Amplicons: Lesson 1 Introduction

Goals

- What are amplicons?
- What are the different pipelines for processing amplicon data?
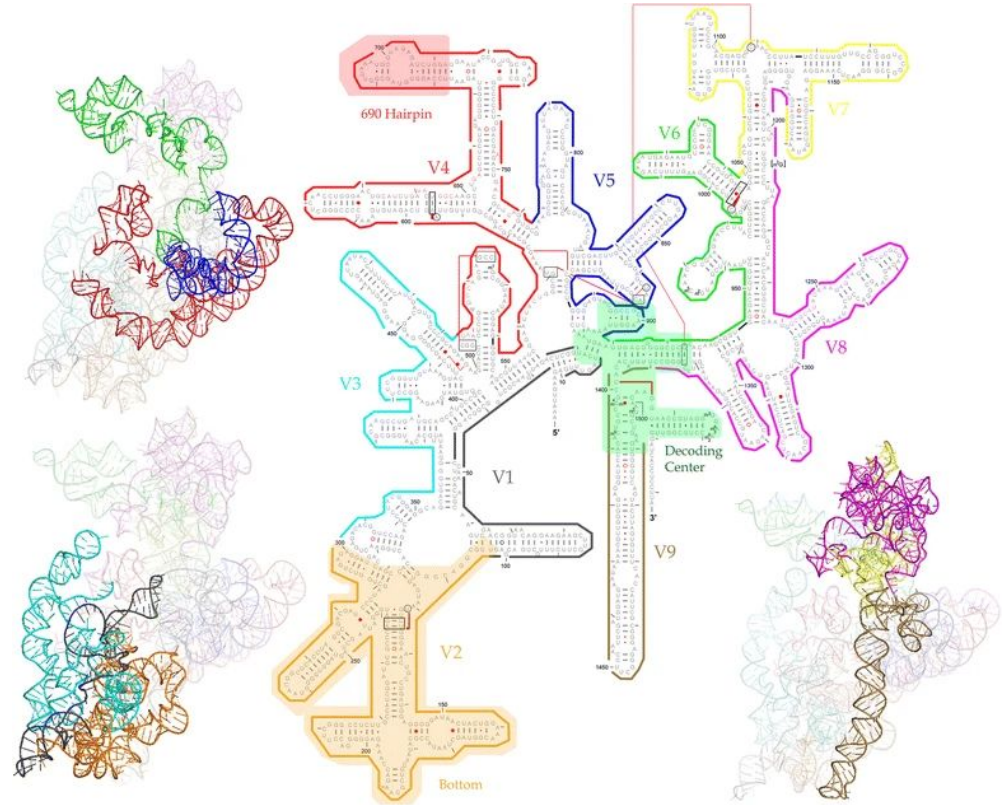- What are the available methods for producing count tables?

