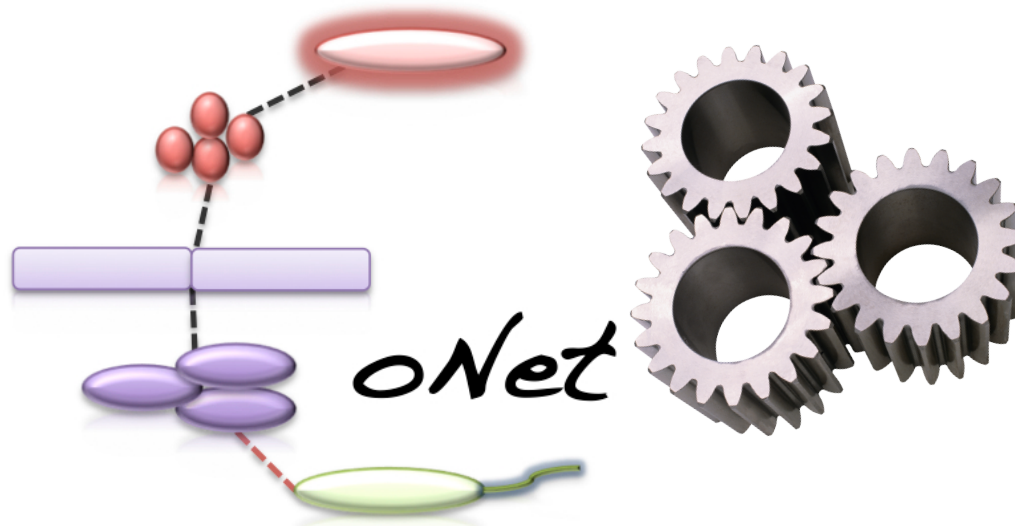
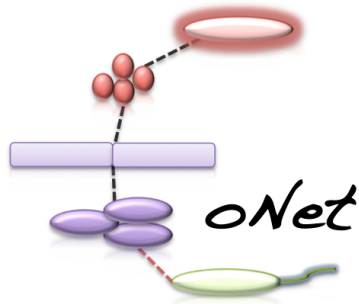


CoNet Demo

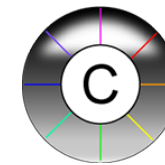




About CoNet



3.X



2.X

- Cytoscape plugin and command line tool to do ensemble-based relevance (similarity, dissimilarity, correlation) network inference
- web pages:
 - <http://systemsbiology.vub.ac.be/conet>
 - <http://apps.cytoscape.org/apps/conet>
- well documented (help pages, manual, tutorials, FAQ)
- support for row groups, presence/absence data, lagged similarities, minet integration, settings loading/saving and more

Demo data

- 52 arctic tundra soil samples collected in 2007 and 2008 by Chu and co-workers
- Roche FLX 454 sequencing using V1V2 region
- Processed with the QIIME pipeline

environmental
microbiology



Environmental Microbiology (2010) 12(11), 2998–3006

doi:10.1111/j.1462-2920.2010.02277.x

Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes

Download demo data from QIIME



The screenshot shows the QIIME website interface. At the top is the QIIME logo with the tagline 'Quantitative Insights Into Microbial Ecology'. Below the logo is a navigation bar with links for 'Home', 'Meta Analysis', 'Tools', 'Help', and 'Log Out'. A note states: 'NOTE: Please read the QIIME-DB Processing Protocol to understand how all uploaded data is handled.' Below this is a section titled 'View Study Details' with a sub-note: 'NOTE: Only qiime studies are available here. If you don't see your study, please check the emp portal for your study.' The 'Available Studies' section contains a scrollable list of study names, with 'Chu_arctic_soils' highlighted. To the right of the list, the text 'ID: Chu_arctic_soils' is displayed, followed by the URL <http://www.microbio.me/qiime/>. Below the list is a 'Study Information' section with the following details: Study ID: 104, Project Name: Chu_arctic_soils, and Study Title: Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes.

