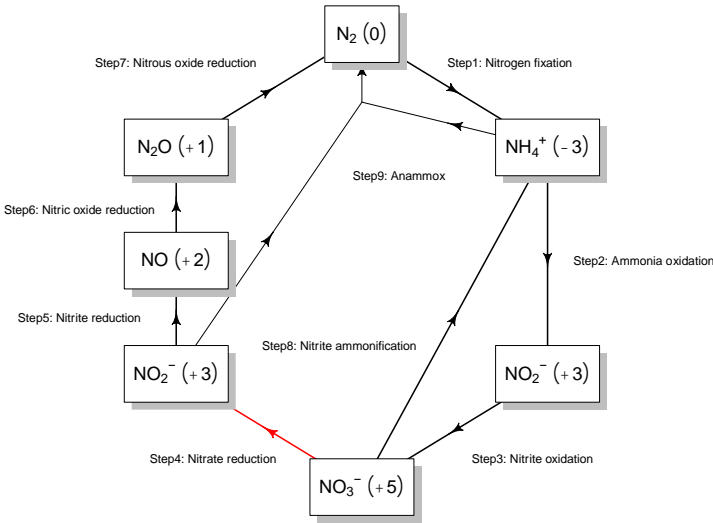
1. Nutrient Cycle Diagram
 1. Organismal-level
 2. Community-level
3. Sequential Metabolic Transformations
4. Functional Networks
5. Metabolic Alluvial Diagram

All visualizations are created as part of the default pipeline, but can also be ran individually using an Rscript.

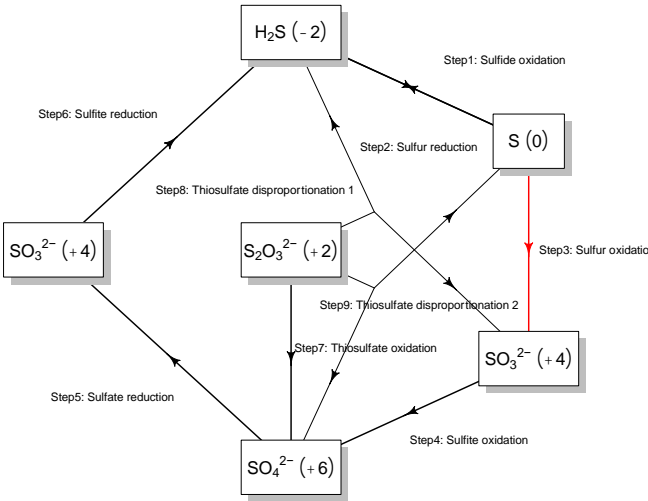
Nutrient cycling diagrams: Organismal-level