SSU rRNA

- 12S- mitochondria
- 16S- prokaryotes
- 18S- eukaryotes

Also used as markers…

- ITS
- MT- CYB

Yang B, et al. 2016. *Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis*. BMC Bioinformatics. 17(1):135. doi:10.1186/s12859-016-0992-y.
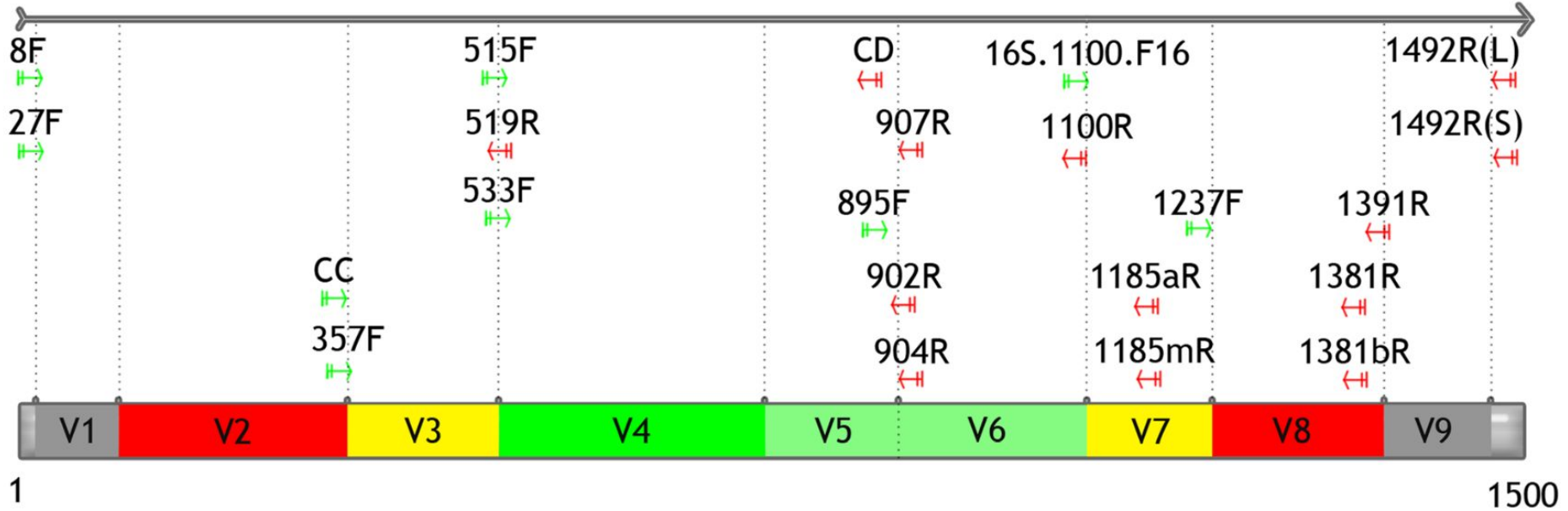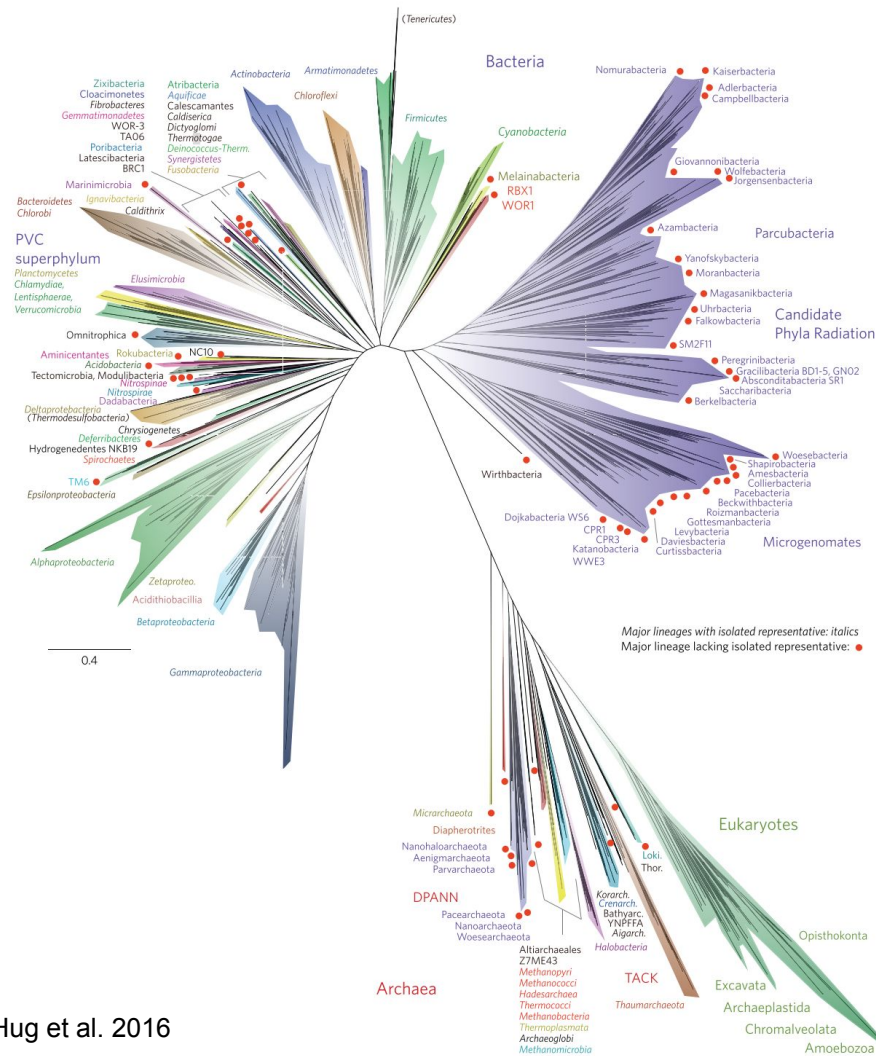
Fig. 3 Illustration of different variable regions. Red regions (V2, V8) have a poor phylogenetic resolution at the phylum level. Green regions (V4, V5, V6) are associated with the shortest geodesic distance, which suggests that they may be the best choice for phylogeny-related analyses and the phylogenetic analysis of novel bacterial phyla. The figure refers to the primer map from Lutzonilab (http://lutzonilab.org/16s-ribosomal-dna/). Use of this information was approved by the original authors of the website

Yang B, et al. 2016. *Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis*. BMC Bioinformatics. 17(1):135. doi:10.1186/s12859-016-0992-y.