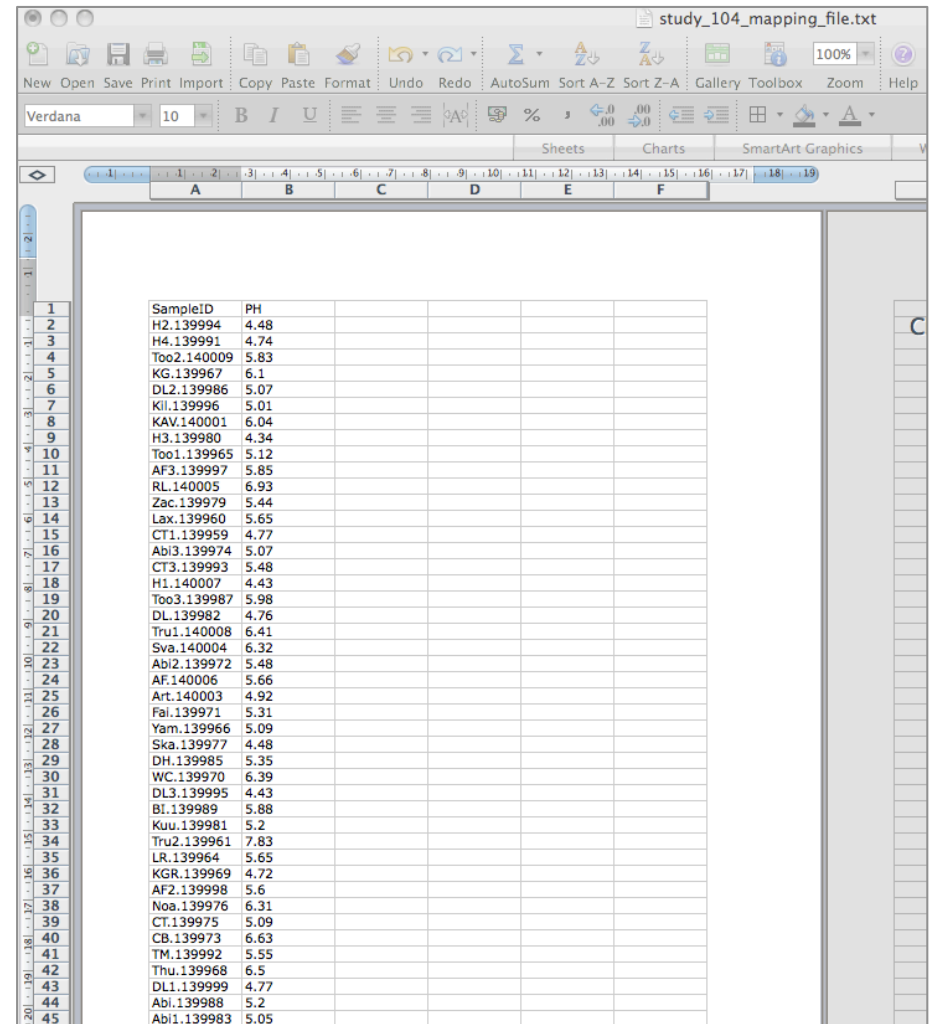
QIIME database
needs
registration,
which is for
free

- convert biom file into classical OTU table using converter from <http://biom-format.org/>
- Command:

```
biom convert -i study_104_closed_reference_otu_table.biom -o arctic_soils.txt -b --header-key taxonomy
```

Extract features from mapping file

- mapping file (study_104_mapping_file.txt) contained in zip file downloaded from the QIIME database
- open mapping file in Excel
- select SampleID (remove #) and PH column
- save selection as tab-delimited file



The screenshot shows an Excel spreadsheet with the following data:

SampleID	PH
H2.139994	4.48
H4.139991	4.74
Too2.140009	5.83
KG.139967	6.1
DL2.139986	5.07
Kil.139986	5.01
KAV.140001	6.04
H3.139980	4.34
Too1.139965	5.12
AF3.139997	5.85
RL.140005	6.93
Zac.139979	5.44
Lax.139960	5.65
CT1.139959	4.77
Abi3.139974	5.07
CT3.139993	5.48
H1.140007	4.43
Too3.139987	5.98
DL.139982	4.76
Tru1.140008	6.41
Sva.140004	6.32
Abi2.139972	5.48
AF.140006	5.66
Art.140003	4.92
Fal.139971	5.31
Yam.139966	5.09
Ska.139977	4.48
DH.139985	5.35
WC.139970	6.39
DL3.139995	4.43
BI.139989	5.88
Kuu.139981	5.2
Tru2.139961	7.83
LR.139964	5.65
KGR.139969	4.72
AF2.139998	5.6
Noa.139976	6.31
CT.139975	5.09
CB.139973	6.63
TM.139992	5.55
Thu.139968	6.5
DL1.139999	4.77
Abi.139988	5.2
Abi1.139983	5.05