Red arrows = present
Black arrows = not present

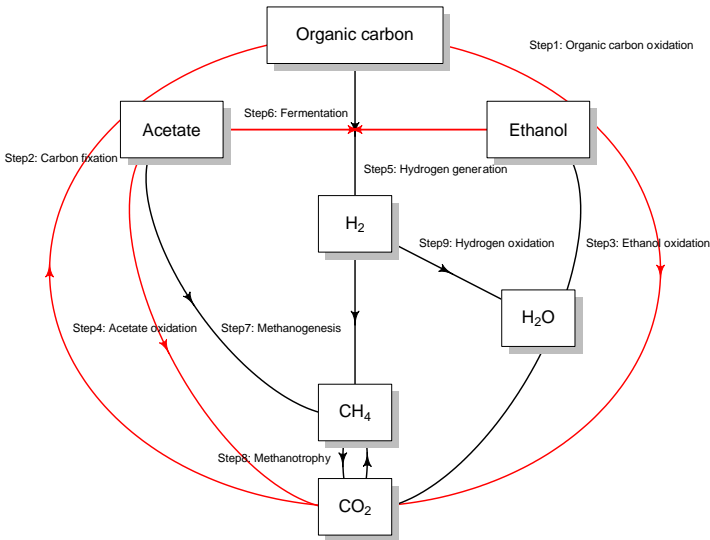
Nitrogen Cycle: Isolate43-LM-B-01-03



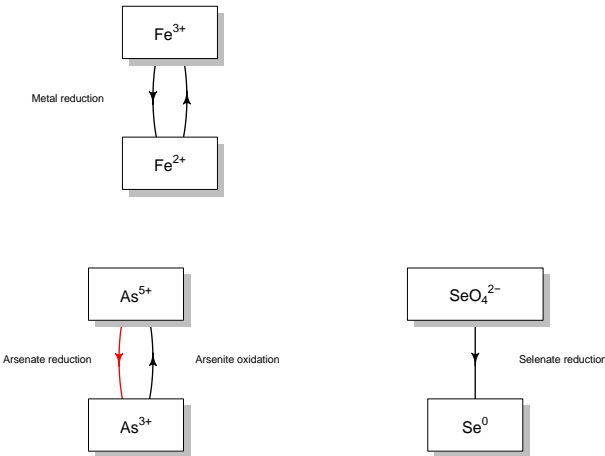
Sulfur Cycle: Isolate43-LM-B-01-03



Carbon Cycle: Isolate43-LM-B-01-03



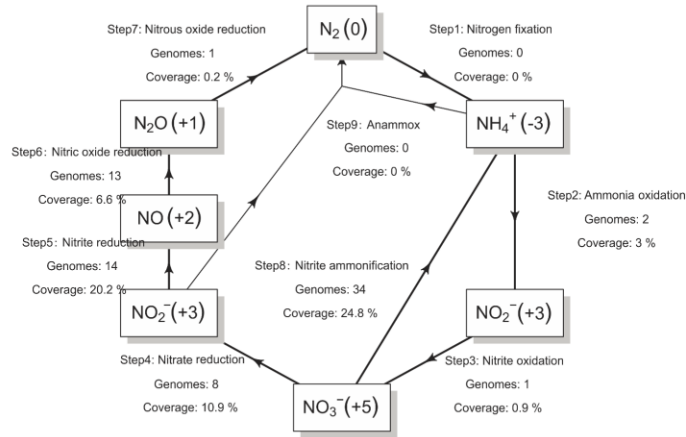
Other cycles Isolate43-LM-B-01-03



Nutrient cycling diagrams: Community-level

