The Metaproteomic Analysis of Arctic Soils with Novel Bioinformatic Methods

Samuel Miller
University of Chicago
Department of the Geophysical Sciences
Dissertation Defense
November 1, 2018

Outline

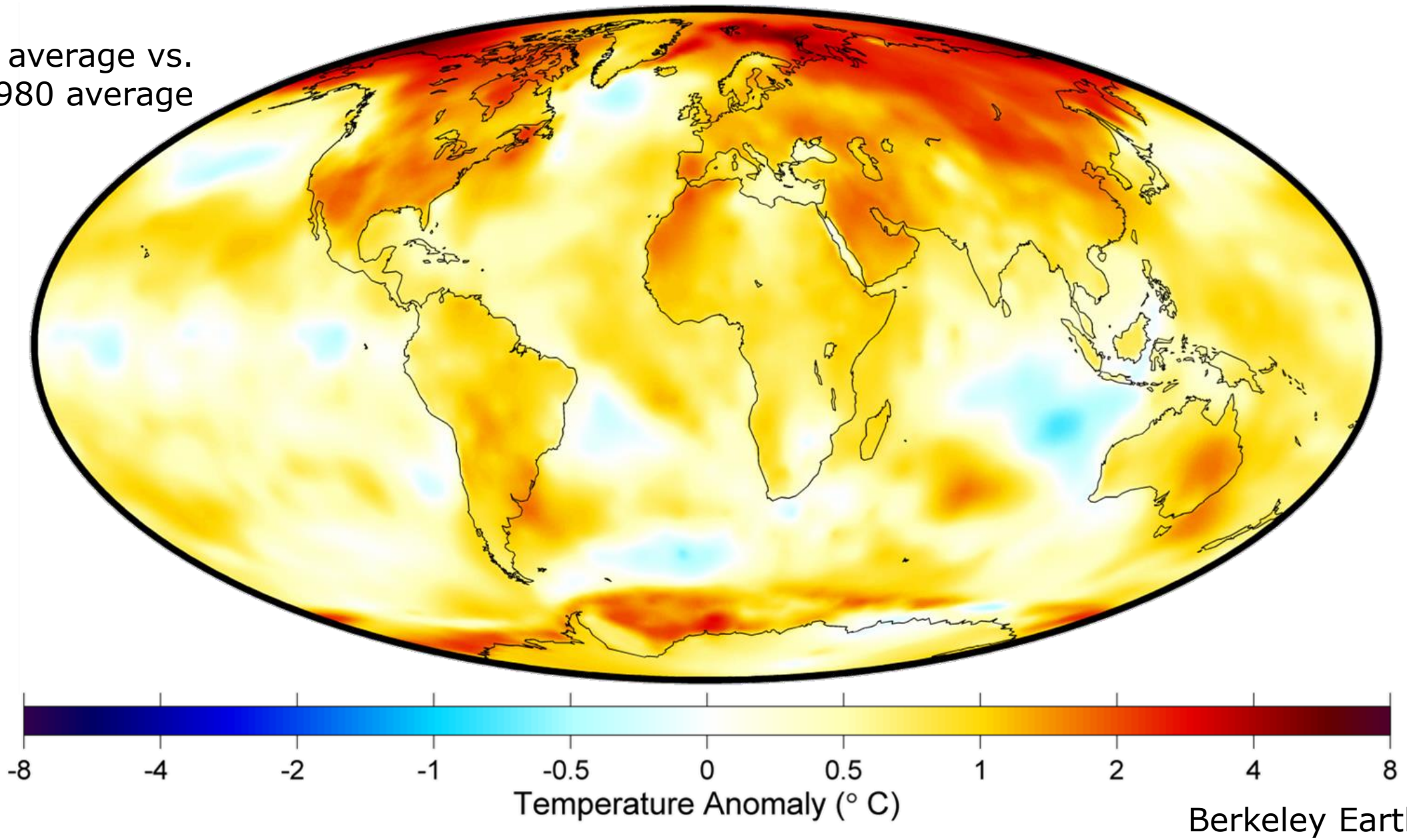
- Introduction
  - Rapid Arctic warming has large effects on plants and soil
  - Models of soil biogeochemistry can be improved with a greater understanding of microbial processes
  - Direct study of proteins clarifies these processes

- Proteomic methods development
  - *Postnovo* improves protein sequence accuracy
  - *ProteinExpress* increases the amount of useful data recovered from complex metaproteomes

- Arctic soil analyses
  - Identification of key biogeochemical processes and which microbes are doing them