Processed Demo Data

- Processed demo data are available from the CoNet web page, tutorial 5
- demo data contain:
 - OTU table (arctic_soils.txt)
 - feature matrix (arctic_soils_features.txt)
 - precomputed permutations (arctic_soils_permutations.txt)
 - precomputed bootstraps (arctic_soils_bootstraps.txt)

Configure OTU table loading

input file location

standard QIIME OTU table (lineages, if present, are parsed directly from this table)

Select your data matrix. Cooccurrence is calculated between rows. Indicate missing values as NaN

Select file

arctic_soils.txt

Second input matrix (optional)

Select file

Parsing options

Change default column delimiter (default: tab)

Transpose data matrix

Table obtained from biom file (use converter from <http://biom-format.org/> to obtain table from biom file)

Matrix type

abundance

count

incidence

Time series

Specify maximum lag 0

Close menu Help Metadata and features (optional) Output (optional)

Higher-level taxa assignment

Metadata and Feature loading menu

Metadata
Load tab-delimited metadata file.

Enter metadata column names separated by a slash.
Optionally specify the group column name (should be one of the metadata columns).
Optionally specify a second group column name (should be one of the metadata columns).

Lineages
 explore links between higher-level taxa
Specify the lineage separator:
--

Row combination filters
Warning: The parent-child relationship filter needs the lineage and taxon attribute. Please see the help first.
 Parent-child exclusion
If a group attribute was specified, the following filter options are available:
 Within-group relationships only
 Between-group relationships only

Features
Load tab-delimited feature file.

 Exclude features from association rule mining
 Only keep Spearman-supported feature links

enable higher-level taxon assignment

exclusion of parent-child relationships

Configure feature loading

The screenshot shows a dialog box titled "Metadata and Feature loading menu" with four main sections: Metadata, Lineages, Row combination filters, and Features. The "Features" section is highlighted with a blue circle around the "Select file" button and the filename "ils_features.txt". Three blue arrows point from text labels on the right to the "Select file" button, the "Transpose" checkbox, and the "Match samples" checkbox.

Metadata
Load tab-delimited metadata file.

Enter metadata column names separated by a slash.
Optionally specify the group column name (should be one of the metadata columns).
Optionally specify a second group column name (should be one of the metadata columns).

Lineages
 explore links between higher-level taxa
Specify the lineage separator:

Row combination filters
Warning: The parent-child relationship filter needs the lineage and taxon attribute. Please see the help first.
 Parent-child exclusion
If a group attribute was specified, the following filter options are available:
 Within-group relationships only
 Between-group relationships only

Features
Load tab-delimited feature file.

ils_features.txt
 Transpose
 Match samples
 Exclude features from association rule mining
 Only keep Spearman-supported feature links

Annotations:
select feature file
transpose feature matrix
match feature samples to OTU table samples

Configure preprocessing

Preprocessing and filtering menu

Input filtering

Methods

- row_minsum
- row_minocc 20.0
- col_minsum
- minzeropairs

Keep sum of filtered rows:

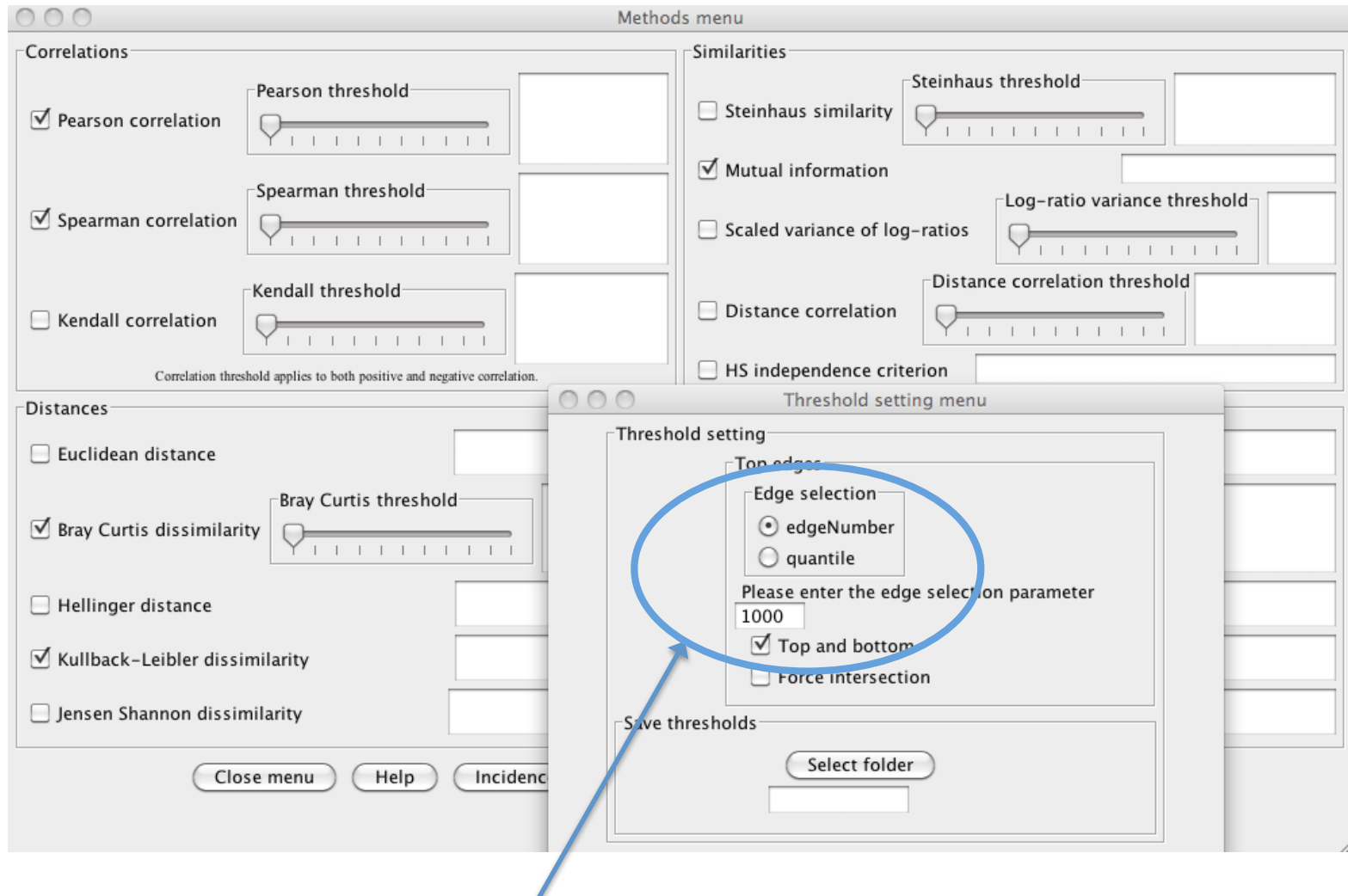
Standardization

- col_norm
- col_downsample
- row_stand
- row_stand_robust
- row_norm
- row_downsample
- log2

filter rows with too many zeros
(zeros bias correlation measures)

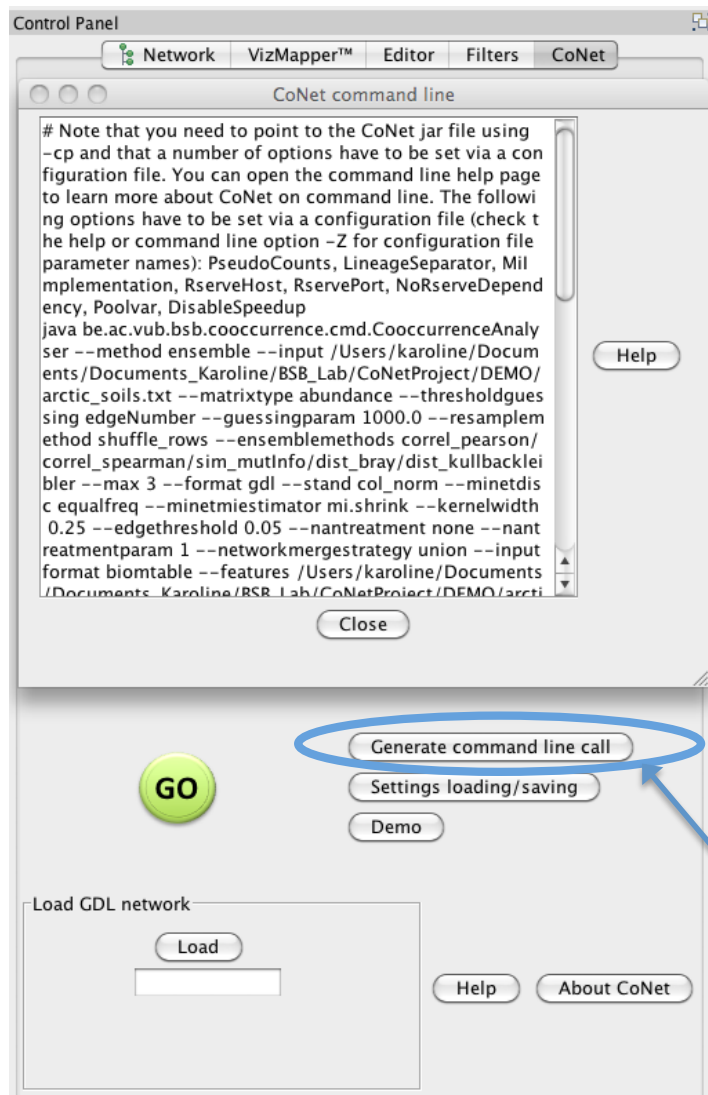
divide entries by the total count of their
corresponding column (takes out effects of
varying sequencing depth)