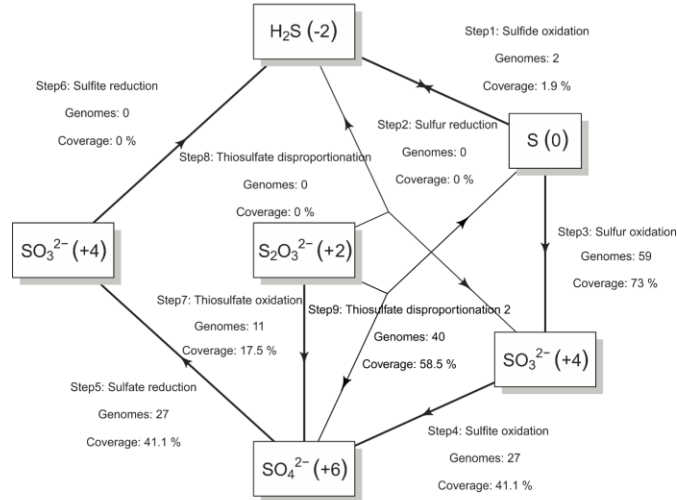
A

Nitrogen Cycle: Summary Figure



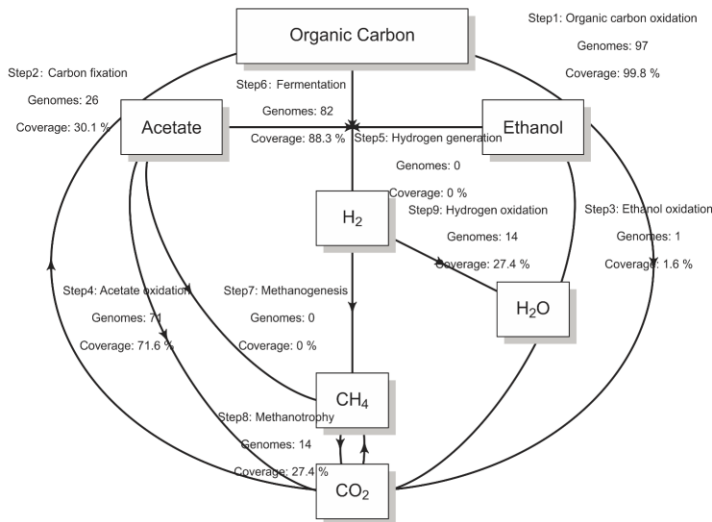
B

Sulfur Cycle: Summary Figure



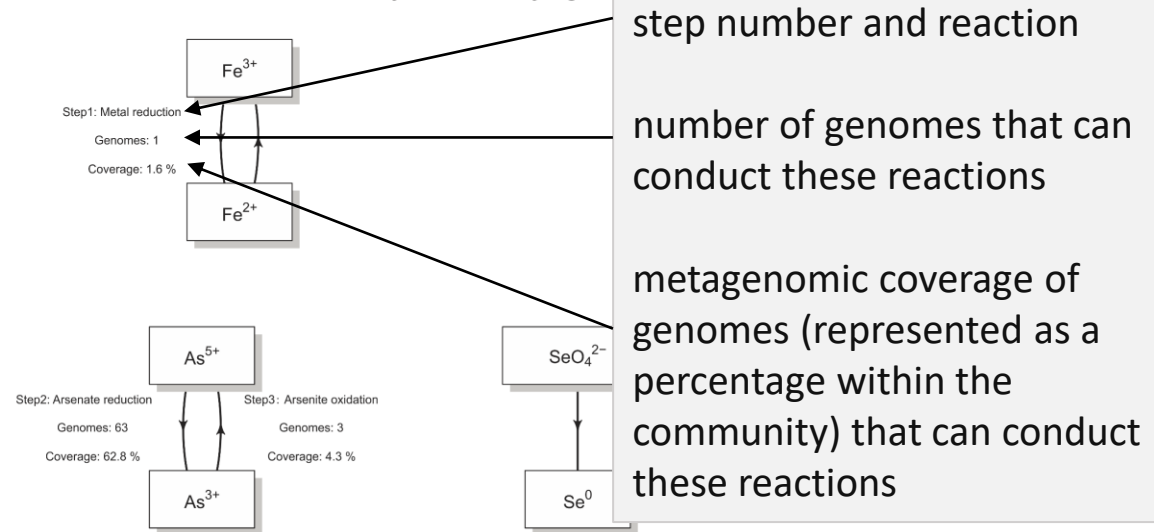
C

Carbon Cycle: Summary Figure



D

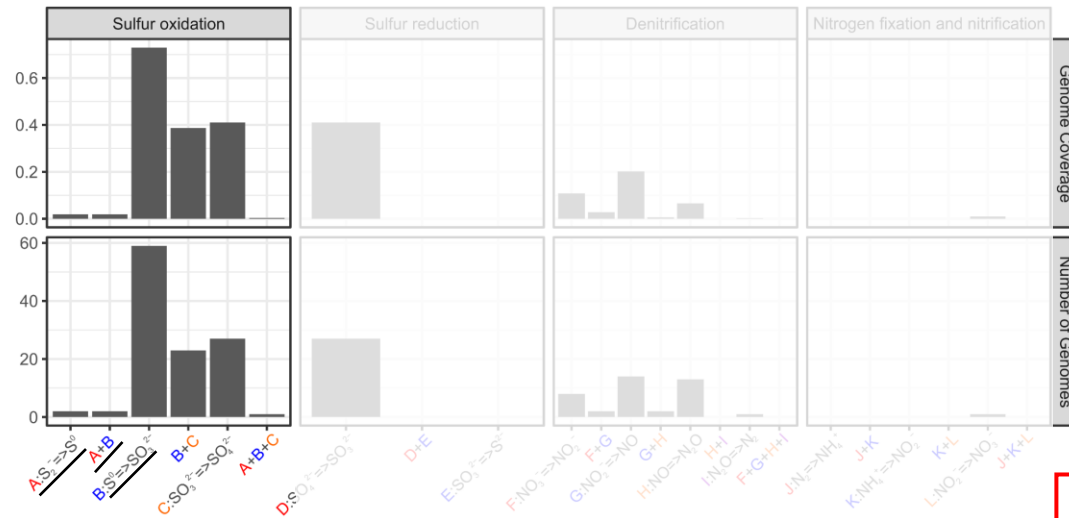
Other cycles: Summary Figure



Sequential metabolic transformations:

To show how *whole* biogeochemical pathways are *broken up*

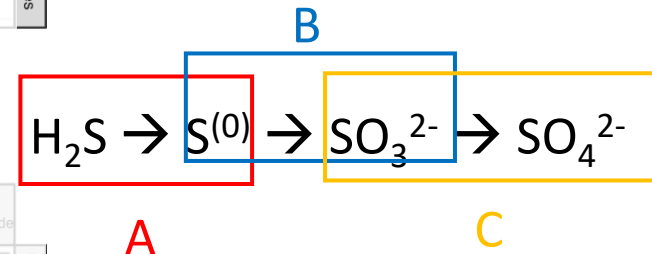
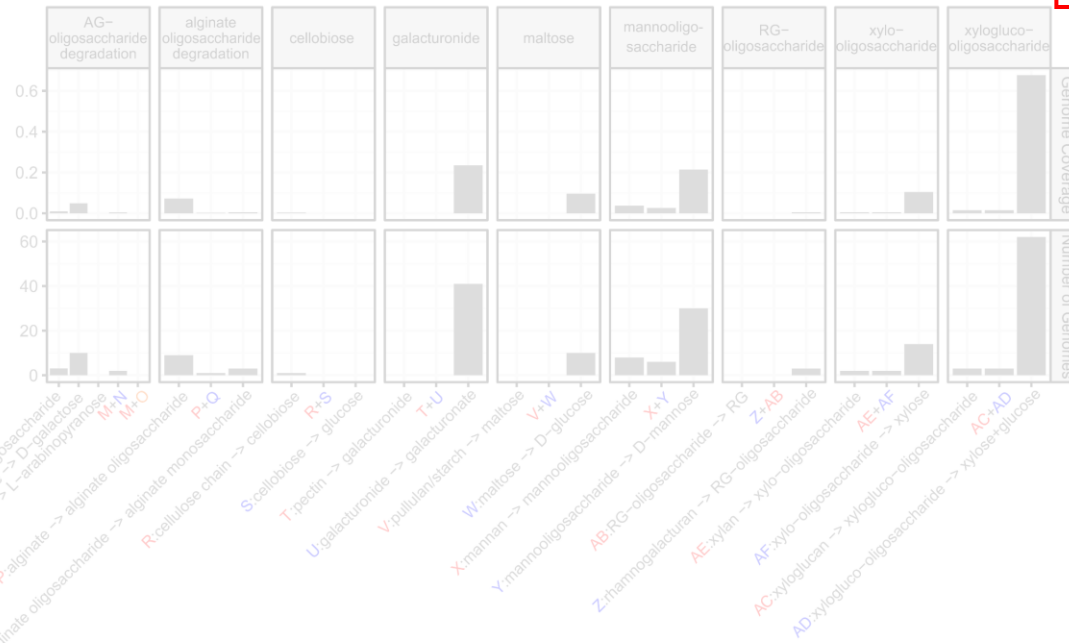
A Inorganics



The genome coverage

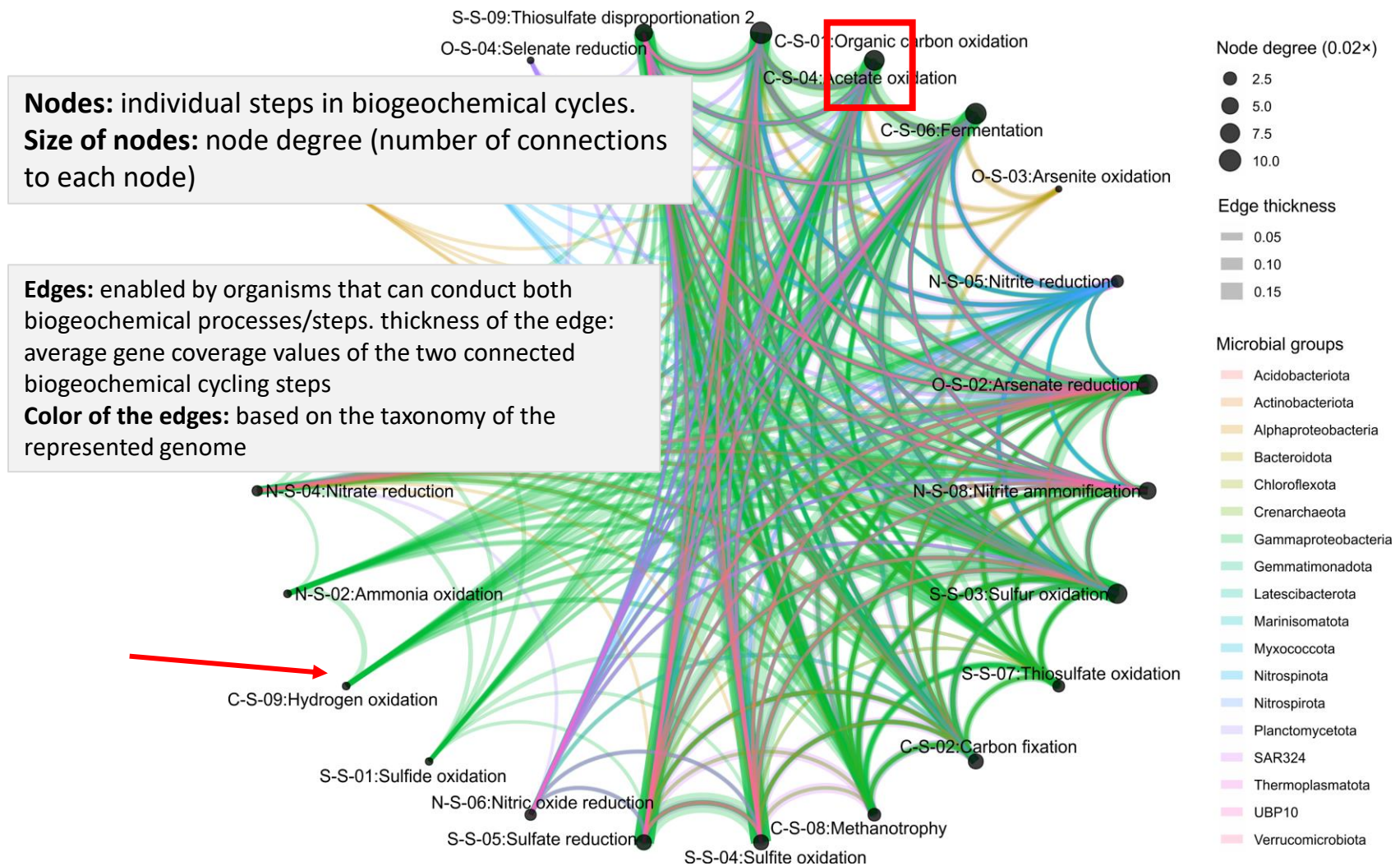
The number of genomes of organisms that are involved in certain sequential metabolic transformations