Arctic warming

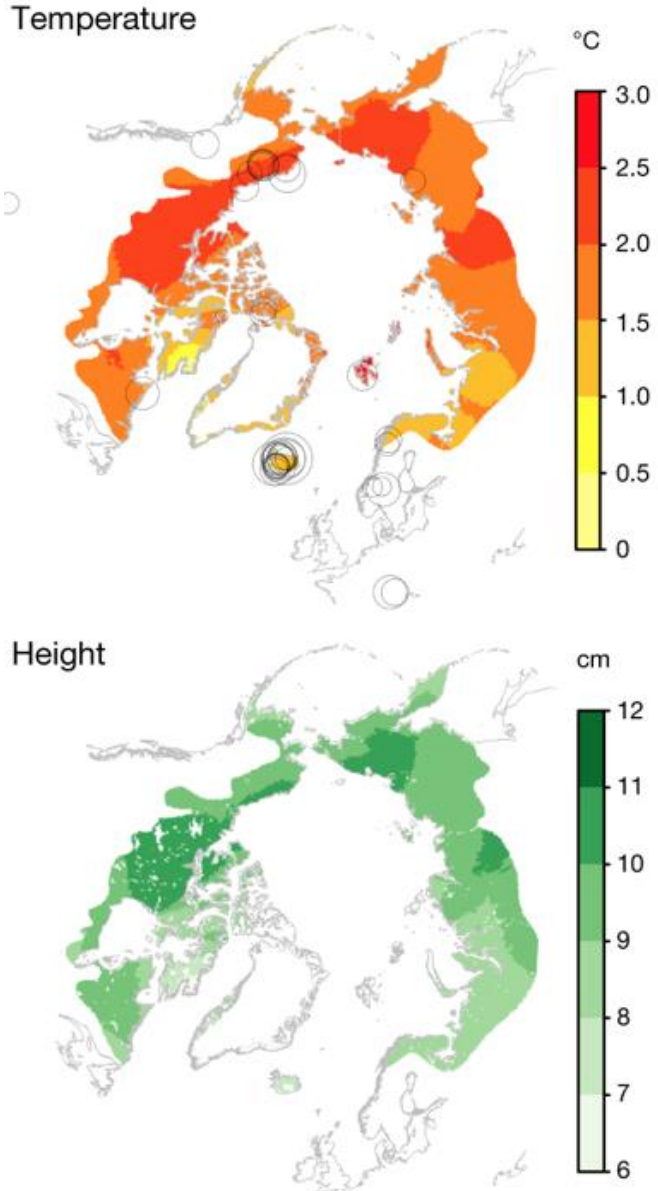
2017 average vs.
1951-1980 average



Berkeley Earth, 2018

Arctic greening

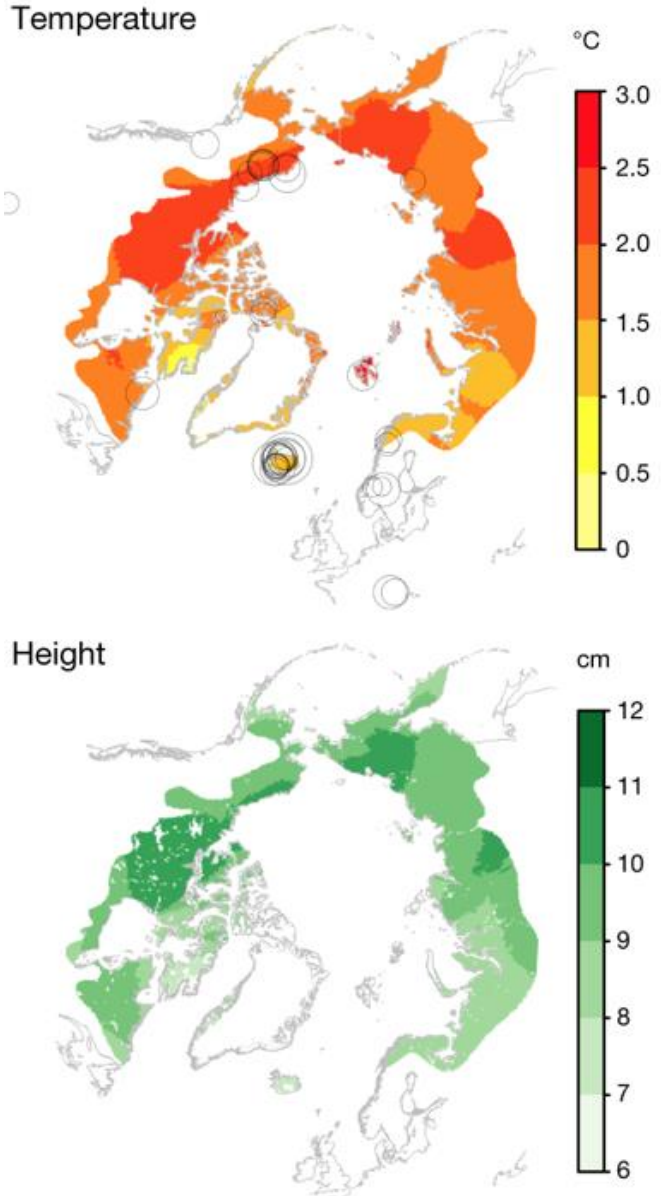
1979-2016



Bjorkman et al., 2018, *Nature*

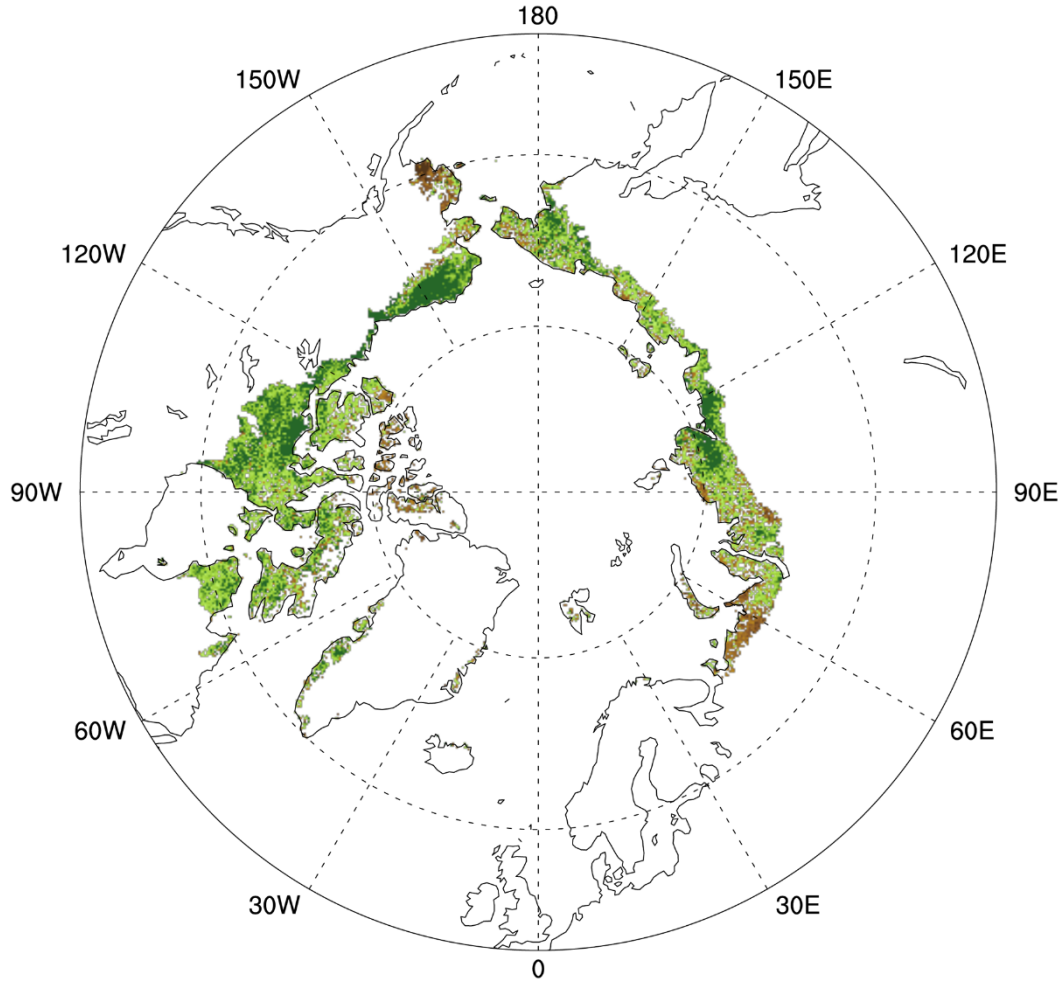
Arctic greening

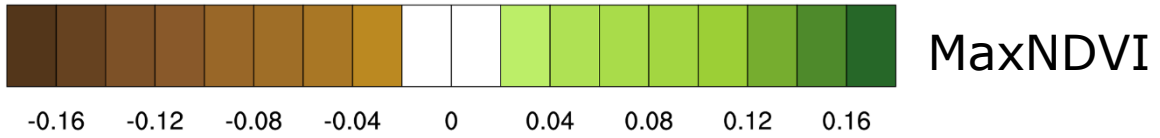
1979-2016



Bjorkman et al., 2018, *Nature*

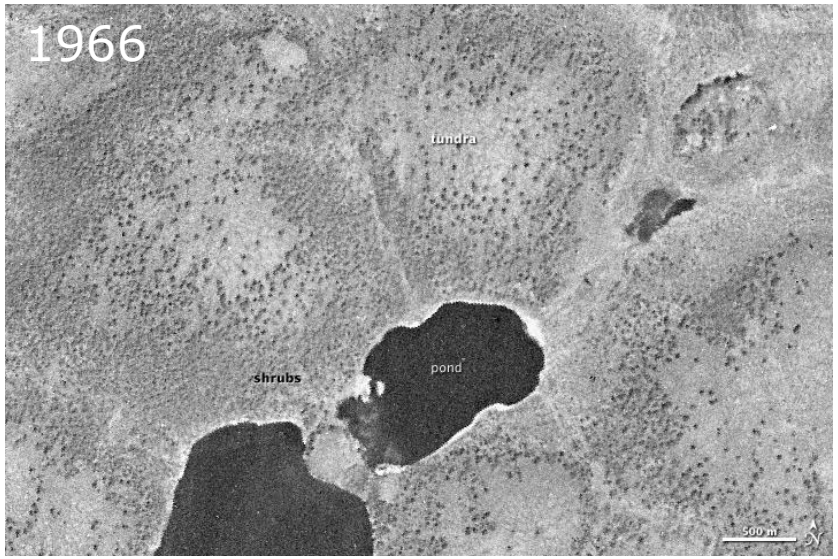
1982-2016

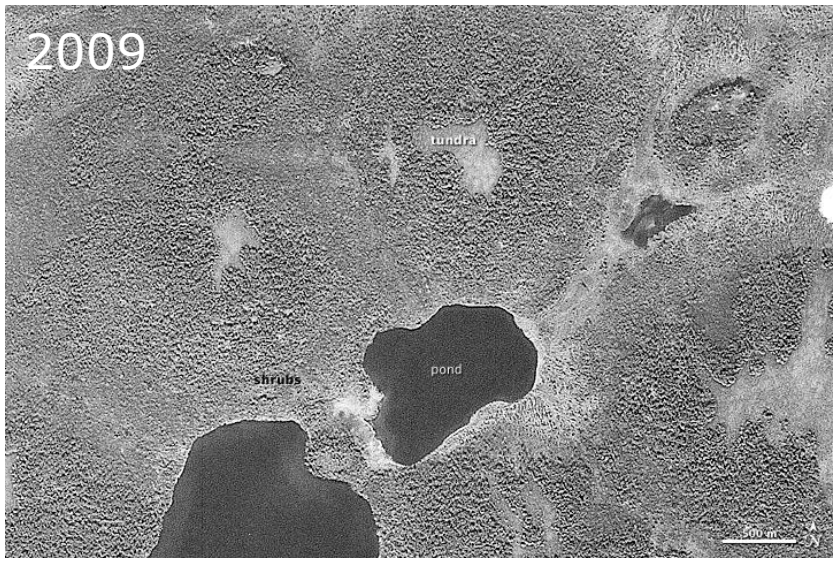




Epstein et al., 2017, *NOAA Arctic Report Card*

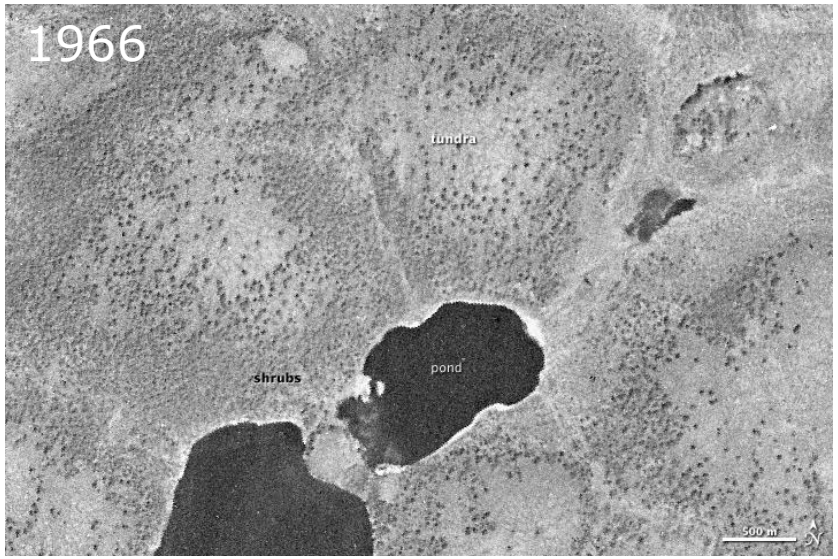
Changing floras

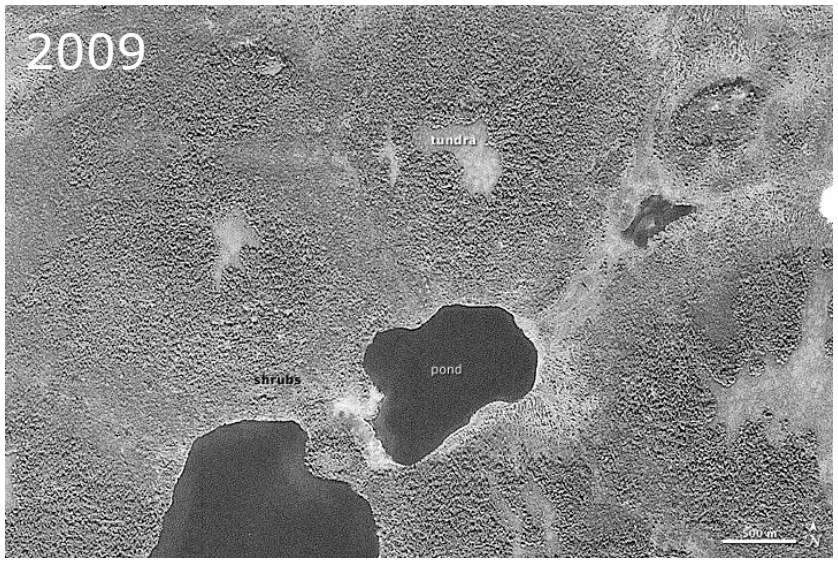




Frost, 2011,
NOAA Arctic Report Card

Changing floras

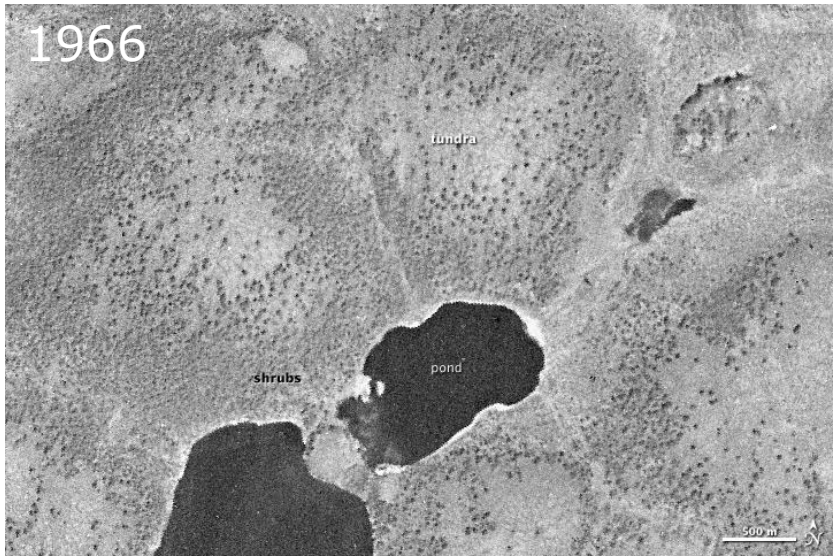


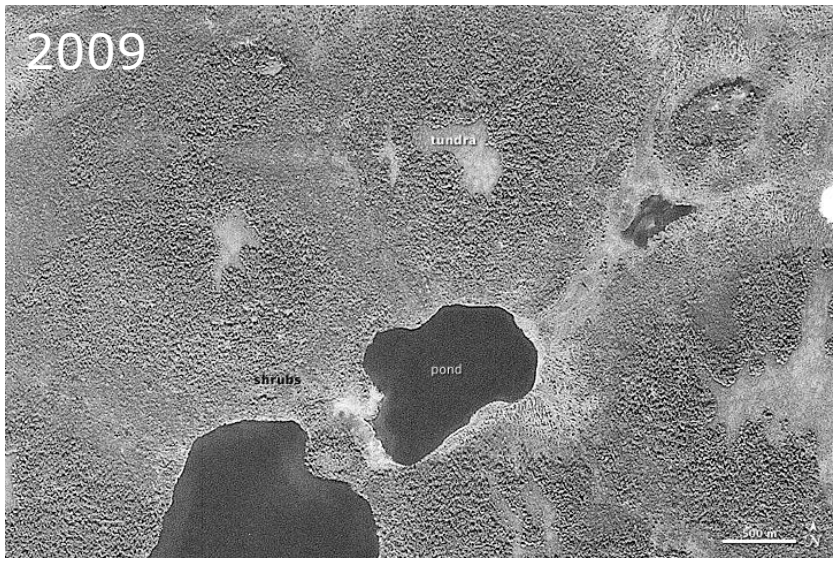


Frost, 2011,
NOAA Arctic Report Card

Myers-Smith, 2017, in *Medium*

Changing floras



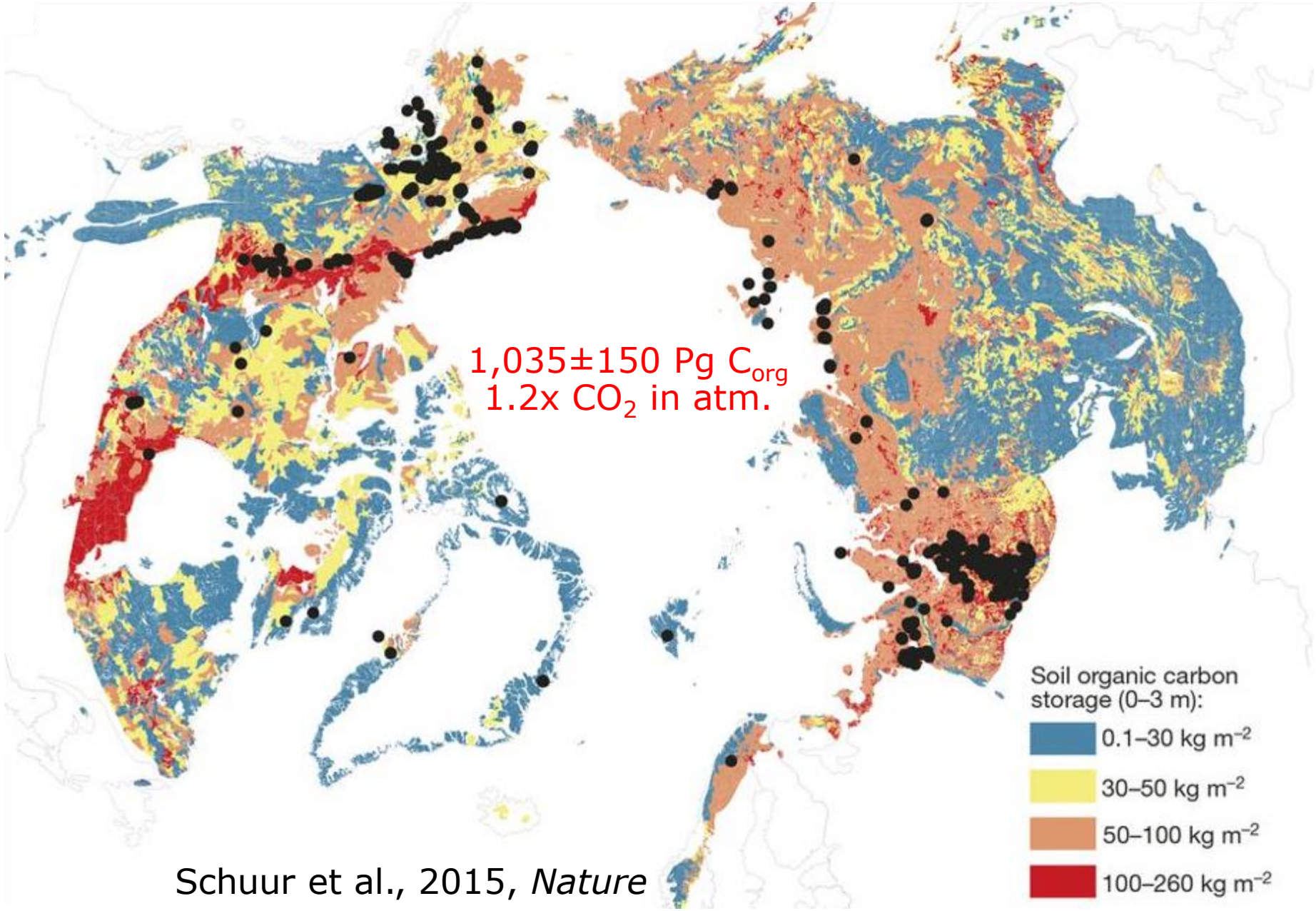


Frost, 2011,
NOAA Arctic Report Card

Myers-Smith, 2017, in *Medium*

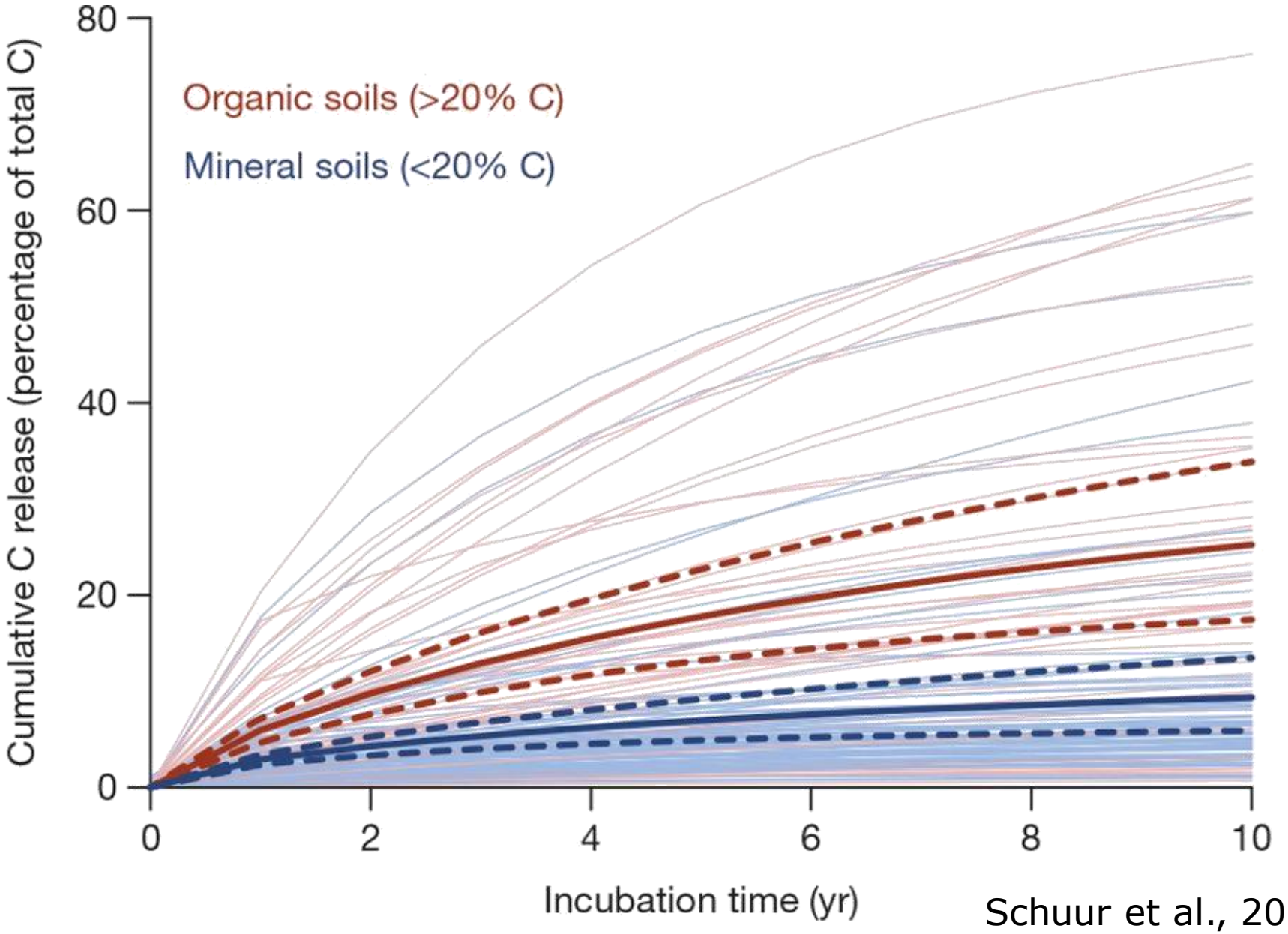
Walker,
Arctic Geobotanical Atlas

Large stock of soil carbon



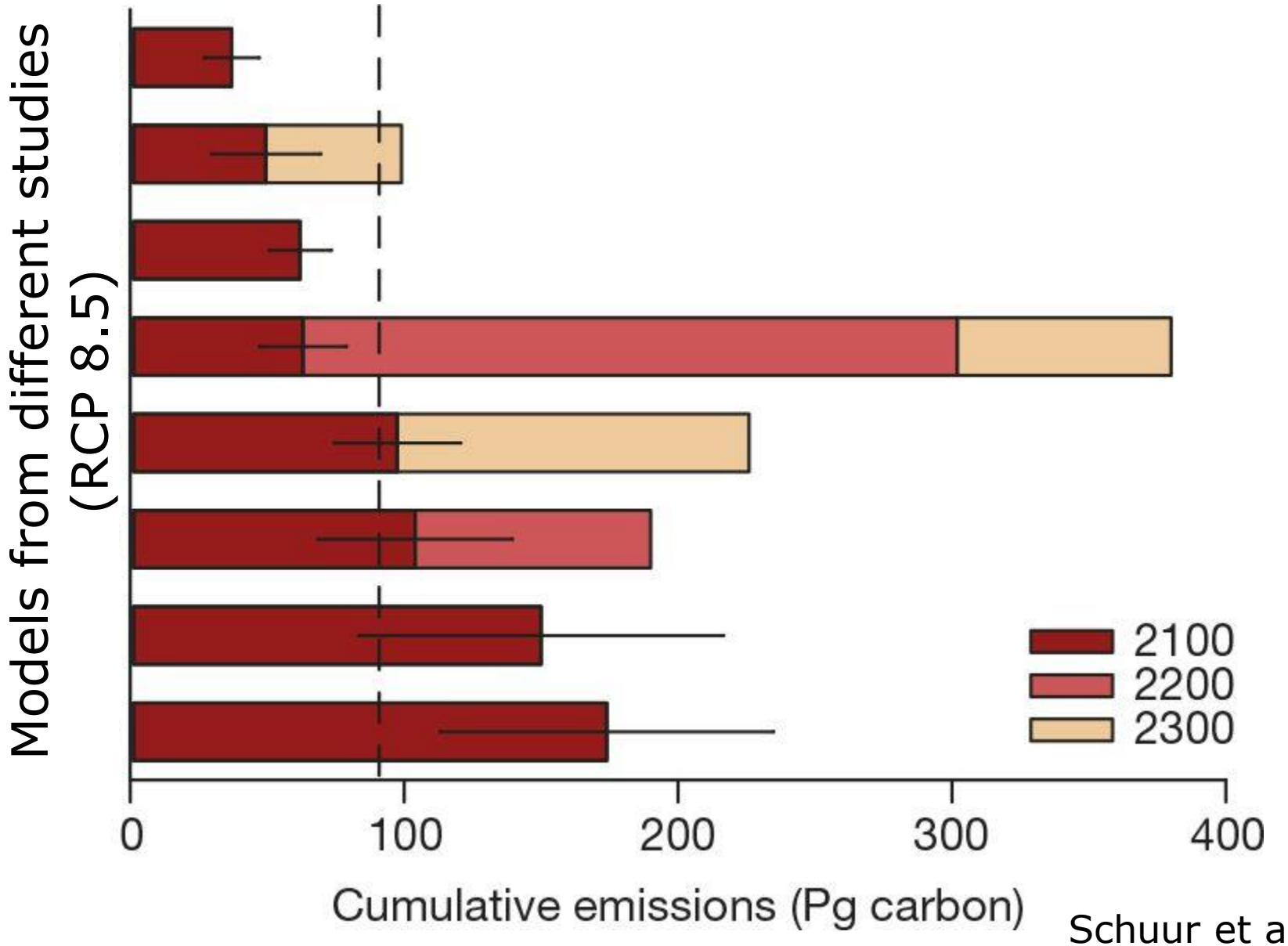
Schuur et al., 2015, *Nature*

Carbon loss in 5°C incubations



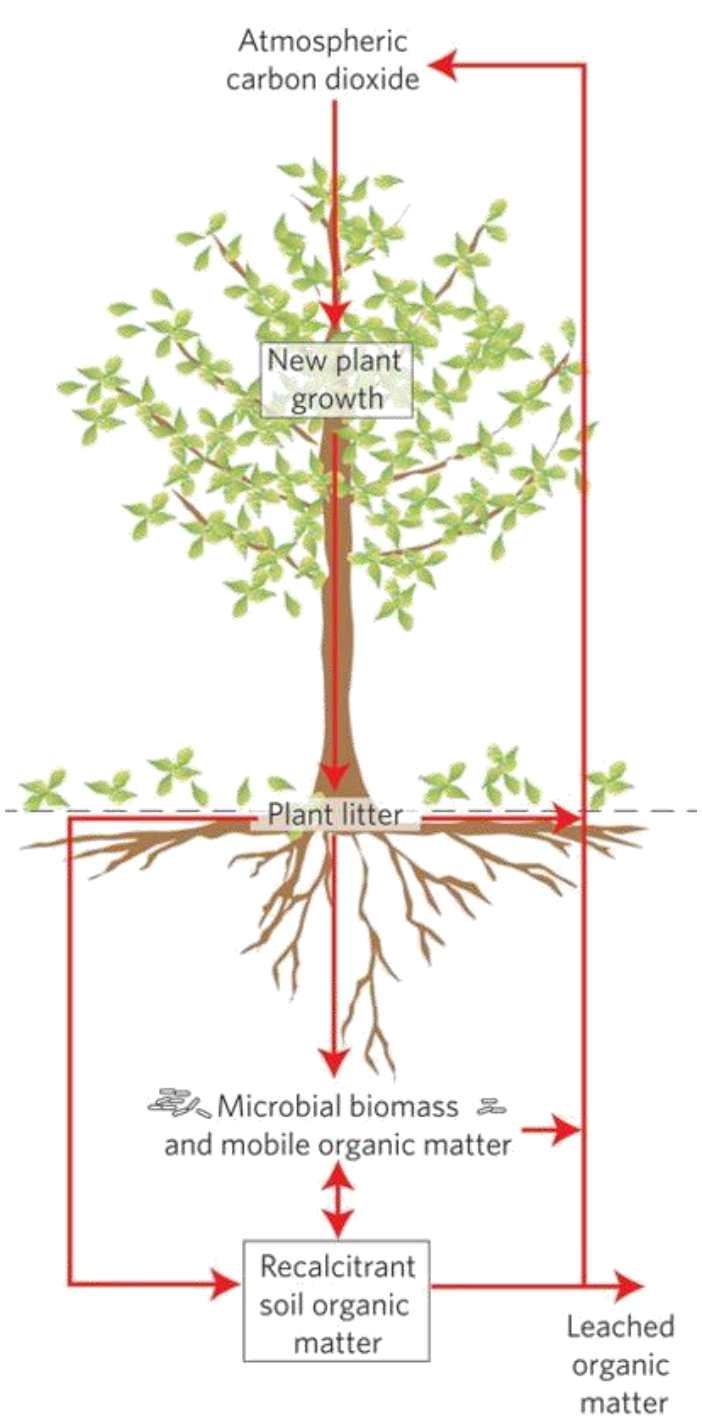
Schuur et al., 2015, *Nature*

Projected carbon loss



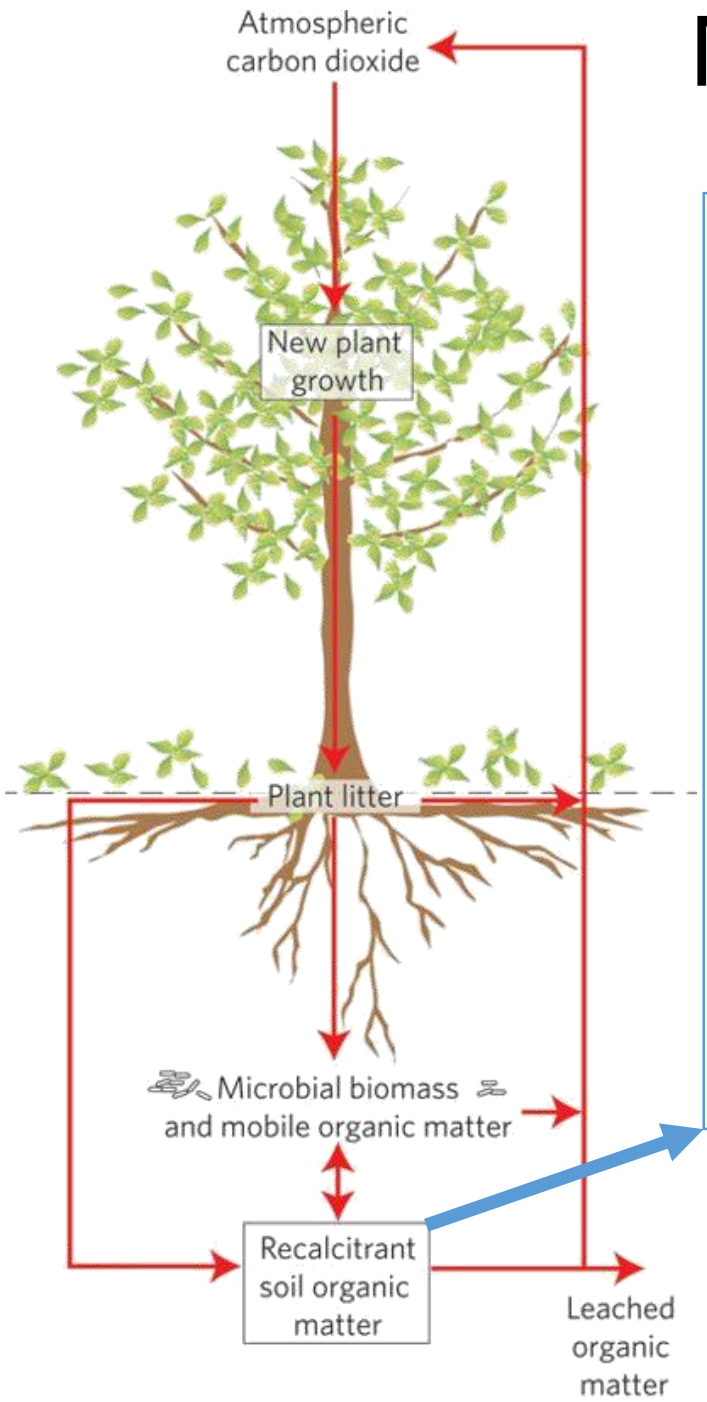
Schuur et al., 2015, *Nature*

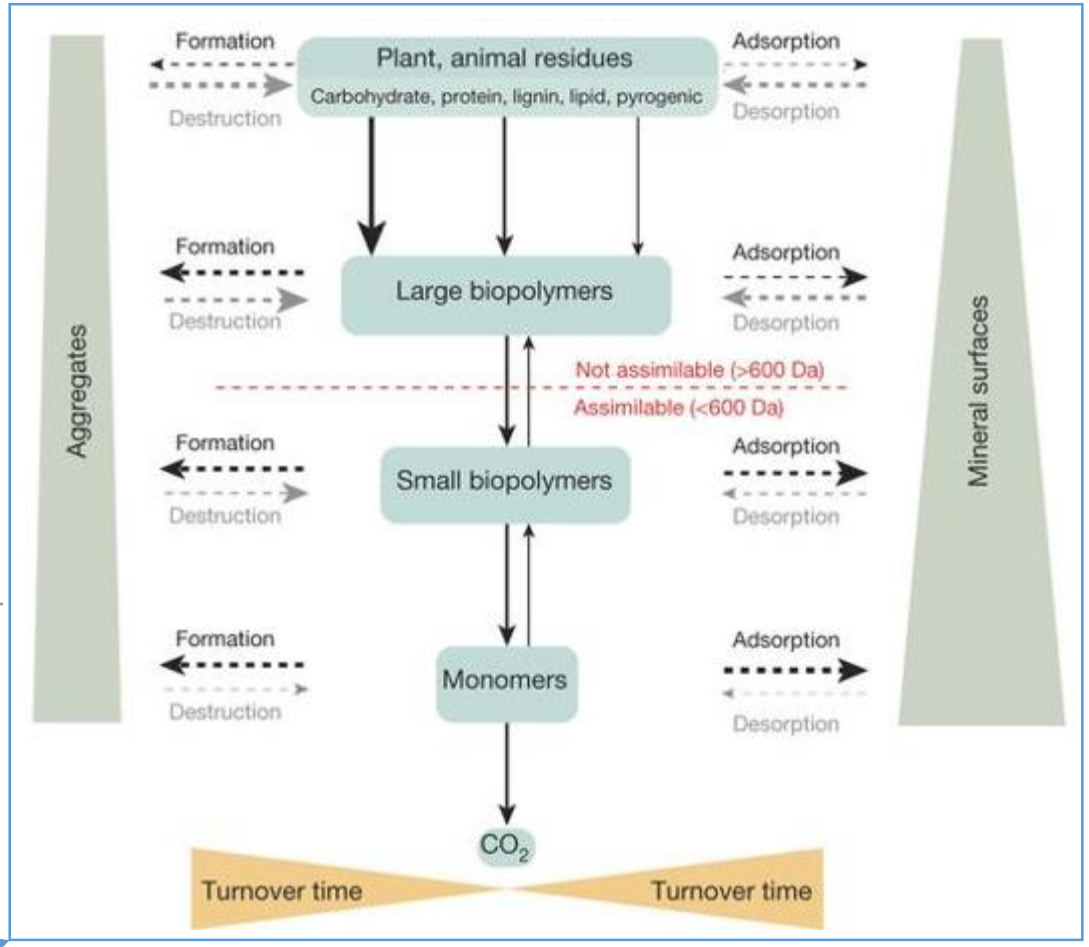
Carbon gain and loss



Talbot and Treseder, 2011, *Nature Climate Change*

Model of organic matter

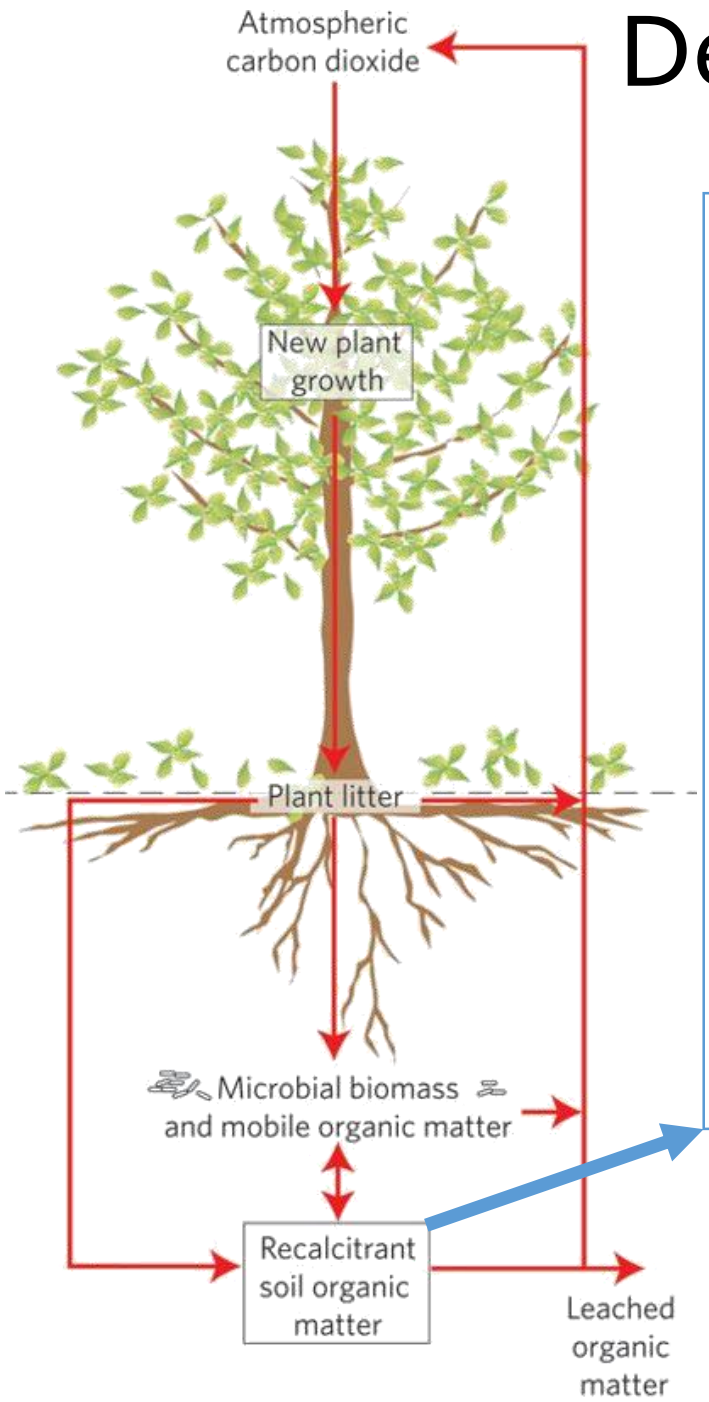




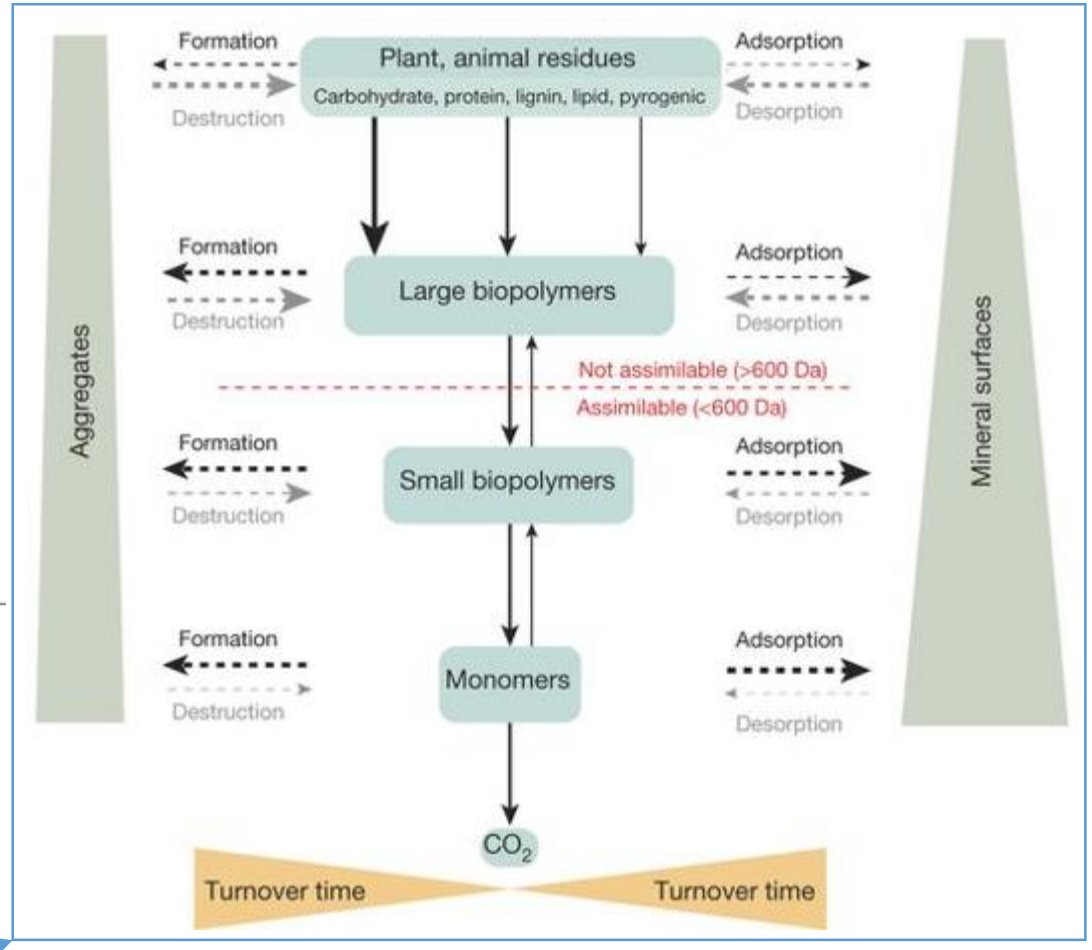
Lehmann and Kleber, 2015, *Nature*

Talbot and Treseder, 2011,
Nature Climate Change

Decomposition mechanism



Talbot and Treseder, 2011, *Nature Climate Change*

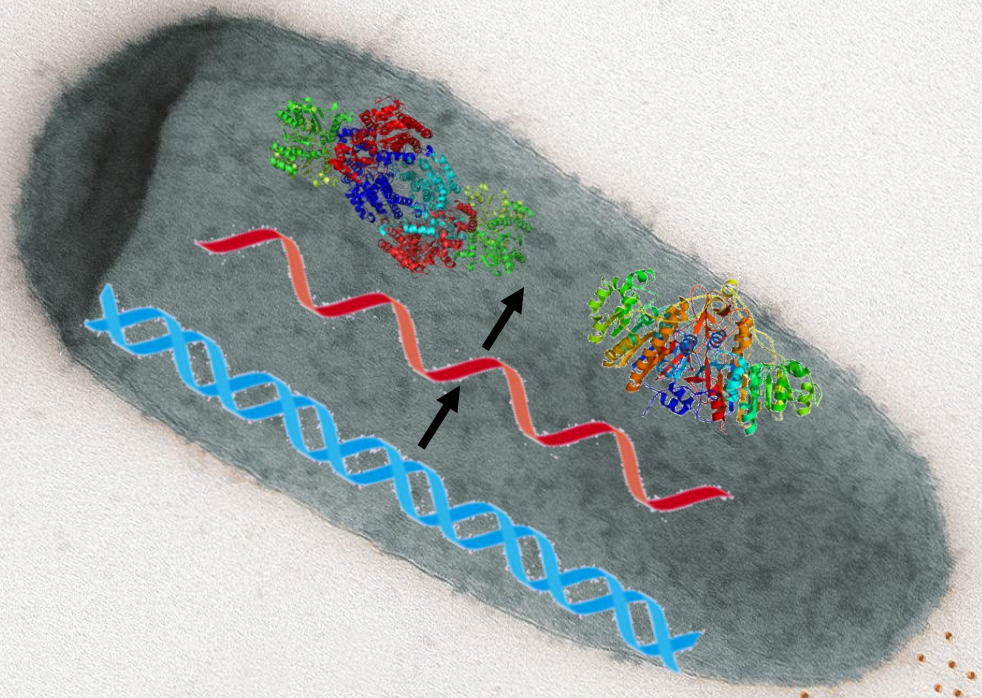


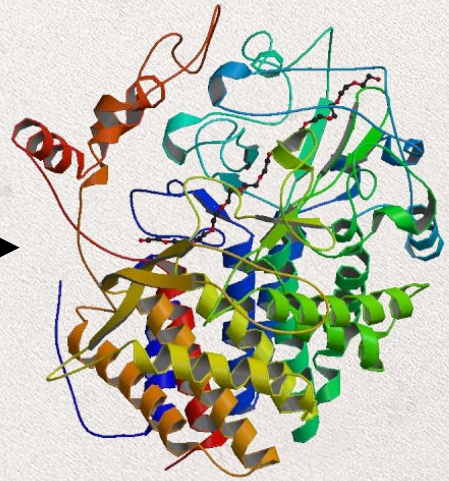
Lehmann and Kleber, 2015, *Nature*

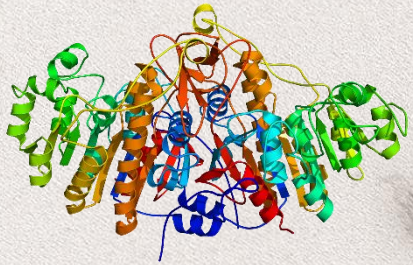
"[T]here are numerous individual processes and taxa associated with the metabolism of the thousands of organic compounds found in soil. This complexity makes it very difficult to predict soil function. If, for example, we want to know the fate of labile carbon compounds in soil (which is important in soil carbon models), information about what taxa are present in a given soil sample is unlikely to be useful."

-Noah Fierer, 2017, *Nature Rev. Microbiol.*

Proteins control decomposition: study the proteins



cellulose →  → glucose

R—PO_4 →  → PO_4^{3-}

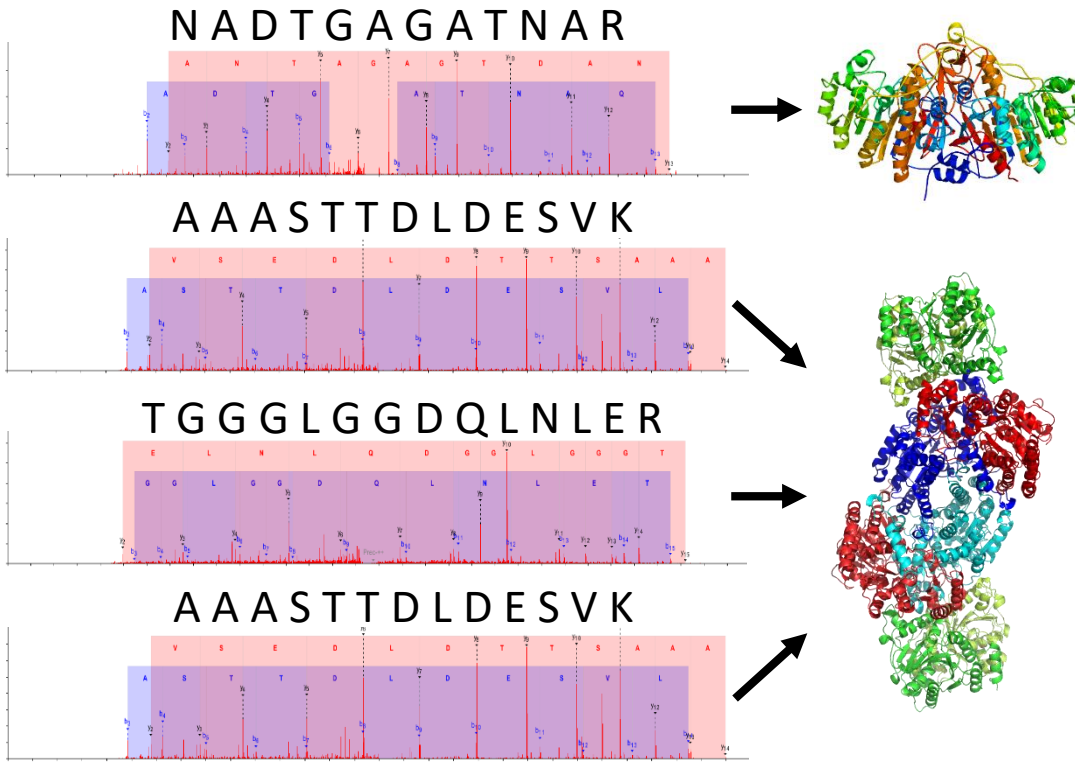
A. Kuijpers, UMC Utrecht

Metaproteomic workflow

Sample

Peptide extract

HPLC-tandem mass spectrometry



Mass spectra
Peptide ID

Protein ID

Postnovo: Miller, Rizzo, Waldbauer, 2018, *J. Proteome Res.*

ProteinExpress: this work, in prep

Peptide identification by database search

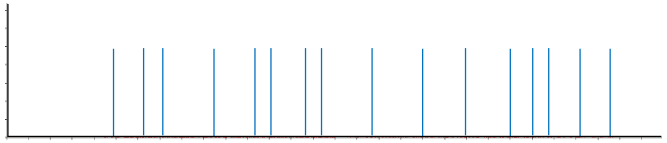
Database → Predicted protease cleavage/ fragment peaks → Compare predicted peptides to spectra with same precursor mass

>VIMSS208926 dnaA-1 chromosomal replication initiator protein DnaA (TIGR)
MLQQALRQHLRQTCSEQR LHQWFDPLDLR DDAQQRLEVRFPHPHFARWFEAQALERFTA
GVRDVVGTSVALVFPEGIKTTDRSTVQPPSQPPLASESVACPFGAAFTFDAFITNRKNQF
PLAAAREAARNGHQRTYNPFVLCGASGNGKTHLLRALANELAALYGTDAVFCGSAEELHD
RYNTEERLAMRRTLCAHRALLLDDLHRLRALPDLREELTALFDHFYDHGKQMAFAYAGRL
SDLDFLEAPLRSRLELGLIVDLKEPDLDVRVRYIHARCAALSLQLAREHVLTLAQRCHEF
RHLSGLLLKVAAYRDMVGRDILDRELEQILRNTDSGPERAVAPDTIISAAAEHFGVTPRD
ILSDKRQQHIVQARQVAMFLCRELIGSSFPALGRMFGGKDHSTAMYAVKKIKKLQYDNKD
MHALVAELKRRCLNHDG

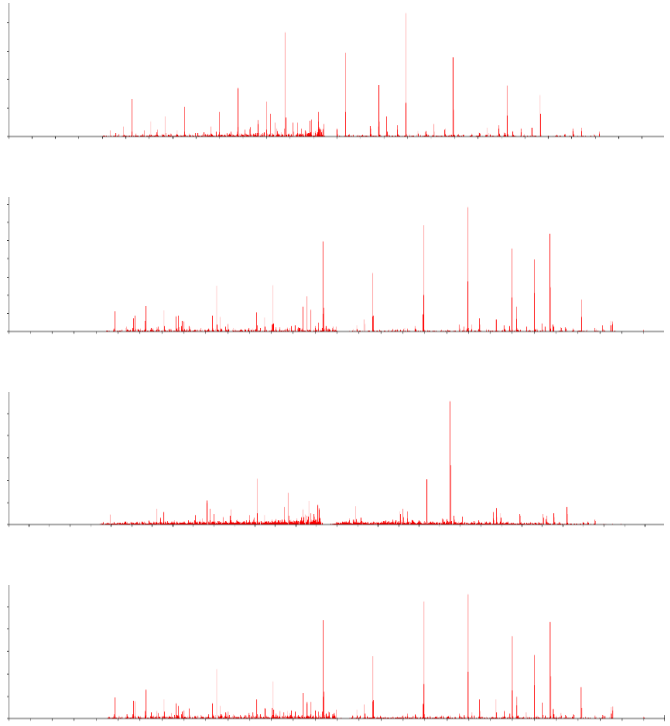
>VIMSS208928 dnaN DNA polymerase III, beta subunit (TIGR)
MYLKVYKEDVIEGLQKAANIIPAKTGAAYLRSIWLNAADGTLSVMATDSNIEFRGTYTAE
VTEPGLAGVQGRAFVDLLRKLPAGEIVLQLDAEANVLRIEQGRRKYKLPVNDPVWFQNFS
DFPADGAVVWSGDFLQELIDRIAFCISDEDAVEAIACLFMKPVAEGRIEACGLNGHQFAM
LRFLHDDLHAKLPAEGVLVQKKYLGELKKWLGADEIELNISEKRMHLRTGDGRETFSLPL
SSYQYPDYMNFIAKLQGEGVSNLEVDRKEGMDALDRLQIFNSDNNRCTYFDLSGGEVVLT
AQGQDVGSASESLEASYDGDIRRIAFPTRNLIDIMNHYQSGRLRLTLTGAEGPCGISGEE
DPEYQVIVMPMKIVEETYYSEEEV