Hug et al. 2016

Overall Workflow for Microbiome Analyses
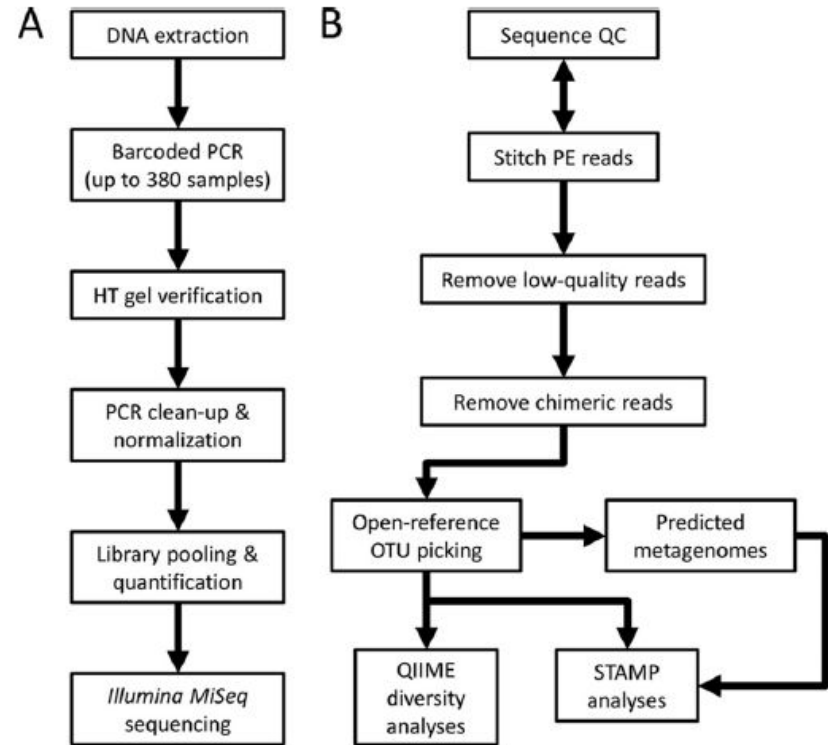
Sample Collection →

DNA extraction →

PCR Amplification →
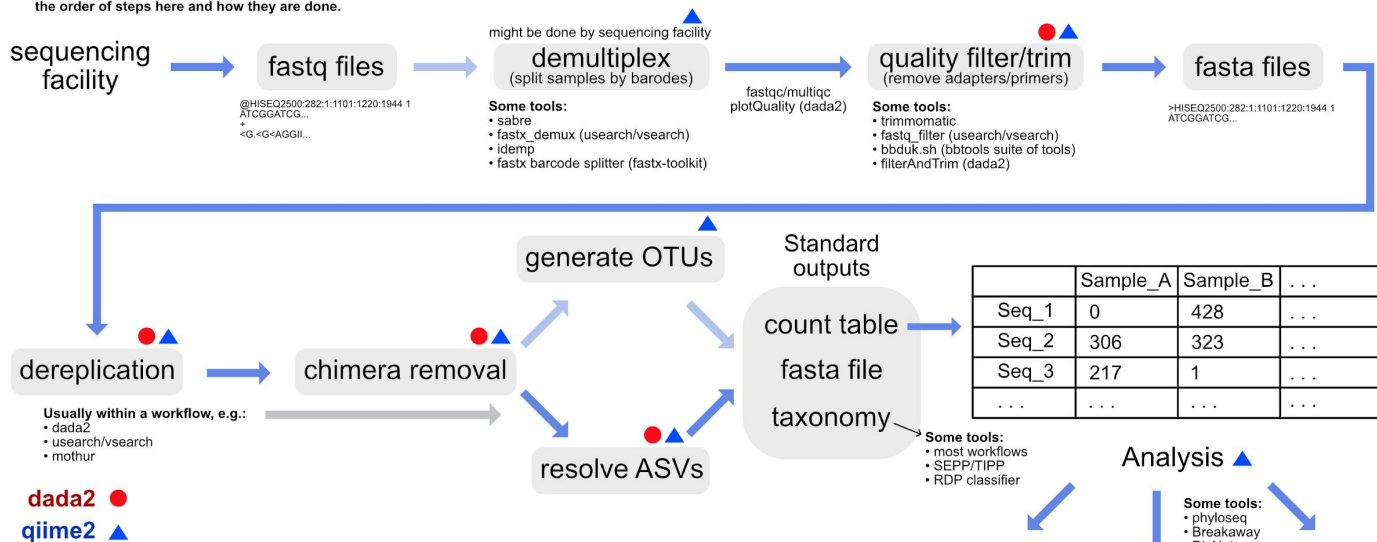
Sequencing →

Bioinformatics →

Statistical Analyses

Overview of generic* amplicon workflow

*This is generic; specific workflows can vary on the order of steps here and how they are done.

When working with your own data you should never follow any pipeline blindly. There can be critical differences based on your data.

sequencing facility → fastq files → demultiplex (split samples by barodes) → quality filter/trim (remove adapters/primers) → fasta files

@HISEQ2500:282:1:1101:1220:1944 1
ATCGGATCG...
+
<G.<G<AGGII...

might be done by sequencing facility

Some tools:
• sabre
• fastx_demux (usearch/vsearch)
• idemp
• fastx barcode splitter (fastx-toolkit)

fastqc/multiqc
plotQuality (dada2)

Some tools:
• trimmomatic
• fastq_filter (usearch/vsearch)
• bbduk.sh (bbtools suite of tools)
• filterAndTrim (dada2)

>HISEQ2500:282:1:1101:1220:1944 1
ATCGGATCG...

dereplication → chimera removal → generate OTUs

Standard outputs

count table
fasta file
taxonomy

Usually within a workflow, e.g.:
• dada2
• usearch/vsearch
• mothur

resolve ASVs

dada2 ●
qiime2 ▲

Some tools:
• most workflows
• SEPP/TIPP
• RDP classifier

	Sample_A	Sample_B	. . .
Seq_1	0	428	. . .
Seq_2	306	323	. . .
Seq_3	217	1	. . .
. . .	. . .	. . .	. . .

