PANAM: Phylogenetic Analysis of Next generation AMplicons

TAIB N, BRONNER G, DEBROAS D.
Laboratoire Microorganismes : Génome et Environnement - UMR 6023 CNRS - FRANCE

Depict the microbial diversity using a phylogenetic approach

PANAM is a pipeline for depicting the microbial diversity within one or several samples. It implements a phylogenetic approach for the annotation of 16S and 18S rRNA genes, and is optimized for the processing of large datasets (Pyrosequencing and Illumina MiSeq runs). PANAM combines the publicly available tools: PANGEA, PANDASEQ, USEARCH, HMMALIGN and FASTTREE with perl scripts dedicated to (i) cleaning raw sequences, (ii) trimming profile alignments to fit the studied region and (iii) processing generated phylogenies to describe clades and phylogenetic diversity indexes.

If you use this package, give this link as reference: http://code.google.com/p/panam-phylogenetic-annotation/

License:

PANAM is free software: you may redistribute it and/or modify its under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or any later version.

PANAM is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
See the GNU General Public License for more details
(http://www.gnu.org/licenses/).

Requirements:

Linux OS. PANAM was tested on ubuntu system 12.4 and 14.04
The following softwares are required by the PANAM package. They need to be downloaded and installed separately from PANAM.
1. Perl 5 or later (www.perl.org)
2. USEARCH (www.drive5.com). Tested versions are v1.1.579q; v3.0.617; v4.0.38 and v5.0.150. Other versions may not be suitable to the running of PANAM.
3. PANDASEQ
4. gcc.
5. make.
6. R.
7. R packages: Vegan; Phyloseq; Picante; Mass.

The following softwares are included in the PANAM package, they will be installed when setting PANAM up. (We do not guarantee the running of PANAM with other versions)
1. FastTree-2.1.3
2. HMMER-2.3.2
3- Bioperl-1.5.2
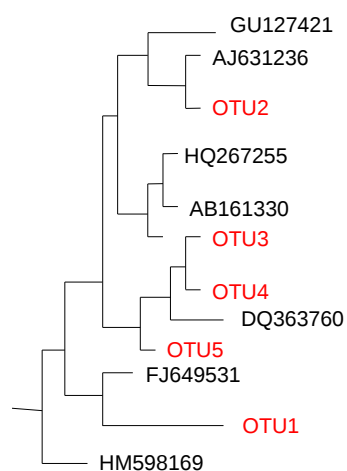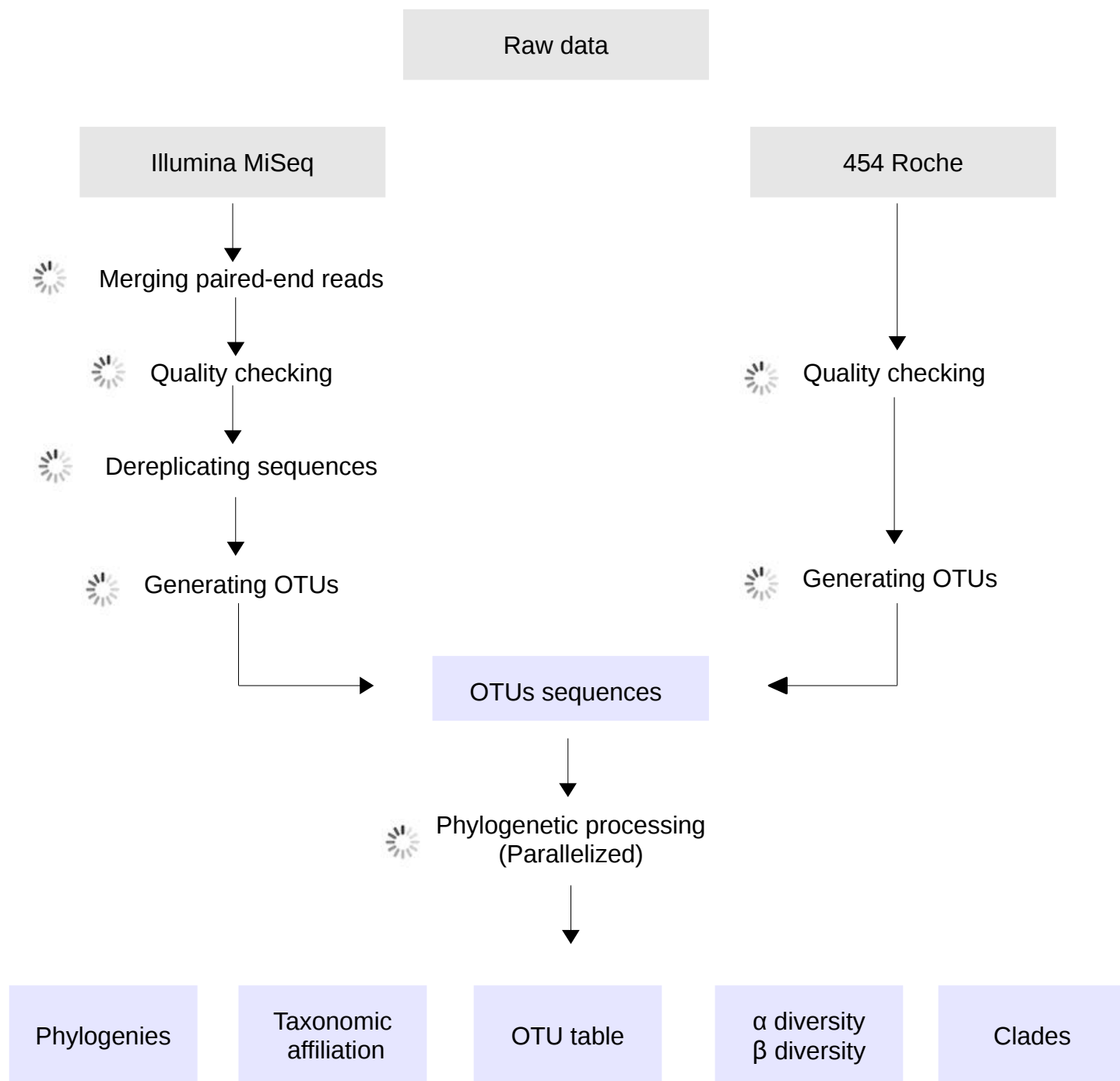

