

AMPLImock_v0.1

Umer Zeeshan Ijaz

AMPLImock is useful for analyzing 16s RNA amplicons for mock communities and serves as a benchmark for answering the following questions:

- Is there any difference in read quality between forward and reverse reads?
- What is the best library preparation strategy?
- What is the best next generation sequencing platform?
- Which 16s RNA region performs better?
- Which primer performs better?
- Is there any evidence of carry-over or barcode switching for the primers used?
- Do DNA starting input and cycling condition matter?
- Do longer paired-end reads overlap and if they do then how well they perform when assembled together?
- What is the proportion of chimeras?

You can run the pipeline without any arguments to get the usage information:

Usage

```
[uzi@quince-srv2 ~]$ ~/AMPLImock/AMPLImock
Usage: AMPLImock [options] filename
```

Options:

```
--version          show program's version number and exit
-h, --help         show this help message and exit
-q QUERY_READS, --query_reads=QUERY_READS
                   specify single or paired-end reads in fasta/fastq
                   format. For each file, use -q, e.g., -q R1.fastq -q
                   R2.fastq
-p PRIMERS, --primers=PRIMERS
                   specify forward and reverse primers, for each primer
                   use -p, e.g., -p CAGTGTCAGTTGA -p
```

TCATATCTAATCTATATCGA. The first primer will be considered the forward primer and the second primer will be considered the reverse primer

```

-d REFERENCE_DATABASE_FORWARD, --reference_database_forward=REFERENCE_DATABASE_FORWARD
    reference database forward
-e REFERENCE_DATABASE_REVERSE, --reference_database_reverse=REFERENCE_DATABASE_REVERSE
    reference database reverse
-l LIST_FILE, --list_file=LIST_FILE
    list file that contains the mapping from multiple 16S
    copies to genome
-f FULL_LENGTH_AMPLICONS_FILE, --full_length_amplicons_file=FULL_LENGTH_AMPLICONS_FILE
    sequence file that contains reconstructed full length
    amplicons using 3rd party software such as emirge
-t QUALITY_TYPE, --quality_type=QUALITY_TYPE
    Type of quality values (solexa (CASAVA < 1.3),
    illumina (CASAVA 1.3 to 1.7), sanger (which is CASAVA
    >= 1.8) (Default:sanger)
-u, --use_quality_trimming
    flag to allow quality trimming for fastq files
    (Default:False)
-n, --eliminate_n
    flag to eliminate all sequences with unknown
    nucleotides in the output of pandaseq (Default:False)
-o OVERLAP_BASE_PAIRS, --overlap_base_pairs=OVERLAP_BASE_PAIRS
    minimum number of overlapping base pairs to consider
    (Default:50)
-s THREADS, --threads=THREADS
    number of threads to use in usearch and pandaseq
    (Default:10)

```

You need:

- A forward reference (16S.fa)
- A reverse-compliment of the forward reference (16S_R.fa)
- A mapping file used to identify 16S copies of a given genome (IDs.txt). You can generate one by taking the headers of forward reference and then manually grouping them together. For example, here are the first few records of my mapping file:

```
[uzi@quince-srv2 ~/AMPLImock/test_AMPLImock]$ head -40 ~/TSBAmplicons/References/IDs.txt
Acidocap1_643768675,Acidobacterium_capsulatum
AkkMuc36273762,Akkermansia_muciniphila
AkkMuc36274338,Akkermansia_muciniphila
AkkMuc36275228,Akkermansia_muciniphila
AnaTherm37409222,Anaerocellum_thermophilum
AnaTherm37409327,Anaerocellum_thermophilum
AnaTherm37409526,Anaerocellum_thermophilum
ArchFull1_638158565,Archaeoglobus_fulgidus
BacTheta5_4983323,Bacteroides_thetaiotaomicron
BacTheta5_c1631800,Bacteroides_thetaiotaomicron
BacTheta5_c2342291,Bacteroides_thetaiotaomicron
BacTheta5_c3036295,Bacteroides_thetaiotaomicron
BacTheta5_c3330547,Bacteroides_thetaiotaomicron
BacVul7_5301198,Bacteroides_vulgatus
BacVul7_5301427,Bacteroides_vulgatus
BacVul7_5302575,Bacteroides_vulgatus
BacVul7_5302661,Bacteroides_vulgatus
BacVul75302965,Bacteroides_vulgatus
BacVul7_5304532,Bacteroides_vulgatus
BacVul7_5304800,Bacteroides_vulgatus
BorBro3_3131222,Bordetella_bronchiseptica
BorBro3_3131255,Bordetella_bronchiseptica
BorBro3_3131264,Bordetella_bronchiseptica
BurXen6_1300517,Burkholderia_xenovorans
BurXen6_1425392,Burkholderia_xenovorans
BurXen6_1555931,Burkholderia_xenovorans
BurXen6_539350,Burkholderia_xenovorans
BurXen6_c4367202,Burkholderia_xenovorans
BurXen6_c4541233,Burkholderia_xenovorans
CalSac35087085,Caldicellulosiruptor_saccharolyticus
CalSac35087220,Caldicellulosiruptor_saccharolyticus
CalSac35087545,Caldicellulosiruptor_saccharolyticus
ChloAur35824787,Chloroflexus_aurantiacus
ChloAur35825654,Chloroflexus_aurantiacus
ChloAur35825939,Chloroflexus_aurantiacus
```