Analysis ▲

Some tools:
• phyloseq
• Breakaway
• DivNet
• CORNCOB
• SpiecEasi
• DESeq2
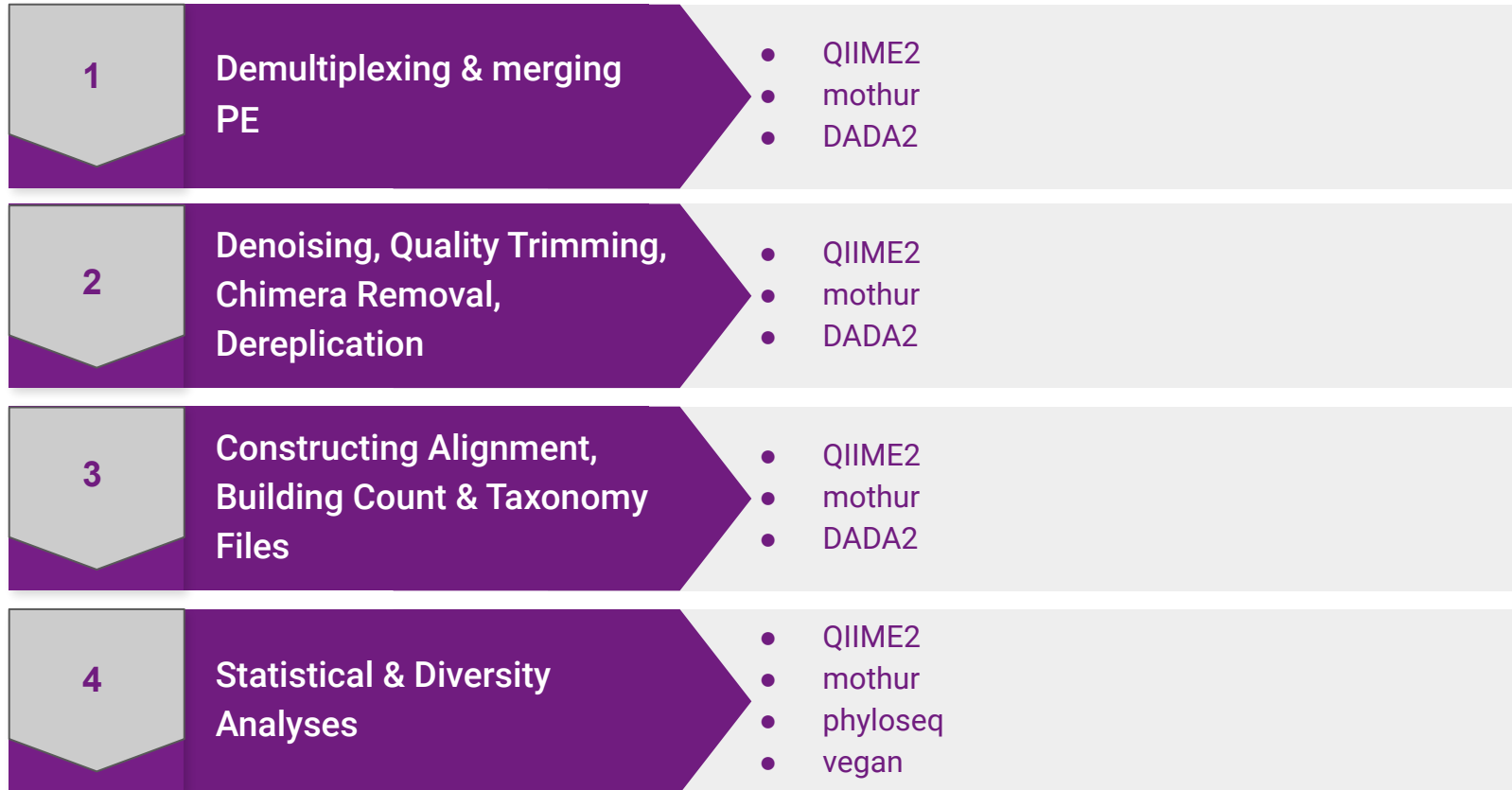
Some tools that provide whole workflows:

dada2 runs within R (ASVs)

usearch/vsearch runs at the command line (ASVs and OTUs)

mothur runs at the command line (OTUs only currently)

qiime2 provides a multi-interface environment that employs processing tools like those above, infrastructure for easily documenting all processing performed, and interactive visualizations

Beta diversity

e.g. dissimilarity metrics
ordination
hierarchical clustering

Alpha diversity

e.g. richness
evenness
diversity

Taxonomic summaries

astrobiomike.github.io

https://astrobiomike.github.io/amplicon/

Bioinformatics Workflow for Amplicons

1 — **Demultiplexing & merging PE**
- QIIME2
- mothur
- DADA2

2 — **Denoising, Quality Trimming, Chimera Removal, Dereplication**
- QIIME2
- mothur
- DADA2

3 — **Constructing Alignment, Building Count & Taxonomy Files**
- QIIME2
- mothur
- DADA2