>VIMSS208929 gyrB DNA gyrase, B subunit (TIGR)
MTTGTGNNYTADSITILEGLSAVRKRPAMYIGSTDARGLHHLVYEVVDNSIDEAMAGFCS
KVVVKLHLDNSVTVSDNGRGIPVDMHPKEGRPAVEVVMTKLHAGGKFDNNAYKVSGGLHG
VGVSCVNALSEWLEVRVKRDGQRYGQRYARGVPQEDLRVLGASEGHGTTVRFKPDEEIFE
TNQFSYETLRKRFEELSYLNRGLEIEFIDERSGESHTFFAEGGIRQFVKDLNSGETGIHP
IVAGEGKTDGVVVDFALQYNAGYKENVLTFANNIRTKEGGTHLAGFKTALTRAINGYVKG
QADLTKKMKGTALSGDDVREGLTAVVSVKLPQPQFEGQTKTKLGNSEIAGIVAGIVYDRL
NVHFEENPKDVRLIIDKAVDAARARDAARRAKELVRRKGALSDNALPGKLADCQSKDPAL
SELFIVEGDSAGGSAKQGRDPSTQAILPLRGKILNTERTRFDKMLANKEVKALITAMGAG
IGEDDTDYDKLRYHKIVIMTDADVDGAHIRTLLLTFFFRQYQKLIESGYLYIAQPPLYRA
HNSRMERFIKDDAELNNFLLTRVSEDVEVRTPGGCSFKGTDIVRLMRSIETVAARIRDVE
TAGISRELFLGFLEYGERITPEWFEQHDGNGLRPWLAGRGYSMTAELESGDEGDRTFVVF
ENTNGHRTRIGAEFFGSRMYRSSWDALEAIRSECGGLSFTLDRKGETVGEGDMFDLQRMV
FEEARRGLNIQRYKGLGEMNPEQLWVTTMNPENRTLLQVSIEDAEEASDAFEQLMGDRVE
PRREFIVRNALAVQELDI

L H Q W F D P L D L R
mass = 1438.736 Da



Experimental fragmentation spectra
precursor masses ≈ 1438.736 Da



Peptide identification by database search

Database → Predicted protease cleavage/ fragment peaks → Compare predicted peptides to spectra with same precursor mass

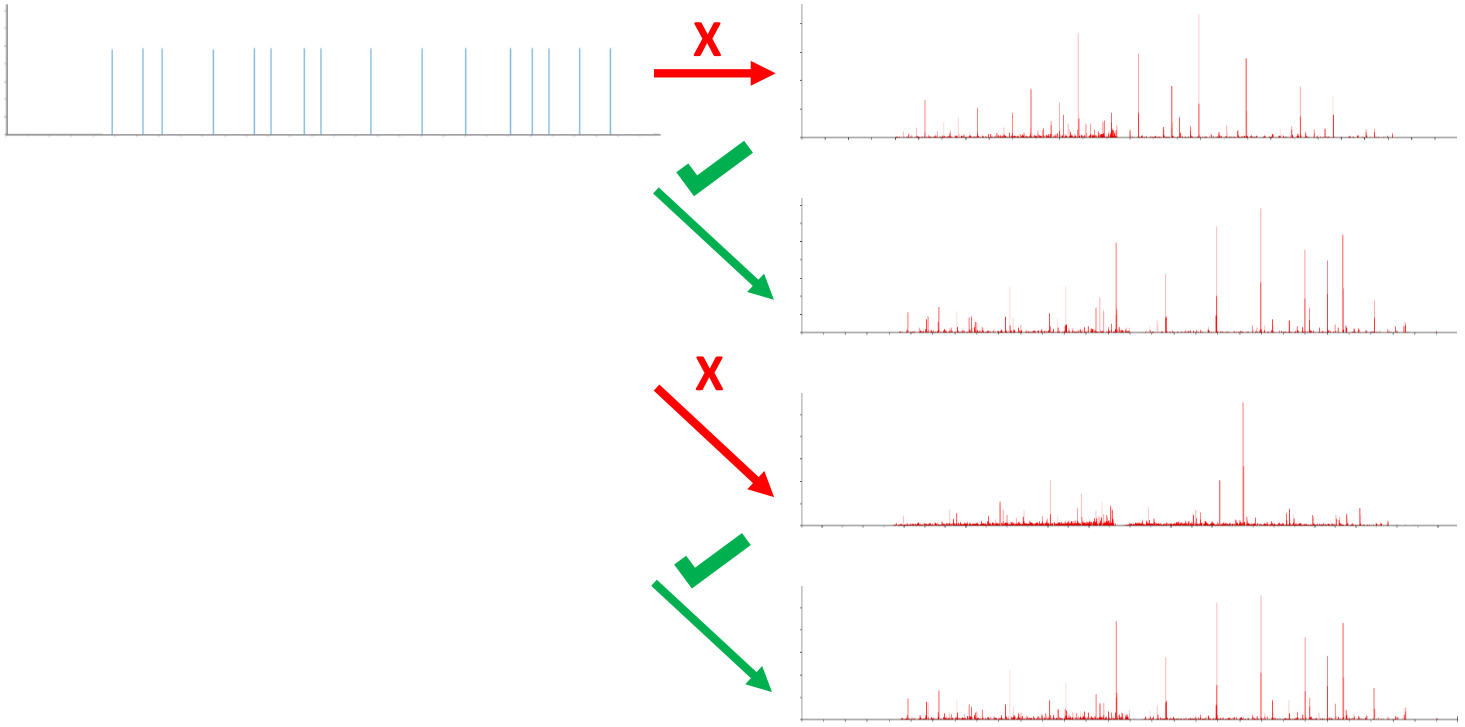
>VIMSS208926 dnaA-1 chromosomal replication initiator protein DnaA (TIGR)
MLQQALRQHLRQTCSEQR LHQWFDPLDLR DDAQQRLEVRFPHPHFARWFEAQALERFTA
GVRDVVGTSVALVFPEGIKTTDRSTVQPPSQPPLASESVACPFGAAFTFDAFITNRKNQF
PLAAAREAARNGHQRTYNPFVLCGASGNGKTHLLRALANELAALYGTDAVFCGSAEELHD
RYNTEERLAMRRTLCAHRALLLDDLHRLRALPDLREELTALFDHFYDHGKQMAFAYAGRL
SDLDFLEAPLRSRLELGLIVDLKEPDLDVRVRYIHARCAALSLQLAREHVLTLAQRCHEF
RHLSGLLLKVAAYRDMVGRDILDRELEQILRNTDSGPERAVAPDTIISAAAEHFGVTPRD
ILSDKRQQHIVQARQVAMFLCRELIGSSFPALGRMFGGKDHSTAMYAVKKIKKLQYDNKD
MHALVAELKRRCLNHDG

>VIMSS208928 dnaN DNA polymerase III, beta subunit (TIGR)
MYLKVYKEDVIEGLQKAANIIPAKTGAAYLRSIWLNAADGTLSVMATDSNIEFRGTYTAE
VTEPGLAGVQGRAFVDLLRKLPAGEIVLQLDAEANVLRIEQGRRKYKLPVNDPVWFQNFS
DFPADGAVVWSGDFLQELIDRIAFCISDEDAVEAIACLFMKPVAEGRIEACGLNGHQFAM
LRFLHDDLHAKLPAEGVLVQKKYLGELKKWLGADEIELNISEKRMHLRTGDGRETFSLPL
SSYQYPDYMNFIAKLQGEGVSNLEVDRKEGMDALDRLQIFNSDNNRCTYFDLSGGEVVLT
AQGQDVGSASESLEASYDGDIRRIAFPTRNLIDIMNHYQSGRLRLTLTGAEGPCGISGEE
DPEYQVIVMPMKIVEETYYSEEEV

>VIMSS208929 gyrB DNA gyrase, B subunit (TIGR)
MTTGTGNNYTADSITILEGLSAVRKRPAMYIGSTDARGLHHLVYEVVDNSIDEAMAGFCS
KVVVKLHLDNSVTVSDNGRGIPVDMHPKEGRPAVEVVMTKLHAGGKFDNNAYKVSGGLHG
VGVSCVNALSEWLEVRVKRDGQRYGQRYARGVPQEDLRVLGASEGHGTTVRFKPDEEIFE
TNQFSYETLRKRFEELSYLNRGLEIEFIDERSGESHTFFAEGGIRQFVKDLNSGETGIHP
IVAGEGKTDGVVVDFALQYNAGYKENVLTFANNIRTKEGGTHLAGFKTALTRAINGYVKG
QADLTKKMKGTALSGDDVREGLTAVVSVKLPQPQFEGQTKTKLGNSEIAGIVAGIVYDRL
NVHFEENPKDVRLIIDKAVDAARARDAARRAKELVRRKGALSDNALPGKLADCQSKDPAL
SELFIVEGDSAGGSAKQGRDPSTQAILPLRGKILNTERTRFDKMLANKEVKALITAMGAG
IGEDDTDYDKLRYHKIVIMTDADVDGAHIRTLLLTFFFRQYQKLIESGYLYIAQPPLYRA
HNSRMERFIKDDAELNNFLLTRVSEDVEVRTPGGCSFKGTDIVRLMRSIETVAARIRDVE
TAGISRELFLGFLEYGERITPEWFEQHDGNGLRPWLAGRGYSMTAELESGDEGDRTFVVF
ENTNGHRTRIGAEFFGSRMYRSSWDALEAIRSECGGLSFTLDRKGETVGEGDMFDLQRMV
FEEARRGLNIQRYKGLGEMNPEQLWVTTMNPENRTLLQVSIEDAEEASDAFEQLMGDRVE
PRREFIVRNALAVQELDI

L H Q W F D P L D L R
mass = 1438.736 Da

Experimental fragmentation spectra
precursor masses ≈ 1438.736 Da



Database search can’t find unexpected sequences

>VIMSS208926 dnaA-1 chromosomal replication initiator protein DnaA (TIGR)
MLQQALRQHLRQTCSEQR LHQWFDPLDLR DDAQQRLEVRFPHPHFARWFEAQALERFTA
GVRDVVGTSVALVFPEGIKTTDRSTVQPPSQPPLASESVACPFGAAFTFDAFITNRKNQF
PLAAAREAARNGHQRTYNPFVLCGASGNGKTHLLRALANELAALYGTDAVFCGSAEELHD
RYNTEERLAMRRTLCAHRALLLDDLHRLRALPDLREELTALFDHFYDHGKQMAFAYAGRL
SDLDFLEAPLRSRLELGLIVDLKEPDLDVRVRYIHARCAALSLQLAREHVLTLAQRCHEF
RHLSGLLLKVAAYRDMVGRDILDRELEQILRNTDSGPERAVAPDTIISAAAEHFGVTPRD
ILSDKRQQHIVQARQVAMFLCRELIGSSFPALGRMFGGKDHSTAMYAVKKIKKLQYDNKD
MHALVAELKRRCLNHDG

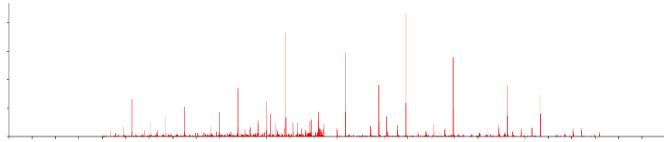
>VIMSS208928 dnaN DNA polymerase III, beta subunit (TIGR)
MYLKVYKEDVIEGLQKAANIIPAKTGAAYLRSIWLNAADGTLSVMATDSNIEFRGTYTAE
VTEPGLAGVQGRAFVDLLRKLPAGEIVLQLDAEANVLRIEQGRRKYKLPVNDPVWFQNFS
DFPADGAVVWSGDFLQELIDRIAFCISDEDAVEAIACLFMKPVAEGRIEACGLNGHQFAM
LRFLHDDLHAKLPAEGVLVQKKYLGELKKWLGADEIELNISEKRMHLRTGDGRETFSLPL
SSYQYPDYMNFIAKLQGEGVSNLEVDRKEGMDALDRLQIFNSDNNRCTYFDLSGGEVVLT
AQGQDVGSASESLEASYDGDIRRIAFPTRNLIDIMNHYQSGRLRLTLTGAEGPCGISGEE
DPEYQVIVMPMKIVEETYYSEEEV

>VIMSS208929 gyrB DNA gyrase, B subunit (TIGR)
MTTGTGNNYTADSITILEGLSAVRKRPAMYIGSTDARGLHHLVYEVVDNSIDEAMAGFCS
KVVVKLHLDNSVTVSDNGRGIPVDMHPKEGRPAVEVVMTKLHAGGKFDNNAYKVSGGLHG
VGVSCVNALSEWLEVRVKRDGQRYGQRYARGVPQEDLRVLGASEGHGTTVRFKPDEEIFE
TNQFSYETLRKRFEELSYLNRGLEIEFIDERSGESHTFFAEGGIRQFVKDLNSGETGIHP
IVAGEGKTDGVVVDFALQYNAGYKENVLTFANNIRTKEGGTHLAGFKTALTRAINGYVKG
QADLTKKMKGTALSGDDVREGLTAVVSVKLPQPQFEGQTKTKLGNSEIAGIVAGIVYDRL
NVHFEENPKDVRLIIDKAVDAARARDAARRAKELVRRKGALSDNALPGKLADCQSKDPAL
SELFIVEGDSAGGSAKQGRDPSTQAILPLRGKILNTERTRFDKMLANKEVKALITAMGAG
IGEDDTDYDKLRYHKIVIMTDADVDGAHIRTLLLTFFFRQYQKLIESGYLYIAQPPLYRA
HNSRMERFIKDDAELNNFLLTRVSEDVEVRTPGGCSFKGTDIVRLMRSIETVAARIRDVE
TAGISRELFLGFLEYGERITPEWFEQHDGNGLRPWLAGRGYSMTAELESGDEGDRTFVVF
ENTNGHRTRIGAEFFGSRMYRSSWDALEAIRSECGGLSFTLDRKGETVGEGDMFDLQRMV
FEEARRGLNIQRYKGLGEMNPEQLWVTTMNPENRTLLQVSIEDAEEASDAFEQLMGDRVE
PRREFIVRNALAVQELDI

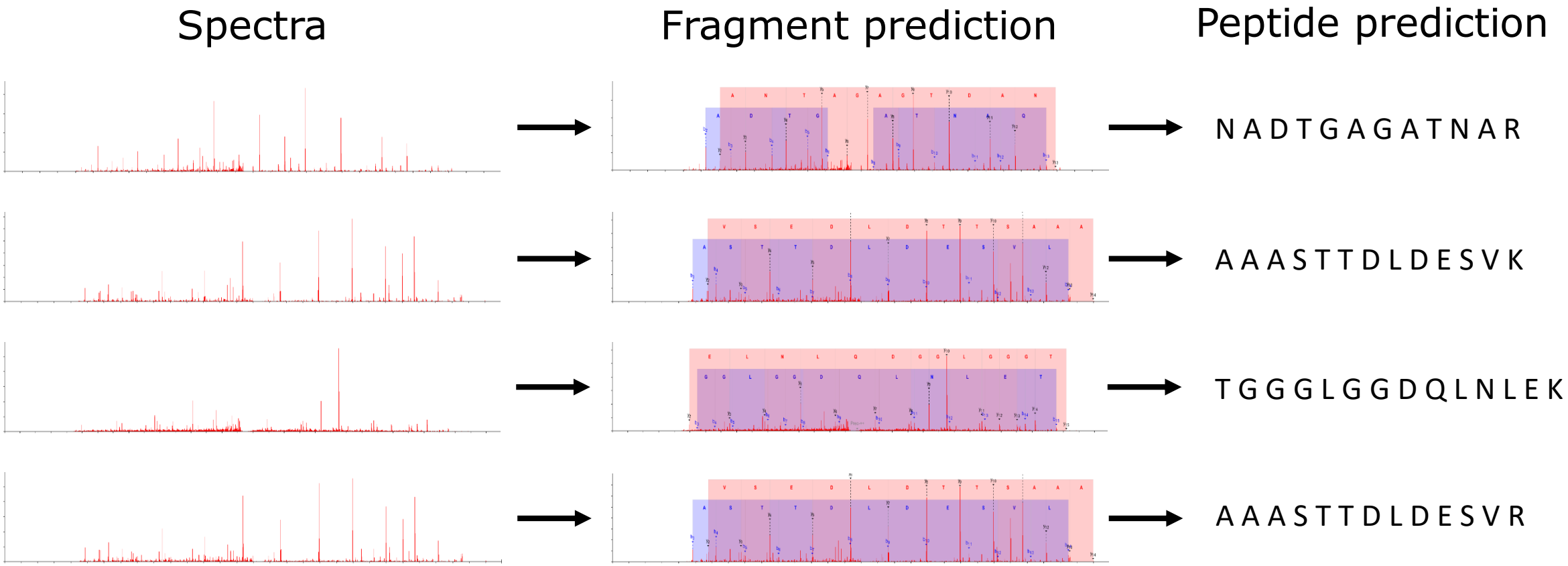
L H Q W F D P L D L R
mass = 1438.736 Da

X

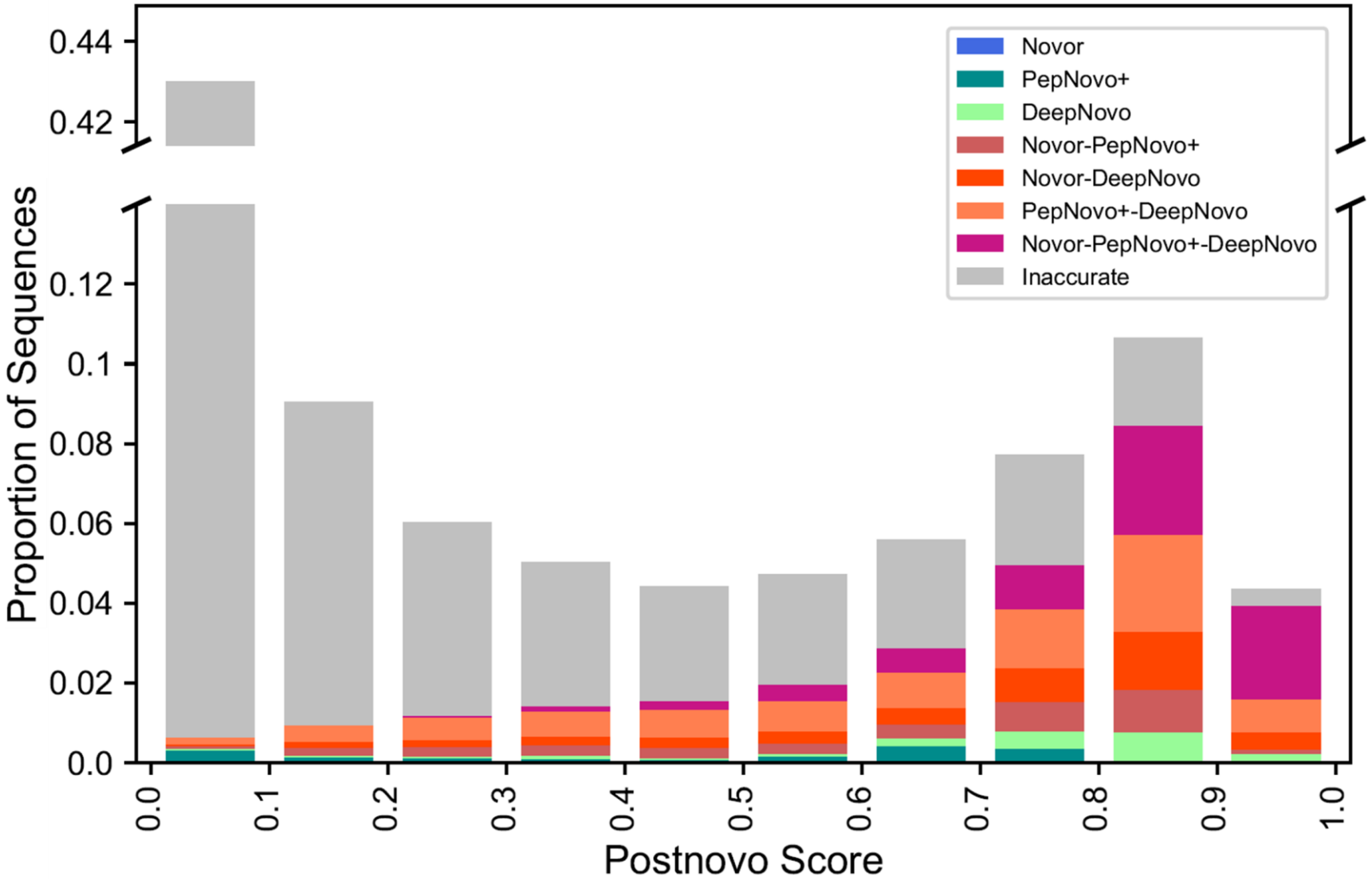
Experimental fragmentation spectra
L H Q W F E P L D L R
mass = 1452.751 Da ≠ 1438.736 Da



An alternate approach: “de novo” sequencing

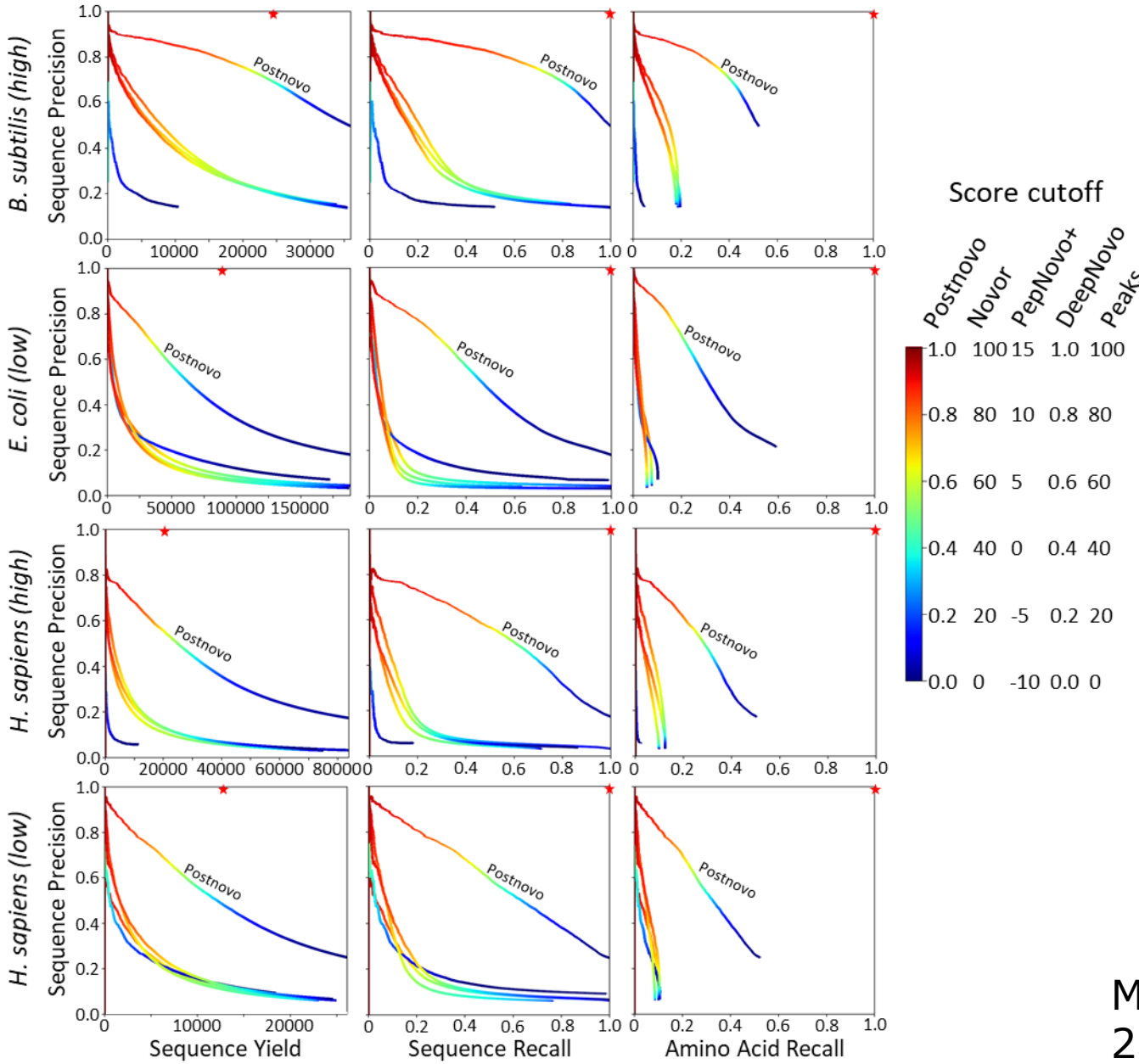


Postnovo refines de novo sequence predictions



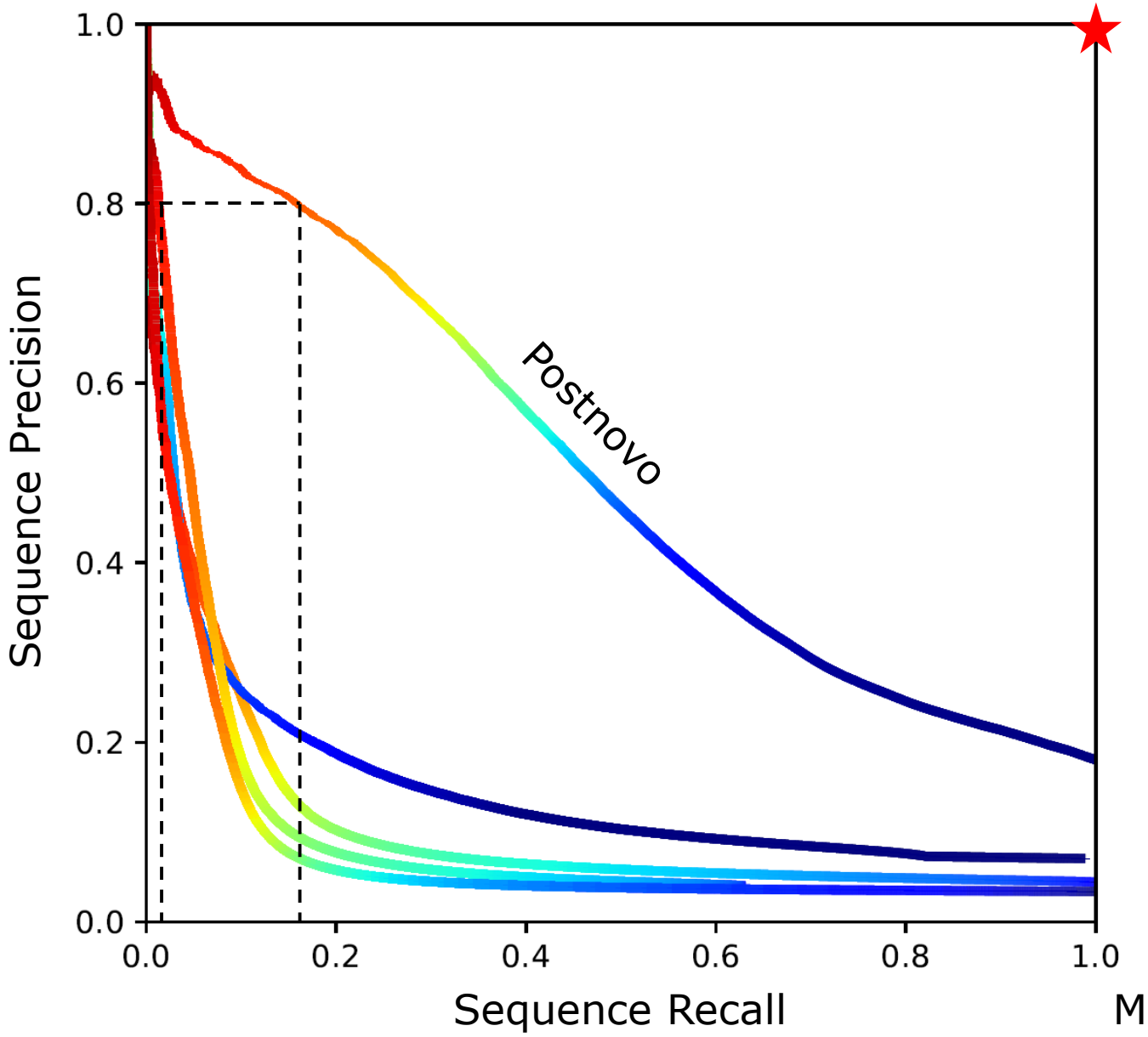
Miller, Rizzo, Waldbauer, 2018, *J. Proteome Res.*

Postnovo validation with simple datasets



Miller, Rizzo, Waldbauer, 2018, *J. Proteome Res.*

Ten-fold increase in yield of accurate sequences



Miller, Rizzo, Waldbauer, 2018, *J. Proteome Res.*

Additional bioinformatic advances

- Taxonomic/functional annotation of short de novo sequences
  - Very short de novo sequences cannot be linked to related sequences in large databases – obvious for a sequence of length 1
  - Sequences >11 amino acids with identical database sequences can be uniquely identified across a range of database sizes
  - Divergence between soil and database sequences has a relatively small effect on identification of database sequences related to soil sequences

Additional bioinformatic advances (continued)

- *ProteinExpress* compares spectra to (pooled) DNA/RNA datasets, identifying metaproteomic peptides in protein coding sequences
  - DNA sequences from related organisms are binned
  - Identified proteins are similarity-searched against taxonomic bins to find which taxa likely express which proteins
  - Proteins are assigned to pathways and 141 functional groups that I defined, e.g., cellulases and amino acid transporters

Sampling at Toolik Field Station, Alaska

Toolik Field Station

Sampling at Toolik Field Station

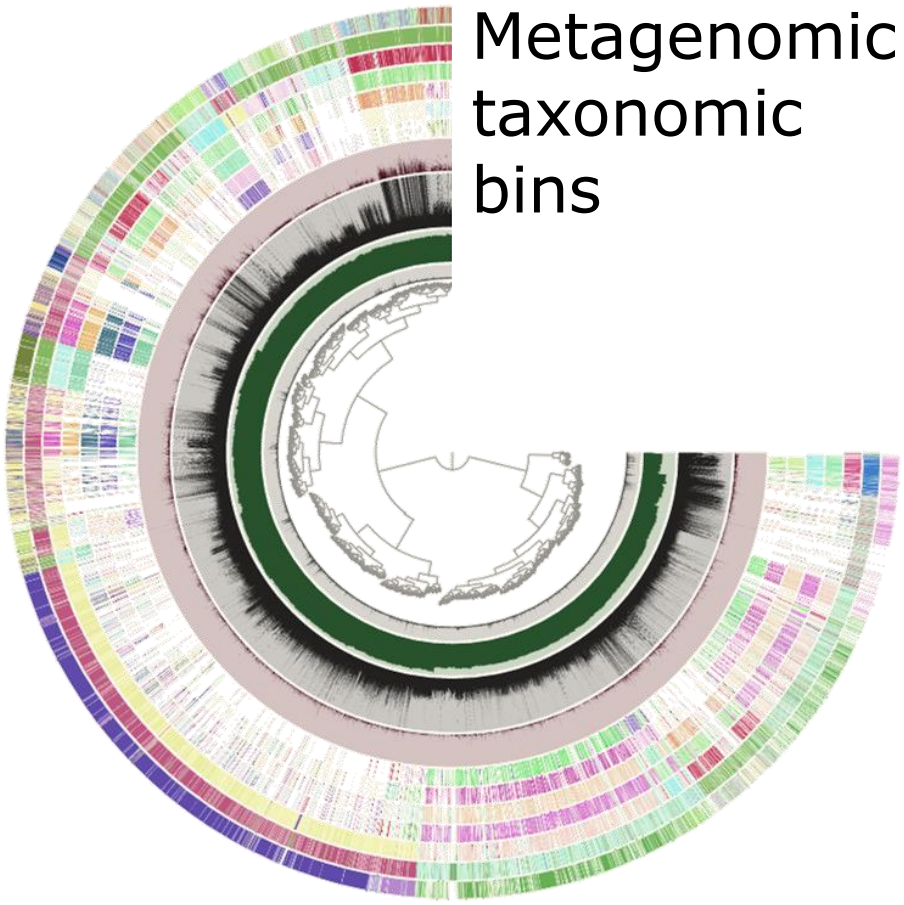
Sequence datasets

- 18 metaproteomes from Toolik and nearby Imnavait Valley
- Database search against 20 metagenomes and 8 metatranscriptomes from Imnavait Valley and Fairbanks area
- Relative abundances of proteins/functions calculated from spectral abundances

