

Introduction to metabarcoding

Antonino Malacrinò

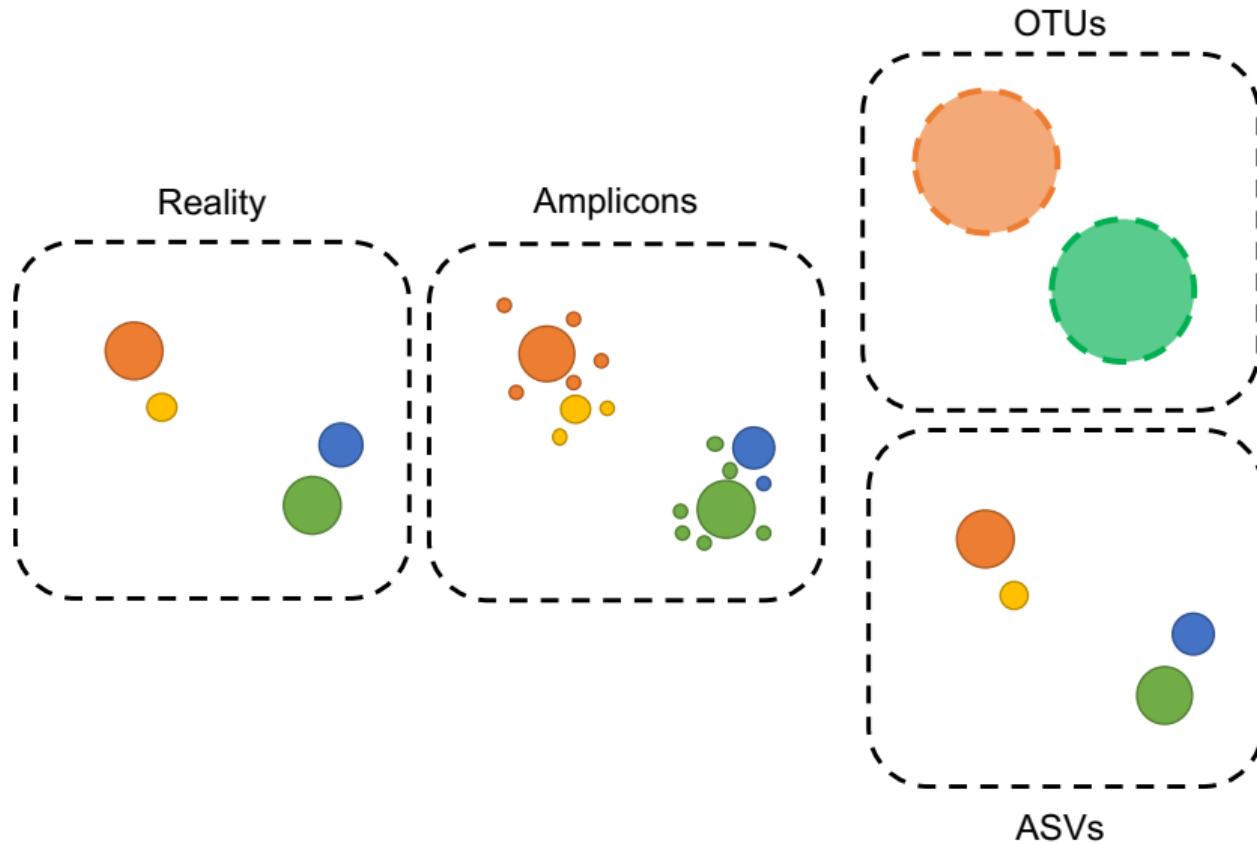
23 September 2021

Data processing

First choice!

- ▶ OTUs (Operational Taxonomic Unit)
- ▶ AVSs (Amplicon Sequence Variant)

OTUs vs ASVs



OTU approach

- ▶ De novo
- ▶ Closed-reference
- ▶ Open-reference
- ▶ Phylogeny-based placing

Hierarchical clustering (e.g. MOTHUR)

- ▶ Exact full-length dereplication of the reads
- ▶ Alignment
 - ▶ Multiple sequence alignment (denovo / reference)
 - ▶ Pairwise sequence alignment
- ▶ Calculate pairwise distance matrix
- ▶ Hierarchical clustering at fixed distance

Seed-based clustering (e.g. VSEARCH)

- ▶ Exact full-length dereplication of the reads
- ▶ Sort sequences by certain criteria (usually abundance)
- ▶ Top sequence in list (most abundant) is placed as first seed and subsequent query reads are aligned against seed
 - ▶ Scenario 1: query matches seed within threshold, abundance information update
 - ▶ Scenario 2: query does not match and is placed as new seed in database

ASVs approach

- ▶ Why binning sequences at all? Can we model PCR and sequencing errors sufficiently well to retrieve true sequence variants?
- ▶ Amplicon Sequence Variants (ASVs)
 - ▶ Needham et al. 2017
 - ▶ Callahan et al. 2017
- ▶ Exact Sequence Variants (ESVs)
 - ▶ Callahan et al. 2017
- ▶ sub-OTUs (sOTUs)
 - ▶ Amir et al. 2017
- ▶ Zero radius OTUs (zOTUs)
 - ▶ Edgar et al. 2017
- ▶ All the same!