4 — **Statistical & Diversity Analyses**
- QIIME2
- mothur
- phyloseq
- vegan

Pipeline/ Approach	Options for Interface	Tutorials & Links
QIIME2	- Command line (locally or on virtual machine) - GUI: qstudio - GUI: DNA Subway Purple Line (Cyverse)	https://qiime2.org/ https://dnasubway.cyverse.org/
mothur	- Command line - GUI?	https://mothur.org/ https://galaxyproject.github.io/training-material/topics/metagenomics/tutorials/mothur-miseq-sop/tutorial.html
DADA2	- R - GUI: DNA Subway Purple Line (Cyverse)	https://benjjneb.github.io/dada2/index.html https://dnasubway.cyverse.org/
phyloseq	R	https://joey711.github.io/phyloseq/
vegan	R	https://cran.r-project.org/web/packages/vegan/vignettes/intro-vegan.pdf https://peat-clark.github.io/BIO381/veganTutorial.html

Reference Databases

Database	Target Amplicon	Link
SILVA	16S & 18S rRNA	https://www.arb-silva.de/
RDP	16S rRNA & 28S rRNA (fungi)	https://rdp.cme.msu.edu/
PR2	18S rRNA	https://pr2-database.org/
Greengenes	16S rRNA (*no longer maintained*)	https://greengenes.secondgenome.com/
UNITE	Fungal ITS	https://unite.ut.ee/

Example Count Table

#OTU ID	C1	W2	C2	S1	A1	K1	G2	I2	A2	Q2	V2	X2	U1	T1	Z2	Y2	W1	G1
e39494332a51f284d0a5b18f77297277	15222	18891	10089	52	113	1755	1795	2098	93	1173	3433	0	18501	2057	3270	7147	18645	2187
d6c3280e0b453e32bb0f5848a8adf5ca	11297	8753	8457	73	2932	3741	3187	857	6345	3552	1478	3021	3566	5826	5671	3883	8989	3728
9404163724571a398aa4869436759308	7206	6119	5933	16	16	33	172	1598	26	77	3985	2621	3530	66	69	676	6696	249
028058ffc2ff7988449b2eda71d267db	7077	2987	5237	0	17	1375	1158	1139	27	870	2720	3660	3085	2779	3393	1767	3194	1379
f22ea8603e4c6e7ed0313dfb71acdb1b	5521	1010	4649	54	20	3103	5345	0	23	694	0	22	2274	3305	2428	4508	1126	5934
b3f916eb895596b0aab5ce67c710cf3d	4586	3420	3310	22	37	383	2483	774	55	759	1160	2120	3413	697	904	2280	3720	3838
d806a7c6845f6c50f890e0b11fe652de	4406	47	3667	17	102	355	664	67	224	32	94	70	335	402	1316	1272	57	837
13239bfd601be5658be2653f6caf8aa3	4153	2440	3353	18	332	2760	1733	57	710	916	50	645	1066	3623	1740	1211	2861	2039
f762fc06a47ae561a60323d8bd6f405a	4091	1675	3241	8	235	810	1319	199	702	779	193	1303	441	1407	1065	1324	1698	1605
f637ea7b3ef52b3415d209c09447f941	3630	3449	2740	19	8	1237	1016	603	38	481	787	332	2301	1932	2533	2588	3234	1512
96714508a6fea480041ace7d01bf2048	3352	2584	2591	13	433	1086	3422	590	1091	895	967	2615	413	1523	1251	1442	2664	3881
ed59e9f608568c35f72e987ca3189dfc	2966	5304	2436	0	58	495	1544	419	148	770	836	6896	3907	661	769	1484	5672	1926
6fd41b87609aa26d165c06c377fe1af2	2609	1501	2111	10	3	853	85	8	2	40	7	0	2124	936	1919	980	1460	91
83b10ccbf2734f44be635405d4afab78	2525	1287	1802	5	4	49	780	1057	3	26	1372	628	552	33	136	915	1258	799
f1cc90a34338cd248d6cde6ff20f6a20	1981	378	1487	158	25	7	5733	66	95	427	80	4476	570	21	6	226	384	7220
a133a24b4a69605f6866addd84018039	1870	771	1361	16	3	1313	256	153	6	148	242	89	965	2183	3071	861	800	352
711cdee9b83a816ef327fea01338957c	1765	597	1445	5	17	47	2695	28	20	750	33	50	548	58	64	1043	594	3006
6ea794aa47071a501015fbad65dd1575	1594	151	1333	0	12	214	39	37	23	73	71	61	240	554	1455	438	131	44
84a7b9b6e6548f9624d2d758e5030b18a	1461	2112	1204	5	3	6	8	0	4	1	0	0	908	1	0	303	2161	12
c7afbd062dae23491d83984fe4386c9c	1434	204	1009	323	468	3255	326	166	1124	563	133	308	431	5420	3853	661	211	449
6f1ef5652fec7baf3341adf218299edb	1394	898	1127	16	76	721	1728	119	130	1359	177	176	1387	1123	1054	689	882	1885
98b4799119283d1389a0151e222832cd	1229	586	1009	62	13	14843	2593	668	38	363	728	87	622	11743	4424	1103	534	2685
f4abd32ff2c89cbbf103dd8a13d63d7c	1203	91	890	3	2	76	686	7	2	15	0	0	256	50	171	530	137	836
91247b8e53a048da3d777777284d8dad	900	182	761	78	13	7010	4085	2037	35	685	2841	156	432	5231	6511	773	167	4574
fbbc649b1c6ec671b3ad2248ffc21563	866	50	796	20	308	2900	1003	54	717	6298	8	264	927	1684	1248	1475	46	1204
e9152853e1d8630f9e0a26870da742a9	831	279	639	0	0	0	0	20	0	0	45	113	112	0	0	476	313	0
5fbd67cda1e78321177bd4c3762d720f	817	189	705	0	0	6	109	0	0	14	152	81	1127	0	0	273	181	2
ec884156bfdaa8333e0a800b35a71d54	796	344	520	7	6	996	137	37	46	366	52	93	566	1825	802	198	378	113
e66893ec9c61f9999656e619c08ae986	781	158	677	6	0	447	19	24	2	793	23	43	15	391	372	344	155	24
b9dc48f4d098a8df739a779de4f256941	769	10	558	567	449	6	4083	53	1295	1983	8	2929	273	20	2	269	67	5023
bb1aede821f34ae2ff572dd104ab9edc	708	219	495	8	2	80	979	16	0	417	37	110	493	130	151	338	190	1173
02ee69101655c2c379be21482892c23c	701	0	674	2	0	0	7	0	4	45	0	3	73	0	2	1124	2	15
352802c9d8be8d8759fb43d289b818f6	674	0	600	0	0	0	0	0	0	4	3	39	114	0	0	700	0	16
df4ae84d1e62e9ba7a0a5fb518d5b126	630	441	460	469	14	1947	189	1060	15	12	1682	504	89	2536	150	140	450	201
132244e7d03cc5f4e29a59d9440488e7	597	67	416	0	9	306	829	17	27	209	24	461	182	416	638	213	75	959
95b599ea0ddc5a65f29d8e807785011c	591	1441	461	6	54	1480	32	309	99	131	867	2304	134	2667	3264	301	1516	17
a5e9e3253c3a817795de811b3b7876ad	513	14	430	0	0	307	223	35	1	145	84	417	304	665	441	268	65	231
a43f1cb4098183bc58828a5f288249e0	490	57	442	9	305	0	5081	17	645	1997	0	0	205	0	0	263	66	5528
fa54247f3b7a2194cf8f626d65d4702a	460	11	259	228	826	0	6330	62	1775	3203	0	2854	160	3	0	127	11	7699
4559f6a9fd835b9be8f0739368d8f53f	445	8	403	12	548	1094	1074	18	1433	1073	3	350	416	584	396	640	12	1376