Myers-Smith, 2017, in *Medium*

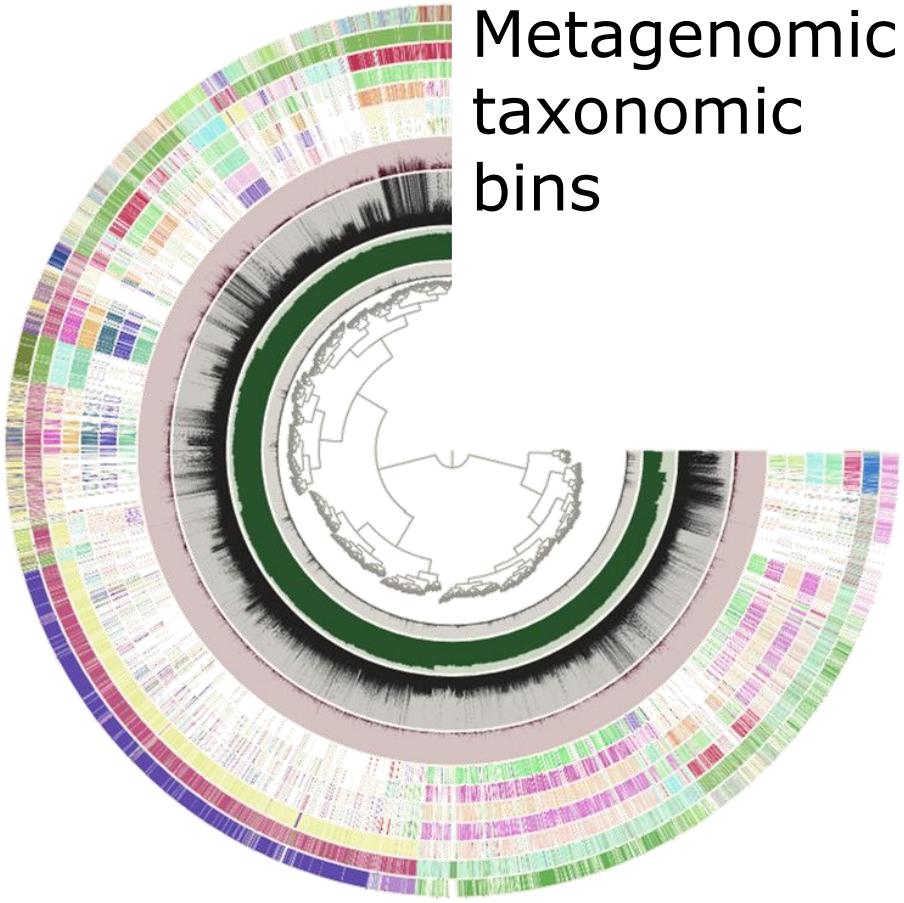
Walker,
Arctic Geobotanical Atlas

Taxonomic diversity of soil bacteria

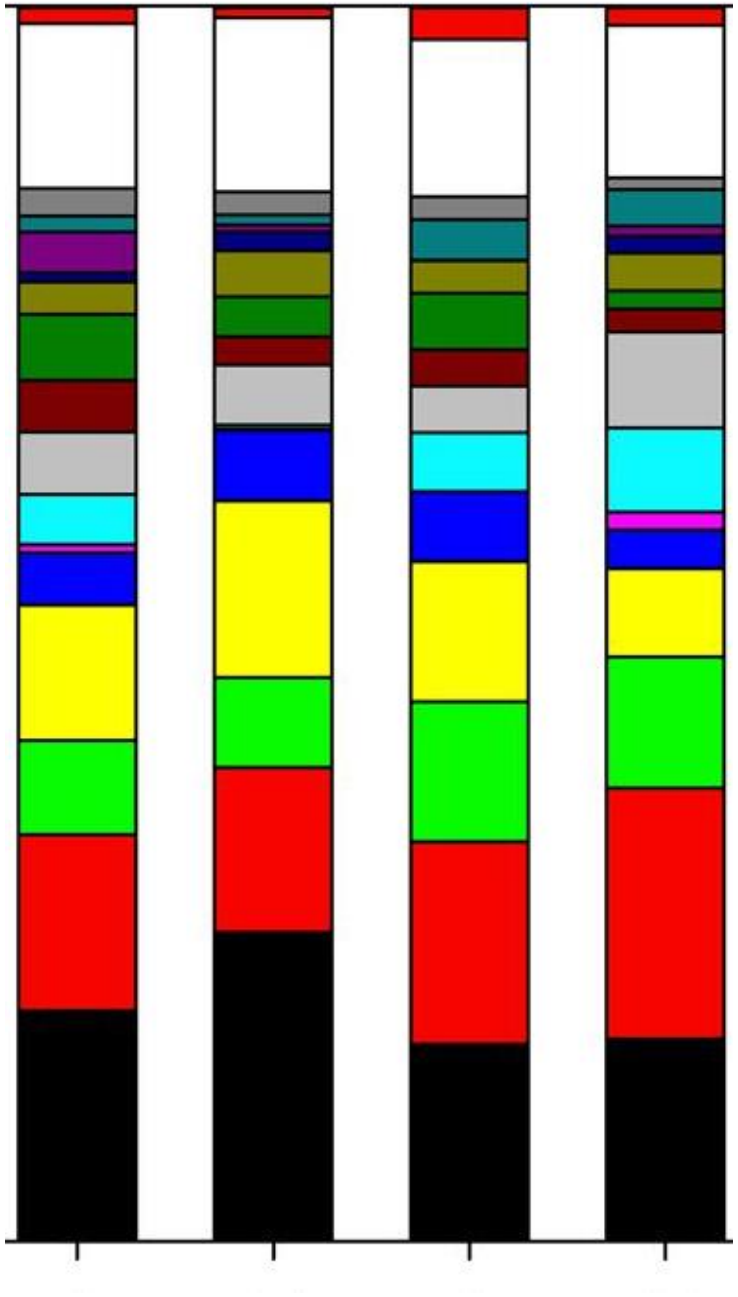


Metagenomic taxonomic bins

Taxonomic diversity of soil bacteria

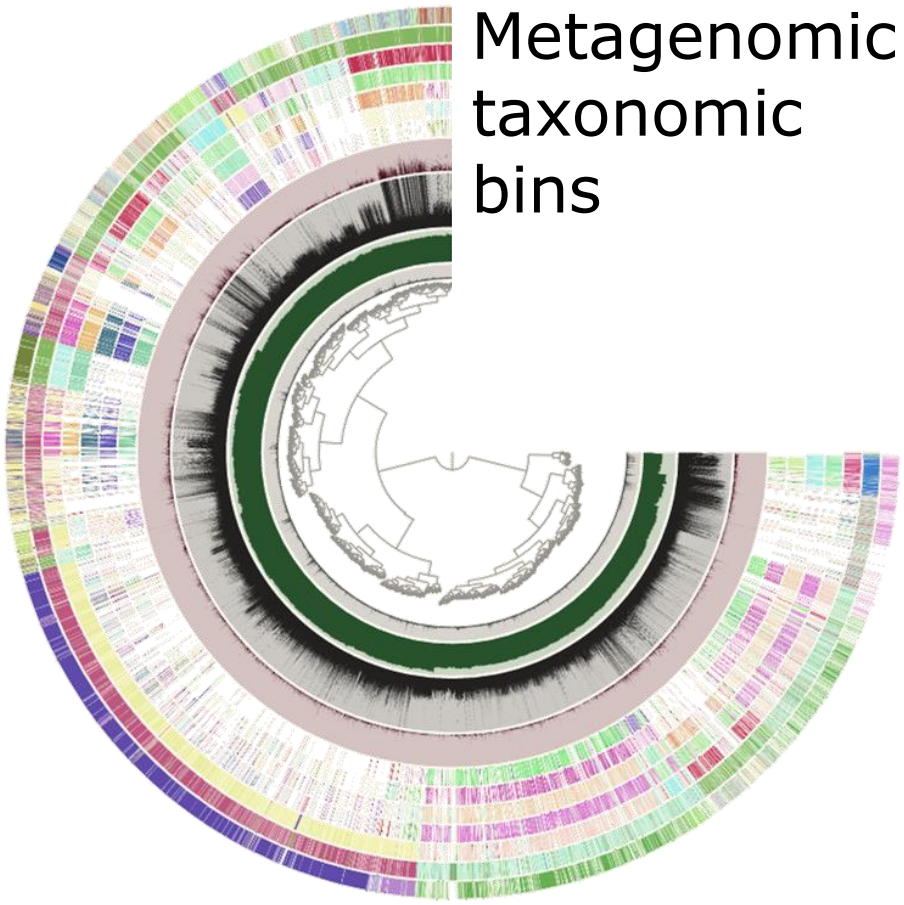


Metagenomic taxonomic bins

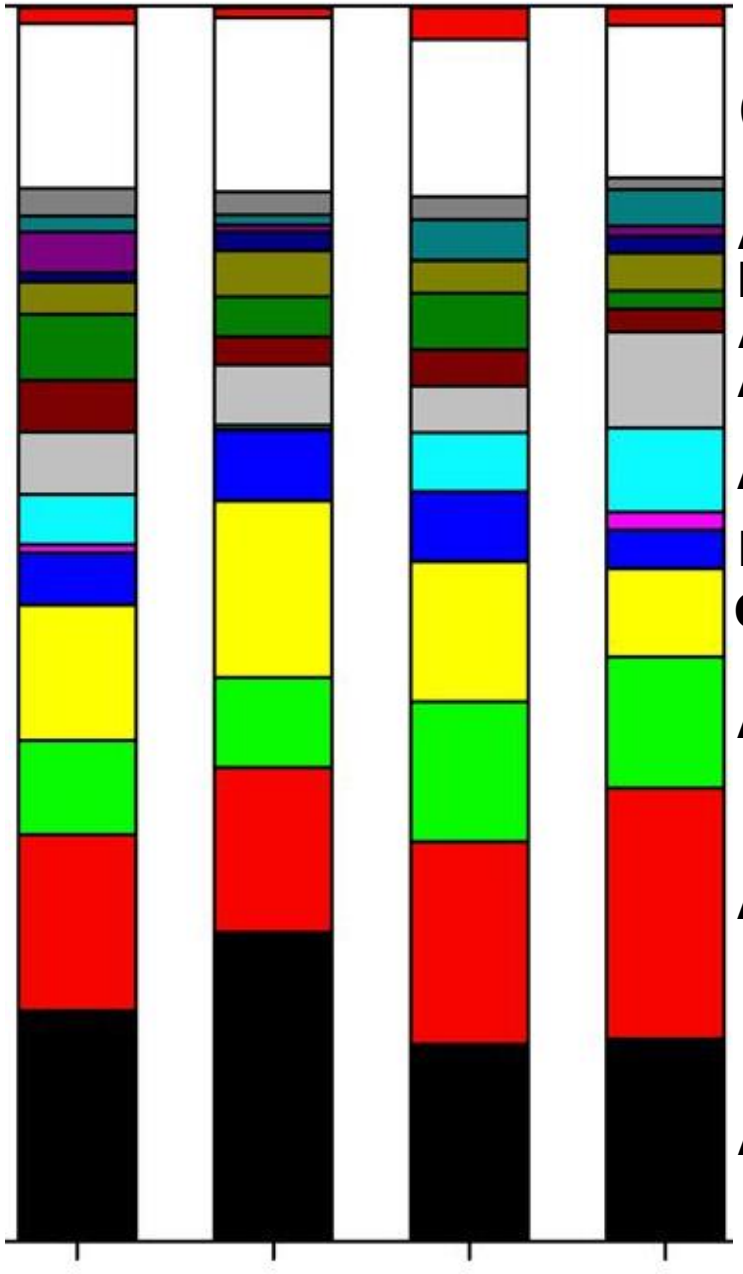


Koyama et al., 2014, *Front. Microbiol.*

Taxonomic diversity of soil bacteria



Metagenomic taxonomic bins



(Bacteroidetes)

Actinobacteria: Acidimicrobiia
Deltaproteobacteria
Acidobacteria: Solibacteres
Actinobacteria: Thermoleophilia

Acidobacteria

Betaproteobacteria
Gammaproteobacteria

Acidobacteria

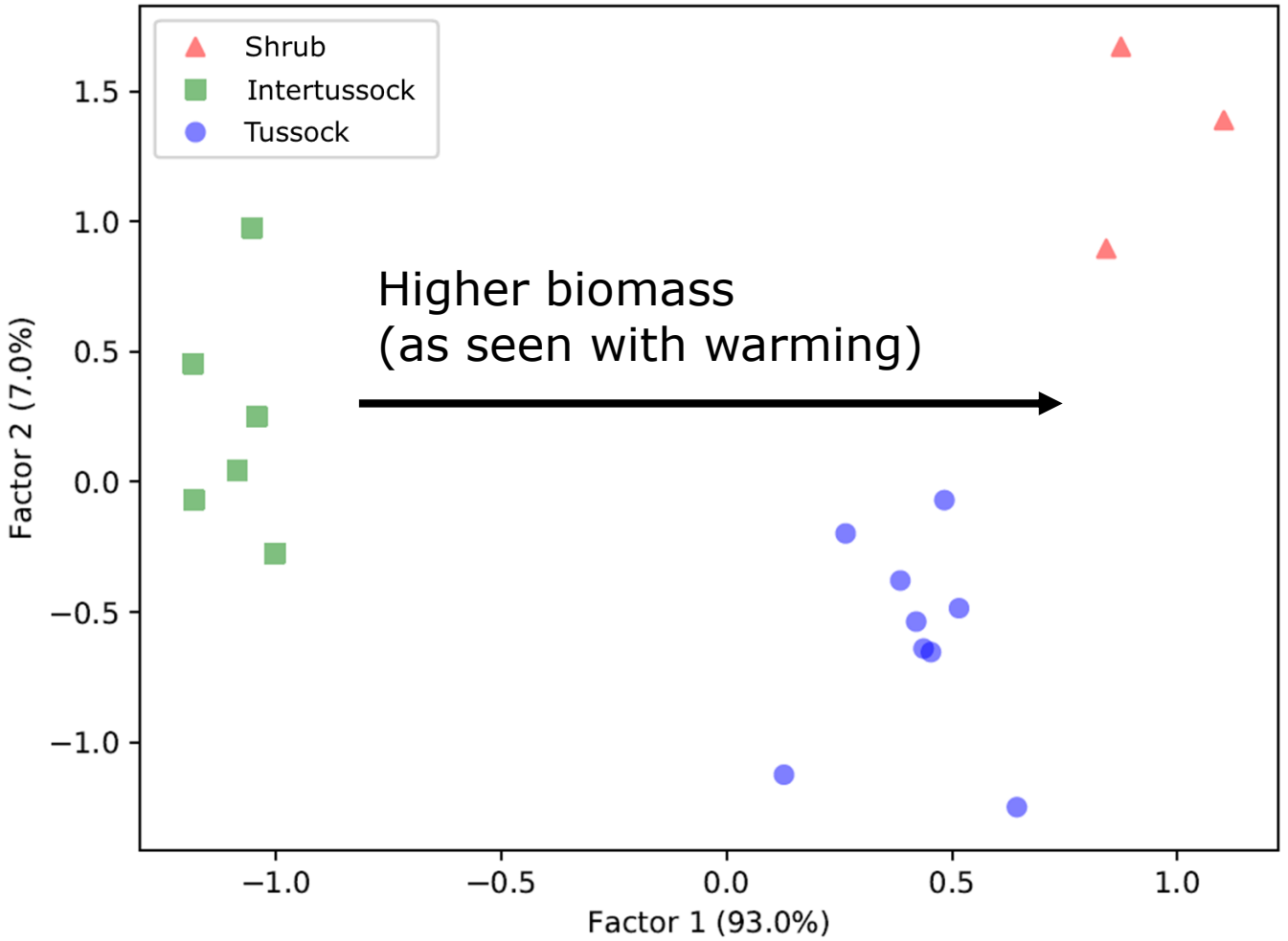
Actinobacteria

Alphaproteobacteria

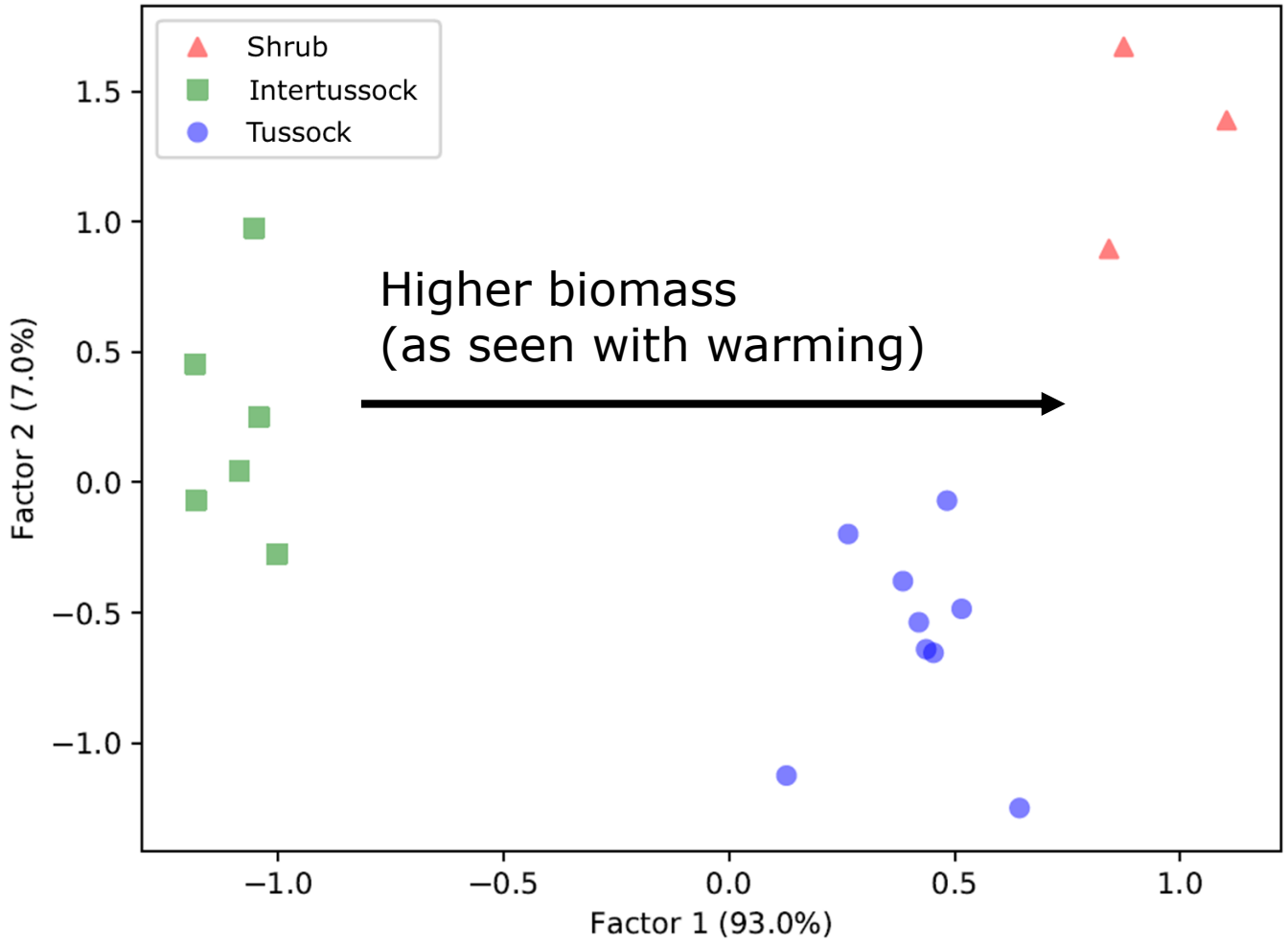
Koyama et al., 2014, *Front. Microbiol.*

Protein expression profile changes with flora

- Based only on the relative abundances of proteins in the 18 soil metaproteomic samples, the samples sort into low and high biomass floras



Protein expression profile changes with flora

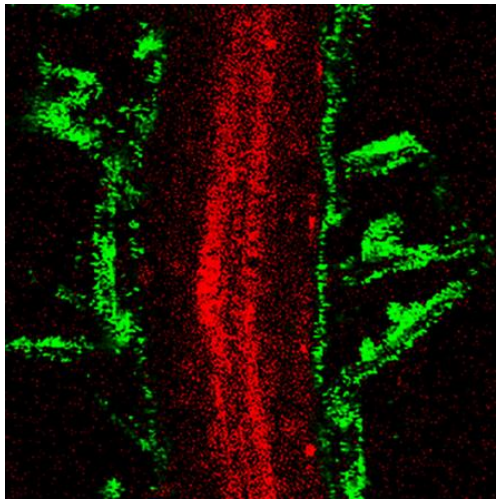


Functions associated with Factor 1 include
sugar transporters and
succinoglycan EPS (slime) production

Protein expression profile changes with flora

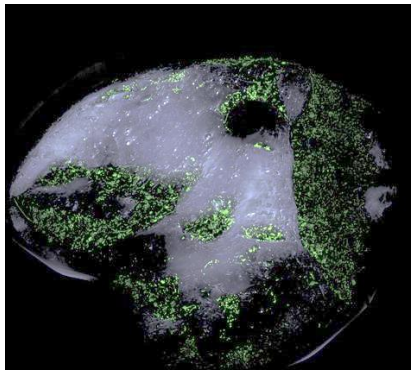
- Hypothesis: Greater plant biomass increases importance of microbe/ plant interactions in biogeochemical cycling

Root

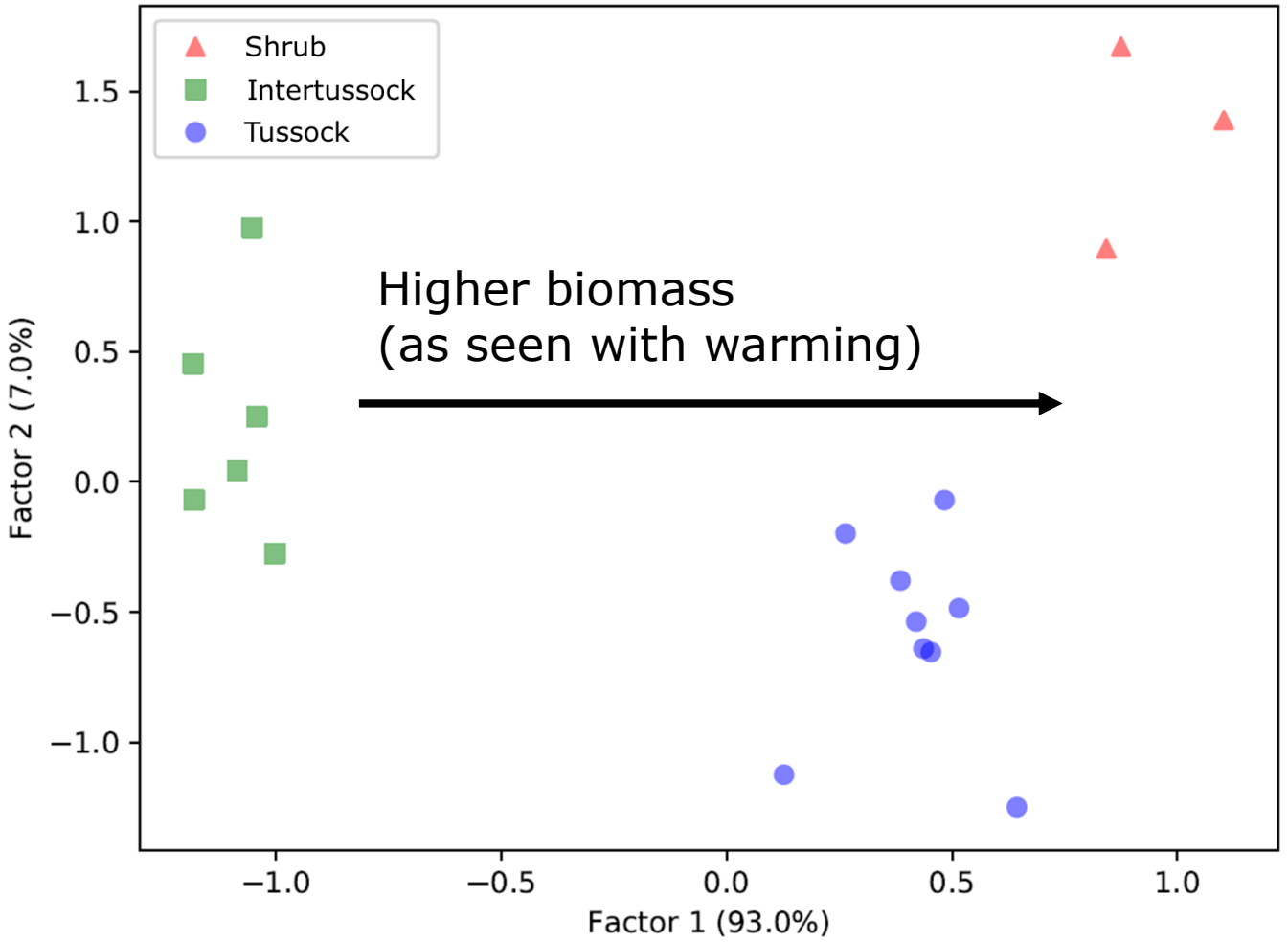


Zhao et al., 2018, *Front. Microbiol.*

Sand grain



Probandt et al., 2018, *ISME J.*



Functions associated with Factor 1 include sugar transporters and succinoglycan EPS (slime) production

Taxonomic protein expression profiles

Taxonomic “fidelity” =
Overall abundance of function *
Relatedness of function to taxon (similarity score)

Fidelity is used as a measure of protein expression by different taxa

Taxonomic protein expression profiles

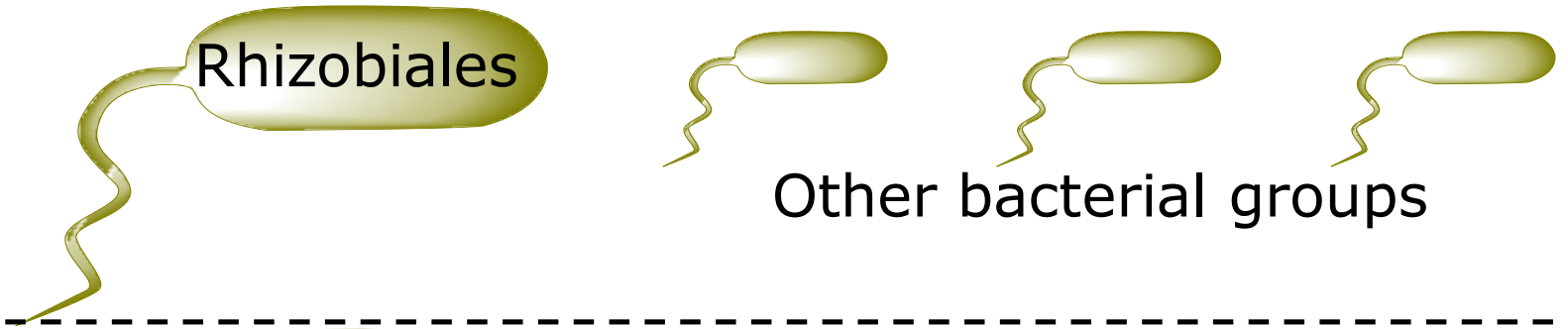
Taxonomic "fidelity" =
Overall abundance of function *
Relatedness of function to taxon (similarity score)

Fidelity is used as a measure of protein expression by different taxa

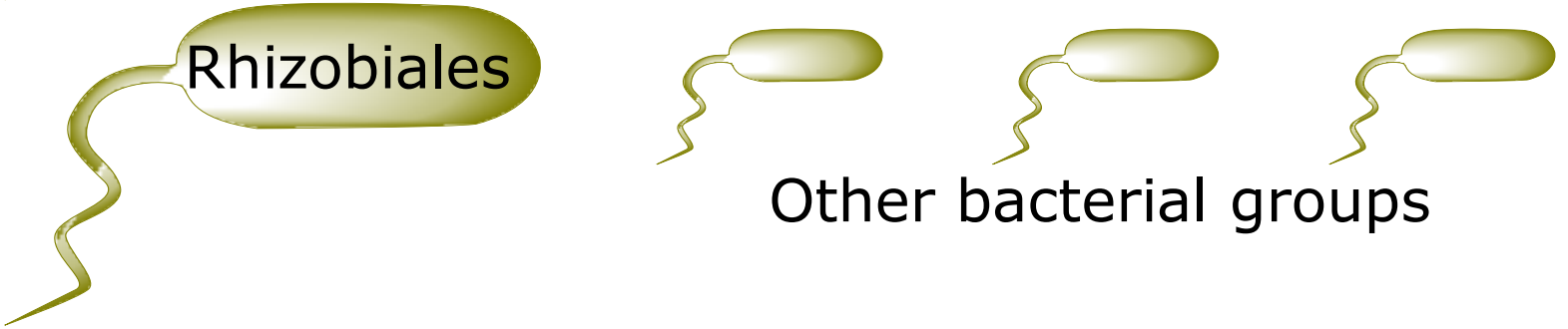
When fidelity scores for a function are **normalized across taxa**, they indicate the relative expression of each function by different taxa

Sugar transporters

Low plant biomass



High plant biomass



Taxonomic protein expression profiles

Taxonomic "fidelity" =
Overall abundance of function *
Relatedness of function to taxon (similarity score)

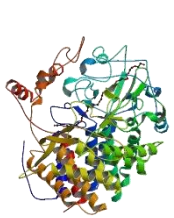
Fidelity is used as a measure of protein expression by different taxa