Select methods



set thresholds such that the 1,000 highest and lowest scoring edges for each method are included in the initial network

Optional: Run CoNet on command line



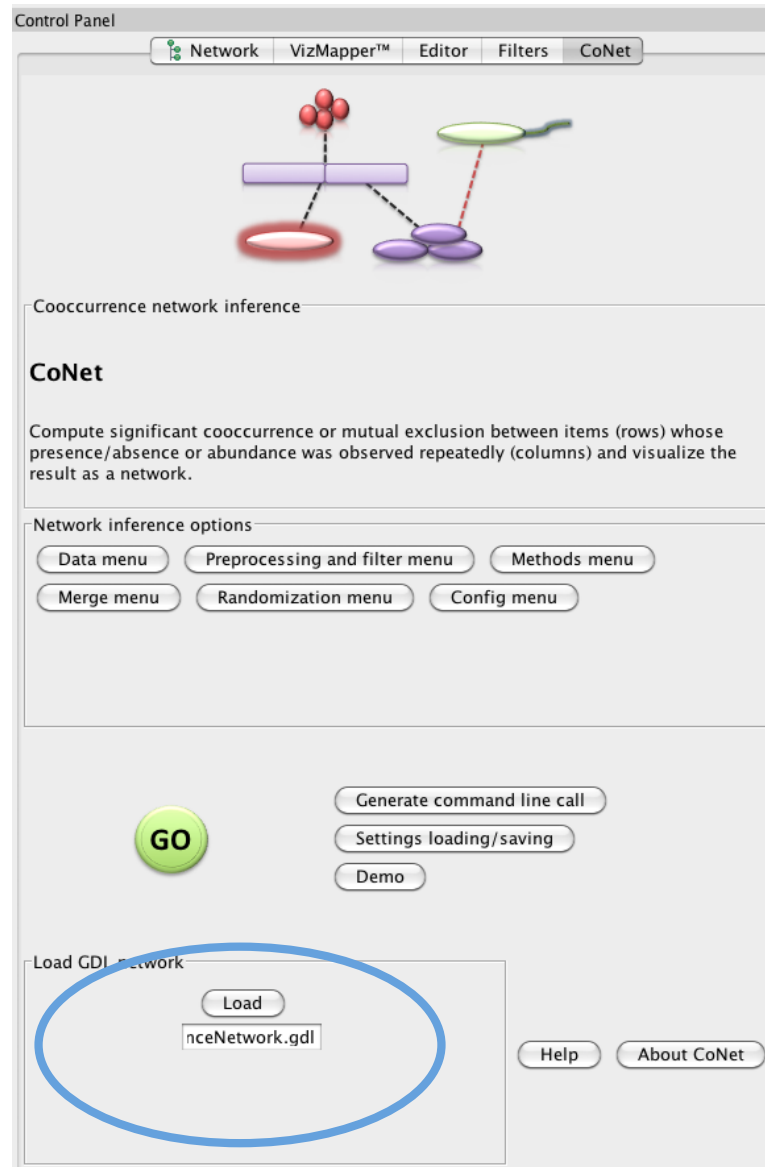
- copy-paste first java command line call (first line starting with java and ending with thresholds.txt) into a text editor, **add the path to the CoNet jar** file via -cp argument, copy the extended command into a shell and enter
- **threshold file** will be created in temporary directory
- repeat the same steps for the second java command line call (second line starting with java)
- **network file** (ending in .gdl) will be generated in current directory