Overview

The PANAM implementation allows a user to complete a full massif sequencing run, including raw data processing, OTU clustering, phylogenetic analysis and diversity indexes computation. The main workflow is represented in the following Figure. The analysis is guided by a configuration file (.ini) and several PERL programs (quality; preprocess; panam; phylodiv).

The quality.pl script takes either an input.fasta and its corresponding input.qual (pyrosequencing data) or an input_R1.fastq and an input_R2.fastq files (Illumina MiSeq data), and a barcodes file. This script is dedicated to discard short and chimeric reads, sequences with mismatches in the primers, and sequences with low quality scores or weak pairings using either PANGEA functionalities or PANDASEQ. It then sorts sequences according to their barcodes into different files and removes barcode sequences from the reads. The resulting file is then run through an OTUs generating step using the clustering threshold defined by the user (preprocess.pl). This script dereplicates the reads for huge data sets (Illumina MiSeq).

The taxonomic affiliation step is processed using perl script panam.pl. The generated OTUs are compared against the reference database of the domain of interest using USEARCH, and sorted according to the higher taxonomy level (e.g., phylum) of their best hits. Files containing the OTUs together with their best hits are then generated. Next, each file is aligned to the reference alignment of the taxonomic group it has been associated to via a HMM profile; alignments are then used to generate phylogenetic trees with 100 bootstraps using FASTTREE2 (Price et *al.*, 2010). Finally, the generated trees are rooted and parsed, taxonomy is then inferred to each OTU depending on its neighboring reference sequences in the tree. This step generates an OTU table file, with the abundance of each OTU across all the samples and their taxonomy. It also generates a file describing the phylogenetic diversity of the OTUs inserted in each profile, and a file depicting the phylogenetic units or clades, i.e. OTUs branching with the same reference sequence, with (i) their taxonomy, (ii) the bootstrap value supporting their branch, (iii) the number of samples they span, (iv) their internal dispersion and (v) their phylogenetic distance regarding to their nearest neighbor.