ASVs approach

- ▶ Why binning sequences at all? Can we model PCR and sequencing errors sufficiently well to retrieve true sequence variants?
- ▶ Amplicon Sequence Variants (ASVs)
 - ▶ Needham et al. 2017
 - ▶ Callahan et al. 2017
- ▶ Exact Sequence Variants (ESVs)
 - ▶ Callahan et al. 2017
- ▶ sub-OTUs (sOTUs)
 - ▶ Amir et al. 2017
- ▶ Zero radius OTUs (zOTUs)
 - ▶ Edgar et al. 2017
- ▶ All the same!

ASVs approach

- ▶ Why binning sequences at all? Can we model PCR and sequencing errors sufficiently well to retrieve true sequence variants?
- ▶ Amplicon Sequence Variants (ASVs)
 - ▶ Needham et al. 2017
 - ▶ Callahan et al. 2017
- ▶ Exact Sequence Variants (ESVs)
 - ▶ Callahan et al. 2017
- ▶ sub-OTUs (sOTUs)
 - ▶ Amir et al. 2017
- ▶ Zero radius OTUs (zOTUs)
 - ▶ Edgar et al. 2017
- ▶ All the same!

ASVs approach

Pro

- ▶ Increased resolution
- ▶ Consistent labels

Cons

- ▶ Intra-genome marker variability
- ▶ Pushing technology beyond precision

ASV approaches are not inherently better, since both approaches have pros and cons.

Workflow

- ▶ De-multiplexing
- ▶ Reads pre-processing
- ▶ Dereplication
- ▶ Clustering of variants
- ▶ Filtering of artefacts
- ▶ Alignment to references
- ▶ Taxonomy annotation
- ▶ Downstream analysis

What do we have?

Raw data

1

Metadata

1	sampleid	Treatment	Years
2	AM-16S-1	maize-mono	1999
3	AM-16S-2	maize-mono	1999
4	AM-16S-3	maize-mono	1999
5	AM-16S-4	push-pull	1999
6	AM-16S-5	push-pull	1999

Which is our goal?

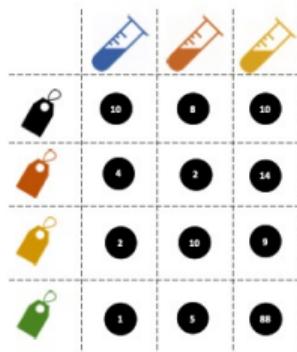
Metadata

1	sampleid	Treatment	Years
2	AM-16S-1	maize-mono	1999
3	AM-16S-2	maize-mono	1999
4	AM-16S-3	maize-mono	1999
5	AM-16S-4	push-pull	1999
6	AM-16S-5	push-pull	1999

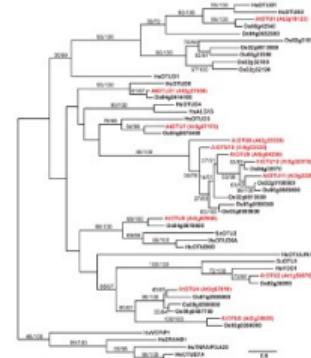
Taxonomy

ASV148428	Bacteria(100);Proteobacteria(100);Gammaproteobacte...
ASV212114	Bacteria(100);Cyanobacteria(100);Cyanobacteria(100);...
ASV9520	Bacteria(100);Proteobacteria(100);Alphaproteobacteri...
ASV147186	Bacteria(100);Proteobacteria(100);Betaproteobacteria(...
ASV89359	Bacteria(100);Proteobacteria(100);Alphaproteobacteri...
ASV1061	Bacteria(100);Proteobacteria(100);Gammaproteobacte...
ASV328581	Bacteria(100);Bacteroidetes(100);Bacteroidia(100);Bac...
ASV86104	Bacteria(100);Proteobacteria(100);Alphaproteobacteri...