ChloLim2_6356066,Chlorobium_limicola
ChloLim26356108,Chlorobium_limicola
ChloPhaba24568764,Chlorobium_phaeobacteroides
ChloPhaba24569699,Chlorobium_phaeobacteroides
ChloPhavi14969771,Chlorobium_phaeovibrioides

[MISEQ paired-end reads](#)

```
[uzi@quince-srv2 ~/AMPLImock/test_AMPLImock]$ ../AMPLImock -q
~/TSB_AMPLICONS/Anaerocellum_Thermophilum_Fusion_Golay/4_TATCGTTGACCA_L001_R1_001.fastq -q
~/TSB_AMPLICONS/Anaerocellum_Thermophilum_Fusion_Golay/4_TATCGTTGACCA_L001_R2_001.fastq -d References/16S.fa -e
References/16S_R.fa -l ~/TSBAmplicons/References/IDs.txt -u -s 20
This is AMPLImock v0.1. Copyright (c) 2013 Computational Microbial Genomics Group, University of Glasgow, UK
Checking for 'sickle': found /home/opt/sickle/sickle
Checking for 'usearch': found /home/opt/usearch_v6.0.307/usearch
Checking for 'pandaseq': found /home/opt/pandaseq/pandaseq
Forward reference database:References/16S.fa
Reverse reference database:References/16S_R.fa
List file containing mapping from multiple 16S genes:/home/uzi/TSBAmplicons/References/IDs.txt
Forward reads:/home/uzi/TSB_AMPLICONS/Anaerocellum_Thermophilum_Fusion_Golay/4_TATCGTTGACCA_L001_R1_001.fastq
Reverse reads:/home/uzi/TSB_AMPLICONS/Anaerocellum_Thermophilum_Fusion_Golay/4_TATCGTTGACCA_L001_R2_001.fastq
Query files are in fastq format
Query files will be trimmed
Processing the data in /home/uzi/AMPLImock/test_AMPLImock
[2013-08-20 09:56:02] Trimming
/home/uzi/TSB_AMPLICONS/Anaerocellum_Thermophilum_Fusion_Golay/4_TATCGTTGACCA_L001_R1_001.fastq and
/home/uzi/TSB_AMPLICONS/Anaerocellum_Thermophilum_Fusion_Golay/4_TATCGTTGACCA_L001_R2_001.fastq using quality type
as 'sanger', quality score of 20, and minimum of 10 bp to keep
[2013-08-20 09:56:04] Generated 4_TATCGTTGACCA_L001_R1_001_trim.fastq ,4_TATCGTTGACCA_L001_R2_001_trim.fastq ,and
4_TATCGTTGACCA_L001_R1_001_singlet.fastq
[2013-08-20 09:56:04] 4_TATCGTTGACCA_L001_R1_001_trim.fastq has following stats:
Total reads:123249
Unique reads:3326
Unique reads percentage:2.6986
```

Most abundant

sequence:TACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGGGTGCCTAGGCGGCCTGGTAAGTTGAGCGTGAAATTTTTGGGCTCAACCCAAAAGGAGCGTTCAA
AACTGCCAGGCTTGAGTGCGGGAGAGGACGGCGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGAAGGCGGCCGTCTGG

Most abundant sequence reads:106975

Most abundant sequence read percentage:86.7958

[2013-08-20 09:56:05] 4_TATCGTTGACCA_L001_R2_001_trim.fastq has following stats:

Total reads:123249

Unique reads:9186

Unique reads percentage:7.4532

Most abundant

sequence:CCTGTTTTGCTCCCCACGCTTTTCGTGCCTCAGCGTCAGTTACGGTCCAGACGGCCGCTTCGCCACTGGTGTTCCTCCCGATATCTACGCATTTACCGCTACACCGG
GAATTCGCGCTCCTCTCCGCACTCAAGCCTGGCAGTTTTGAACGCTCCTTTTTGGGTTGAGCCAAAAATTTACGCTCAACTTACCAGGCCGCTACGCA

Most abundant sequence reads:62524

Most abundant sequence read percentage:50.7298

[2013-08-20 09:56:15] Plotting PHRED quality scores for 4_TATCGTTGACCA_L001_R1_001_trim.fastq

[2013-08-20 09:56:16] Generated 4_TATCGTTGACCA_L001_R1_001_trim_phred-quality.png

[2013-08-20 09:56:16] Plotting PHRED quality scores for 4_TATCGTTGACCA_L001_R2_001_trim.fastq

[2013-08-20 09:56:16] Generated 4_TATCGTTGACCA_L001_R2_001_trim_phred-quality.png

[2013-08-20 09:56:16] Converting 4_TATCGTTGACCA_L001_R1_001_trim.fastq to 4_TATCGTTGACCA_L001_R1_001_trim.fasta

[2013-08-20 09:56:17] Generated 4_TATCGTTGACCA_L001_R1_001_trim.fasta

[2013-08-20 09:56:17] Converting 4_TATCGTTGACCA_L001_R2_001_trim.fastq to 4_TATCGTTGACCA_L001_R2_001_trim.fasta