B Organics



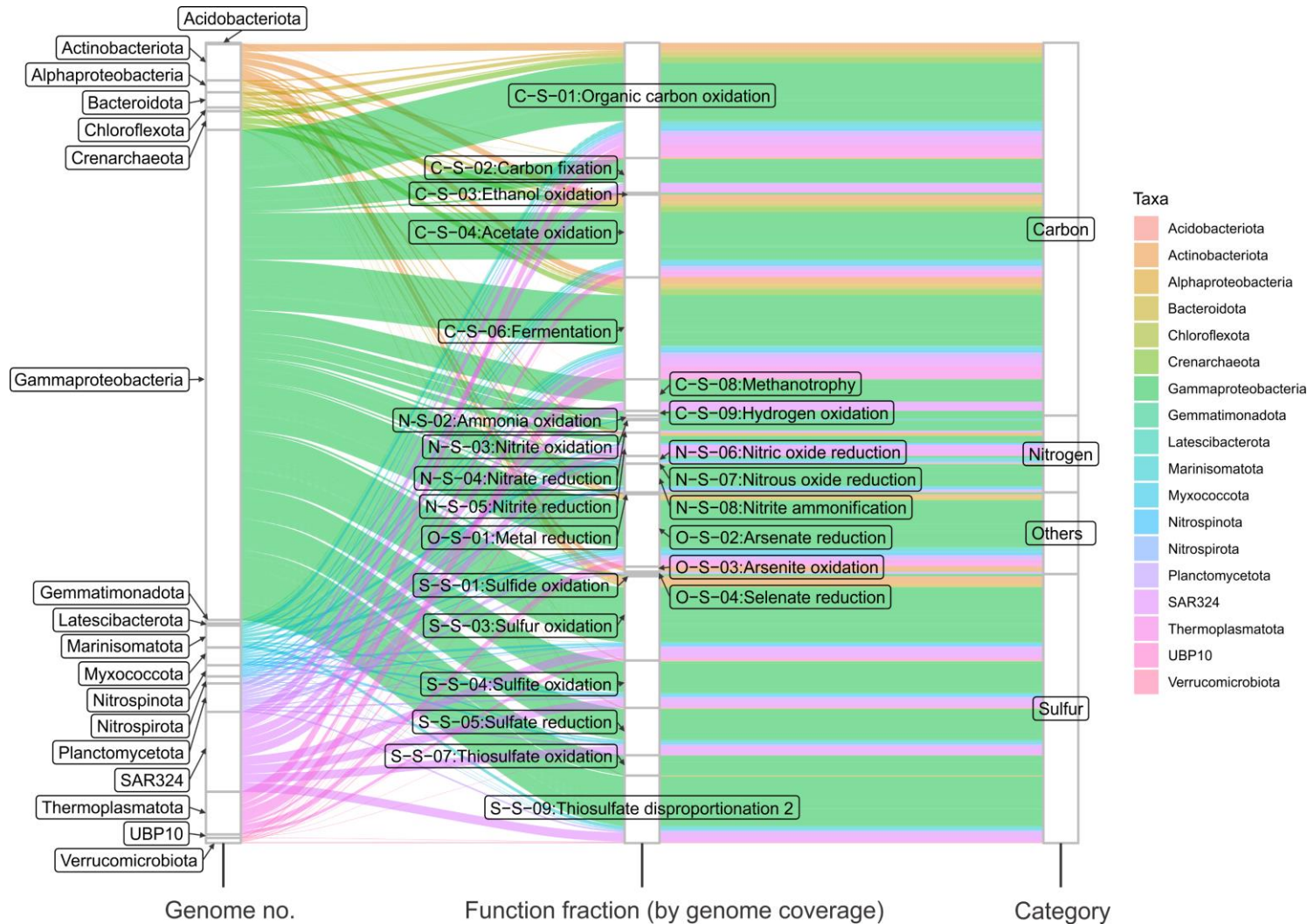
Functional networks:

To visualize *functional networks* and which pathways are more often associated with each other than others



Metabolic Alluvial diagram:

To see how *groups of taxa* are associated with *which biogeochemical reactions/categories*

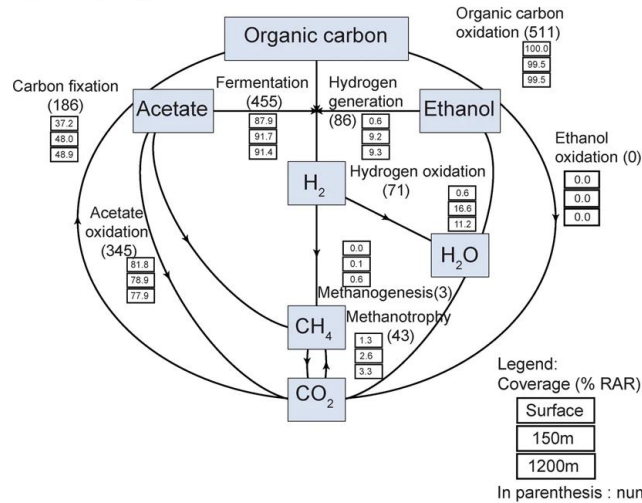


Width of curved line: from a specific microbial group to a given functional trait indicates their corresponding proportional contribution to a specific metabolism

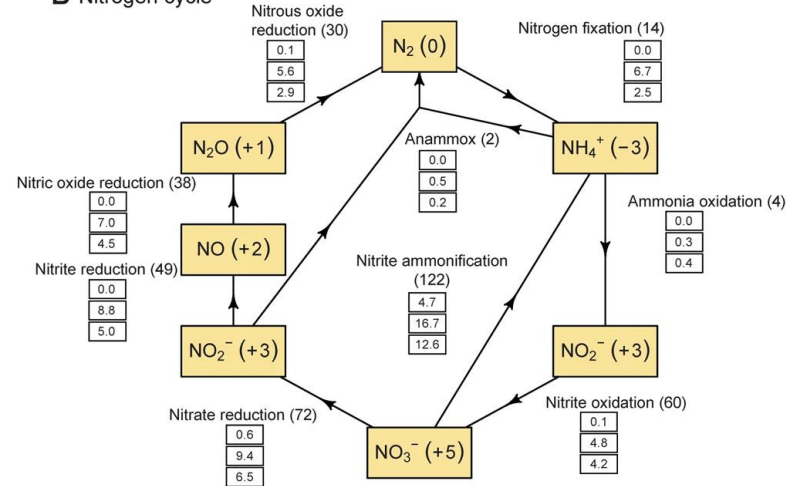
Applications of METABOLIC to different datasets: Freshwater

To compare across spatial depths: C common throughout, S and N under the oxycline