OTU table



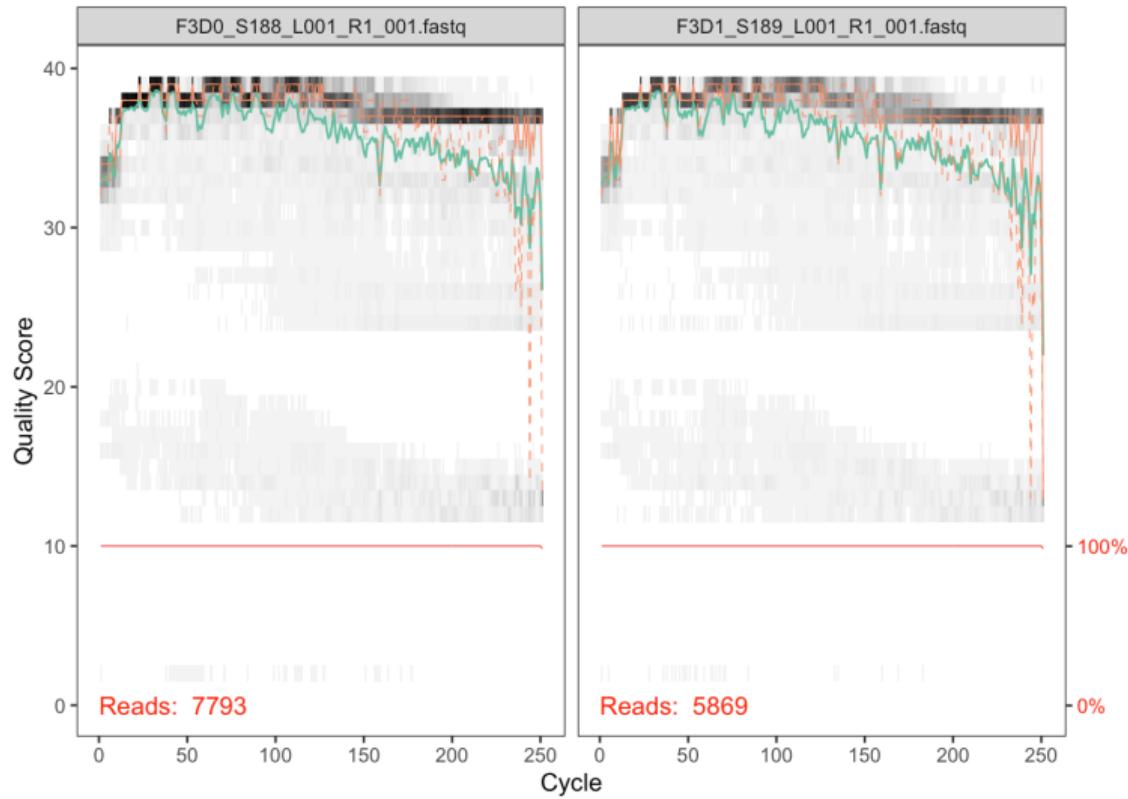
Phylogenetic tree



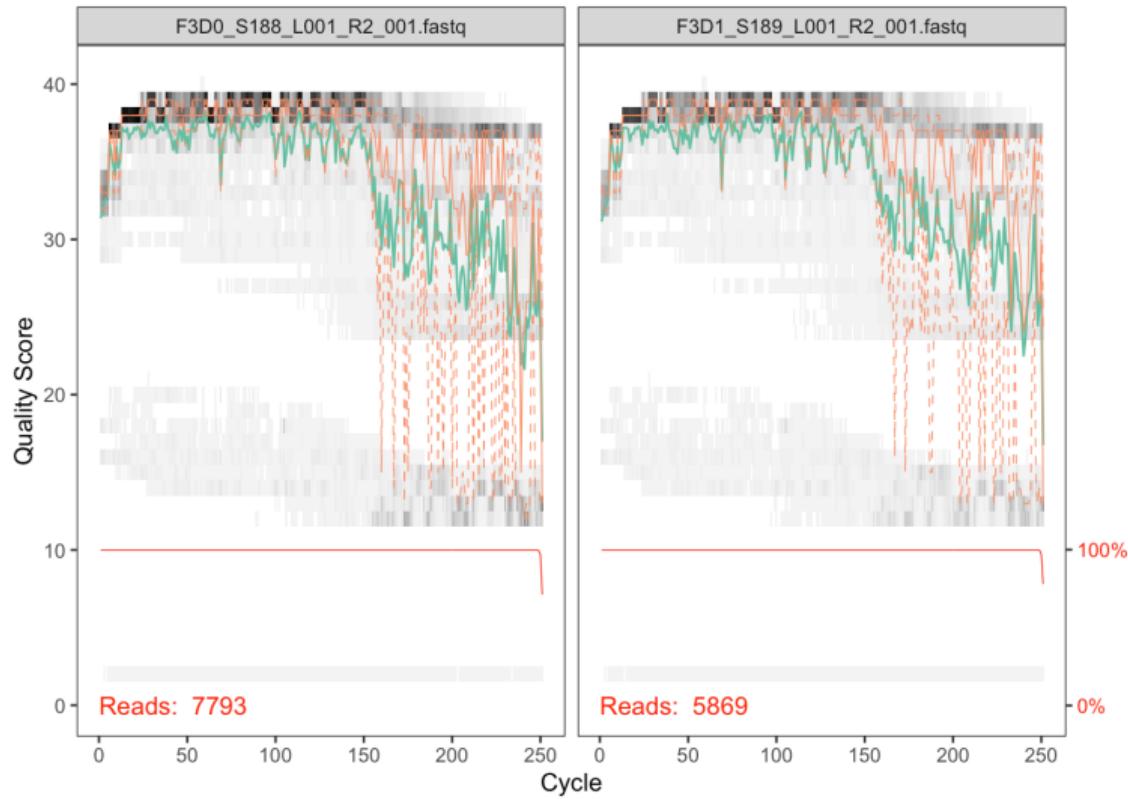
*.fastq files

Identifier	@SRR566546.970 HWUSI-EAS1673_11067_FC7070M:4:1:2299:1109 length=50
Sequence	TTGCCTGCCTATCATTAGTGCCTGTGAGGTGGAGATGTGAGGATCAGT
'+' sign	+
Quality scores	hhhhhhhhhhghhhhhhhfhhhhfffffe'ee[‘X]b[d[ed‘[Y[^Y
Identifier	@SRR566546.971 HWUSI-EAS1673_11067_FC7070M:4:1:2374:1108 length=50
Sequence	GATTGTATGAAAGTATAACAACTAAAAGTGCAGGTGGATCAGAGTAAGTC
'+' sign	+
Quality scores	hhhhgfhhcghghggfcffdhfehhhhcehdchhdhahehffffde‘bVd

Sequence quality



Sequence quality



Merging PE

	C	A	T	T	G	A	C	A	
	32	34	20	20	28	16	14	10	
Reverse read	T	A	G	A	C	A	T	T	Base calls
	2	5	4	8	12	20	38	40	Q scores
	C	A	T	T	G	A	C	A	
	32	34	22	16	35	28	30	34	
									Consensus
									Posterior Qs

↑ Mismatch Merged read

Remove suspicious reads

```
>GQY1XT001A6MUA
AATGGTACCCGTCAATTCATTTGATCTTGCGGTCGTTACGGCGTGGACTACCAGTCGACTCCAGTCATACAGTTCCAATG
>GQY1XT001BTRWS
AATGGTACCCGTCAATTCTTGATCTTGCGGCCGTTACGGCGTGGACTACCAGTCGACTCGAGCTGCACAGTTCAAAGCAGTTCCGGGTTGGG
>GQY1XT001AK4J0
TCTAGCCGACAGTTCAAAAGCACTCCCAGGGTT
>GQY1XT001BBPBR
AATGGTACCCGTCAATTCATTTGACGTTGCCCCGTTACTGTGCGGACTACCAGTCGACTCAAGGCCCCCAGTTCAACGG
>GQY1XT001BDDE9
AATGGTACCCGTCAATTCTTAATCTTGCGGTCGTTACGGCGTGGACTACCAGTCGACTCCAGTTACACAGTTCCAGAG