Example Taxonomy File

	Feature ID	Taxon	Confidence
1	Feature ID	Taxon	Confidence
2	e39494332a51f284d0a5b18f77297277	D_0__Bacteria;D_1__Actinobacteria;D_2__Actinobacteria;D_3__Frankiales;D_4__Sporichthyaceae;D_5__hgcI clade	0.9999917
3	d6c3280e0b453e32bb0f5848a8adf5ca	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Betaproteobacteriales;D_4__Burkholderiaceae;D_5__RS62 marine group;D_6__uncultured bacterium	0.93596318
4	9404163724571a398aa4869436759308	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__SAR11 clade;D_4__Clade III	1
5	028058ffc2ff7988449b2eda71d267db	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Betaproteobacteriales;D_4__Methylophilaceae;D_5__OM43 clade	0.97353215
6	f22ea8603e4c6e7ed0313dfb71acdb1b	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Synechococcales;D_4__Cyanobiaceae;D_5__Cyanobium PCC-6307;D_6__uncultured cyanobacterium	0.9357934
7	b3f916eb895596b0aab5ce67c710cf3d	D_0__Bacteria;D_1__Actinobacteria;D_2__Acidimicrobiia;D_3__Microtrichales;D_4__Ilumatobacteraceae;D_5__CL500-29 marine group	1
8	d806a7c6845f6c50f890e0b11fe652de	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Flavobacteriales;D_4__Crocinitomicaceae;D_5__Fluviicola;D_6__uncultured Flexibacter sp.	0.92119022
9	13239bfd601be5658be2653f6caf8aa3	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Cytophagales;D_4__Cyclobacteriaceae;D_5__Algoriphagus;D_6__uncultured Hongiella sp.	0.77940565
10	f762fc06a47ae561a60323d8bd6f405a	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Flavobacteriales;D_4__Flavobacteriaceae;D_5__NS3a marine group;D_6__marine metagenome	0.74257502
11	f637ea7b3ef52b3415d209c09447f941	D_0__Bacteria;D_1__Actinobacteria;D_2__Acidimicrobiia;D_3__Microtrichales;D_4__Ilumatobacteraceae;D_5__CL500-29 marine group;D_6__uncultured bacterium	0.99939711
12	96714508a6fea480041ace7d01bf2048	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Flavobacteriales;D_4__Cryomorphaceae;D_5__uncultured;D_6__uncultured bacterium	0.94924388
13	ed59e9f608568c35f72e987ca3189dfc	D_0__Bacteria;D_1__Actinobacteria;D_2__Actinobacteria;D_3__Frankiales;D_4__Sporichthyaceae;D_5__hgcI clade	0.99995253
14	6fd41b87609aa26d165c06c377fe1af2	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Chitinophagales;D_4__Chitinophagaceae;D_5__Terrimonas	0.998042
15	83b10ccbf2734f44be635405d4afab78	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Sphingobacteriales;D_4__NS11-12 marine group;D_5__uncultured marine bacterium;D_6__uncultured marine bacterium	0.99814135
16	f1cc90a34338cd248d6cde6ff20f6a20	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__SAR11 clade;D_4__Clade III;D_5__uncultured bacterium;D_6__uncultured bacterium	0.99999703
17	a133a24b4a69605f6866addd84018039	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Gammaproteobacteria Incertae Sedis;D_4__Unknown Family;D_5__Acidibacter;D_6__uncultured marine bacterium	0.9639967
18	711cdee9b83a816ef327fea01338957c	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Gammaproteobacteria Incertae Sedis;D_4__Unknown Family;D_5__Acidibacter	0.72892947
19	6ea794aa47071a501015fbad65dd1575	D_0__Bacteria;D_1__Tenericutes;D_2__Mollicutes;D_3__Acholeplasmatales;D_4__Acholeplasmataceae;D_5__Acholeplasma;D_6__uncultured marine bacterium	0.96739961
20	84a79b9e6548f9624d2d758e5030b18a	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodospirillales;D_4__AEGEAN-169 marine group	1
21	c7afbd062dae23491d83984fe4386c9c	D_0__Bacteria;D_1__Actinobacteria;D_2__Actinobacteria;D_3__Micrococcales;D_4__Microbacteriaceae;D_5__Candidatus Aquiluna	0.99421108
22	6f1ef5652fec7baf3341adf218299edb	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Betaproteobacteriales;D_4__Burkholderiaceae	0.999984