Example of a (simplified) command line call with jar file included (do not copy-paste this example into a shell):

```
java be.ac.vub.bsb.cooccurrence.cmd.CooccurrenceAnalyser -cp /Applications/Cytoscape_v2.8.0/plugins/CoNet.jar --method ensemble --input arctic_soils.txt --matrixtype abundance --thresholdguessing edgeNumber --guessingparam 1000.0 --ensemblemethods correl_pearson/correl_spearman/sim_mutInfo/dist_bray/dist_kullbackleibler --stand col_norm --minetdisc equalfreq --inputformat biomtable --features arctic_soils_features.txt --multigraph --higherleveltaxa --transposefeatures --matchfeaturesamples --topbottom --filter row_minocc/noinclusivetaxalinks --filterparameter 20.0 --output thresholds.txt
```

click here to generate command line call

Optional: Load command line result into Cytoscape



select gdl network file to load

Click



Compute permutations

row-wise permutation with 100 iterations

The screenshot shows a 'Randomization menu' window with several sections:

- Randomization routine:** Includes an 'Iterations' input field set to '100'. Below it are two columns of radio buttons: 'Routine' (none, edgeScores, lallich) and 'Resampling' (shuffle_rows, shuffle_cols, shuffle_both, bootstrap). A blue oval highlights the 'edgeScores' and 'shuffle_rows' options.
- Randomization options:** A 'P-value merge*' section with radio buttons (none, brown, fisher, simes, mean, geometric_mean, median, max). A blue arrow points from the 'Renormalize' checkbox to the text 'enable renormalization'.
- P-value options:** A 'Multiple test correction' section with radio buttons (none, compute_eval, bonferroni, benjaminihochberg). A 'P-value threshold*' input field is set to '0.05'. A note explains that thresholds refer to p-values, except for E-value computation.
- Load randomizations:** Two sections for loading files, each with an 'Open file' button and an empty input field.
- Save (excludes load):** A 'Select folder' button above an input field containing ':tProject/DEMO'. Below is a 'Specify file name' input field with 'demo-permutations.txt' and a checked 'Save randomizations to file' checkbox. A red warning note is at the bottom: '*If a file of same name exists in selected directory, it will be overwritten!'.
- Bottom right:** 'Close menu' and 'Help' buttons, and a green 'GO' button.

save permutation scores to a file

enable renormalization

Click



Compute bootstraps

100
bootstrap
iterations

Randomization menu

Randomization routine

Iteration: 100

Routine

- none
- edgeScores
- lallich

Resampling

- shuffle_rows
- shuffle_cols
- shuffle_both
- bootstrap

Randomization options

P-value merge*

- none
- brown
- fisher
- simes
- mean
- geometric_mean
- median
- max

* p-value merge is available for multigraphs only

Filter unstable edges (bootstrap only)

Renormalize (edgeScore routine only)

Pool variances (only for p-values combining permutation and bootstrap)

P-value options

Multiple test correction

- none
- compute_eval
- bonferroni
- benjaminihochberg

P-value threshold* 0.05

*Thresholds refer to p-values, except for E-value computation, where they refer to significances (-log10(E-val)).

Load randomizations

Load rand file

Load randomization file

Load null distributions

Load null distributions

Save (excludes load)

Select folder

!tProject/DEMO

Specify file name

demo-bootstrap.txt

Save randomizations to file

*If a file of same name exists in selected directory, it will be overwritten.

save bootstrap
scores to a file

Click



Build final network

Randomization menu

Randomization routine

Iterations: 100

Routine: none, edgeScores, lallich

Resampling: shuffle_rows, shuffle_cols, shuffle_both, bootstrap

Randomization options

P-value merge*: none, brown, fisher, simes, mean, geometric_mean, median, max

Filter unstable edges (bootstrap only)

Renormalize (edgeScore routine only)

Pool variances (only for p-values combining permutation and bootstrap)

P-value options

Multiple test correction: none, compute_eval, bonferroni, benjaminihochberg

P-value threshold*: 0.05

*Thresholds refer to p-values, except for E-value computation, where they refer to significances (-log10(E-val)).