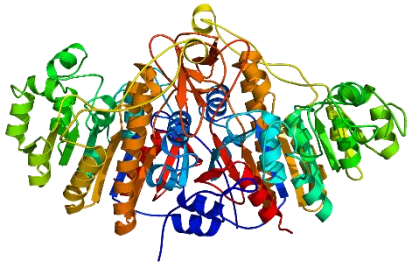
When fidelity scores for a taxon are **normalized across functions**, they indicate the relative expression by each taxon of different functions

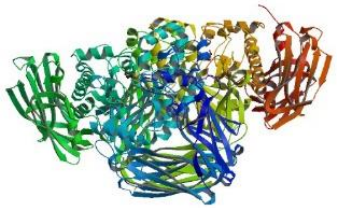
Rhizobiales

Low plant biomass

Sugar transporters

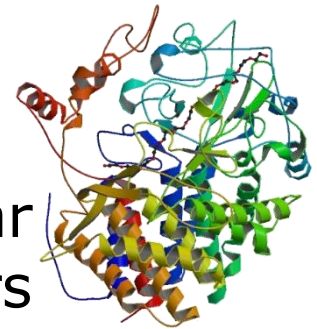


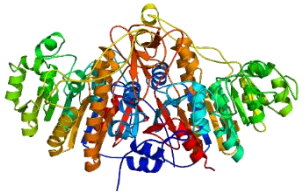


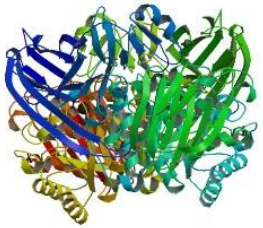


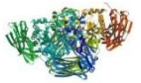
High plant biomass

Sugar transporters

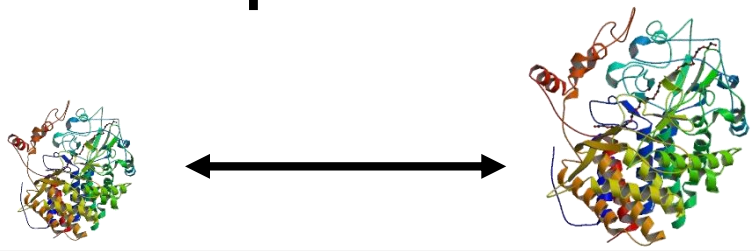


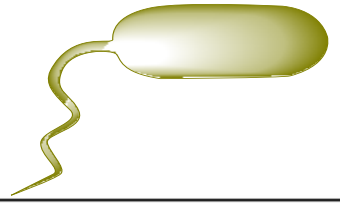


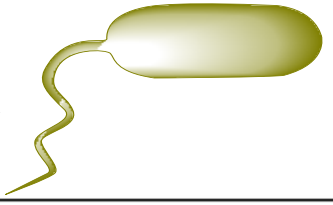


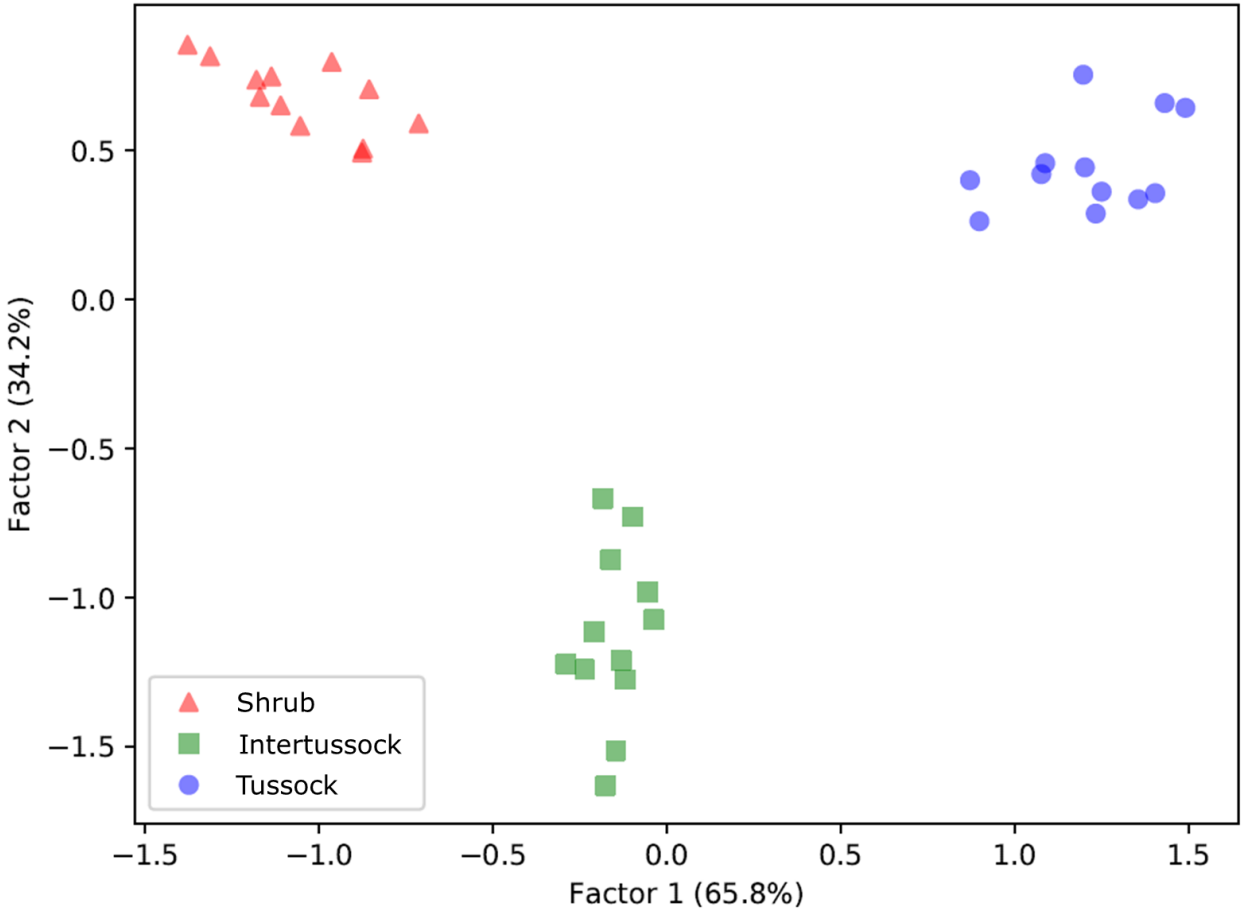


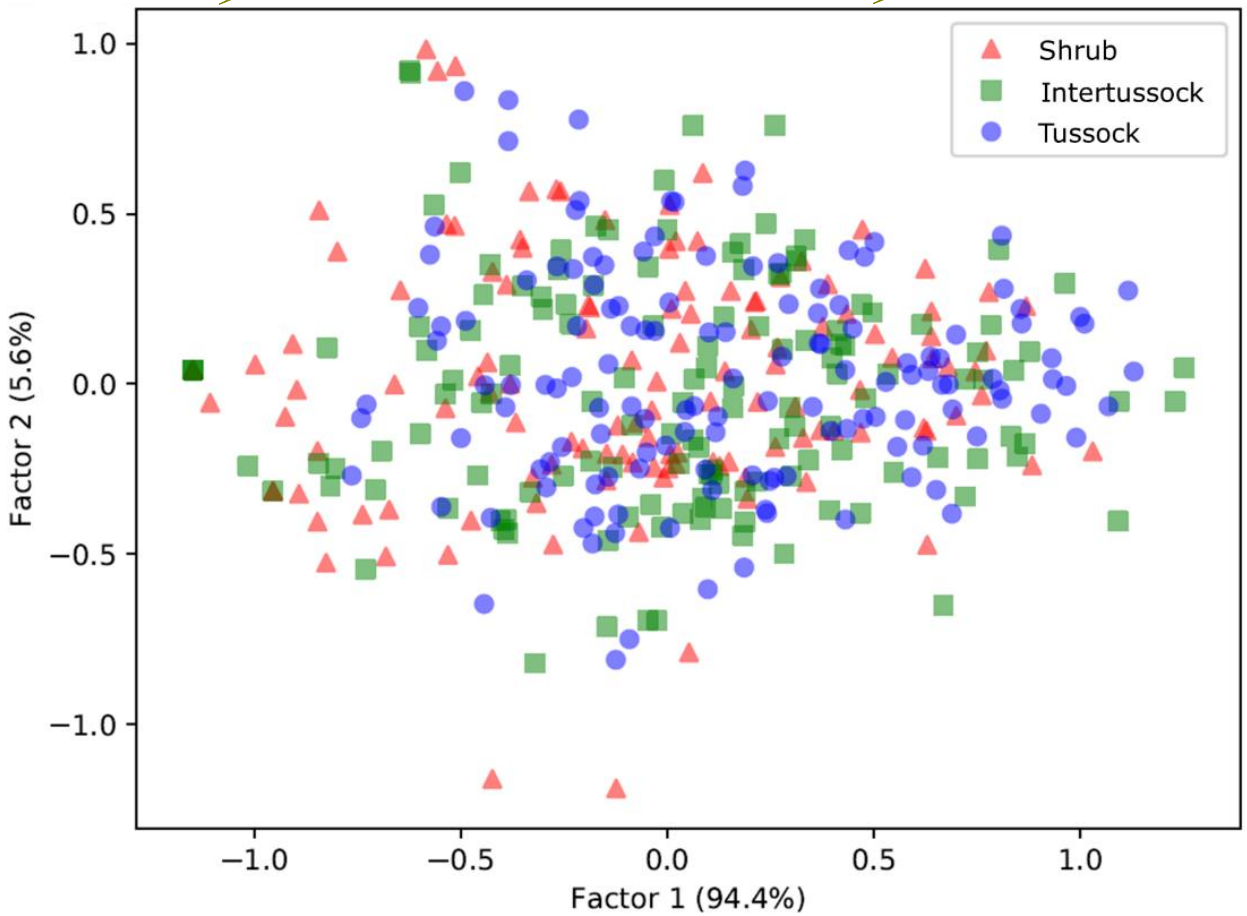
Flexible expression of functions by the same taxa





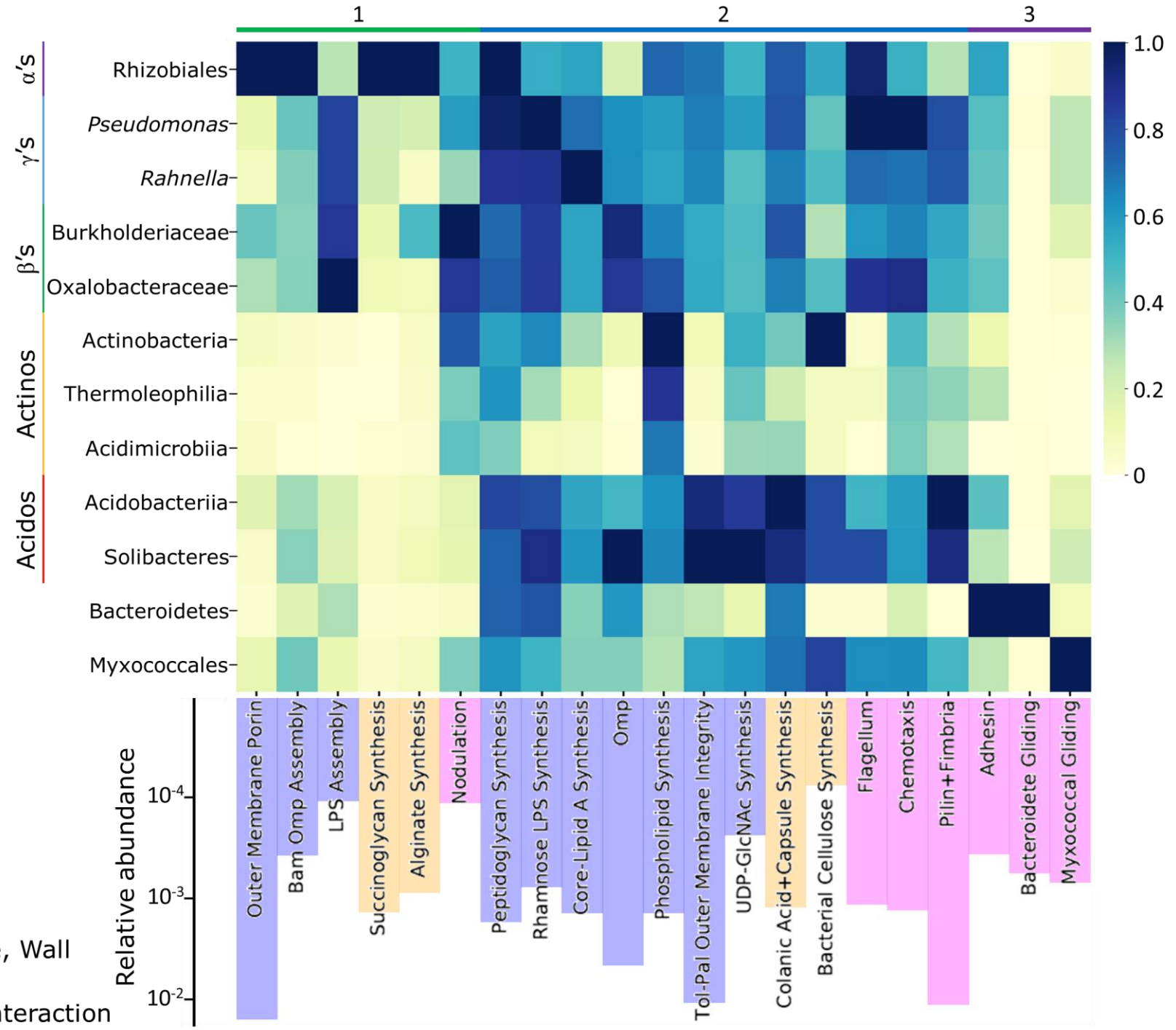






Heatmap:
Taxonomic fidelity

Bars:
Overall abundance

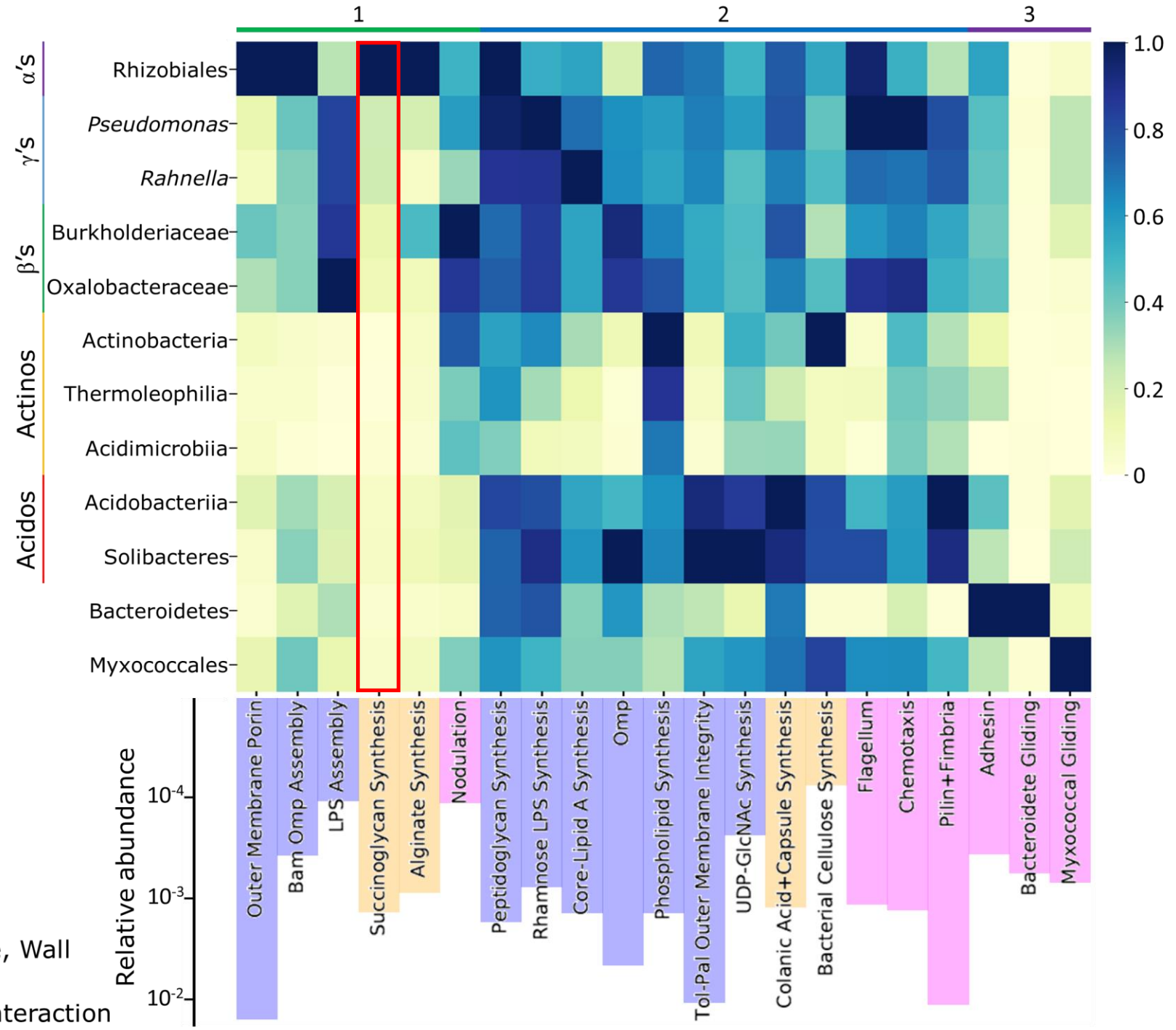


Cell Envelope

Membrane, Wall
EPS
Motility, Interaction

Heatmap:
Taxonomic fidelity

Bars:
Overall abundance

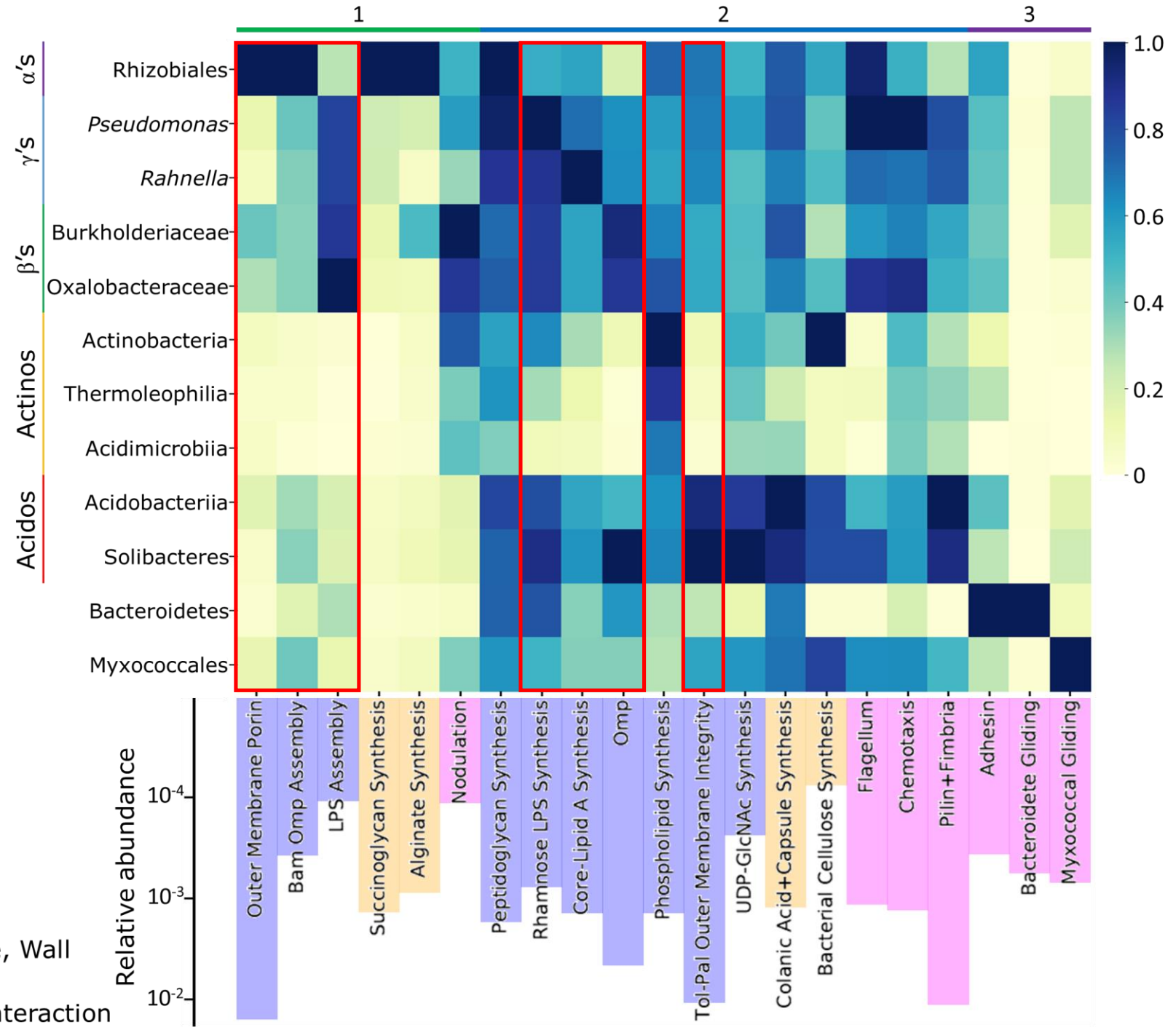


Cell Envelope

Membrane, Wall
EPS
Motility, Interaction

Validation of similarity score as a measure of taxonomic relatedness for calculation of fidelity

Outer membrane-related proteins have very low scores in Gram-positives



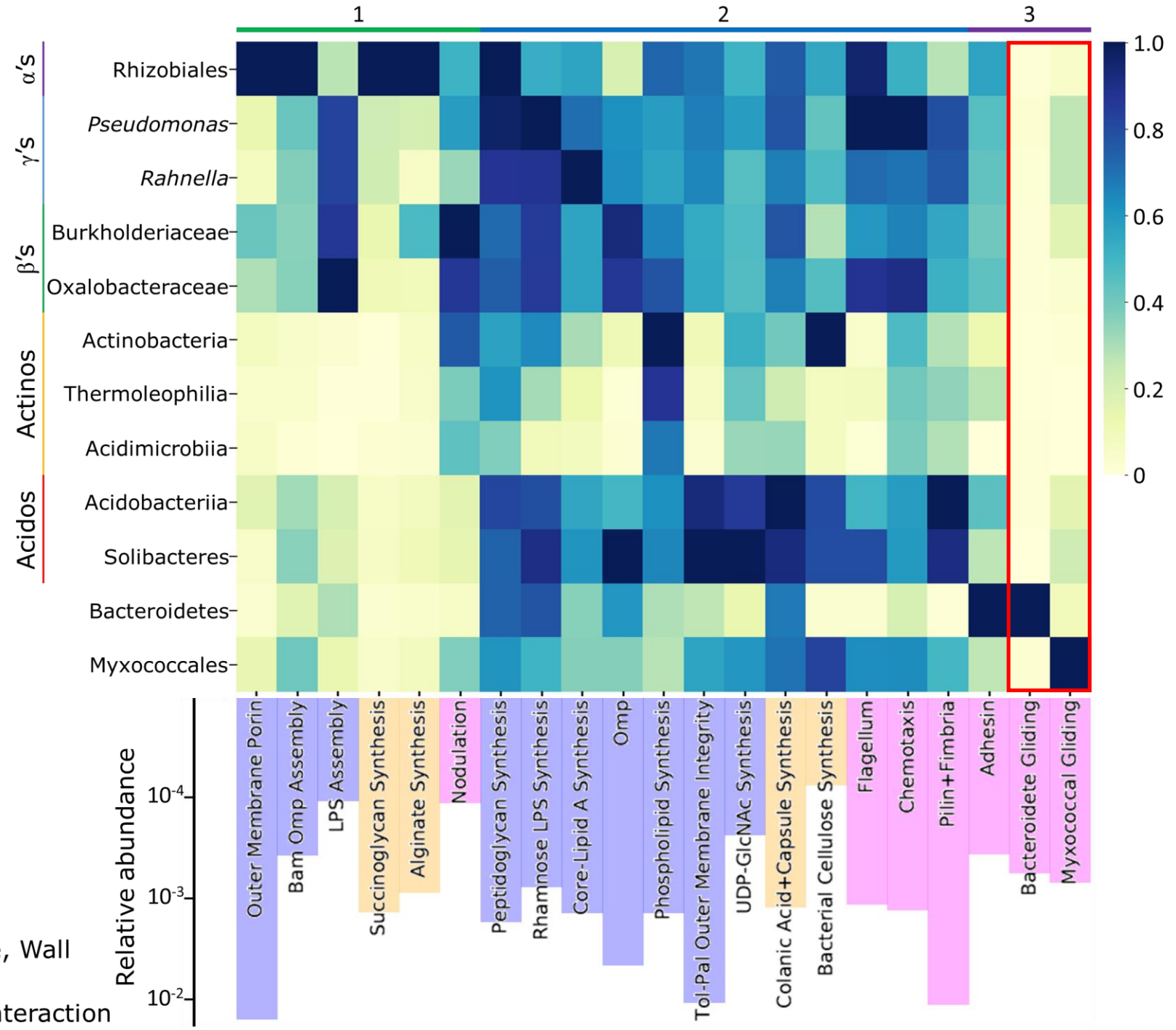
Cell Envelope

Membrane, Wall
EPS
Motility, Interaction

Validation of similarity score as a measure of taxonomic relatedness for calculation of fidelity

Outer membrane-related proteins have very low scores in Gram-positives

Bacteroidete and myxococcal gliding proteins have very high scores in Bacteroidetes and Myxococcales



Cell Envelope

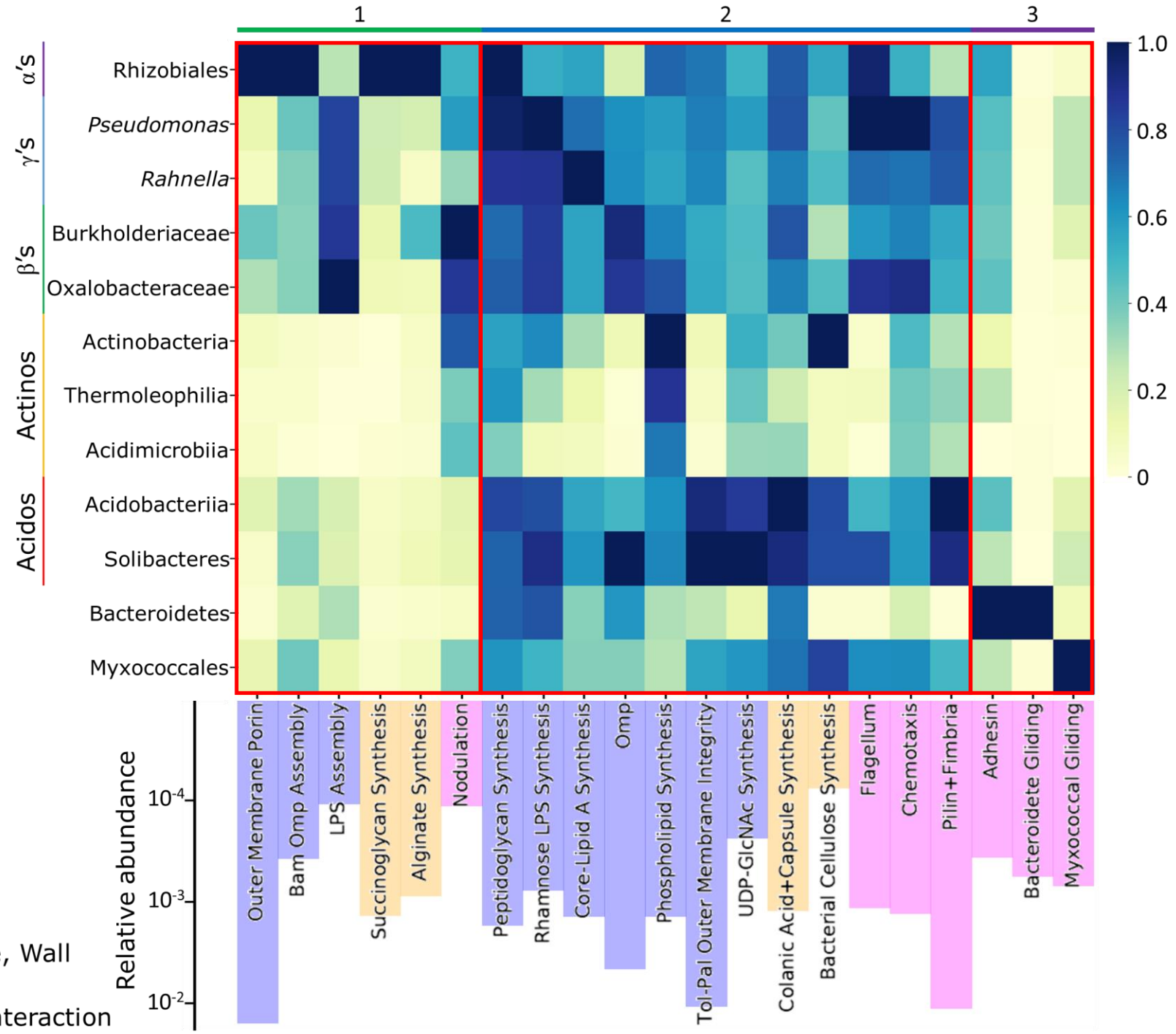
Membrane, Wall
EPS
Motility, Interaction