After phylogenies building, rarefaction curves and rank-abundance curves are generated for each sample to check the sampling effort (post_process.pl). Next, taxonomy-based indices (Chao1 and Shannon) are calculated in order to compare the global richness and evenness between all the samples, and to set the diversity within taxonomic groups at the phylum and class levels across the samples. These indices are displayed for both the whole data set in a sample and for a normalized size library of sequences that are randomly picked in the sample, for an accurate comparison between the samples by avoiding biases associated with different sampling depths.

A set of phylogeny-based indices is also implemented in the script phylodiv.pl to draw the evolutionary history of community members. These indices investigate the phylogenies based on the branches' length of the leaves, and use phylogenetic relatedness among taxa as an indicator of their ecological similarity. They employ either a minimum spanning path to measure the phylogenetic diversity (PD); or a pairwise distance approach to assess the phylogenetic structure (MPD, MNND); or the phylogenetic clustering of the individuals (NRI, NTI).

Raw data

Illumina MiSeq

454 Roche

Merging paired-end reads

Quality checking

Quality checking

Dereplicating sequences

Generating OTUs

Generating OTUs

OTUs sequences

Phylogenetic processing
(Parallelized)

Phylogenies

Taxonomic
affiliation

OTU table

α diversity
β diversity

Clades

GU127421

AJ631236

OTU2

HQ267255

AB161330

OTU3

OTU4

DQ363760

OTU5

FJ649531

OTU1

HM598169

	Sample 1	Sample 2	Sample 3	Sample 4	Taxonomy
OTU1	1	2	7	0	Archaea;Crenarchaea;Miscellaneous Crenarchaeotic Group;
OTU2	0	0	4	2	Archaea;Euryarchaea;Halobacteria;Halobacteriales;SM1K20;
OTU3	15	22	3	12	Archaea;Crenarchaea;Miscellaneous Crenarchaeotic Group;
OTU4	0	1	1	0	Archaea;Crenarchaea;Miscellaneous Crenarchaeotic Group;

Sample 3

Sample 2

Sample 1

Sample 4

Sample 7

Sample 6

3
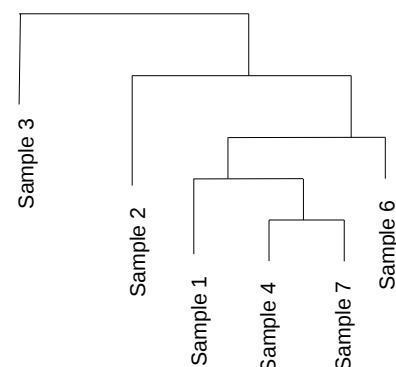
Installing in a terminal:

1. Download panam_v4.tar.gz into the user specified directory
2. Unpack panam_v4.tar.gz

```
tar -xvzf panam_v4.tar.gz
```

3. Go to the directory panam_v4/

```
cd panam_v4
```

4. Install panam with the script setup.pl.
This step require to know the full path of USEARCH.

```
perl setup.pl -USEARCH_path /user_path_to_usearch/usearchX.X.X
```

Package contents:

1. Reference/: contains the reference database restricted to the 18S rRNA sequences of microeukaryotes in fasta format (Reference_base.fasta); the taxonomy file (Taxonomy) and the profiles of 24 phyletic groups for full length sequences, in fasta and HMM formats (Profiles/).
2. bin/: contains FastTree-2.1.3, HMMER-2.3.2 and Bioperl-1.5.2.
3. perl scripts:
     - **setup.pl:** PANAM setting.
     - **quality.pl:** processes raw sequences upon quality criteria defined by the user.
     - **preprocess.pl:** dereplicates and clusters sequences into reads.
     - **panam.pl**: trims alignment profiles; aligns experimental sequences; builds phylogenies, infers taxonomies and clades.
     - **postprocess.pl:** computes taxonomy-based diversity indices and rarefaction curves.
     - **phylodiv.pl:** computes phylogeny-based alpha- and beta- diversity indices.
4. panam.ini: configuration file where you should set running options and the paths to the input and output files to run PANAM.
5. test/: a dataset to test PANAM. Contains a raw sequences file (Sequences.fasta); a quality sequences file (Sequences.qual) and a bar code file (barcode.txt). This data corresponds to 18S amplicons obtained with primers NSF573 and NSR951 from 3 lacustrine ecosystems.