[2013-08-20 09:56:19] Generated 4_TATCGTTGACCA_L001_R2_001_trim.fasta

[2013-08-20 09:56:19] Plotting sequences lengths histogram for 4_TATCGTTGACCA_L001_R1_001_trim.fasta

[2013-08-20 09:56:21] Generated 4_TATCGTTGACCA_L001_R1_001_trim_length-histogram.png

[2013-08-20 09:56:21] Plotting sequences lengths histogram for 4_TATCGTTGACCA_L001_R2_001_trim.fasta

[2013-08-20 09:56:23] Generated 4_TATCGTTGACCA_L001_R2_001_trim_length-histogram.png

[2013-08-20 09:56:23] Plotting GC percentage plot for 4_TATCGTTGACCA_L001_R1_001_trim.fasta

[2013-08-20 09:56:29] Generated 4_TATCGTTGACCA_L001_R1_001_trim_GC-percentage.png

[2013-08-20 09:56:29] Plotting GC percentage plot for 4_TATCGTTGACCA_L001_R2_001_trim.fasta

[2013-08-20 09:56:34] Generated 4_TATCGTTGACCA_L001_R2_001_trim_GC-percentage.png

[2013-08-20 09:56:34] Searching 4_TATCGTTGACCA_L001_R1_001_trim.fasta against References/16S.fa using usearch on 20 threads and matching identity of 0.95

[2013-08-20 09:56:39] Generated 4_TATCGTTGACCA_L001_R1_001_trim.aln, 4_TATCGTTGACCA_L001_R1_001_trim.ualn, 4_TATCGTTGACCA_L001_R1_001_trim_matched.fasta, and 4_TATCGTTGACCA_L001_R1_001_trim_notmatched.fasta

[2013-08-20 09:56:39] 4_TATCGTTGACCA_L001_R1_001_trim_matched.fasta has 121561 reads

[2013-08-20 09:56:39] 4_TATCGTTGACCA_L001_R1_001_trim_matched.fasta has 121561 reads

[2013-08-20 09:56:39] Searching 4_TATCGTTGACCA_L001_R2_001_trim.fasta against References/16S_R.fa using usearch on 20 threads and matching identity of 0.95
[2013-08-20 09:56:44] Generated 4_TATCGTTGACCA_L001_R2_001_trim.aln, 4_TATCGTTGACCA_L001_R2_001_trim.ualn, 4_TATCGTTGACCA_L001_R2_001_trim_matched.fasta, and 4_TATCGTTGACCA_L001_R2_001_trim_notmatched.fasta
[2013-08-20 09:56:44] 4_TATCGTTGACCA_L001_R2_001_trim_matched.fasta has 121652 reads
[2013-08-20 09:56:44] 4_TATCGTTGACCA_L001_R2_001_trim_matched.fasta has 121652 reads
[2013-08-20 09:56:44] Computing usearch statistics for 4_TATCGTTGACCA_L001_R1_001_trim.ualn
[2013-08-20 09:56:45] Generated 4_TATCGTTGACCA_L001_R1_001_trim_freq.csv. Mean identity is 99.9347405829, simpson index is 0.0156141077397, inverse simpson is 1.01586177521, entropy is 0.0713192291713
[2013-08-20 09:56:45] Computing usearch statistics for 4_TATCGTTGACCA_L001_R2_001_trim.ualn
[2013-08-20 09:56:46] Generated 4_TATCGTTGACCA_L001_R2_001_trim_freq.csv. Mean identity is 99.7777011475, simpson index is 0.0163994383953, inverse simpson is 1.016672864, entropy is 0.0733682411447
[2013-08-20 09:56:46] 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fastq has following stats:
Total reads:119982
Unique reads:2193
Unique reads percentage:1.82777
Most abundant
sequence:TACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGGGTGCCTAGCGCGCCTGGTAAGTTGAGCGTGAATTTTTGGGCTCAACCCAAAAGGAGCGTTCAA
AACTGCCAGGCTTGAGTGCGGGAGAGGACGGCGGAATTCGCGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGAAGGCGGCCGTCTGGACCGTAACTGACGC
TGAGGCACGAAAGCGTGGGGAGCAAACAGG
Most abundant sequence reads:103928
Most abundant sequence read percentage:86.6197
[2013-08-20 09:56:46] Running pandaseq on 20 threads without eliminating undetermined nucleotide sequences in overlap and minimum overlap of 50bps
[2013-08-20 09:56:52] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fastq
[2013-08-20 09:56:52] Converting 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fastq to 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fasta
[2013-08-20 09:56:54] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fasta
[2013-08-20 09:56:55] Plotting sequences lengths histogram for 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fasta
[2013-08-20 09:56:57] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap_length-histogram.png
[2013-08-20 09:56:57] Plotting GC percentage plot for 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fasta
[2013-08-20 09:57:03] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap_GC-percentage.png
[2013-08-20 09:57:03] Searching 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fasta against References/16S.fa using usearch on 20 threads and matching identity of 0.95
[2013-08-20 09:57:08] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap.aln, 4_TATCGTTGACCA_L001_R1_001_trim_overlap.ualn, 4_TATCGTTGACCA_L001_R1_001_trim_overlap_matched.fasta, and 4_TATCGTTGACCA_L001_R1_001_trim_overlap_notmatched.fasta