Three functional clusters

Cluster 1:
High Proteobacteria,
Low Non-Proteobacteria

Cluster 2:
Relatively even
Proteobacteria and
Non-Proteobacteria

Cluster 3:
Low Proteobacteria,
High Non-Proteobacteria



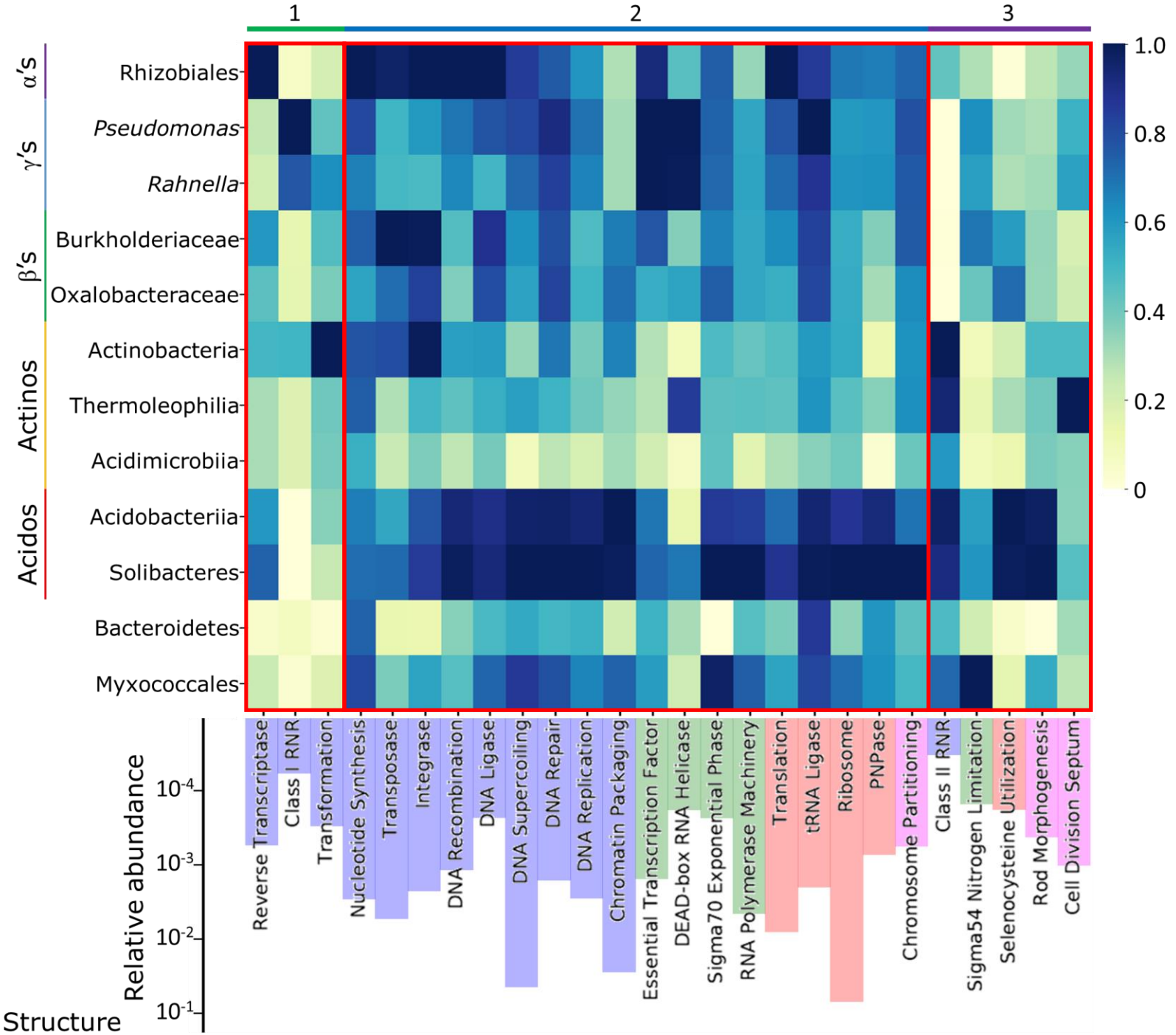
Cell Envelope

Three functional clusters

Cluster 1:
High Proteobacteria,
Low Non-Proteobacteria

Cluster 2:
Relatively even
Proteobacteria and
Non-Proteobacteria

Cluster 3:
Low Proteobacteria,
High Non-Proteobacteria



Core Cellular Functions

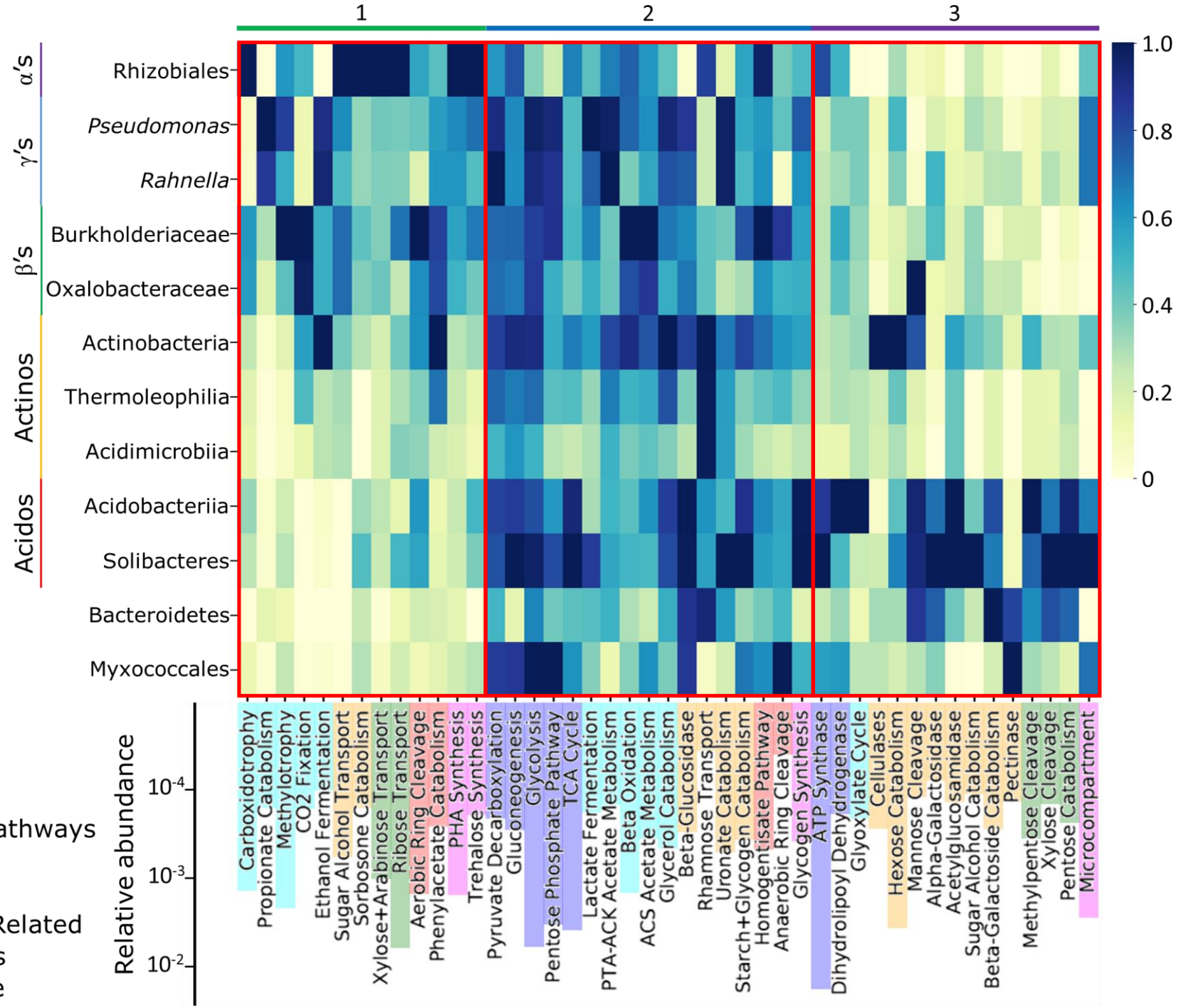
DNA
RNA
Protein
Division, Structure

Three functional clusters

Cluster 1:
High Proteobacteria,
Low Non-Proteobacteria

Cluster 2:
Relatively even
Proteobacteria and
Non-Proteobacteria

Cluster 3:
Low Proteobacteria,
High Non-Proteobacteria

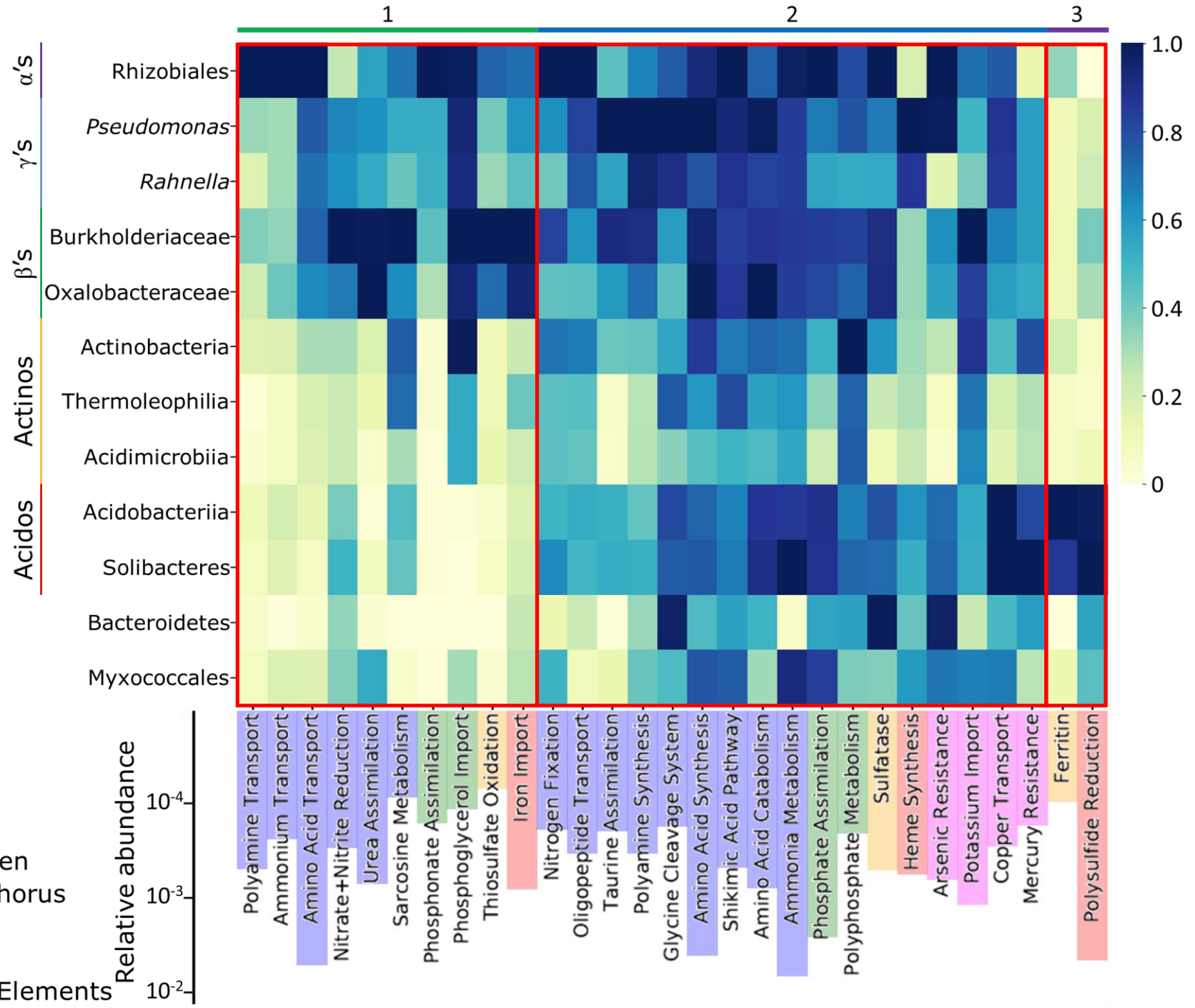


Three functional clusters

Cluster 1:
High Proteobacteria,
Low Non-Proteobacteria

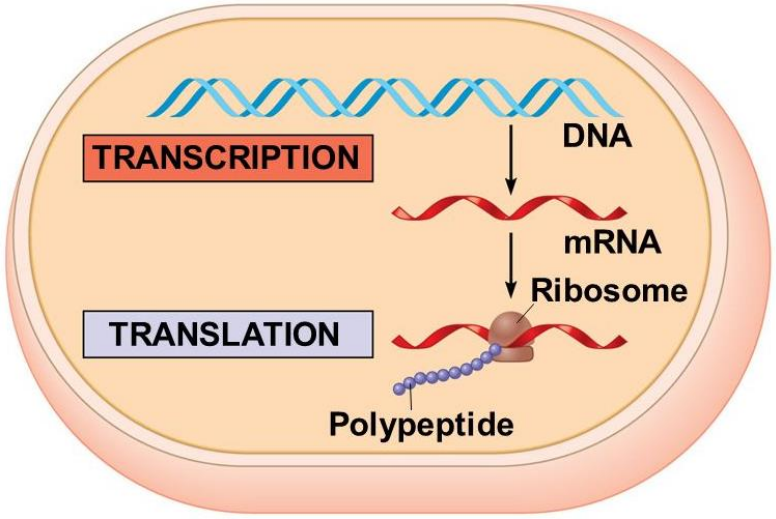
Cluster 2:
Relatively even
Proteobacteria and
Non-Proteobacteria

Cluster 3:
Low Proteobacteria,
High Non-Proteobacteria

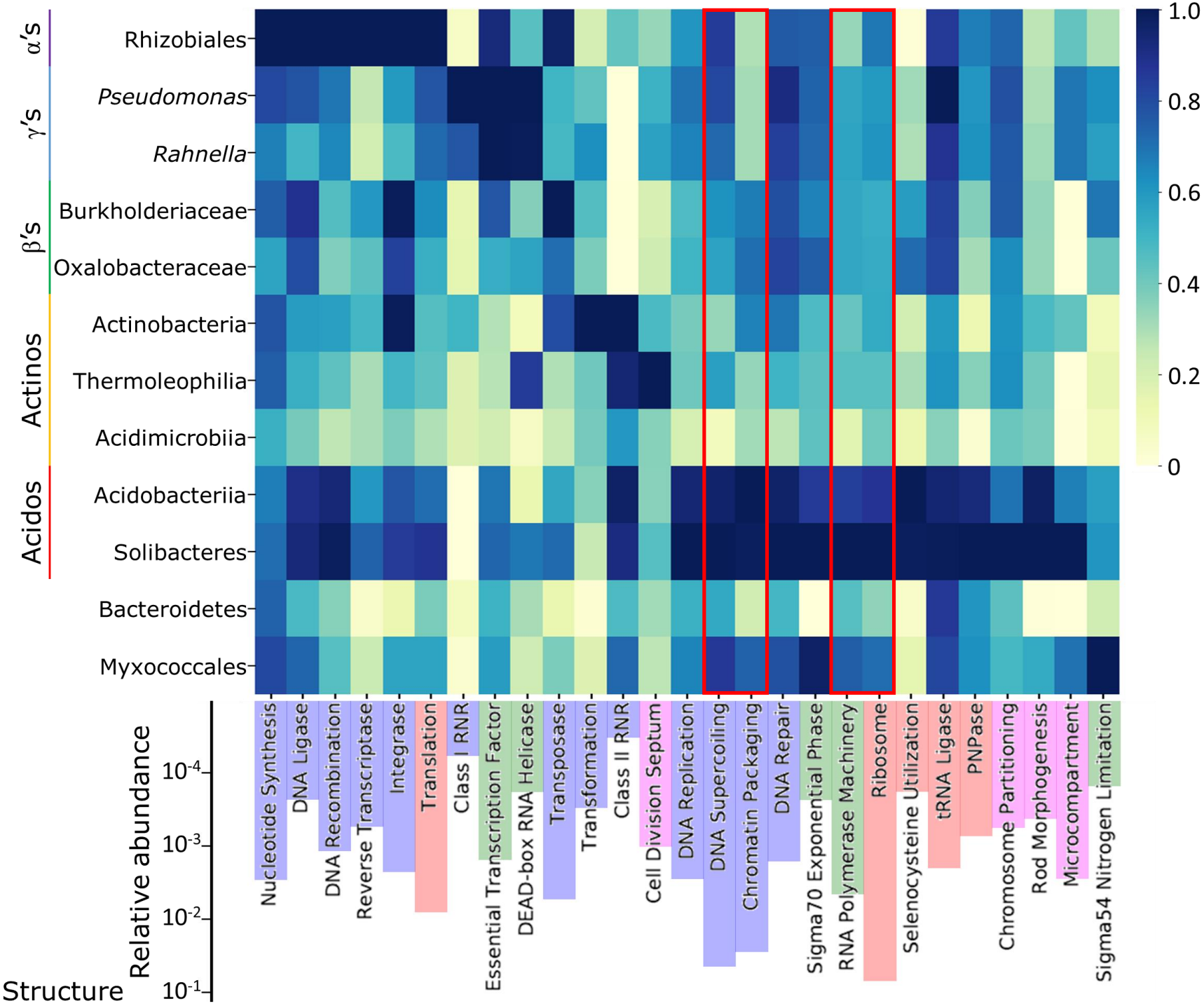


Nutrient Metabolism

Acidobacteria are the most active group, contrary to expectation from DNA abundance



Pearson Education

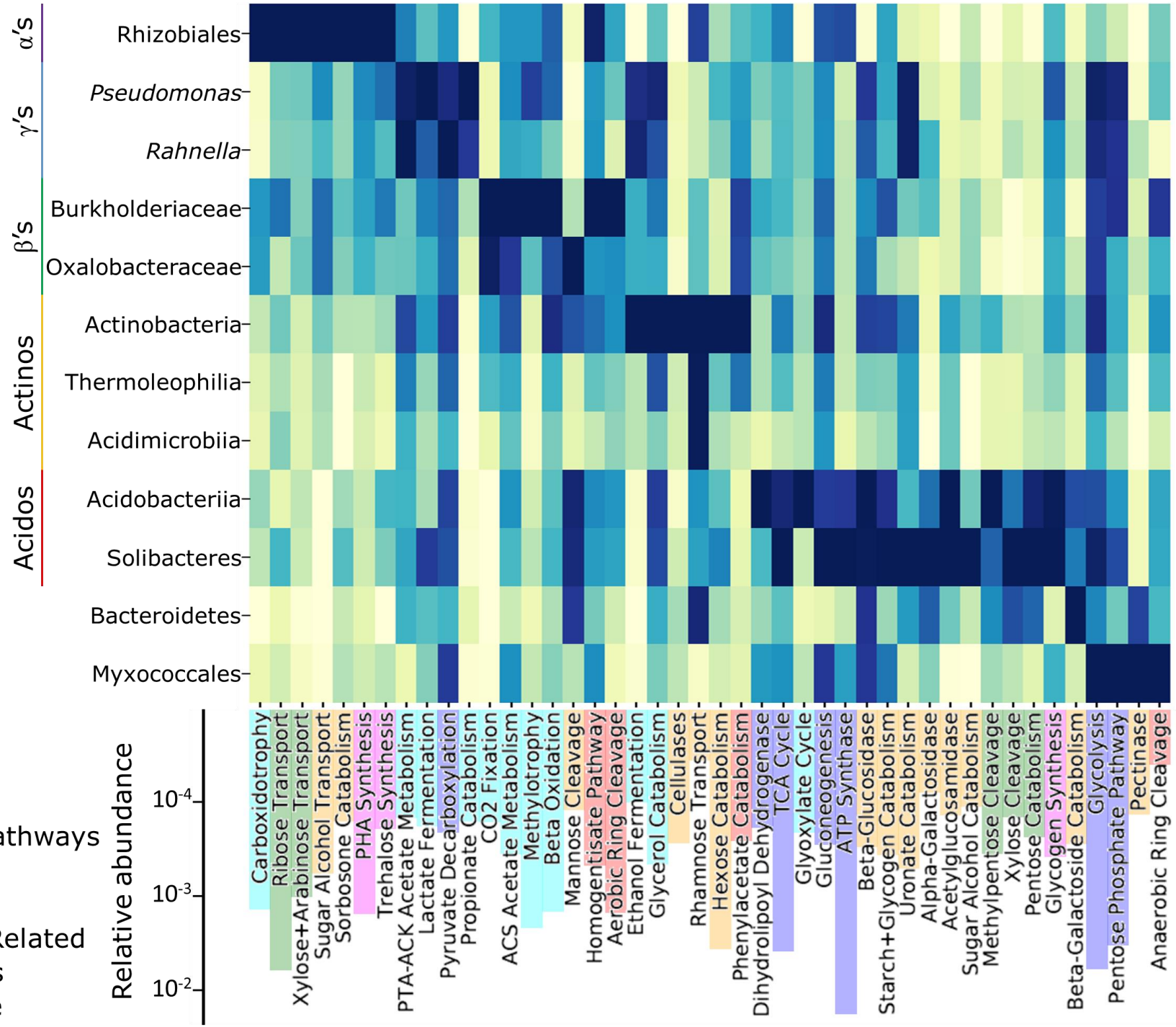


Core Cellular Functions

DNA
RNA
Protein
Division, Structure

"[T]here are numerous individual processes and taxa associated with the metabolism of the thousands of organic compounds found in soil. This complexity makes it very difficult to predict soil function..."

Hypothesis: Pathways of organic matter decomposition are expressed by taxa in proportion to overall activity. (No resource partitioning)

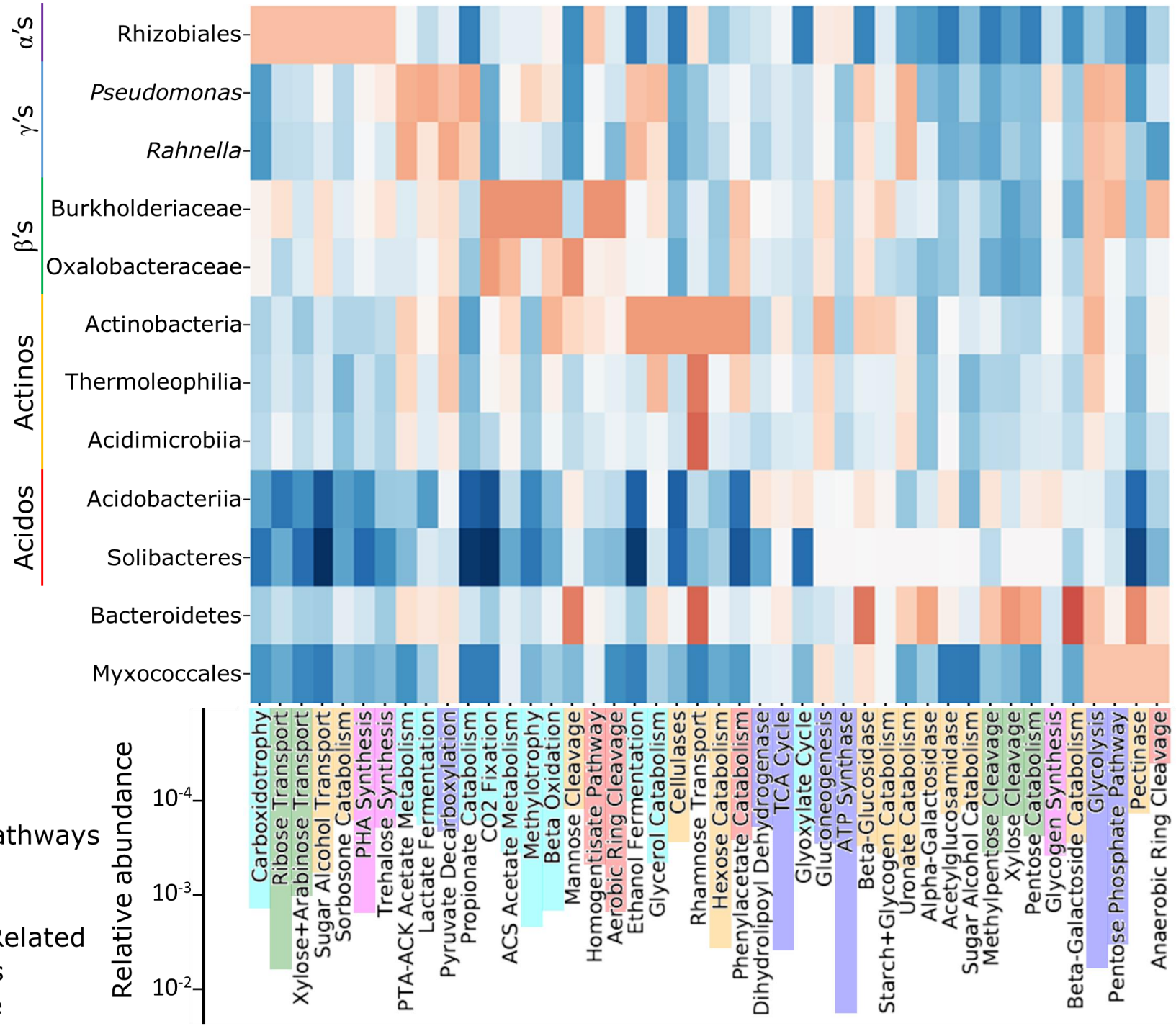


Carbon Metabolism

- Central Pathways
- ≤3-C
- 5-C
- 6-C and Related
- Aromatics
- C Storage

"[T]here are numerous individual processes and taxa associated with the metabolism of the thousands of organic compounds found in soil. This complexity makes it very difficult to predict soil function..."

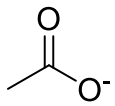
Hypothesis: Pathways of organic matter decomposition are expressed by taxa in proportion to overall activity. (No resource partitioning) **FALSE**



Carbon Metabolism

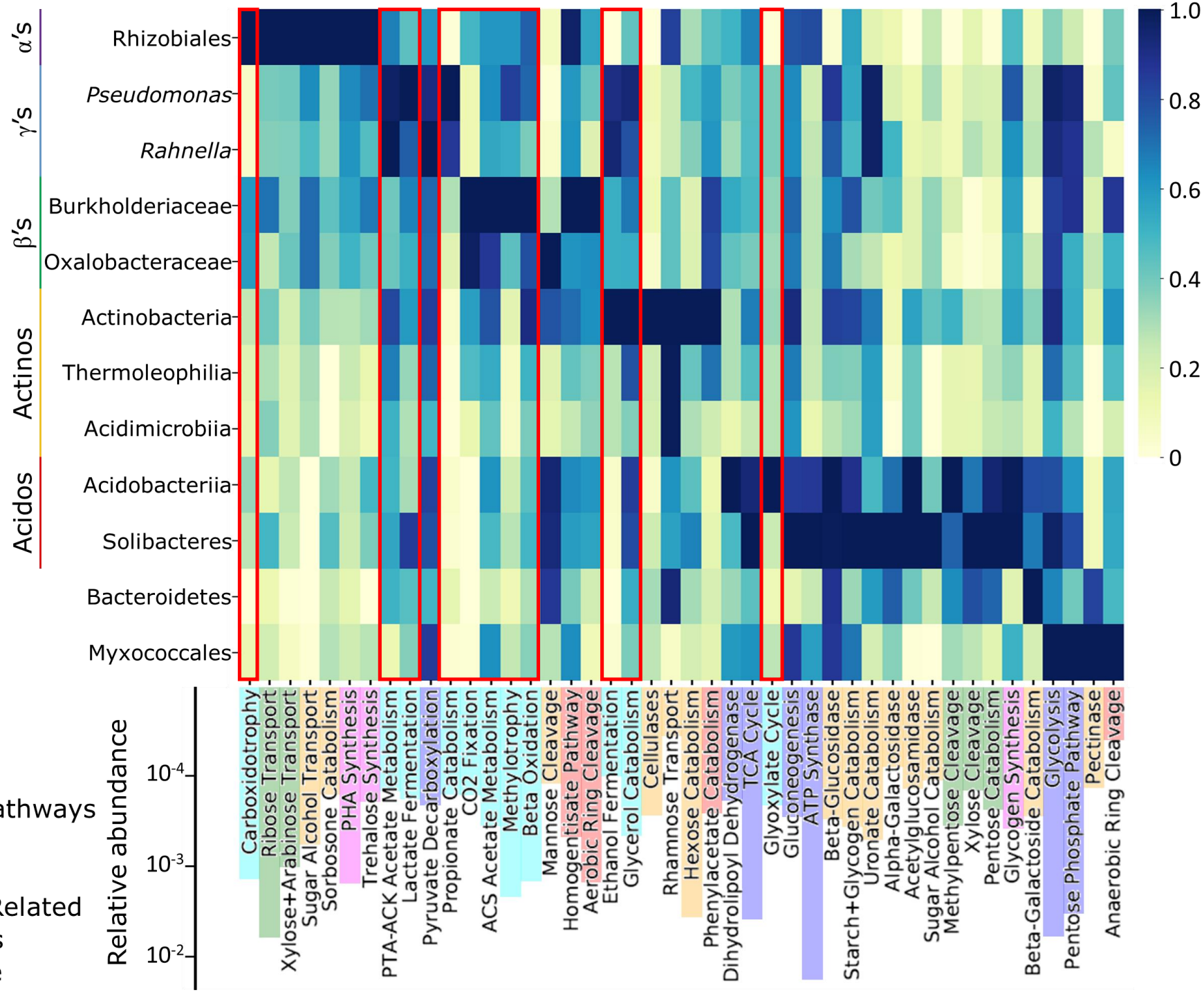
- Central Pathways
- ≤3-C
- 5-C
- 6-C and Related
- Aromatics
- C Storage

Proteobacteria dominate the metabolism of small compounds



H_3C-OH

$C\equiv O$

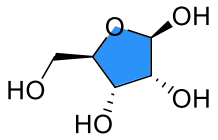


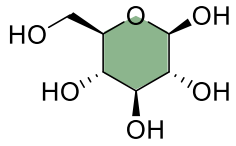
Carbon Metabolism

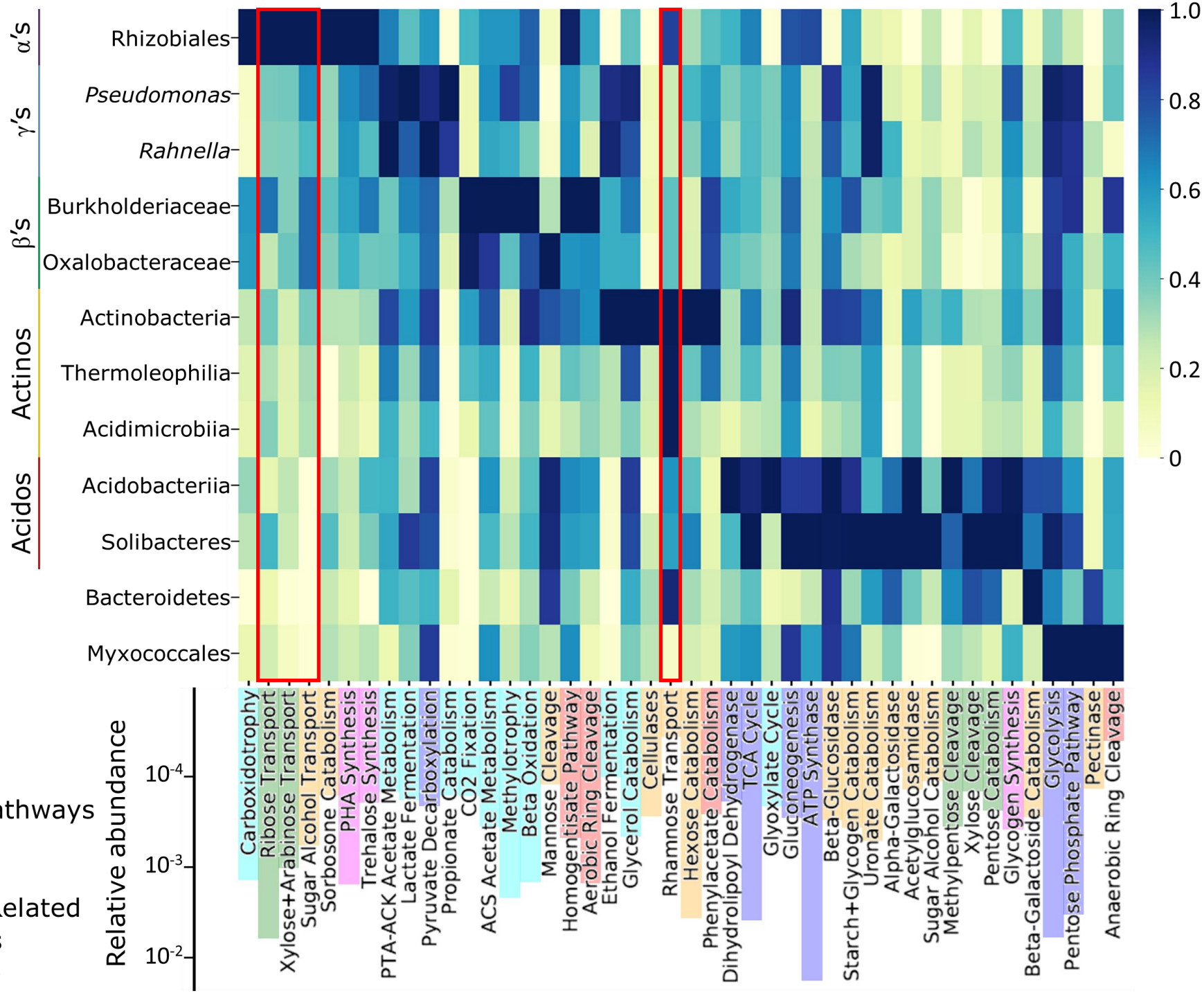
- Central Pathways
- ≤3-C
- 5-C
- 6-C and Related
- Aromatics
- C Storage