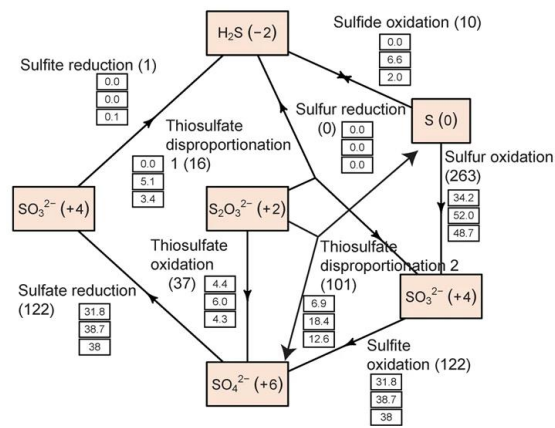
A Carbon cycle



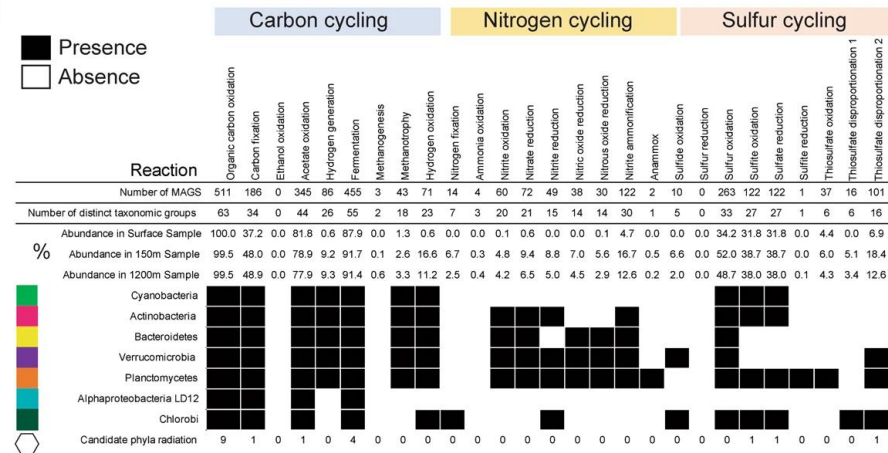
B Nitrogen cycle



C Sulfur cycle

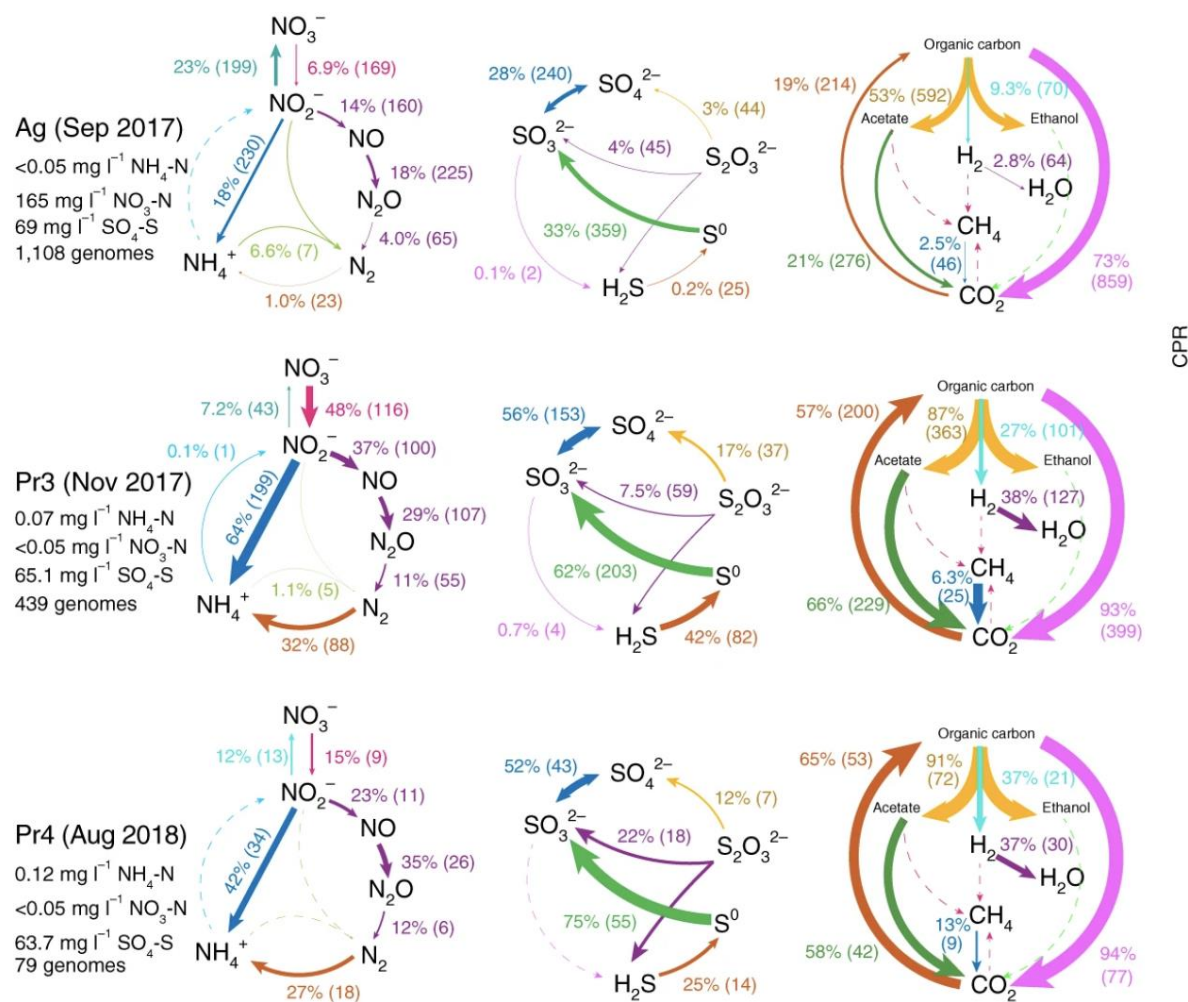


D



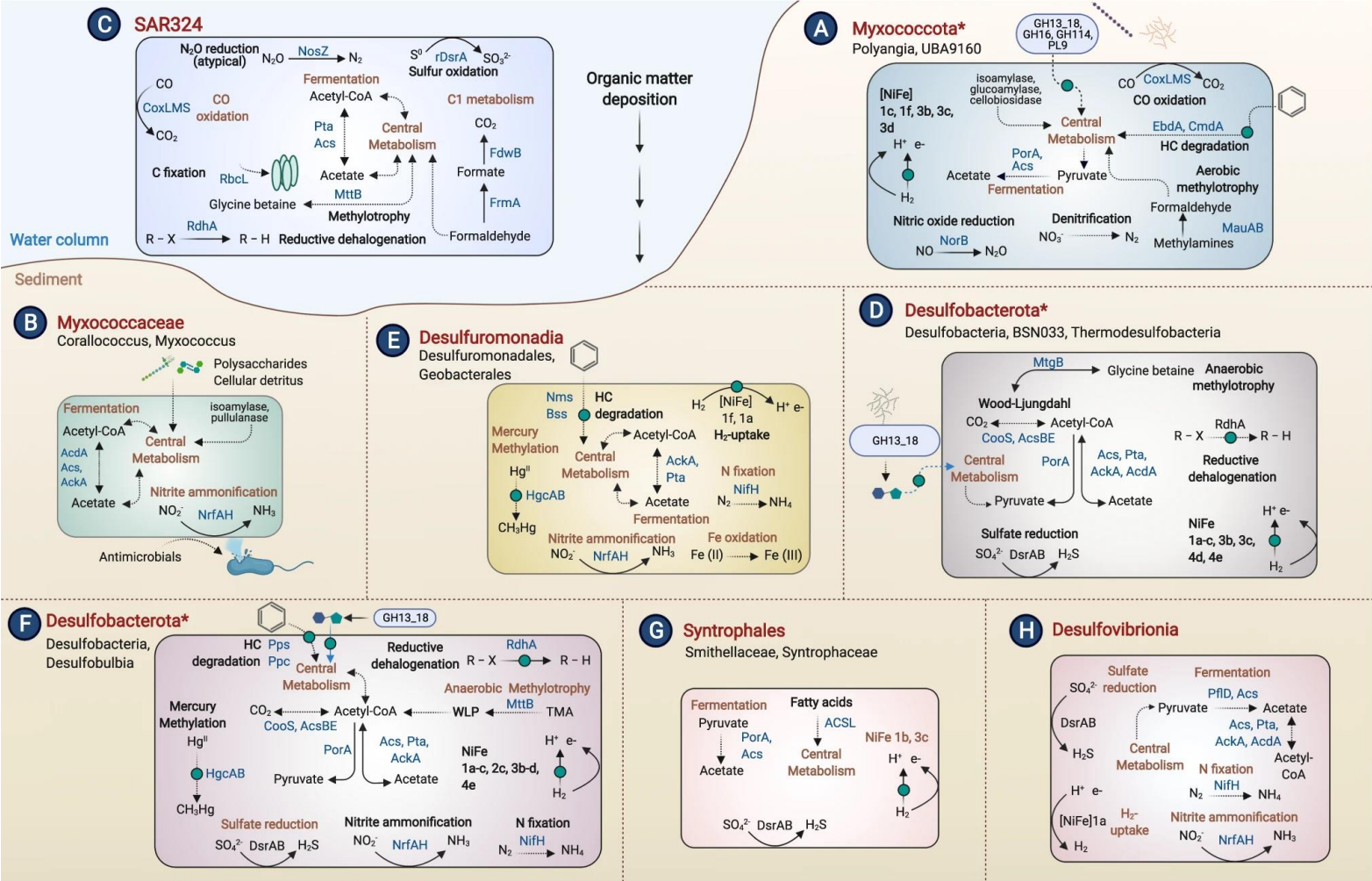
Applications of METABOLIC to different datasets: Groundwater

To compare across 2 environmental types (agricultural & pristine), over time



Applications of METABOLIC to different datasets: Phylum-level comparative genomics

To leverage the textual output to create their own user-made cellular maps



Langwig, M.V., De Anda, V., Dombrowski, N. *et al.* Large-scale protein level comparison of Deltaproteobacteria reveals cohesive metabolic groups. *ISME J* **16**, 307–320 (2022).

- METABOLIC is an open-source, free tool that aims to facilitate the annotations of genomes, or MAGS, with a focus on genes involved in biogeochemical cycling
- Focused on added annotation validation steps, multiple ways to explore the output data, and visualizations
- Developed an M-W score which can be used for standardized, reproducible comparisons between communities & samples
- Can be applied to diverse environments and utilized at different scales of analysis (genome or community).