>GQY1XT001CIUF3
AATGGTACCCGTCAATTCTTGATCTTGCGGCCGTTACGGCGTGGACTACCAGGCCCCTCAGCCGGCAGTTCCAGTGCAGTCCGGGTT
>GQY1XT001BKRP5
AATGGTACCCGTCAATTCATTTAATCTCTCCCCCTTCCCCCCCCCCCCCCCCCTTCCCCCCCCCCCCCTTCCCCCCCCCCCC
>GQY1XT001B44ZE
AATGGTACCCGTCAATTCATTTAACTTGCGGTTTACCGCGTGGACTACCAGGCCCTAAGAACAGTTGAACGCAGCTATGGGTT
>GQY1XT001CIW3P
AATGGTACCCGTCAATTCATTTGACGTTGCCTCTGTTACTGCGTGGACTACCAGTCGACTCAAGGCCCCA
>GQY1XT001A731D
AATGGTACCCGTCAATTCATTTAACTTGCGGCCGTTACTGCGTGGACTACCAGGGCAATCAAGACTGCCA
```

Dereplication

>GQY1XT001A6MUA
AATGGTACCCGTCAATTGATCTTCGGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA
>GQY1XT001BTRWS
AATGGTACCCGTCAATTCTTGGATCTTCGGGCCGTTACGGCGTGGACTACCAGTCGCACTCGAGCTGCA
>GQY1XT001BBPBR
AATGGTACCCGTCAATTGATCTTCGGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA
>GQY1XT001BDDE9
AATGGTACCCGTCAATTGATCTTCGGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA
>GQY1XT001CIUF3
AATGGTACCCGTCAATTCTTGGATCTTCGGGCCGTTACGGCGTGGACTACCAGTCGCACTCGAGCTGCA
>GQY1XT001B44ZE
AATGGTACCCGTCAATTGATCTTCGGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA
>GQY1XT001CIW3P
AATGGTACCCGTCAATTCTTGGATCTTCGGGCCGTTACGGCGTGGACTACCAGTCGCACTCGAGCTGCA
>GQY1XT001A731D
AATGGTACCCGTCAATTGATCTTCGGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA

Dereplication

>GQY1XT001A6MUA DEPTH = 5

AATGGTACCCGTCAATTGATCTTCGGTTCGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA

>GQY1XT001BTRWS DEPTH = 3

AATGGTACCCGTCAATTCTTGTATCTTCGGGCCGTTACGGCGTGGACTACCAGTCGCACTCGAGCTGCA

~~>GQY1XT001BBPBR~~

~~AATGGTACCCGTCAATTGATCTTCGGTTCGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA~~

~~>GQY1XT001BDDE9~~

~~AATGGTACCCGTCAATTGATCTTCGGTTCGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA~~

~~>GQY1XT001CIUF3~~

~~AATGGTACCCGTCAATTCTTGTATCTTCGGGCCGTTACGGCGTGGACTACCAGTCGCACTCGAGCTGCA~~

~~>GQY1XT001B44ZE~~

~~AATGGTACCCGTCAATTGATCTTCGGTTCGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA~~

~~>GQY1XT001CIW3P~~

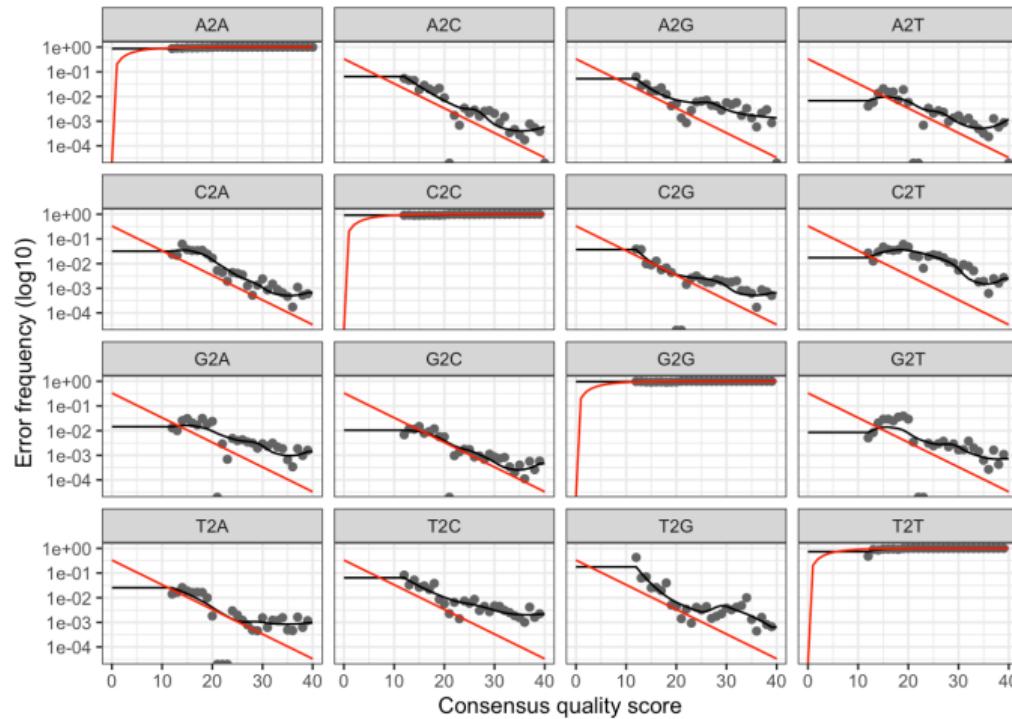
~~AATGGTACCCGTCAATTCTTGTATCTTCGGGCCGTTACGGCGTGGACTACCAGTCGCACTCGAGCTGCA~~

~~>GQY1XT001A731D~~

~~AATGGTACCCGTCAATTGATCTTCGGTTCGTTACGGCGTGGACTACCAGTCGCACTCCAGTCATA~~

Error modelling

Algorithms aim at modelling and correcting substitutions errors originating from PCR and sequencing by using a probabilistic approach (and including Q-scores).



Cluster variants

```
>*S16-0000006
TACGTTTATCGCGTT-AGCTTCGCCAAGCACAGCATTGCGCTTAGCCAACGTACATCGTTAGGGTGTGGACTAA
>#S16-0000046
TACGTTTATCGCGTTTAGCTTCGCCAAGCACAGCATTGCGCTTAGCCAACGTACATCGTTAGGGTGTGGACTAA
>#S16-0000241
TACGTTTATCGCGTT-AGCTTCGCCAAGCACAGCATTGCGCTTAGCCAACGTACATCGT-TAGGGTGTGGACTAA
>#S16-0000375
TACGTTTATCGCATT-AGCTTCGCCAAGCACAGCATTGCGCTTAGCCAACGTACATCGTTAGG-TGTGGACTAA
>*S16-0000001
GGCACTTAAAGCGTTAGCTACGGCGCAGAAACCACGGGTGG-CCCCCACACCTAGTGCCAACGTTACAGCGTGGT
>#S16-0000209
GGCACTTAAAGCGTTAGCTACGGCGCAGAAACCACGGGTGGTCCCCCACACCTAGTGCCAACGTTACAGCGTGGG
>#S16-0000667
GGCACTTAAAGCGTTAGCTACGGCGCAGAAACCACGGGTGG-CCCCCACACCTAGTGC-CAACGTTACAGCGTGGT
>*S16-0000004
TCGACTTAAACCGTTAGCTCCGGAAGCCACGCCCTCAAGG-GCACAAACCTCCAAGTCGACATCGTTACGGCGTGGAT
>#S16-0000625
TCGACTTAAACCGTTAGCTCCGGAAGCCACGCCCTCAAGG-GCACAAACCTCCAAGTCGACATCGT-TACGGCGTGGAT
>#S16-0000673
TCGACTTAAACCGTTAGCTCCGGAAGCCACGCCCTCAAGG-GGCACAAACCTCCAAGTCGACATCGTTACGGCGTGGAT
```