Proteobacteria dominate the metabolism of small compounds

Proteobacteria, especially Rhizobiales, dominate the expression of simple sugar transporters



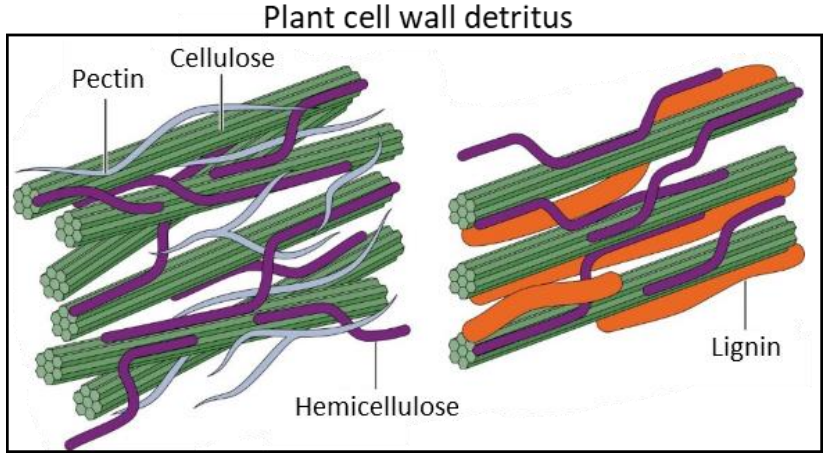


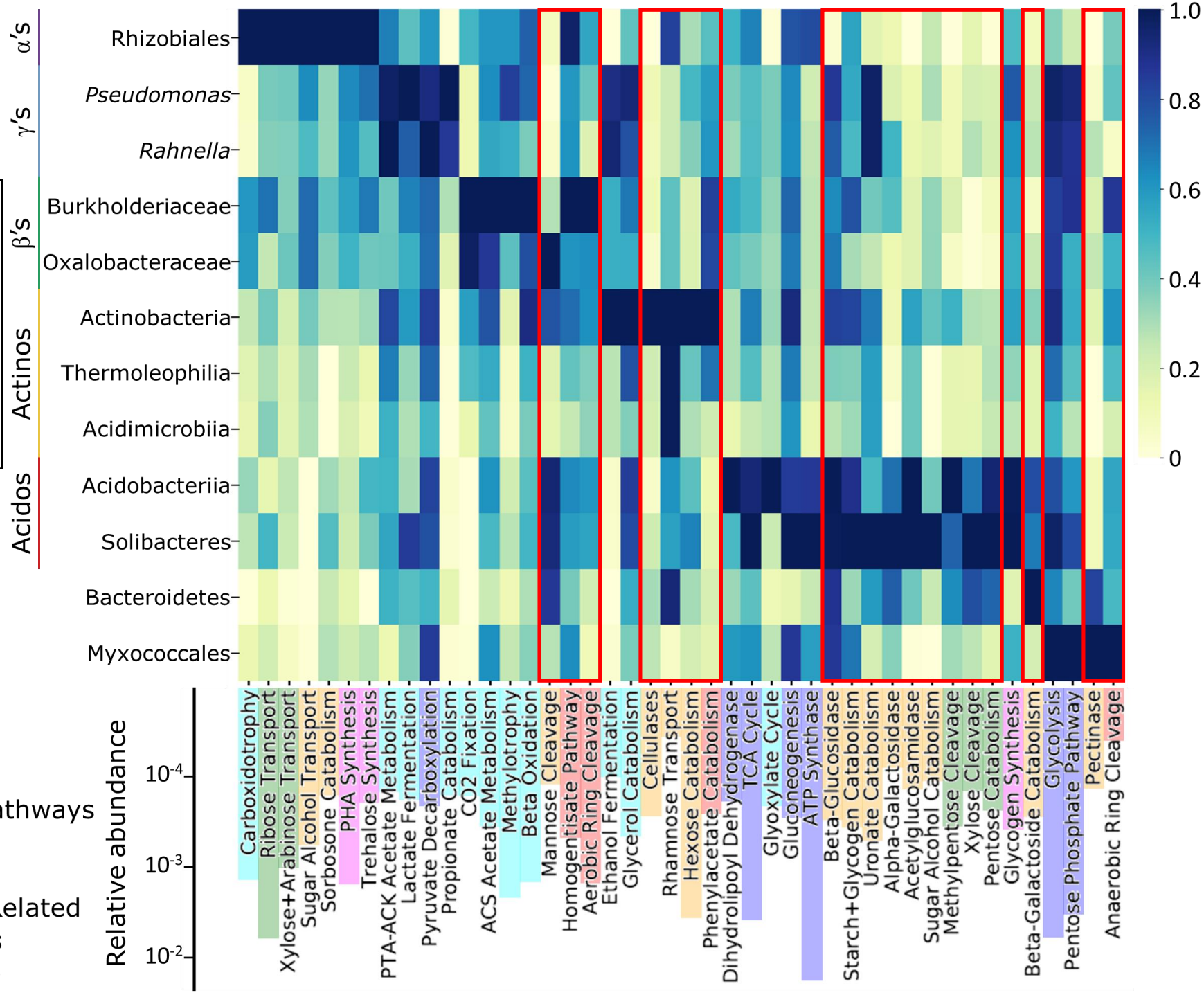


Carbon Metabolism

- Central Pathways
- ≤3-C
- 5-C
- 6-C and Related
- Aromatics
- C Storage

Non-Proteobacteria dominate polymer degradation



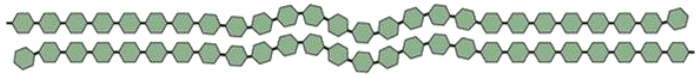


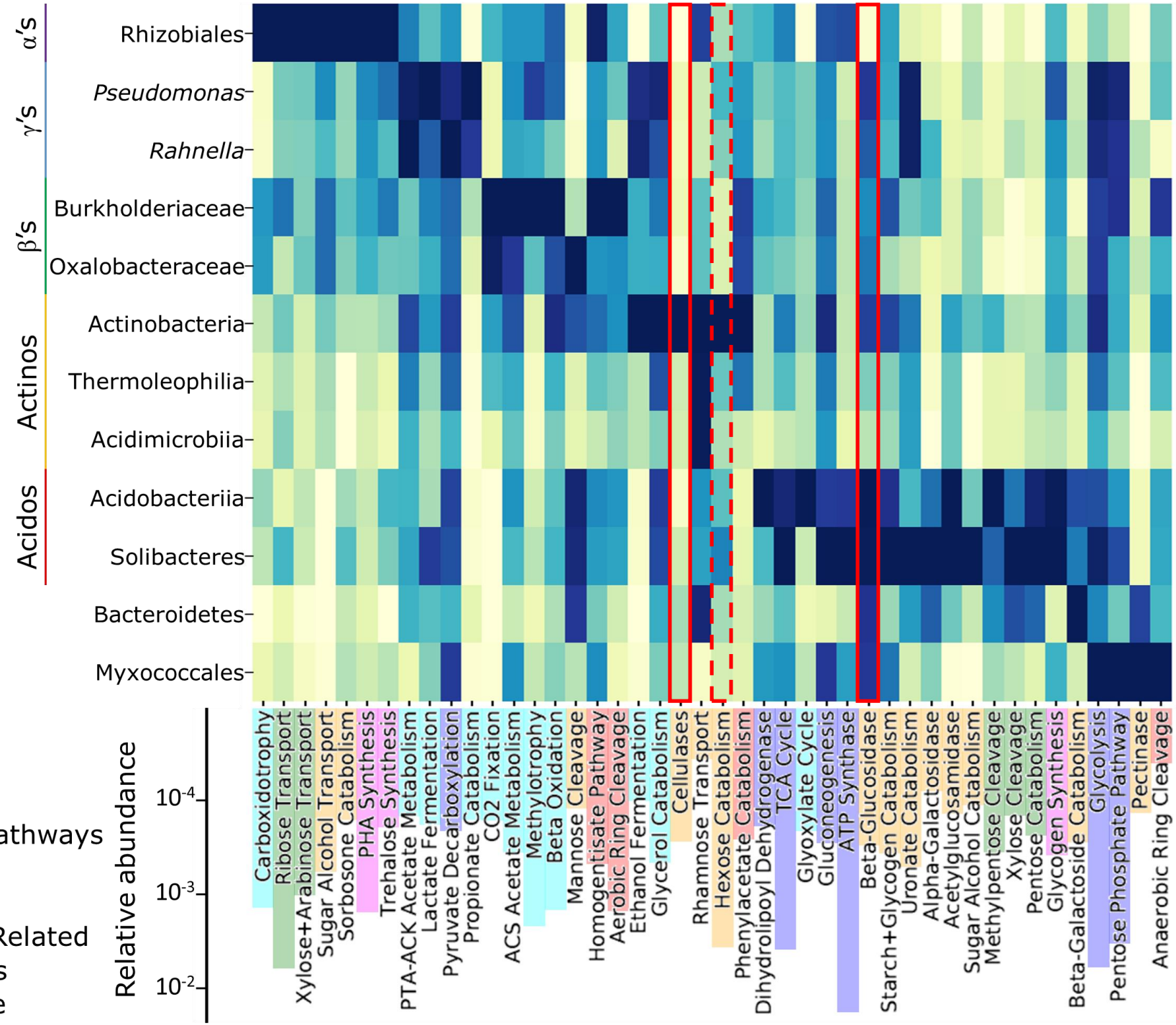
Carbon Metabolism

- Central Pathways
- ≤3-C
- 5-C
- 6-C and Related
- Aromatics
- C Storage

Non-Proteobacteria dominate polymer degradation

Actinobacteria dominate cellulose degradation





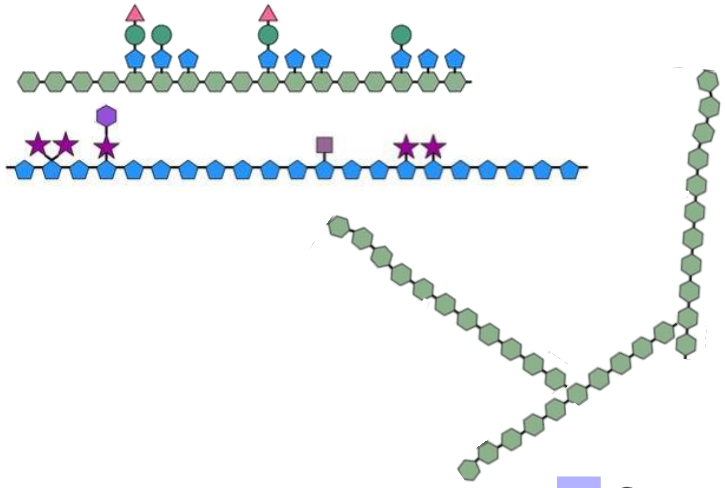
Carbon Metabolism

- Central Pathways
- ≤3-C
- 5-C
- 6-C and Related
- Aromatics
- C Storage

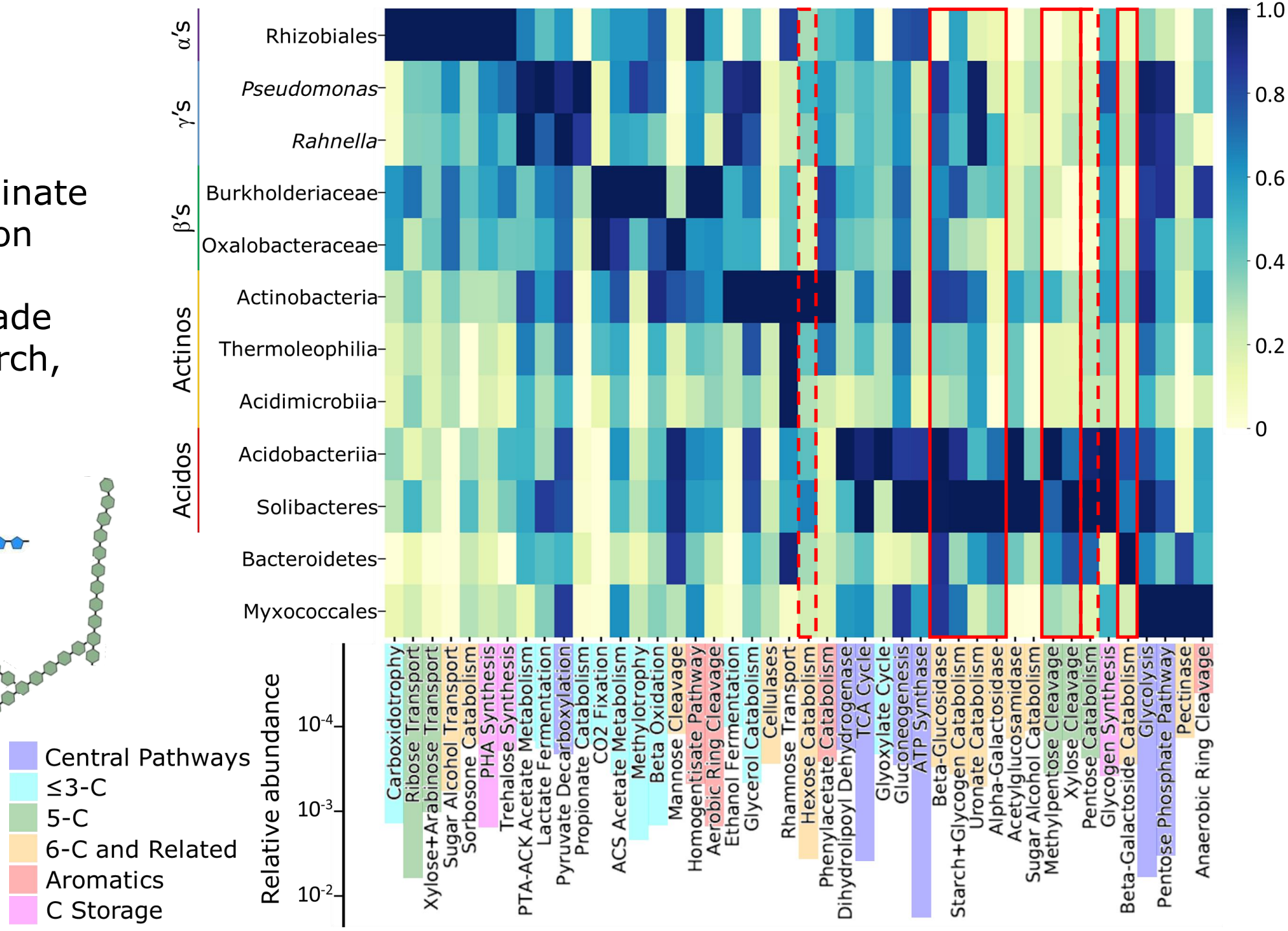
Non-Proteobacteria dominate polymer degradation

Actinobacteria dominate cellulose degradation

Acidobacteria degrade hemicelluloses, starch, and glycogen



Carbon Metabolism

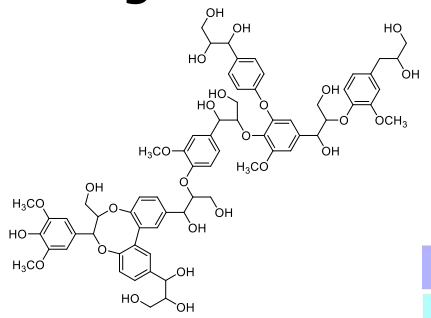


Non-Proteobacteria dominate polymer degradation

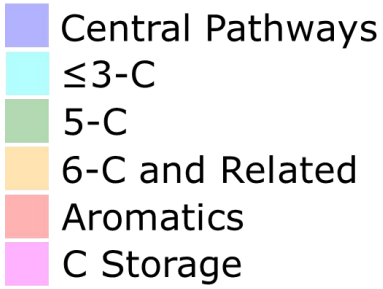
Actinobacteria dominate cellulose degradation

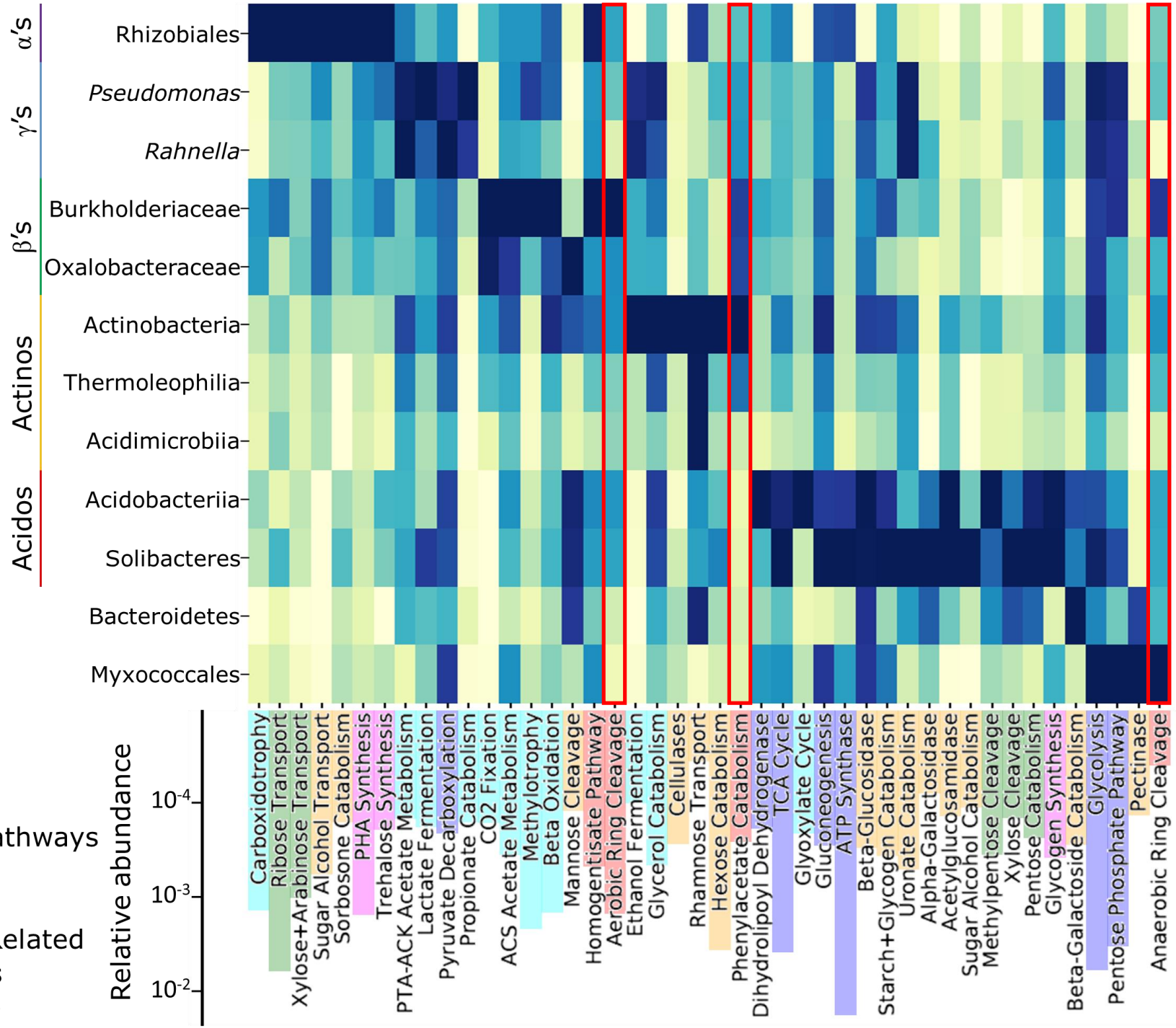
Acidobacteria degrade hemicelluloses, starch, and glycogen

Burkholderiaceae and Actinobacteria dominate lignin degradation

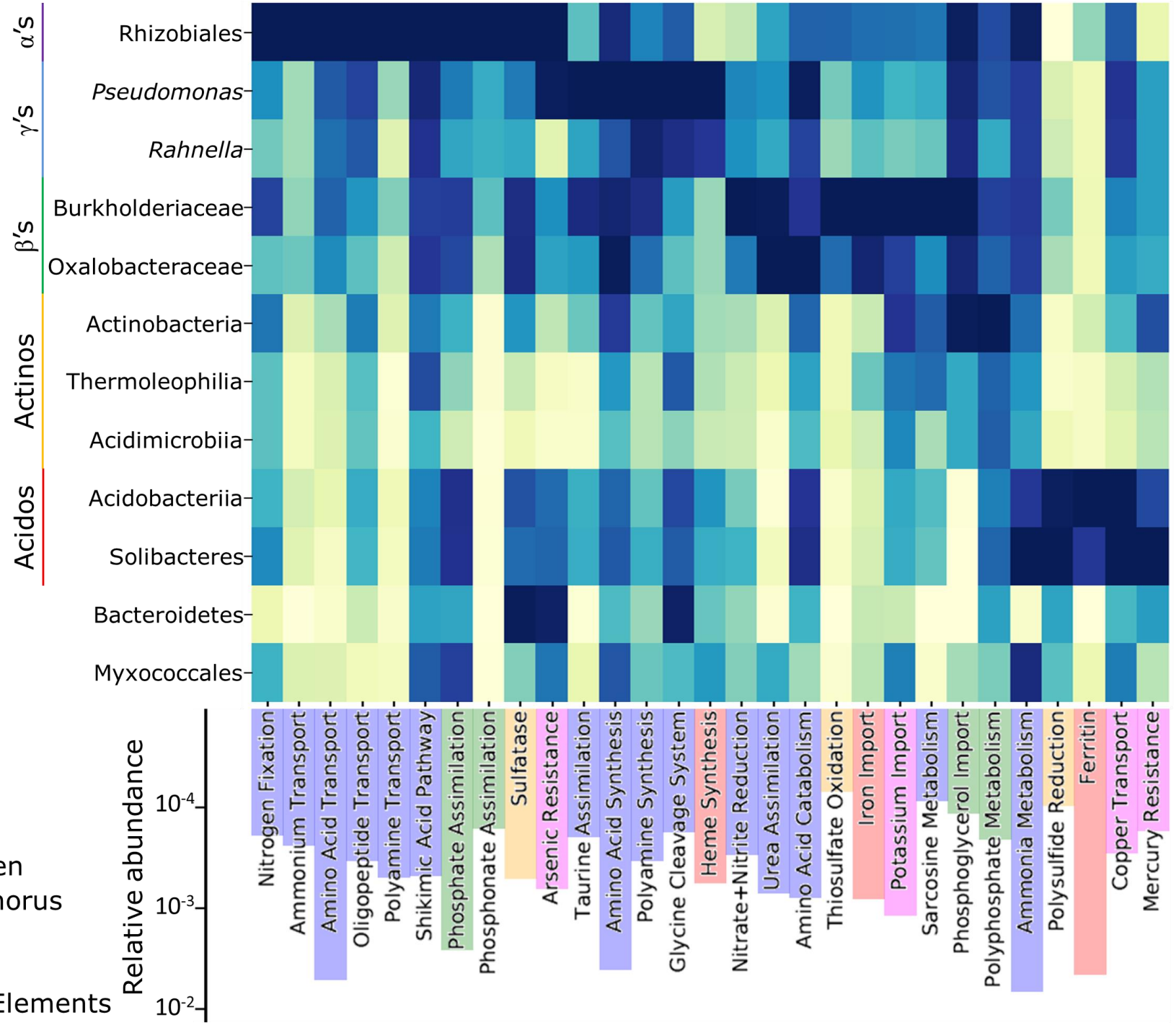


Carbon Metabolism





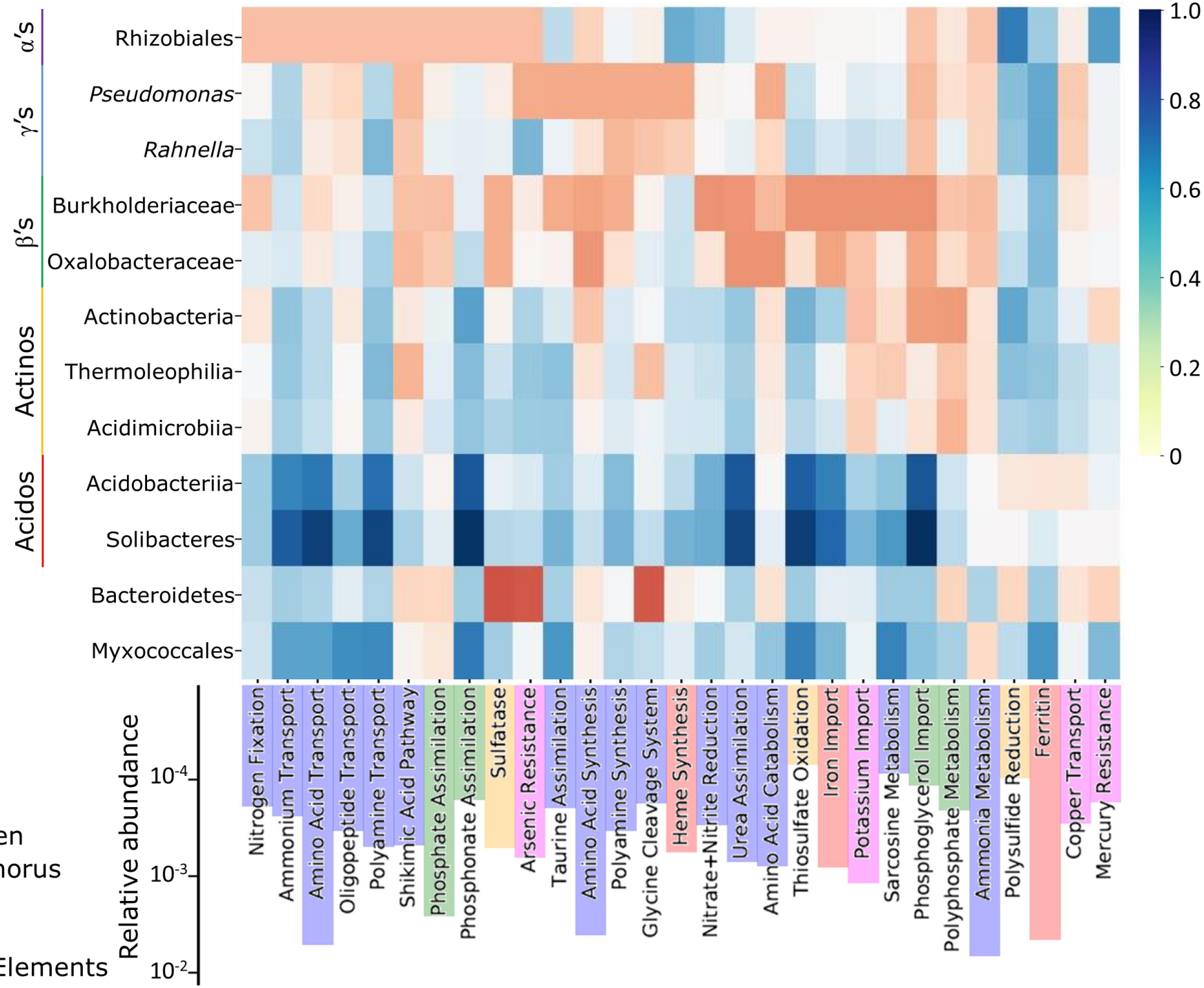
Hypothesis: Proteins involved in nutrient usage are expressed by taxa in proportion to overall activity, especially for the common limiting nutrient, nitrogen. (No resource partitioning)



Nutrient Metabolism

- Nitrogen
- Phosphorus
- Sulfur
- Iron
- Trace Elements

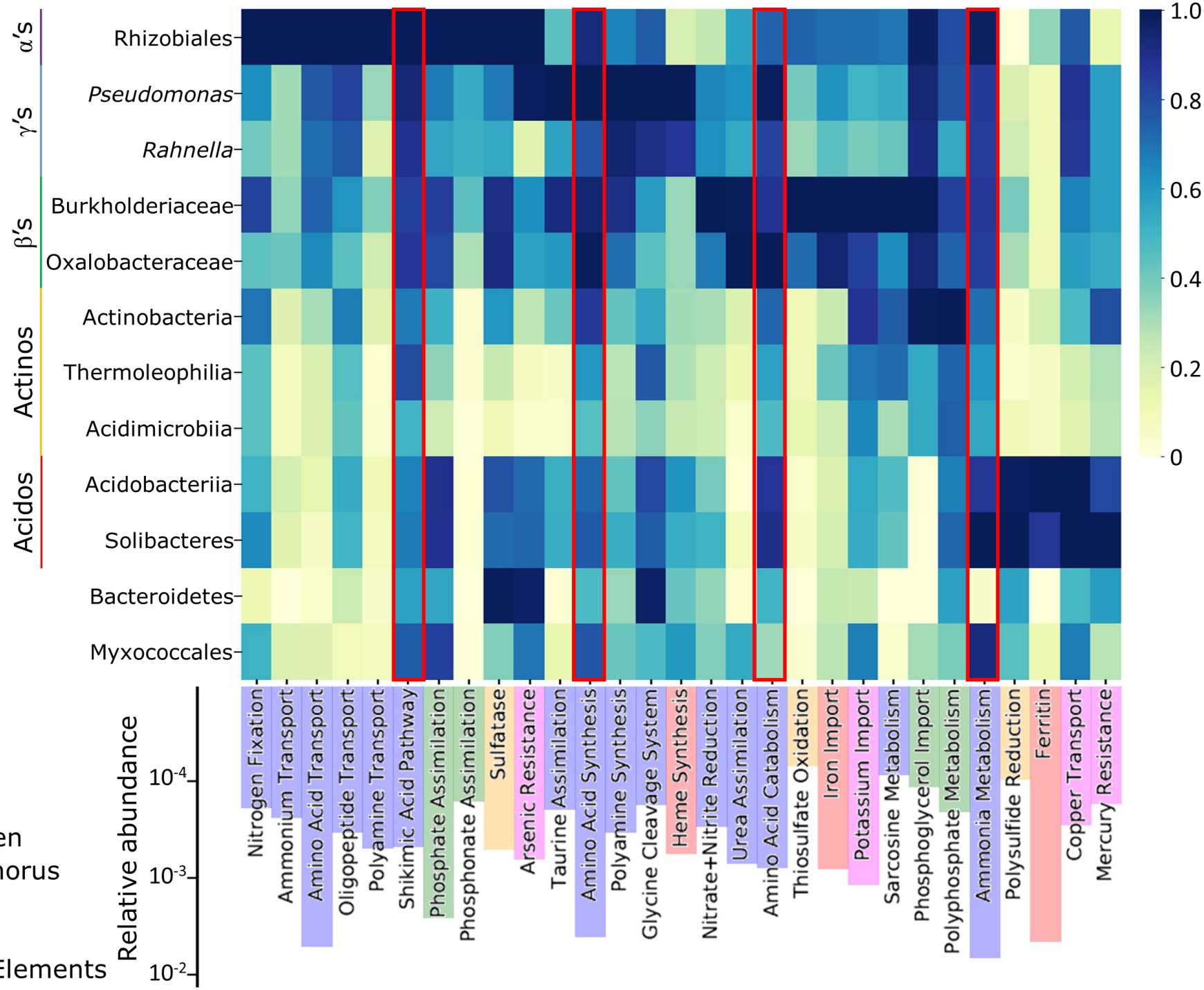
Hypothesis: Proteins involved in nutrient usage are expressed by taxa in proportion to overall activity, especially for the common limiting nutrient, nitrogen. (No resource partitioning) **FALSE**