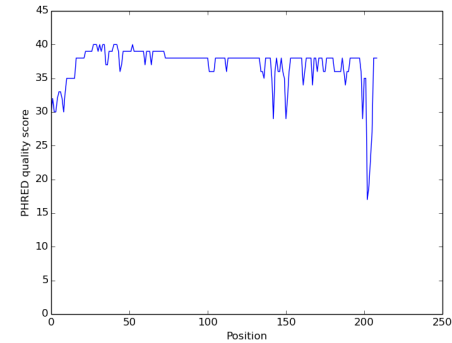
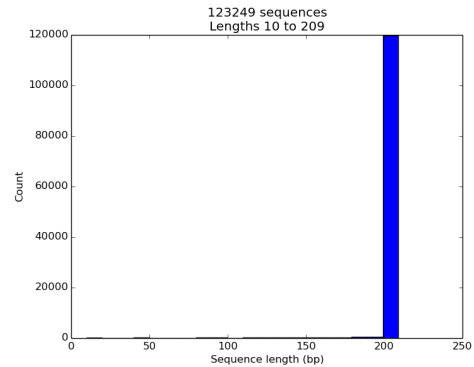
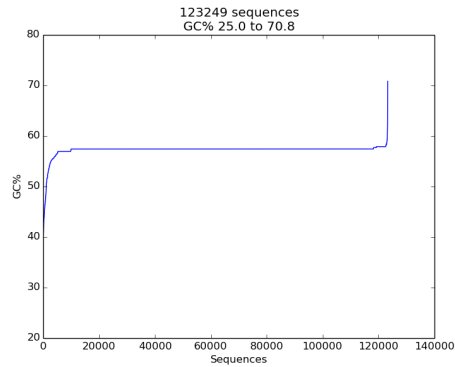
[2013-08-20 09:57:08] 4_TATCGTTGACCA_L001_R1_001_trim_overlap_matched.fasta has 119407 reads
[2013-08-20 09:57:08] 4_TATCGTTGACCA_L001_R1_001_trim_overlap_matched.fasta has 119407 reads
[2013-08-20 09:57:08] Computing usearch statistics for 4_TATCGTTGACCA_L001_R1_001_trim_overlap.ualn
[2013-08-20 09:57:09] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap_freq.csv. Mean identity is 99.9396953277, simpson index is 0.0122880451042, inverse simpson is 1.01244091969, entropy is 0.0585260872077
[2013-08-20 09:57:09] Dereplicating duplicate sequences in 4_TATCGTTGACCA_L001_R1_001_trim_overlap.fasta using usearch on 20 threads, annotating with cluster sizes, and sorting by cluster size
[2013-08-20 09:57:10] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep.fasta
[2013-08-20 09:57:10] Sorting reads in 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep.fasta by abundance using usearch
[2013-08-20 09:57:10] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted.fasta
[2013-08-20 09:57:10] Detecting De novo chimeras in 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted.fasta using UCHIME
[2013-08-20 09:57:12] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera.fasta and 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_chimera.fasta
[2013-08-20 09:57:12] Searching 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera.fasta against References/16S.fa using usearch on 20 threads and matching identity of 0.97
[2013-08-20 09:57:12] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera.aln, 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera.ualn, 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera_matched.fasta, and 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera_notmatched.fasta
[2013-08-20 09:57:12] Computing usearch statistics for 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera.ualn
[2013-08-20 09:57:12] Generated 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera_freq.csv. Mean identity is 99.9437454408, simpson index is 0.0117360687788, inverse simpson is 1.01187543976, entropy is 0.0561794959249
[2013-08-20 09:57:12] 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera.fasta has 119931 reads
[2013-08-20 09:57:12] 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_chimera.fasta has 30 reads
[2013-08-20 09:57:12] 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera_matched.fasta has 119265 reads
[2013-08-20 09:57:12] 4_TATCGTTGACCA_L001_R1_001_trim_overlap_derep_sorted_nonchimera_notmatched.fasta has 666 reads
[2013-08-20 09:57:12] Saving pipeline statistics in AMPLImock_stats.csv. Here is the summary of results:
FORWARD_MATCHED_TOTAL_READS,121561
FORWARD_NOTMATCHED_TOTAL_READS,121561
FORWARD_READS_ENTROPY,0.0713192291713
FORWARD_READS_INVERSE_SIMPSON,1.01586177521
FORWARD_READS_MEAN_IDENT,99.9347405829