Cluster variants

```
>*S16-0000006 DEPTH + 3
TACGTTTATCGCGTT-AGCTTCGCCAAGCACAGCATTGCGCTTAGCCAACGTACATCGTTAGGGTGTGGACTAA
>*S16-0000001 DEPTH + 2
GGCACTTAAAGCGTTAGCTACGGCGCAGAAACCACGGGTGG-CCCCCACACCTAGTGCCCAACGTTACAGCGTGGT
>*S16-0000004 DEPTH + 2
TCGACTTAACCGCGTTAGCTCCGGAAGCCACGCCCTCAAGG-GCACAAACCTCCAAGTCGACATCGTTACGGCGTGGAT
```

>#S16-0000046
TACGTTTATCGCGTTTAGCTTCGCCAAGCACAGCATTGCGCTTAGCCAACGTACATCGTTAGGGTGTGGACTAA
>#S16-0000241
TACGTTTATCGCGTT-ACCTTCGCCAAGCACAGCATTGCGCTTAGCCAACGTACATCGT-TAGGGTGTGGACTAA
>#S16-0000375
TACGTTTATCGCATT-AGCTTCGCCAAGCACAGCATTGCGCTTAGCCAACGTACATCGTTAGG-TGTGGACTAA
>#S16-0000209
GGCACTTAAAGCGTTAGCTACGGCGCAGAAACCACGGGTGG-CCCCCACACCTAGTGCCCAACGTTACAGCGTGGG
>#S16-0000667
GGCACTTAAAGCGTTAGCTACGGCGCAGAAACCACGGGTGG-CCCCCACACCTAGTGC-CAACGTTACAGCGTGGT
>#S16-0000625
TCGACTTAACCGCGTTAGCTCCGGAAGCCACGCCCTCAAGG-GCACAAACCTCCAAGTCGACATCGT-TACGGCGTGGAT
>#S16-0000673
TCGACTTAACCGCGTTAGCTCCGGAAGCCACGCCCTCAAGG-GCACAAACCTCCAAGTCGACATCGTTACGGCGTGGAT

www.sixthsense.com

Chimera removal

- ▶ ~ 1-5% of reads
- ▶ Shorter amplicons show similar formation rates as formation mostly depends on the strength of the chimeric breakpoint
- ▶ Chimera formation between divergent species is not rare and occurs reproducibly
- ▶ Chimeras differ from all other errors since chimeric sequences can be of high quality. Therefore, detection should be applied when confidence in data is highest

A	CCTTGGTAGGCCGtTGCCCTGCCAACTAGCTAATCAGACGCgggtCCATCtcaCACCAccggAgtTTTtcTCaCTgTacc
Q	CCTTGGTAGGCCGtTGCCCTGCCAACTAGCTAATCAGACGCATCCCCATCCATACCGATAAAATCTTAATCTCTTTCAG
B	TCTTGGTgGGCCGtTaCCCCGCCAACaAGCTAATCAGACGCATCCCCATCCATACCGATAAAATCTTAaCTCTTTCAG
Diffs	A A p A A A BBBB BBB BBBB BB BBa B B BBB
Votes	+ + 0 + + + +++ ++++ + + +! + + +++
Model	AAAAAAAAAAAAAAAAAAAAAxxxxxxxxxxxxxBB