Nutrient Metabolism

- Nitrogen
- Phosphorus
- Sulfur
- Iron
- Trace Elements

Key proteins for intracellular N cycling are relatively even across taxa

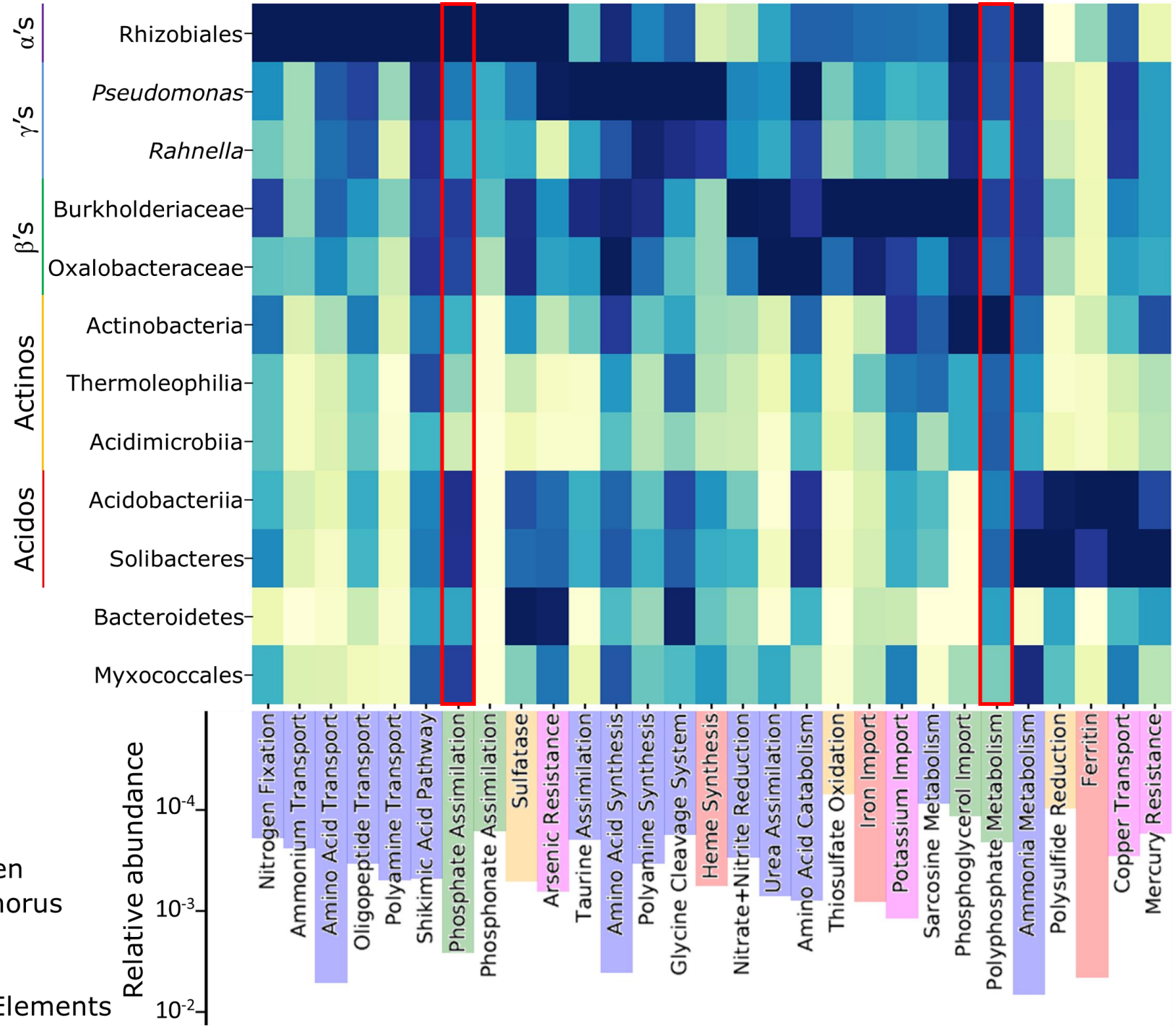


Nutrient Metabolism

- Nitrogen
- Phosphorus
- Sulfur
- Iron
- Trace Elements

Key proteins for intracellular N cycling are relatively even across taxa

Key proteins for phosphate assimilation are relatively even across taxa



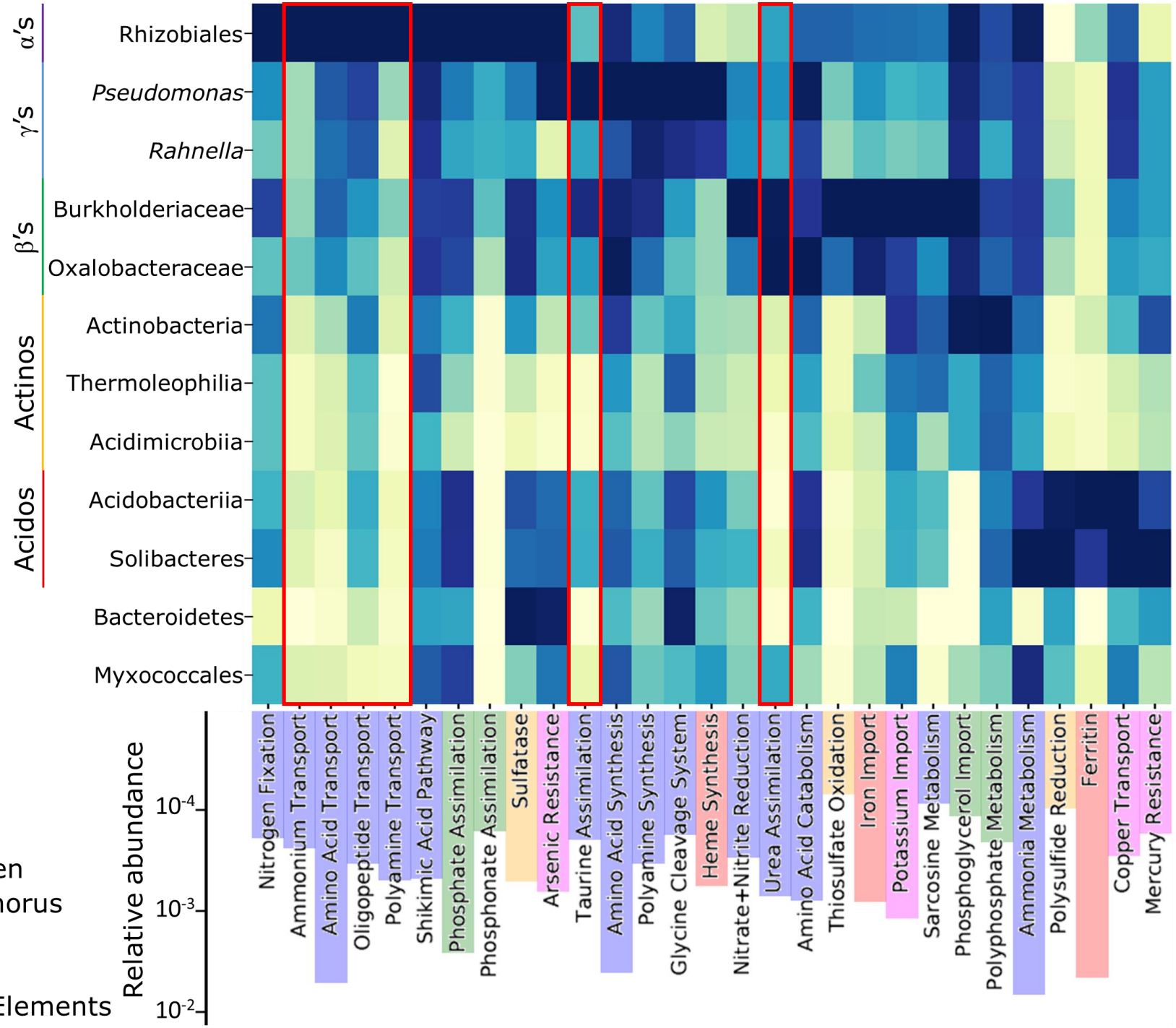
Nutrient Metabolism

- Nitrogen
- Phosphorus
- Sulfur
- Iron
- Trace Elements

Key proteins for intracellular N cycling are relatively even across taxa

Key proteins for phosphate assimilation are relatively even across taxa

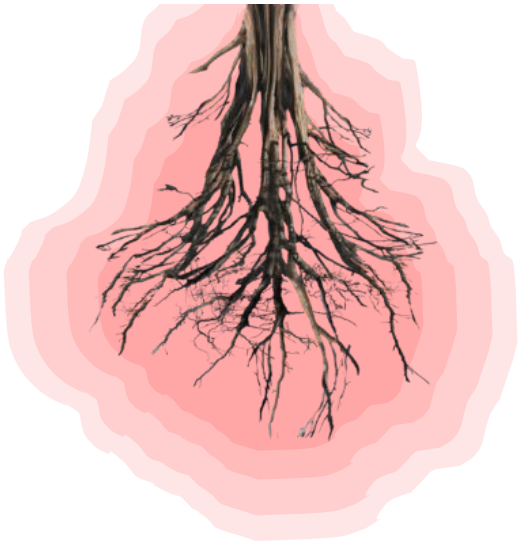
Proteobacteria – especially Rhizobiales – dominate the expression of N transporters, which mostly target organic N

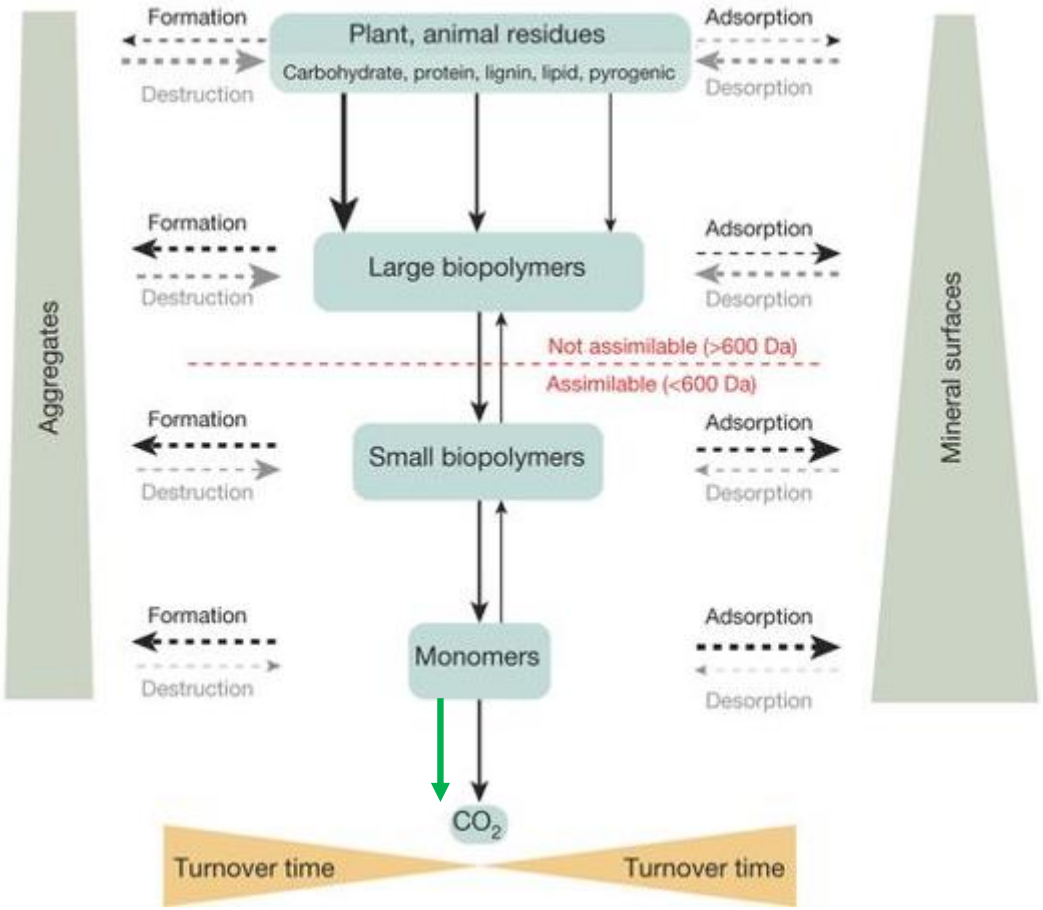


Nutrient Metabolism

- Nitrogen
- Phosphorus
- Sulfur
- Iron
- Trace Elements

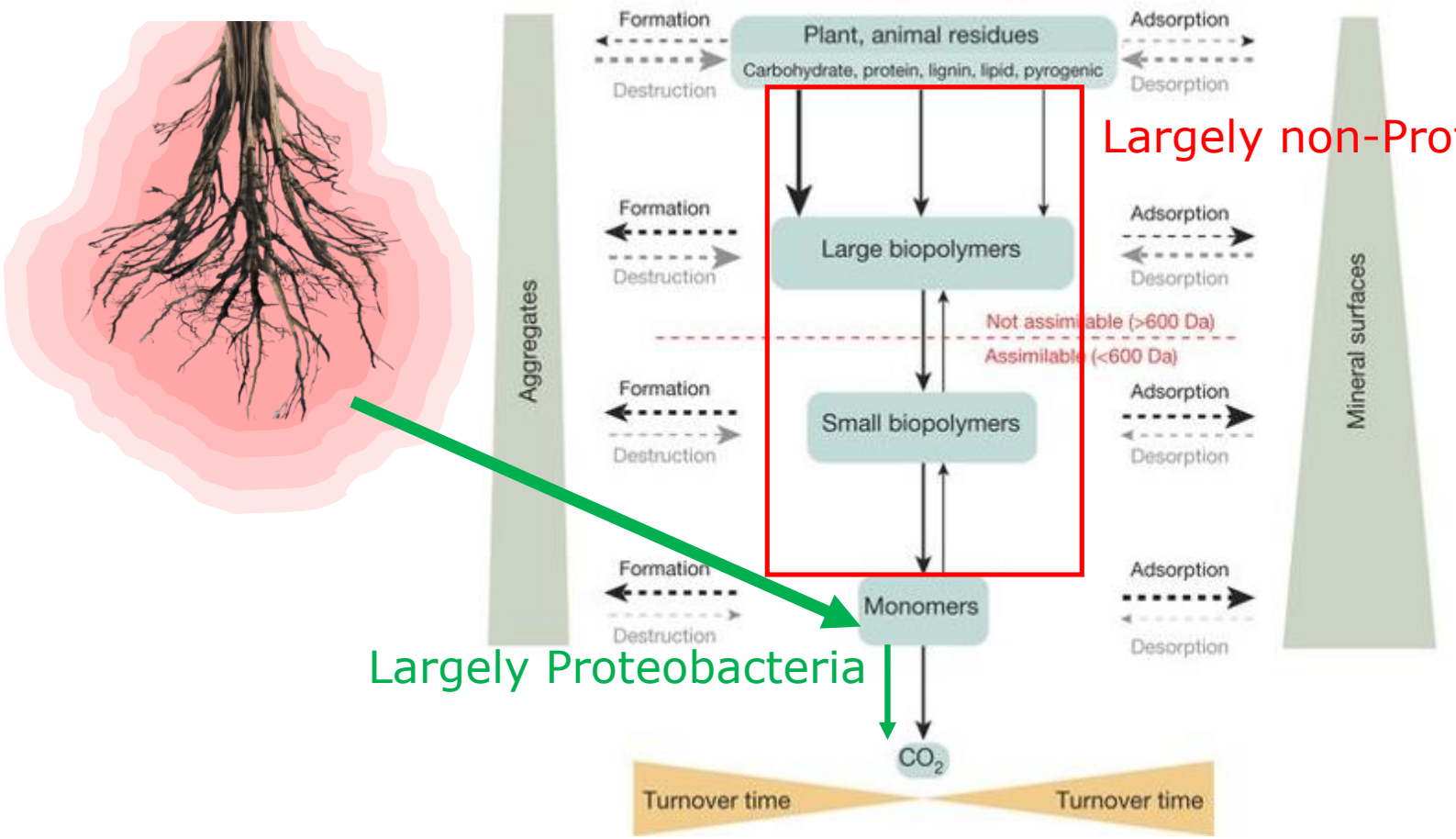
Greater N limitation around roots





Lehmann and Kleber, 2015, *Nature*

Carbon resource partitioning

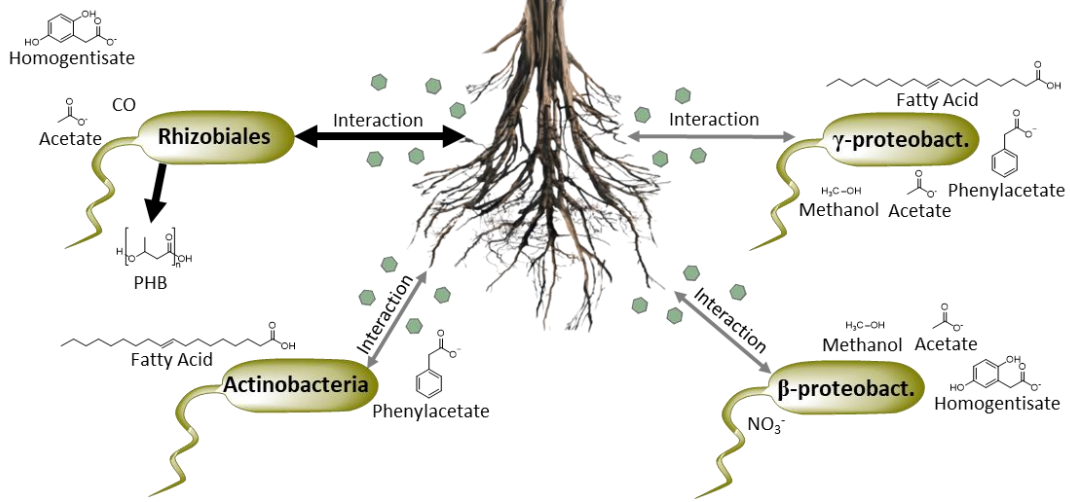


Lehmann and Kleber, 2015, *Nature*

Taxonomic niche partitioning relatively invariant among floras

Proteobacteria consume small solutes

Rhizobiales functions increase with greening, likely involved in exudate consumption and root biofilms

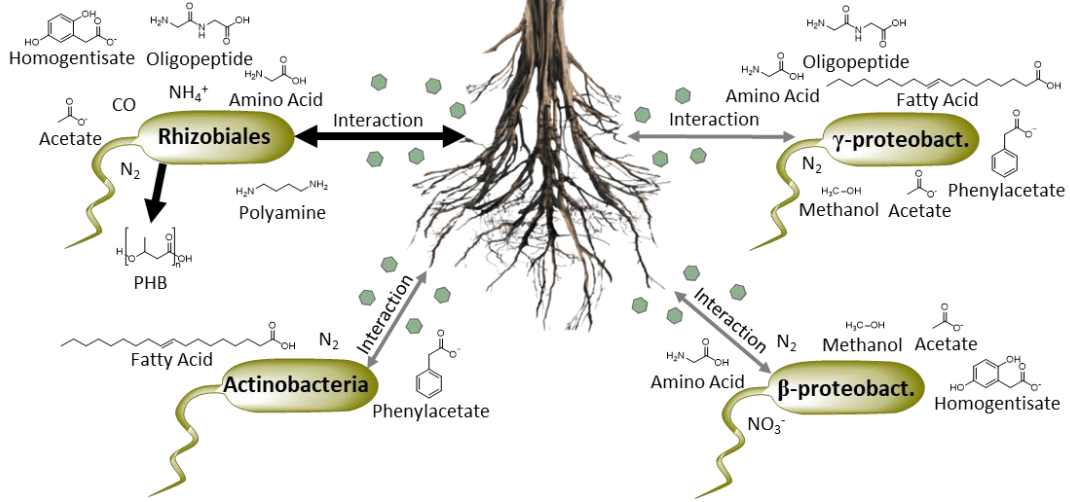


Taxonomic niche partitioning relatively invariant among floras

Proteobacteria consume small solutes

Rhizobiales functions increase with greening, likely involved in exudate consumption and root biofilms

Proteobacteria compete with plants for scarce N, mainly in organic forms



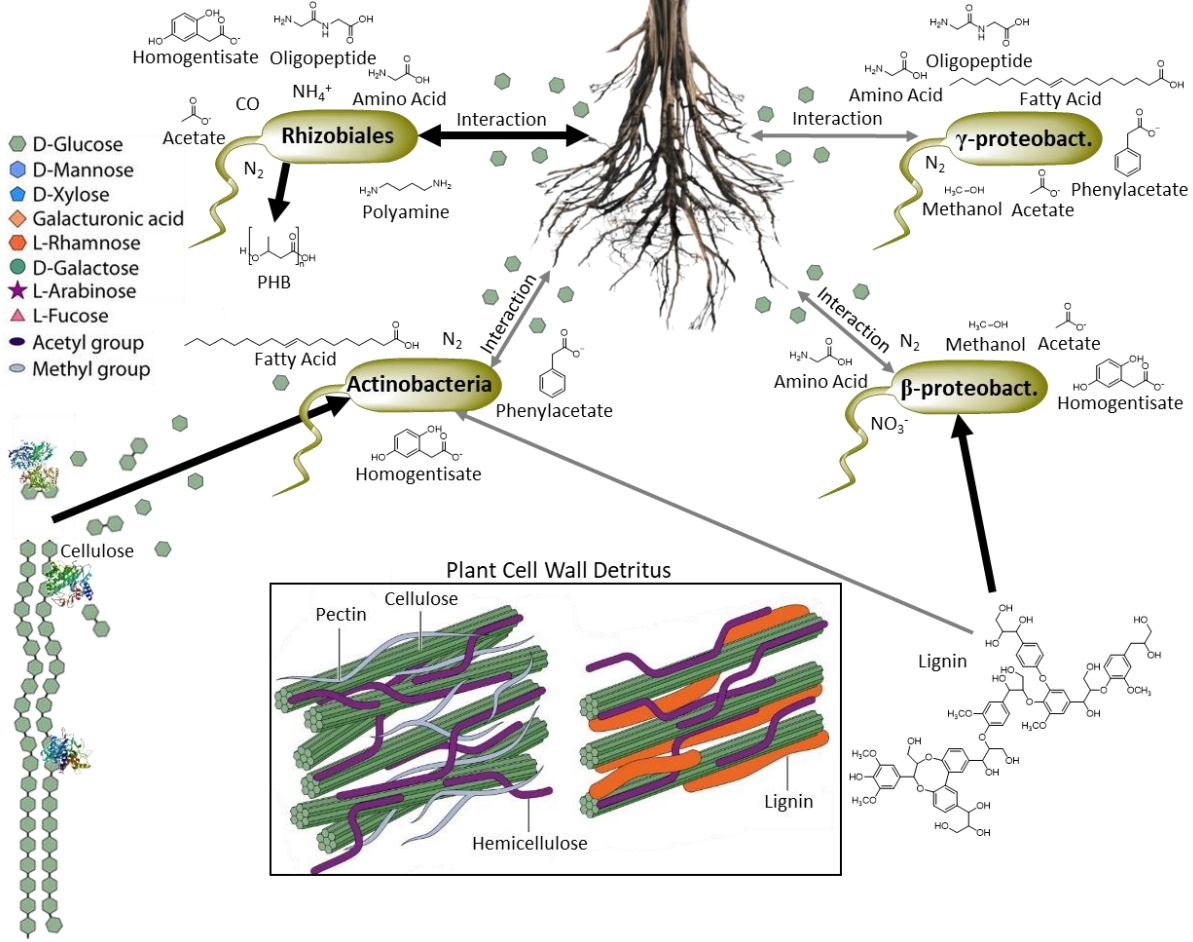
Taxonomic niche partitioning relatively invariant among floras

Proteobacteria consume small solutes

Rhizobiales functions increase with greening, likely involved in exudate consumption and root biofilms

Proteobacteria compete with plants for scarce N, mainly in organic forms

Actinobacteria and Burkholderiaceae (β) span rhizosphere and bulk soil, degrading cellulose and lignin



Taxonomic niche partitioning relatively invariant among floras

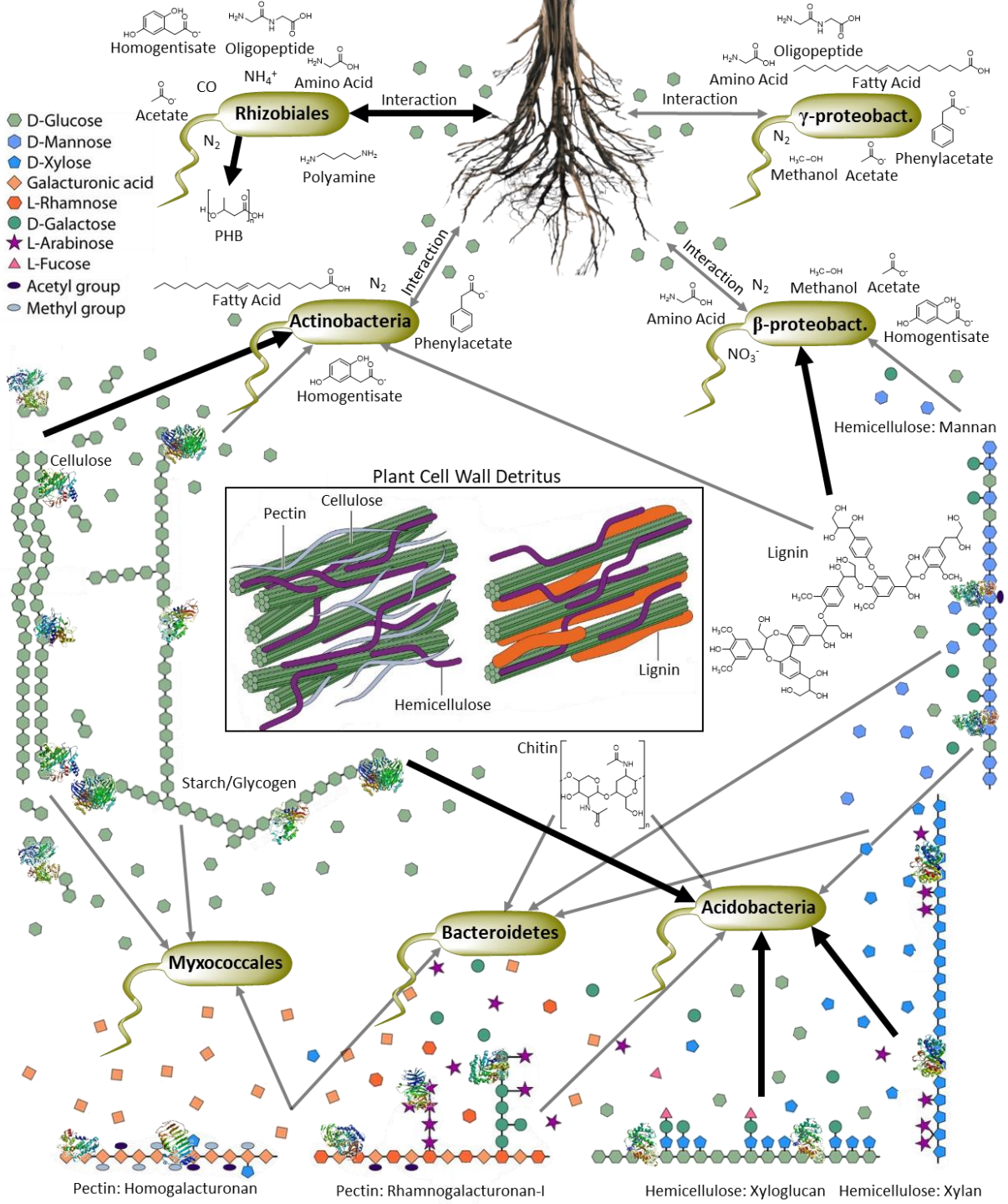
Proteobacteria consume small solutes

Rhizobiales functions increase with greening, likely involved in exudate consumption and root biofilms

Proteobacteria compete with plants for scarce N, mainly in organic forms

Actinobacteria and Burkholderiaceae (β) span rhizosphere and bulk soil, degrading cellulose and lignin

Acidobacteria are most active, degrading labile polysaccharides



Conclusions and Future Directions

- *Postnovo* extends the application of de novo sequencing
- *ProteinExpress* leverages a reference database of pooled DNA/RNA sequences for complex metaproteomes
- Taxonomic “fidelity” metric is useful for understanding which taxa express which functions
- Metaproteomics should now be pursued further in Arctic soils and other complex environments
- How does activity change seasonally and with secular warming?
- Do patterns of niche partitioning hold in other soils?

Acknowledgments

Jacob Waldbauer, Maureen Coleman, Michael Foote, David Archer

Waldbauer/Coleman Lab Groups, Adriana Rizzo, Lichun Zhang, Waldbauer Lab Funding Sources (Moore and Simons Foundations)

Gerard Olack, Albert Colman, Mark Anderson

Alissa Sherman Miller, Jan and Richard Miller, Mary and Gene Sherman