FORWARD_READS_MOST_ABUNDANT_SEQUENCE,TACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGCCTGGTAAGTTGAGCGTGAAATTT
TTGGGCTCAACCCAAAAGGAGCGTTCAAACCTGCCAGGCTTGAGTGCGGGAGAGGACGGCGGAATTCGCCGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGA
AGGCGGCCGTCTGG
FORWARD_READS_MOST_ABUNDANT_SEQUENCE_READS,106975
FORWARD_READS_MOST_ABUNDANT_SEQUENCE_READS_PERCENTAGE,86.7958
FORWARD_READS_SIMPSON,0.0156141077397
FORWARD_READS_TOTAL_READS,123249
FORWARD_READS_UNIQUE_READS,3326
FORWARD_READS_UNIQUE_READS_PERCENTAGE,2.6986
OVERLAP_DEREP_CHIM_TOTAL_READS,30
OVERLAP_DEREP_NONCHIM_MATCHED_TOTAL_READS,119265
OVERLAP_DEREP_NONCHIM_NOTMATCHED_TOTAL_READS,666
OVERLAP_DEREP_NONCHIM_READS_ENTROPY,0.0561794959249
OVERLAP_DEREP_NONCHIM_READS_INVERSE_SIMPSON,1.01187543976
OVERLAP_DEREP_NONCHIM_READS_MEAN_IDENT,99.9437454408
OVERLAP_DEREP_NONCHIM_READS_SIMPSON,0.0117360687788
OVERLAP_DEREP_NONCHIM_TOTAL_READS,119931
OVERLAP_MATCHED_TOTAL_READS,119407
OVERLAP_NOTMATCHED_TOTAL_READS,119407
OVERLAP_READS_ENTROPY,0.0585260872077
OVERLAP_READS_INVERSE_SIMPSON,1.01244091969
OVERLAP_READS_MEAN_IDENT,99.9396953277
OVERLAP_READS_MOST_ABUNDANT_SEQUENCE,TACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGCCTGGTAAGTTGAGCGTGAAATTT
TTGGGCTCAACCCAAAAGGAGCGTTCAAACCTGCCAGGCTTGAGTGCGGGAGAGGACGGCGGAATTCGCCGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGA
AGGCGGCCGTCTGGACCGTAACTGACGCTGAGGCACGAAAGCGTGGGGAGCAAACAGG
OVERLAP_READS_MOST_ABUNDANT_SEQUENCE_READS,103928
OVERLAP_READS_MOST_ABUNDANT_SEQUENCE_READS_PERCENTAGE,86.6197
OVERLAP_READS_SIMPSON,0.0122880451042
OVERLAP_READS_TOTAL_READS,119982
OVERLAP_READS_UNIQUE_READS,2193
OVERLAP_READS_UNIQUE_READS_PERCENTAGE,1.82777
REVERSE_MATCHED_TOTAL_READS,121652
REVERSE_NOTMATCHED_TOTAL_READS,121652
REVERSE_READS_ENTROPY,0.0733682411447
REVERSE_READS_INVERSE_SIMPSON,1.016672864
REVERSE_READS_MEAN_IDENT,99.7777011475

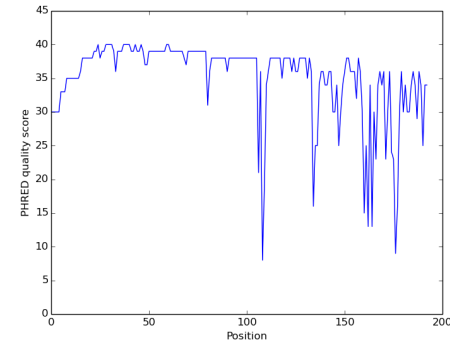
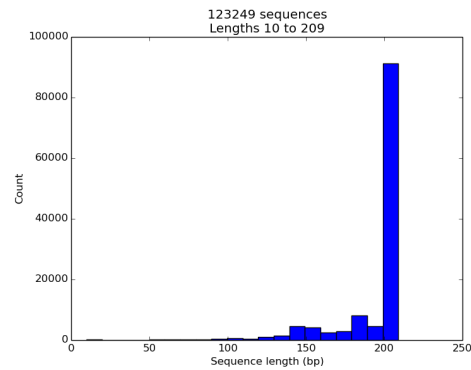
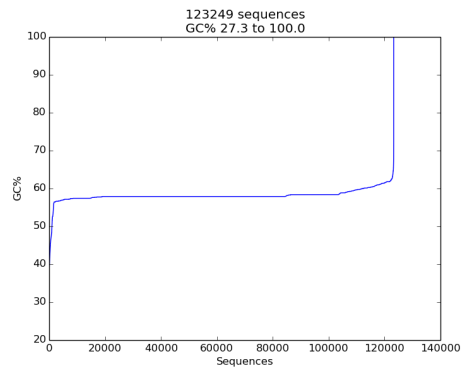

```
REVERSE_READS_MOST_ABUNDANT_SEQUENCE,CCTGTTTGTCTCCCCACGCTTTCGTGCCTCAGCGTCAGTTACGGTCCAGACGGCCGCCTTCGCCACTGGTGTTCCTCCCG
ATATCTACGCATTTTACCCTACACCGGGAATTCGCCGTCTCTCCCGCACTCAAGCCTGGCAGTTTTGAACGCTCCTTTTGGGTTGAGCCCAAAAATTTACAGCTCAACTTACC
AGGCCGCCTACGCA
```

```
REVERSE_READS_MOST_ABUNDANT_SEQUENCE_READS,62524
REVERSE_READS_MOST_ABUNDANT_SEQUENCE_READS_PERCENTAGE,50.7298
REVERSE_READS_SIMPSON,0.0163994383953
REVERSE_READS_TOTAL_READS,123249
REVERSE_READS_UNIQUE_READS,9186
REVERSE_READS_UNIQUE_READS_PERCENTAGE,7.4532
[uzi@quince-srv2 ~/AMPLImock/test_AMPLImock]$
```