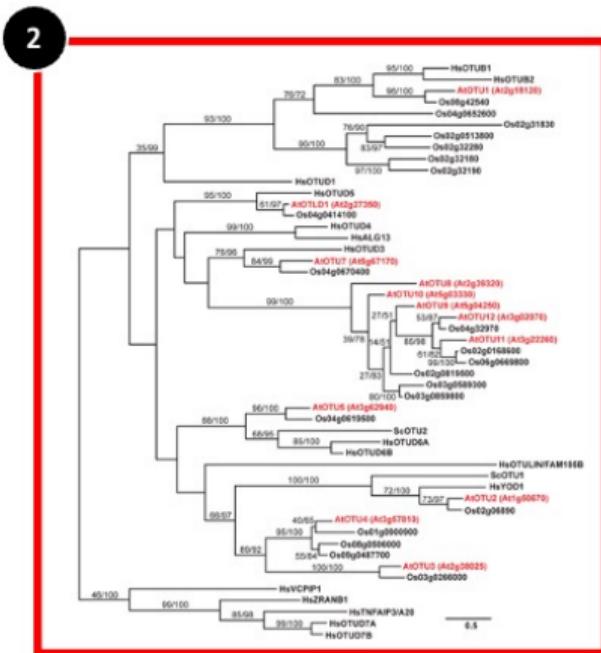
Filtering artefacts

```
>*16S-0000011 | depth=44 | freq=2.42
TTCAGTCGCTCCCTAGCTTCGCACTTCAGCGTCAGTTGCCGTCCAGTGAACATCTTCATCATCGGCATT
CCTGCACATATCTACGAATTCACTCTACTCGTGCAGTTCCGTCACCTCTCCAGCACTTAGCCAACAG
>*16S-0000076 | depth=33 | freq=1.82
TTCAATGTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAAGCCAGAGAGCCGCTTCGCCACCGGT
GTTCCCTCCATAATCTACGCAATTACCGCTACACATGGAATTCCACTCTCCCTCTTGCACTCAAGTTAAA
>*16S-0000052 | depth=32 | freq=1.76
TTCACGATAACCGCACCTTCGACCTTAAGCGTCAGTGGCGCTCCCGTCAGCTGCCCTCGCAATCGGAGTTCT
TCGTCAATCTAACGATTTACCGCTACACGACGAATTCCGCAACGTTGTGCGTACTCAAGGAAACCAGTA
>*16S-0000141 | depth=15 | freq=0.83
TTCAACGTTCGCTCCCTGGCTTCGCGCCTCAGCGTCAGTTTCGTCCAGAAAAGTCGCCCTCGCCACTGGT
GTTCTCTTAATATCTACGCAATTACCGCTACACTAGGAATTCCACTTCCCTCTCGATACTC
>#16S-0000058 | depth=12 | freq=0.66
TTCAGTCGCTCCCTAGCTTCGCACTTCAGCGTCAGTTTCGCTTCTACGACGAAAGCCAGAGAGCCGCTTCGCCACCGGT
GTTCCCTCCATAATCTACGCAATTACCGCTACACATGGAATTCCACTCTCCCTCTTGCACTCAAGTTAAA
>*16S-0000098 | depth=10 | freq=0.55
TTTAGTCCTGTTCGCTCCACGCTTTCGCTCAGCGTCAGTAACGGCCCAGAGACCCGCCCTCGCCACC
GGTGTCTTCCCTGATATCTGCGATTCCACCGCTACACCAGGAGTTCCAGGCCCTCC
>#16S-0000295 | depth=2 | freq=0.11
TTCACGATAACCCACGCTTCGAGCATCAGCGTCAGTGCCTACAGTAAGCTGCCCTGCCAATCGGAGTTCT
TCGTGATATCTAACGATTTACCGCTACACGACGAATTCCGCTACCTTCCGCGCAGTCAAGCCCCCAGTT
>#16S-0000021 | depth=1 | freq=0.06
TTCAACGTTCGCTCCCTGGCTTCGCGCCTCAGCGTCAGTTTCGTCCAGAAAAGTCGCCCTCGCCACTGGT
```

Chimera

Contaminations

Phylogenetic tree!



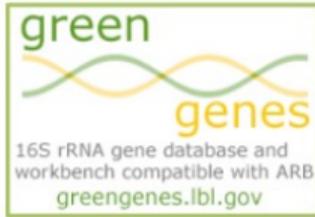
OTU table!

3			
	10	8	10
	4	2	14
	2	10	9
	1	5	88

Align to reference

```
>*16S-0000002 | depth=42 | freq=2.31
TTCACACCTTGCCTGGTCACTCCCCAGGCGGAGTGCTTAATGCGTTAGCTGCAGCCTAAACACCTAGCACTCATCGTT
TACGGCGTGGACTACCAGGGTATCTAATCTGTGCTCCCACGCTTCGAGCCTCAGCGTCAGTTACAAGCCAGAGGCCGCTTCGCCACCG
GTGTTCTCCATATCTACCGATTTCACCGTACACATGGAATTCCACTCTCCCTTGCACTCAAGTTAACAGTTCAAAGCGTACTATG
GTAAAGCCACAGCCTTAACCTCAGACTTATCT
>*16S-0000019 | depth=12 | freq=0.66
TTCAGCCTTGCCTGGTCACTCCCCAGGCGGAGTGCTTAATCGCATTGCTCGGCACAGACAGTCTCCTGCCACACCCAGTAATCATCGTTAC
GCCGGGACTACCAGGGTATCTAATCTGTGCTCCCCGGCTTCGACTTCAGCGTCAGTTACCGTCCAGTGAACATCTTCATCATCGGCA
TTCCCTGCACATATCTACGAATTTCACCTCTACTCGTGCAGTTCCGTCCACCTCTCCGGTACTCCAGCCTATCAGTTCAAAGGCAGGCCTGCCGT
TGAGCCGAGGTTTCACCCCTGACTTGAAAGG
```

VS.