Running PANAM:

PANAM can be used for amplicon quality check and phylogenetic analysis, the two scripts are independent.

1. Configure the parameters file **panam.ini** (Necessary to launch quality.pl, preprocess.pl or panam.pl).
By default, panam.ini is configured to process the file Sequences.fasta in directory test/.

   In panam.ini, PARAMETERS are in upper case letter and parameter_values have to be inserted after one tabulation.
   In this ReadMe, a [PARAMETER_VALUE] is represented in [ ].

   To analyze your data set, please set the paths to your own dataset, and choose your options.

4

2. For raw sequences quality checking, use the script **quality_v4.pl**

> perl quality_v4.pl panam.ini

3. For samples clustering use the script **preprocess_v4.pl**

> perl preprocess_v4.pl panam.ini

4. For phylogenetic analysis use the script **panam_v4.pl**

> perl panam_v4.pl panam.ini

5. For taxonomy-based diversity indices, use the script **postprocess_v4.pl**

> perl postprocess_v4.pl

5. For phylogeny-based diversity indices, use the script **phylodiv.pl**

> perl phylodiv.pl

<u>Example:</u>

panam.ini is configured to process the file Sequences.fasta in directory test/. These sequences were generated on three ecosystems with primers NSF573 and NSR951 located at the position 573 and 951 on Saccharomyces.

All the results are in the [454_RUN_IDENTIFIER] folder located in the PANAM directory.

With the default panam.ini, results are stored in the directory panam_v4/test_Results/.
- preprocess.pl generates a preprocess_ouput directory plus 3 files
- panam.pl generates a panam_ouput directory

---- Outputs of preprocess.pl (in preprocess_ouput directory)

1. preprocess_ouput/ contains:
       - Sequences that passed quality control sorted by bar codes in fasta format.
       - For each bar code clustering results and OTUs sequences at the threshold specified by the user.
2. index_[Clst]: diversity description for each bar code.
3. pooled_sample_SEQ_OTU: OTUs sequences seeds from all bar codes.
4. seqAll_pooledSample.fasta: sequences that passed quality control from all bar codes.

---- Outputs of panam.pl (in panam_ouput directory)

1. Similarity annotation/: contains the results of the primary annotation (seq_results.uc, best_hit_uc) and the sequences sorted by profile (Sorted_Sequences/)
2. Alignment/: contains for each file in Sorted_Sequences/, an alignment in fasta format.
3. Phylogeny/: contains for each file in Alignment/ a phylogenetic tree and a rooted phylogenetic tree. You can use NJplot to view the trees (http://pbil.univ-lyon1.fr/software/njplot.html)
4. PANAM_Affiliation.txt: Phylogenetic annotation of each processed sequence.
5. PANAM_Clades.txt: Description of monophyletic groups with bootstrap value and the taxonomy of its closest reference sequence (nn) / node (lca). Description is limited to the sequences processed by PANAM.
6. seq_Taxonomic_distribution.txt: the number of sequences and OTUs for each bar code by taxonomic group.
7. index_Taxonomic_distribution.txt: diversity indices for each bar code by taxonomic group.
8. Taxonomic_distribution.txt: the number of sequences and OTUs and the schao index for each bar code by taxonomic group.
Taxonomic_distribution.txt is generated ONLY for sequences data that have passed through preprocess.pl - Please pay attention to the USER_OWN_FILE parameter in the panam.ini.