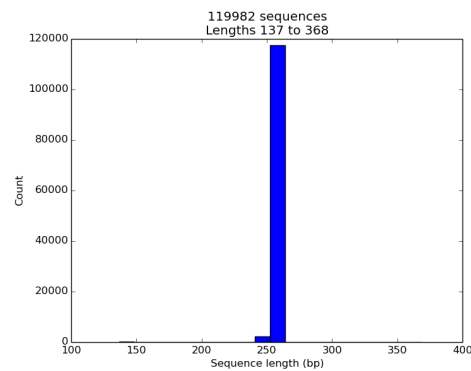
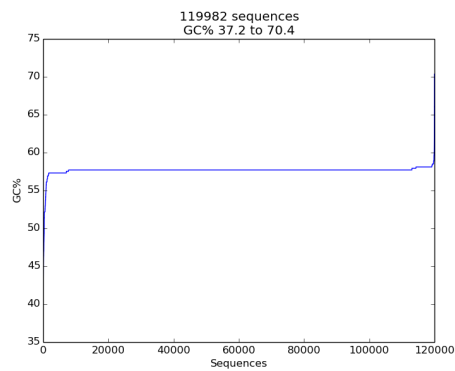
R1:



R2:



Overlapped reads:



[MISEQ paired-end reads when you have assembled full length 16S amplicons using third-party software such as EMIRGE](#)

```
[uzi@quince-srv2 ~/AMPLImock/test_AMPLImock]$ ../AMPLImock -q  
/home/shared1/TSB/Amplicons/Mircea_Balanced_Nextera_Full16S/Raw/16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001.fastq -q  
/home/shared1/TSB/Amplicons/Mircea_Balanced_Nextera_Full16S/Raw/16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001.fastq -d  
References/16S.fa -e References/16S_R.fa -l ~/TSBAmplicons/References/IDs.txt -u -f  
References/M_bal_Nex_full16S.fasta -s 20
```

This is AMPLImock v0.1. Copyright (c) 2013 Computational Microbial Genomics Group, University of Glasgow, UK
Checking for 'sickle': found /home/opt/sickle/sickle
Checking for 'usearch': found /home/opt/usearch_v6.0.307/usearch
Checking for 'pandaseq': found /home/opt/pandaseq/pandaseq
Forward reference database:References/16S.fa
Reverse reference database:References/16S_R.fa
List file containing mapping from multiple 16S genes:/home/uzi/TSBAmplicons/References/IDs.txt
Forward reads:/home/shared1/TSB/Amplicons/Mircea_Balanced_Nextera_Full16S/Raw/16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R1_001.fastq
Reverse reads:/home/shared1/TSB/Amplicons/Mircea_Balanced_Nextera_Full16S/Raw/16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R2_001.fastq
Query files are in fastq format
Query files will be trimmed
Processing the data in /home/uzi/AMPLImock/test_AMPLImock
[2013-08-20 10:23:53] Trimming
/home/shared1/TSB/Amplicons/Mircea_Balanced_Nextera_Full16S/Raw/16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R1_001.fastq and
/home/shared1/TSB/Amplicons/Mircea_Balanced_Nextera_Full16S/Raw/16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R2_001.fastq
using quality type as 'sanger', quality score of 20, and minimum of 10 bp to keep
[2013-08-20 10:24:11] Generated 16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R1_001_trim.fastq ,16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R2_001_trim.fastq ,and 16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R1_001_singlet.fastq
[2013-08-20 10:24:17] 16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R1_001_trim.fastq has following stats:
Total reads:1077945
Unique reads:353086
Unique reads percentage:32.7555
Most abundant
sequence:TCCTTAGAGTGCCAGCCGAAGTCTGCTGGCAACTAAGGATGTGGGTTGCGCTCGTTGCCGGACTTAACCGAACATCTCACGACACGAGCTGACGACAGCCATGCAGCA
CCTGTCACTGCGTCACCGAAGTGAACGCCGATCTCTCGGGTTAGCACAGGATGTCAAGGGTTGGTAAGGTTCTGCGCGTTGCTTCAATTAACCACATGCTCCACCGCTTGTGC
GGGCCCCGTCAATTCCTTTGAGTTTT
Most abundant sequence reads:1029
Most abundant sequence read percentage:0.0954594
[2013-08-20 10:24:22] 16SMIRCEABAL_TAAGGCCA-TAGATCGC_L001_R2_001_trim.fastq has following stats:
Total reads:1077945
Unique reads:338070
Unique reads percentage:31.3625
Most abundant
sequence:TCCTTAGAGTGCCAGCCGAAGTCTGCTGGCAACTAAGGATGTGGGTTGCGCTCGTTGCCGGACTTAACCGAACATCTCACGACACGAGCTGACGACAGCCATGCAGCA

CCTGTCACTGCGTCACCGAAGTGAACGCCCGATCTCTCGGGTTAGCACAGGATGTCAAGGGTTGGTAAGGTTCTGCGCGTTGCTTCGAATTAACCACATGCTCCACCGCTTGTGC
GGGCCCCCGTCAATTCCTTTGAGTTTT