23	98b4799119283d1389a0151e222832cd	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Synechococcales;D_4__Cyanobiaceae;D_5__Cyanobium PCC-6307;D_6__uncultured marine bacterium	0.7685911
24	f4abd32ff2c89cbbf103dd8a13d63d7c	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Sphingobacteriales;D_4__NS11-12 marine group;D_5__uncultured bacterium;D_6__uncultured bacterium	0.9995189
25	91247b8e53a048da3d777777284d8dad	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Synechococcales;D_4__Cyanobiaceae;D_5__Cyanobium PCC-6307;D_6__uncultured cyanobacterium	0.95864016
26	fbbc649b1c6ec671b3ad2248ffc21563	D_0__Bacteria;D_1__Planctomycetes;D_2__OM190;D_3__uncultured bacterium;D_4__uncultured bacterium;D_5__uncultured bacterium;D_6__uncultured bacterium	0.99323341
27	e9152853e1d8630f9e0a26870da742a9	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rhodobacterales;D_4__Rhodobacteraceae	0.9995242
28	5fbd67cda1e78321177bd4c3762d720f	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Flavobacteriales;D_4__Flavobacteriaceae;D_5__Aquibacter;D_6__uncultured bacterium	0.95083556
29	ec884156bfdaa8333e0a800b35a71d54	D_0__Bacteria;D_1__Verrucomicrobia;D_2__Verrucomicrobiae;D_3__Opitutales;D_4__Opitutaceae;D_5__Cephaloticoccus;D_6__uncultured bacterium	0.99766947
30	e66893ec9c61f9999656e619c08ae986	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Flavobacteriales;D_4__Crocinitomicaceae;D_5__Fluviicola	0.99925229
31	b9dc48f4d98a8df739a779de4f256941	D_0__Bacteria;D_1__Actinobacteria;D_2__Actinobacteria;D_3__Frankiales;D_4__Sporichthyaceae;D_5__hgcI clade;D_6__uncultured Candidatus Planktophila sp.	0.72125853
32	bb1aede821f34ae2ff572dd104ab9edc	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Betaproteobacteriales;D_4__TRA3-20	0.99995542
33	02ee69101655c2c379be21482892c23c	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Chloroplast;D_4__Cymbomonas tetramitiformis;D_5__Cymbomonas tetramitiformis;D_6__Cymbomonas tetramitiformis	0.99044829
34	352802c9d8be8d8759fb43d289b818f6	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Chloroplast;D_4__Hemiselmis virescens;D_5__Hemiselmis virescens;D_6__Hemiselmis virescens	0.99908686
35	df4ae84d1e62e9ba7a05fb518d5b126	D_0__Bacteria;D_1__Epsilonbacteraeota;D_2__Campylobacteria;D_3__Campylobacteriales;D_4__Arcobacteraceae;D_5__Arcobacter;D_6__Arcobacter cryaerophilus	0.92352329
36	132244e7d03cc5f4e29a59d9440488e7	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Betaproteobacteriales;D_4__Burkholderiaceae	0.99999106
37	95b599ea0ddc5a65f29d8e807785011c	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Oceanospirillales;D_4__Pseudohongiellaceae;D_5__Pseudohongiella	0.9999962
38	a5e9e3253c3a817795de811b3b7876ad	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Oceanospirillales;D_4__Pseudohongiellaceae;D_5__Blyi10	0.77015288
39	a43f1cb4098183bc58828a5f288249e0	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Betaproteobacteriales;D_4__Burkholderiaceae;D_5__Curvibacter	0.8017833
40	fa54247f3b7a2194cf8f626d65d4702a	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Betaproteobacteriales;D_4__Methylophilaceae;D_5__Candidatus Methylopumilus;D_6__uncultured bacterium	0.88606964
41	4559f6a9fd835b9be8f0739368d8f53f	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Chloroplast;D_4__uncultured Cryptomonadaceae;D_5__uncultured Cryptomonadaceae;D_6__uncultured Cryptomonadaceae	0.93253631
42	8f32458565405e1b1289023dcc7a7a30	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Cytophagales;D_4__Microscillaceae;D_5__uncultured;D_6__marine metagenome	0.92948358
43	1669e6d156bac4e9fcbd417984f0fa4e	D_0__Bacteria;D_1__Planctomycetes;D_2__Planctomycetacia;D_3__Pirellulales;D_4__Pirellulaceae;D_5__uncultured;D_6__marine metagenome	0.78292091