5

PANAM outputs on the test dataset:

PANAM_Affiliation.txt:

>Read: Sample1_M00987-47-000000000-A8JNB-1-1108-10133-22123-1   Bootstrap:1.000
Nearest neighbor: FN665733   Bacteria;Proteobacteria;Alphaproteobacteria;SAR11 clade;LD12
freshwater group;
Lowest node: Node_407 Bacteria;Proteobacteria;Alphaproteobacteria;SAR11 clade;LD12 freshwater group;

>Read: Sample2_M00987-47-000000000-A8JNB-1-2117-13590-12357-1   Bootstrap:0.940
Nearest neighbor: HQ330674   Archaea;Euryarchaea;Halobacteria;Halobacteriales;Miscellaneous
Euryarcheotic Group(MEG);
Lowest node: Node_8   Archaea;Euryarchaea;Halobacteria;Halobacteriales;Miscellaneous Euryarcheotic
Group(MEG);

>Read: Sample2_M00987-47-000000000-A8JNB-1-2118-18368-13420-1   Bootstrap:0.850
Nearest neighbor: HQ330674   Archaea;Euryarchaea;Halobacteria;Halobacteriales;Miscellaneous
Euryarcheotic Group(MEG);
Lowest node: Node_9   Archaea;Euryarchaea;Halobacteria;Halobacteriales;Miscellaneous Euryarcheotic
Group(MEG);

>Read: Sample3_M00987-47-000000000-A8JNB-1-2115-20839-9626-1   Bootstrap:0.900
Nearest neighbor: AB243796   Archaea;Euryarchaea;Halobacteria;Halobacteriales;Deep Sea Hydrothermal
Vent Gp 6(DHVEG-6);Candidatus Parvarchaeum;
Lowest node: Node_591 Archaea;Euryarchaea;Halobacteria;Halobacteriales;Deep Sea Hydrothermal Vent Gp
6(DHVEG-6);Candidatus Parvarchaeum;

>Read: Sample2_M00987-47-000000000-A8JNB-1-2110-16224-5687-1   Bootstrap:0.900
Nearest neighbor: AB243796   Archaea;Euryarchaea;Halobacteria;Halobacteriales;Deep Sea Hydrothermal
Vent Gp 6(DHVEG-6);Candidatus Parvarchaeum;
Lowest node: Node_591 Archaea;Euryarchaea;Halobacteria;Halobacteriales;Deep Sea Hydrothermal Vent Gp
6(DHVEG-6);Candidatus Parvarchaeum;

PANAM_Clades.txt:

> AM997480   Bacteria;TM6;unclassified TM6;
          Phylogeny: Acidobacteria_cp_bacteria_1_rooted.newick   Bootstrap:0.430

      8 neighbors: Sample2_M00987-47-000000000-A8JNB-1-2108-10170-8228-1, Sample2_M00987-47-
000000000-A8JNB-1-1106-19940-13063-1, Sample2_M00987-47-000000000-A8JNB-1-1109-25619-16084-1,
Sample4_M00987-47-000000000-A8JNB-1-2104-9945-15024-1, Sample2_M00987-47-000000000-A8JNB-1-2103-
20714-17147-1, Sample3_M00987-47-000000000-A8JNB-1-2104-18039-9354-1, Sample2_M00987-47-000000000-
A8JNB-1-1118-24747-18979-1, Sample2_M00987-47-000000000-A8JNB-1-1114-16903-13181-1,

> HQ330674   Archaea;Euryarchaea;Halobacteria;Halobacteriales;Miscellaneous Euryarcheotic
Group(MEG);
          Phylogeny: Euryarchaea_archaea_1_rooted.newick   Bootstrap:0.940

      6 neighbors: Sample4_M00987-47-000000000-A8JNB-1-1103-13504-20105-1, Sample4_M00987-47-
000000000-A8JNB-1-2119-13536-20523-1, Sample2_M00987-47-000000000-A8JNB-1-1103-8076-10778-1,
Sample2_M00987-47-000000000-A8JNB-1-2108-20485-11543-1, Sample1_M00987-47-000000000-A8JNB-1-1116-
22967-20420-1, Sample2_M00987-47-000000000-A8JNB-1-1114-10665-5945-1,