Most abundant sequence reads:1280

Most abundant sequence read percentage:0.118744

[2013-08-20 10:24:33] Plotting PHRED quality scores for 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.fastq

[2013-08-20 10:24:33] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim_phred-quality.png

[2013-08-20 10:24:33] Plotting PHRED quality scores for 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.fastq

[2013-08-20 10:24:34] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim_phred-quality.png

[2013-08-20 10:24:34] Converting 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.fastq to 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.fasta

[2013-08-20 10:24:46] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.fasta

[2013-08-20 10:24:46] Converting 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.fastq to 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.fasta

[2013-08-20 10:24:58] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.fasta

[2013-08-20 10:24:58] Plotting sequences lengths histogram for 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.fasta

[2013-08-20 10:25:18] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim_length-histogram.png

[2013-08-20 10:25:18] Plotting sequences lengths histogram for 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.fasta

[2013-08-20 10:25:37] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim_length-histogram.png

[2013-08-20 10:25:37] Plotting GC percentage plot for 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.fasta

[2013-08-20 10:26:26] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim_GC-percentage.png

[2013-08-20 10:26:26] Plotting GC percentage plot for 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.fasta

[2013-08-20 10:27:19] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim_GC-percentage.png

[2013-08-20 10:27:19] Searching 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.fasta against References/16S.fa using usearch on 20 threads and matching identity of 0.95

[2013-08-20 10:28:06] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.aln, 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.ualn, 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim_matched.fasta, and 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim_notmatched.fasta

[2013-08-20 10:28:06] 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim_matched.fasta has 607078 reads

[2013-08-20 10:28:06] 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim_matched.fasta has 607078 reads

[2013-08-20 10:28:06] Searching 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.fasta against References/16S_R.fa using usearch on 20 threads and matching identity of 0.95

[2013-08-20 10:28:52] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.aln, 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.ualn, 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim_matched.fasta, and 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim_notmatched.fasta

[2013-08-20 10:28:52] 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim_matched.fasta has 602332 reads

[2013-08-20 10:28:52] 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim_matched.fasta has 602332 reads

[2013-08-20 10:28:52] Computing usearch statistics for 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim.ualn

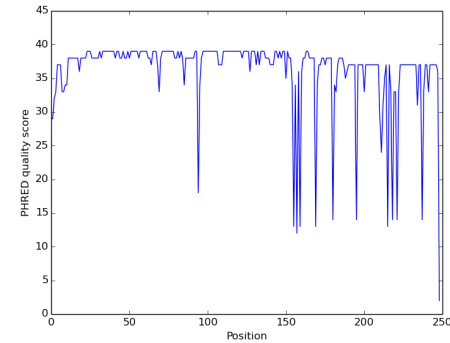
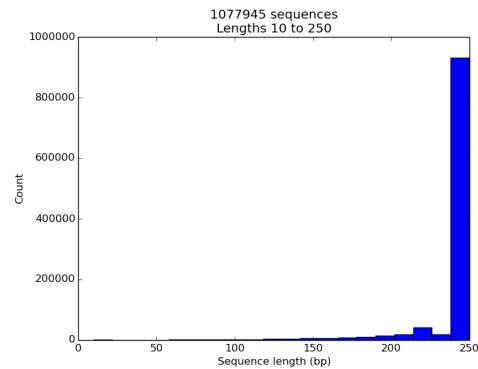
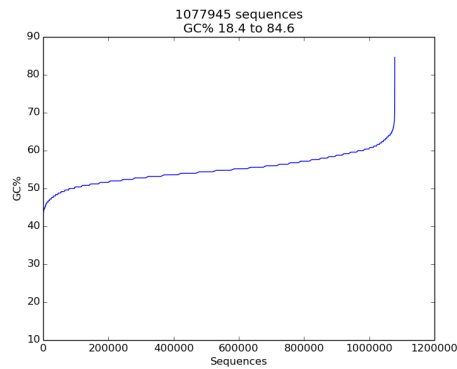
[2013-08-20 10:28:55] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R1_001_trim_freq.csv. Mean identity is 99.7509519372, simpson index is 0.961111834025, inverse simpson is 25.7147637318, entropy is 3.46208601643
[2013-08-20 10:28:55] Computing usearch statistics for 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim.ualn
[2013-08-20 10:28:58] Generated 16SMIRCEABAL_TAAGGCGA-TAGATCGC_L001_R2_001_trim_freq.csv. Mean identity is 99.7560904951, simpson index is 0.961457534643, inverse simpson is 25.9454082853, entropy is 3.46928575527
[2013-08-20 10:28:58] Plotting sequences lengths histogram for References/M_bal_Nex_full16S.fasta
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S_length-histogram.png
[2013-08-20 10:28:59] Plotting GC percentage plot for References/M_bal_Nex_full16S.fasta
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S_GC-percentage.png
[2013-08-20 10:28:59] Searching References/M_bal_Nex_full16S.fasta against References/16S.fa using usearch on 20 threads and matching identity of 0.95
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S.aln, M_bal_Nex_full16S.ualn, M_bal_Nex_full16S_matched.fasta, and M_bal_Nex_full16S_notmatched.fasta
[2013-08-20 10:28:59] M_bal_Nex_full16S_matched.fasta has 45 reads
[2013-08-20 10:28:59] M_bal_Nex_full16S_matched.fasta has 45 reads
[2013-08-20 10:28:59] Computing usearch statistics for M_bal_Nex_full16S.ualn
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S_freq.csv. Mean identity is 98.6022222222, simpson index is 0.972839506173, inverse simpson is 36.8181818182, entropy is 3.65262978298
[2013-08-20 10:28:59] Dereplicating duplicate sequences in References/M_bal_Nex_full16S.fasta using usearch on 20 threads, annotating with cluster sizes, and sorting by cluster size
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S_derep.fasta
[2013-08-20 10:28:59] Sorting reads in M_bal_Nex_full16S_derep.fasta by abundance using usearch
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S_derep_sorted.fasta
[2013-08-20 10:28:59] Detecting De novo chimeras in M_bal_Nex_full16S_derep_sorted.fasta using UCHIME
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S_derep_sorted_nonchimera.fasta and M_bal_Nex_full16S_derep_sorted_chimera.fasta
[2013-08-20 10:28:59] Searching M_bal_Nex_full16S_derep_sorted_nonchimera.fasta against References/16S.fa using usearch on 20 threads and matching identity of 0.97
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S_derep_sorted_nonchimera.aln, M_bal_Nex_full16S_derep_sorted_nonchimera.ualn, M_bal_Nex_full16S_derep_sorted_nonchimera_matched.fasta, and M_bal_Nex_full16S_derep_sorted_nonchimera_notmatched.fasta
[2013-08-20 10:28:59] Computing usearch statistics for M_bal_Nex_full16S_derep_sorted_nonchimera.ualn
[2013-08-20 10:28:59] Generated M_bal_Nex_full16S_derep_sorted_nonchimera_freq.csv. Mean identity is 98.9829268293, simpson index is 0.975609756098, inverse simpson is 41.0, entropy is 3.7135720667
[2013-08-20 10:28:59] M_bal_Nex_full16S_derep_sorted_nonchimera.fasta has 88 reads
[2013-08-20 10:28:59] M_bal_Nex_full16S_derep_sorted_chimera.fasta has reads
[2013-08-20 10:28:59] M_bal_Nex_full16S_derep_sorted_nonchimera_matched.fasta has 41 reads