Load randomizations

Load rand file: Load randomization file: -bootstraps.txt [Open file]

Load null distributions: Load null distributions: rmutations.txt [Open file]

Save (excludes load)

Select folder: []

Specify file name: []

Save randomizations to file

*If a file of same name exists in selected directory, it will be overwritten!

Close menu Help

GO

merge
method-
specific p-
values

optional:
filter edges
with scores
outside the
bootstrap
distribution

load
random
score files

multiple
testing
correction

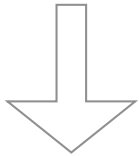
Click



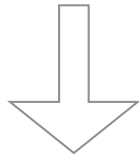
Beautify the network

1. Layout the network with yFiles
2. Color phyla with the VizMapper

Layout



yFiles



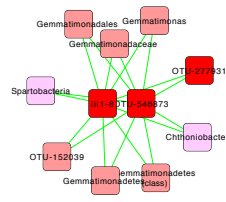
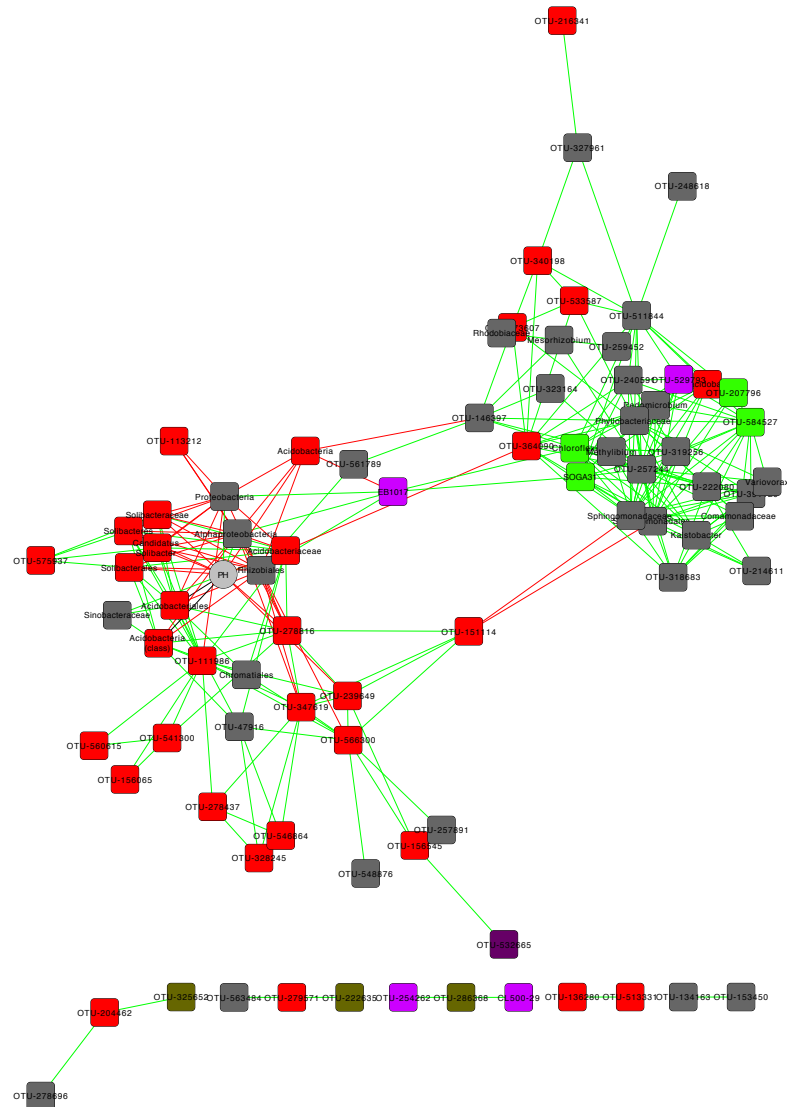
Organic












The screenshot shows the VizMapper software interface. At the top, there are tabs for 'Network', 'VizMapper™', 'Editor', 'Filters', and 'CoNet'. Below the tabs, the 'Current Visual Style' is set to 'cooccurrence'. A 'Defaults' section shows a simple network diagram with two nodes and one edge. The 'Visual Mapping Browser' section is expanded to show 'Node Color' mapping, which is set to 'phylum'. A table lists various phyla with corresponding color swatches:

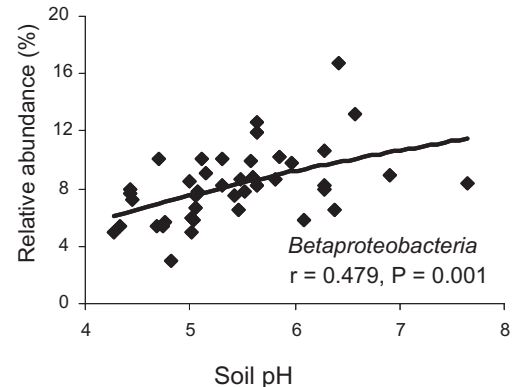
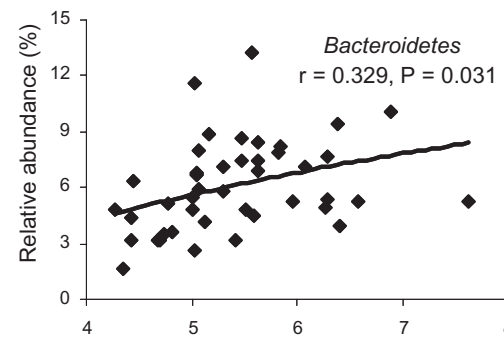
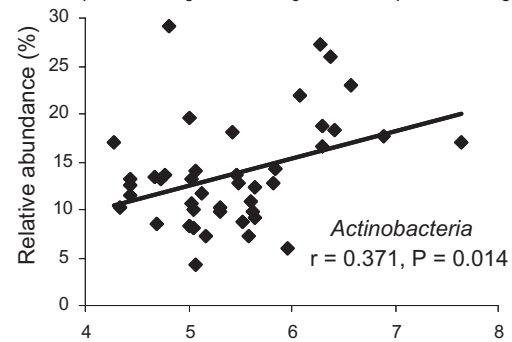
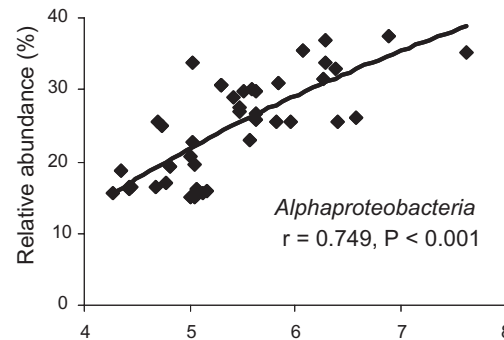
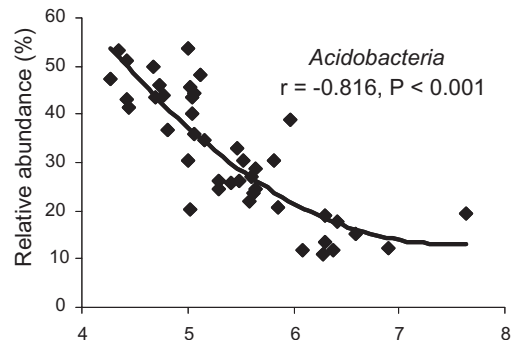
Phylum	Color
Acidobacteria	Red
Actinobacteria	Purple
Bacteroidetes	Olive Green
Chloroflexi	Bright Green
Gemmatimonadetes	Pink
Proteobacteria	Grey
SPAM	Blue
TM7	Dark Purple
Verrucomicrobia	Light Pink

Result network



Node Color	phylum
Mapping Type	Discrete Mapping
Acidobacteria	
Actinobacteria	
Bacteroidetes	
Chloroflexi	
Gemmatimonadetes	
Proteobacteria	
SPAM	
TM7	
Verrucomicrobia	

Interpretation of results



Anti-correlation between Proteobacteria OTUs and Acidobacteria OTUs is pH-driven.

environmental microbiology



Environmental Microbiology (2010) 12(11), 2998–3006

doi:10.1111/j.1462-2920.2010.02277.x

Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes

Next steps

- analyze network properties such as modularity, diameter, average cluster coefficient and node degree distribution using Cytoscape plugin “Network analysis”
- cluster network with Cytoscape plugins “CommFinder”, “clusterMaker” or “MCODE”