> FJ269337   Bacteria;Acidobacteria;Acidobacteria;Subgroup 11;
          Phylogeny: Acidobacteria_cp_bacteria_1_rooted.newick   Bootstrap:0.970

      3 neighbors: Sample2_M00987-47-000000000-A8JNB-1-1117-18768-9825-1, Sample2_M00987-47-
000000000-A8JNB-1-2104-20220-19709-1, Sample2_M00987-47-000000000-A8JNB-1-1111-13673-4400-1,

> EF492913   Bacteria;Acidobacteria;Acidobacteria;Subgroup 25;
          Phylogeny: Acidobacteria_cp_bacteria_1_rooted.newick   Bootstrap:0.990

      5 neighbors: Sample2_M00987-47-000000000-A8JNB-1-1103-15712-5767-1, Sample3_M00987-47-
000000000-A8JNB-1-2102-18638-18574-1, Sample2_M00987-47-000000000-A8JNB-1-2116-14531-9551-1,
Sample1_M00987-47-000000000-A8JNB-1-1111-25438-17687-1, Sample3_M00987-47-000000000-A8JNB-1-2117-
23365-5983-1,

6

Taxononomic_distribution.txt

		Sample2					Sample1					Sample3					Sample4				
		#seq	#OTUs	Schao1	Shannon	Coverage	#seq	#OTUs	Schao1	Shannon	Coverage	#seq	#OTUs	Schao1	Shannon	Coverage	#seq	#OTUs	Schao1	Shannon	Coverage
Acidobacteria		65	41	146.6	3.33	49.23	38	6	12	0.81	89.47	3	3	6	1.1	0	7	6	11	1.75	28.57
	Acidobacteria	54	32	92	3.04	53.7	38	6	12	0.81	89.47	3	3	6	1.1	0	7	6	11	1.75	28.57
	Holophagae	8	6	16	1.67	37.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Subgroup 22	3	3	6	1.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actinobacteria		282	63	251	2.56	82.98	1173	56	196.25	1.89	97.1	678	27	40	1.11	97.94	437	47	93.43	2.24	94.05
	Acidimicrobiia	7	6	11	1.75	28.57	39	9	15	1.8	89.74	7	4	4.5	1.28	71.43	15	6	7	1.41	80
	Actinobacteria	243	35	197.5	2	89.3	1113	41	122.25	1.66	97.66	670	22	31.17	1.04	98.36	406	30	52.75	1.94	96.55
	Coriobacteriia	9	9	45	2.2	0	1	1	1	0	0	0	0	0	0	0	8	7	14.5	1.91	25
	Thermoleophilia	23	13	20	2.34	65.22	20	5	8	1.05	85	1	1	1	0	0	8	4	7	1.07	62.5
Armatimonadetes		4	4	10	1.39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Armatimonadia	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	unclassified Armatimonadetes	3	3	6	1.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BD1-5		2	2	3	0.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	unclassified BD1-5	2	2	3	0.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bacteroidetes		4172	222	583.8	2.42	96.76	2837	111	248.75	2.36	97.96	1679	64	181.17	2.35	97.74	1655	119	231.89	2.34	96.01
	BD2-2	3	3	6	1.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	BSV13	1	1	1	0	0	2	2	3	0.69	0	0	0	0	0	0	2	1	1	0	100
	Bacteroidia	56	27	90.33	2.8	64.29	2	2	3	0.69	0	2	2	3	0.69	0	14	9	16.5	2.07	57.14
	Cytophagia	850	23	53	0.41	98.12	23	4	4	0.84	95.65	41	1	1	0	100	137	15	26.25	0.78	92.7
	Flavobacteria	2871	73	139.6	1.6	98.71	456	28	58.33	2.16	96.93	1044	28	63	1.49	98.56	1054	41	60.43	1.26	98.39
	SB-1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0
	Sphingobacteriia	385	92	232	3.52	85.45	2352	73	147	1.81	98.43	592	33	85.5	1.87	96.45	447	52	114	2.04	92.84
	WCHB1-32	4	1	1	0	100	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
	vadinHA17	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Candidate division JS1		3	1	1	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	unclassified Candidate division JS1	3	1	1	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Candidate division OD1		24	23	138.5	3.12	8.33	7	7	28	1.95	0	0	0	0	0	0	7	6	11	1.75	28.57
	unclassified Candidate division OD1	24	23	138.5	3.12	8.33	7	7	28	1.95	0	0	0	0	0	0	7	6	11	1.75	28.57
Candidate division OP11		9	9	45	2.2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
	unclassified Candidate division OP11	9	9	45	2.2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
Candidate division SR1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	3	0.69	0
	unclassified Candidate division SR1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	3	0.69	0
Candidate division TM7		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	3	0.69	0
	unclassified Candidate division TM7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	3	0.69	0
Candidate division WS3		1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
	unclassified Candidate division WS3	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Chlamydiae		2	2	3	0.69	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
	Chlamydiae	2	2	3	0.69	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
Chlorobi		6	6	21	1.79	0	4	3	3.5	1.04	50	0	0	0	0	0	0	0	0	0	0
	Chlorobia	4	4	10	1.39	0	4	3	3.5	1.04	50	0	0	0	0	0	0	0	0	0	0
	Ignavibacteria	2	2	3	0.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Clade_Results