Assign taxonomy

AY053482.1;tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Streptococcaceae,g:Streptococcus,s:pseudopneumoniae

Sequence ID: Icl|Query_210570 Length: 1429 Number of Matches: 1

Range 1: 565 to 882 [Graphics](#)

Score	Expect	Identities	Gaps	Strand
588 bits(318)	7e-172	318/318(100%)	0/318(0%)	Plus/Minus

Taxonomy!

4

ASV148428	Bacteria(100);Proteobacteria(100);Gammaproteobacte...
ASV212114	Bacteria(100);Cyanobacteria(100);Cyanobacteria(100);...
ASV9620	Bacteria(100);Proteobacteria(100);Alphaproteobacteri...
ASV147186	Bacteria(100);Proteobacteria(100);Betaproteobacteria(...
ASV89359	Bacteria(100);Proteobacteria(100);Alphaproteobacteri...
ASV1061	Bacteria(100);Proteobacteria(100);Gammaproteobacte...
ASV328581	Bacteria(100);Bacteroidetes(100);Bacteroidia(100);Bac...
ASV86104	Bacteria(100);Proteobacteria(100);Alphaproteobacteri...

Data analysis

What do we need to start?

```
phyloseq-class experiment-level object
otu_table()  OTU Table:      [ 43879 taxa and 289 samples ]
sample_data() Sample Data:    [ 289 samples by 9 sample variables ]
tax_table()   Taxonomy Table: [ 43879 taxa by 7 taxonomic ranks ]
phy_tree()    Phylogenetic Tree: [ 43879 tips and 43878 internal nodes ]
```

otu_table()

	16S.SOI.41	16S.SOI.48	16S.SOI.18	16S.SOI.46	16S.SOI.59	16S.SOI.34	16S.ROO.57	16S.ROO.43	16S.ROO.58
denovo7709	1	1	0	0	0	0	0	0	0
denovo7708	0	0	1	1	1	0	0	0	0
denovo22216	0	0	0	0	0	1	0	0	0

sample_data()

	A	B	C	D	E	F	G	H
1	SampleID	Community	Category	Sample_type	Genotype	Soil	Aphid	H_defensa
2	16S.APH.1	Bacterial	Experimental	Aphid	TBR	WHS	Present	Absent
3	16S.APH.2	Bacterial	Experimental	Aphid	TBR	WHS	Present	Absent
4	16S.APH.3	Bacterial	Experimental	Aphid	TBR	WHS	Present	Absent
5	16S.APH.4	Bacterial	Experimental	Aphid	TBR	WHS	Present	Absent
6	16S.APH.5	Bacterial	Experimental	Aphid	TBR	WHS	Present	Absent
7	16S.APH.6	Bacterial	Experimental	Aphid	TBR	WHS	Present	Present
8	16S.APH.7	Bacterial	Experimental	Aphid	TBR	WHS	Present	Present
9	16S.APH.8	Bacterial	Experimental	Aphid	TBR	WHS	Present	Present
10	16S.APH.9	Bacterial	Experimental	Aphid	TBR	WHS	Present	Present
11	16S.APH.10	Bacterial	Experimental	Aphid	TBR	WHS	Present	Present
12	16S.APH.11	Bacterial	Experimental	Aphid	TBR	MICROB	Present	Absent
13	16S.APH.12	Bacterial	Experimental	Aphid	TBR	MICROB	Present	Absent
14	16S.APH.13	Bacterial	Experimental	Aphid	TBR	MICROB	Present	Absent
15	16S.APH.14	Bacterial	Experimental	Aphid	TBR	MICROB	Present	Absent
16	16S.APH.15	Bacterial	Experimental	Aphid	TBR	MICROB	Present	Absent
17	16S.APH.16	Bacterial	Experimental	Aphid	TBR	MICROB	Present	Present
18	16S.APH.17	Bacterial	Experimental	Aphid	TBR	MICROB	Present	Present
19	16S.APH.18	Bacterial	Experimental	Anhrid	TBR	MICROB	Present	Present

tax_table()

	Rank1	Rank2
denovo7709	"D_0__Bacteria"	"D_1__Proteobacteria"
denovo7708	"D_0__Bacteria"	"D_1__Proteobacteria"
denovo22216	"D_0__Bacteria"	"D_1__Bacteroidetes"
denovo11322	"D_0__Bacteria"	"D_1__Bacteroidetes"
denovo44859	"D_0__Bacteria"	"D_1__Chloroflexi"

`phy_tree()`