Example Tree File
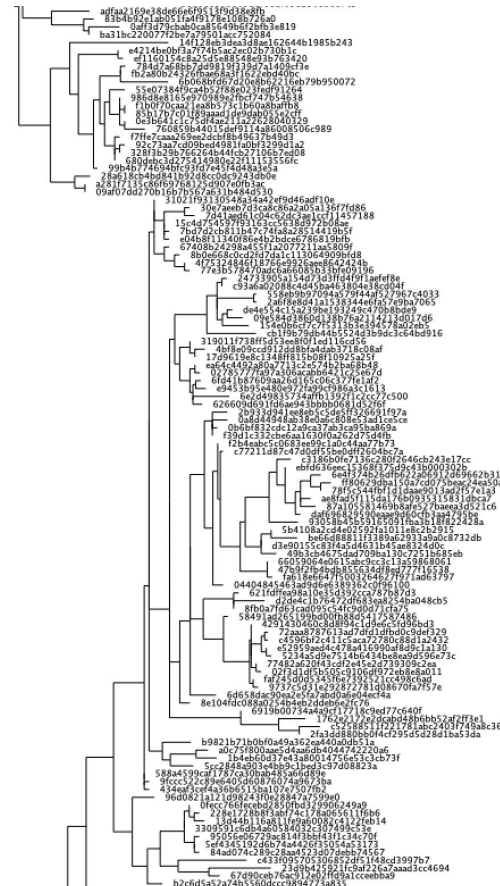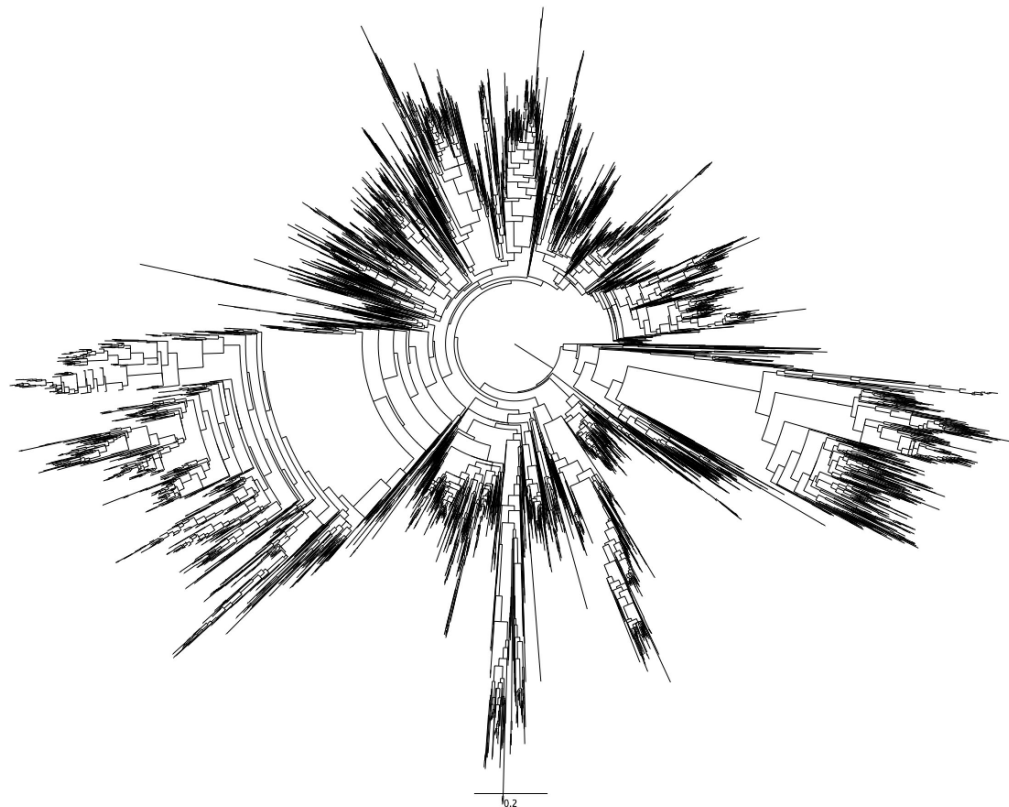
Upcoming

Caveats of amplicon analyses

Removing Contamination

Tutorials

Statistical & diversity analyses and plotting