Clade	Bootstrap p	#OTUs	#Sequences	#Singletons	MNND	depth_deepest	Profile	NN	Taxonomy
clade_1	0.86	4	4	4	0.4278	0.2924	Chloroflexi_bacteria.1_rooted.newick	AY193295	Bacteria;Candidate division OP11;unclassified Candidate division OP11;
clade_2	0.78	3	94	2	0.0738	0.0769	Chloroflexi_bacteria.1_rooted.newick	DQ814518	Bacteria;Chloroflexi;KD4-96;
clade_3	1	4	26	1	0.087	0.1821	Chloroflexi_bacteria.1_rooted.newick	EU703171	Bacteria;Chloroflexi;Chloroflexia;Chloroflexales;Roseiflexaceae;Roseiflexus;
clade_4	0.99	3	29	2	0.1835	0.1197	Chloroflexi_bacteria.1_rooted.newick	HQ828001	Bacteria;Chloroflexi;Caldilineae;Caldilineales;Caldilineaceae;uncultured;
clade_5	1	4	7	1	0.1127	0.0845	Firmicutes_cp_bacteria.1_rooted.newick	AB198443	Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Incertae Sedis;
clade_6	0.97	15	15	15	0.1503	0.1099	Proteobacteria_cp_bacteria.1_rooted.newick	AY592599	Bacteria;Actinobacteria;Coriobacteriia;Coriobacteriales;Coriobacteriaceae;uncultured;
clade_7	0.73	3	8	1	0.1297	0.0769	Proteobacteria_cp_bacteria.1_rooted.newick	JF429120	Bacteria;Proteobacteria;Gammaproteobacteria;NKB5;
clade_8	0.69	3	41	0	0.1177	0.0714	Proteobacteria_cp_bacteria.1_rooted.newick	FN668031	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Paucibacter;
clade_9	0.9	9	138	5	0.0981	0.0769	Proteobacteria_cp_bacteria.1_rooted.newick	AY515390	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Paucibacter;
clade_10	0.86	3	4	2	0.3709	0.2313	Proteobacteria_cp_bacteria.1_rooted.newick	FJ155003	Bacteria;Proteobacteria;Deltaproteobacteria;Bdellovibrionales;Bacteriovoracaceae;uncultured;
clade_11	0.32	3	6	1	0.1131	0.0663	Proteobacteria_cp_bacteria.1_rooted.newick	EF659431	Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Azoarcus;
clade_12	0.9	7	761	2	0.1276	0.0943	Proteobacteria_cp_bacteria.1_rooted.newick	FJ849216	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Polaromonas;
clade_13	0.94	3	5	1	0.1043	0.073	Proteobacteria_cp_bacteria.1_rooted.newick	JF177167	Bacteria;Proteobacteria;Gammaproteobacteria;Legionellales;Coxiellaceae;Rickettsiella;
clade_14	0.89	6	70	0	0.0989	0.0614	Proteobacteria_cp_bacteria.1_rooted.newick	DQ337069	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus;
clade_15	0.9	3	14	0	0.0912	0.0633	Proteobacteria_cp_bacteria.1_rooted.newick	EF540406	Bacteria;Proteobacteria;Gammaproteobacteria;Legionellales;Coxiellaceae;Aquicella;
clade_16	0.94	6	16	3	0.1204	0.094	Proteobacteria_cp_bacteria.1_rooted.newick	JF715446	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Massilia;
clade_17	0.23	4	80	1	0.0868	0.0648	Proteobacteria_cp_bacteria.1_rooted.newick	AY131213	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Alcaligenaceae;Kerstersia;
clade_18	0.1	4	18	2	0.0874	0.0613	Proteobacteria_cp_bacteria.1_rooted.newick	GQ472360	Bacteria;Proteobacteria;Betaproteobacteria;Hydrogenophilales;Hydrogenophilaceae;Thiobacillus;
clade_19	0.82	5	5	5	0.2207	0.1815	Proteobacteria_cp_bacteria.1_rooted.newick	EU801038	Bacteria;Actinobacteria;Actinobacteria;Frankiales;Sporichthyaceae;hgcl clade;
clade_20	0.88	4	12	2	0.129	0.1049	Proteobacteria_cp_bacteria.1_rooted.newick	AY360530	Bacteria;Proteobacteria;Betaproteobacteria;Methylophilales;Methylophilaceae;uncultured;
clade_21	0.75	4	4	4	0.2885	0.2335	Proteobacteria_cp_bacteria.1_rooted.newick	JF168221	Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Candidatus Arthromitus;
clade_22	0.86	19	240	9	0.1159	0.0804	Proteobacteria_cp_bacteria.1_rooted.newick	EU800923	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Limnohabitans;
clade_23	0.89	5	13	3	0.1149	0.0746	Proteobacteria_cp_bacteria.1_rooted.newick	FR667462	Bacteria;Proteobacteria;Betaproteobacteria;Methylophilales;Methylophilaceae;uncultured;
clade_24	0.68	3	8	2	0.1125	0.0671	Proteobacteria_cp_bacteria.1_rooted.newick	JX898172	Bacteria;Proteobacteria;Betaproteobacteria;SC-I-84;
clade_25	0.66	3	4	2	0.1822	0.1298	Proteobacteria_cp_bacteria.1_rooted.newick	FN552695	Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Family Incertae Sedis;Candidatus Anadelfobacter;
clade_26	0.02	4	6	2	0.0896	0.0694	Proteobacteria_cp_bacteria.1_rooted.newick	AF448514	Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Cellvibrio;
clade_27	0.42	8	10	7	0.1472	0.1033	Proteobacteria_cp_bacteria.1_rooted.newick	HM481303	Bacteria;Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;uncultured;
clade_28	0.17	4	4	4	0.1531	0.0956	Proteobacteria_cp_bacteria.1_rooted.newick	EU409117	Bacteria;Proteobacteria;Gammaproteobacteria;Legionellales;Legionellaceae;Legionella;
clade_29	0.93	3	4	2	0.1718	0.1214	Proteobacteria_cp_bacteria.1_rooted.newick	EF520565	Bacteria;Proteobacteria;Gammaproteobacteria;Legionellales;Legionellaceae;Legionella;
clade_30	0.37	4	7	3	0.1789	0.1205	Proteobacteria_cp_bacteria.1_rooted.newick	FJ833268	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Alcaligenaceae;Sutterella;
clade_31	0.58	3	4	2	0.0878	0.0579	Proteobacteria_cp_bacteria.1_rooted.newick	JX105586	Bacteria;Proteobacteria;Gammaproteobacteria;Legionellales;Coxiellaceae;Aquicella;