[2013-08-20 10:28:59] M_bal_Nex_full16S_derep_sorted_nonchimera_notmatched.fasta has 47 reads
[2013-08-20 10:28:59] Saving pipeline statistics in AMPLImock_stats.csv. Here is the summary of results:

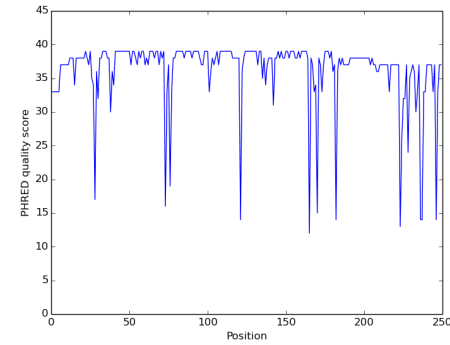
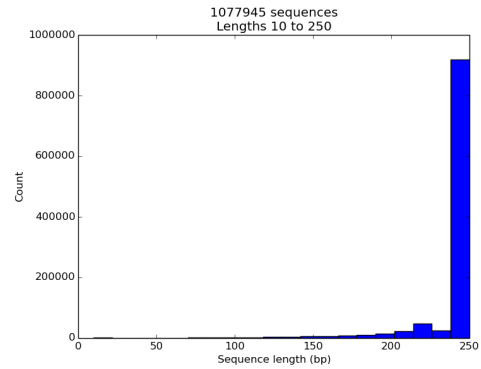
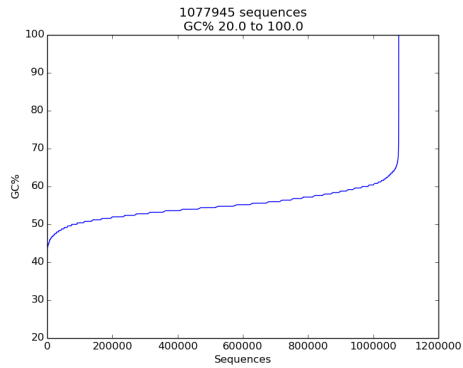
FORWARD_MATCHED_TOTAL_READS,607078
FORWARD_NOTMATCHED_TOTAL_READS,607078
FORWARD_READS_ENTROPY,3.46208601643
FORWARD_READS_INVERSE_SIMPSON,25.7147637318
FORWARD_READS_MEAN_IDENT,99.7509519372
FORWARD_READS_MOST_ABUNDANT_SEQUENCE,TCCTTAGAGTGCCAGCCGAAGTCTGCGCTCGTTGCCGGACTTAACCGAACATCTCAC
GACACGAGCTGACGACAGCCATGCAGCACCTGTCCTGCGTCACCGAAGTGAACGCCCGATCTCTCGGGTTAGCACAGGATGTCAAGGGTTGGTAAGGTTCTGCGCGTTGCTTCGA
ATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTT
FORWARD_READS_MOST_ABUNDANT_SEQUENCE_READS,1029
FORWARD_READS_MOST_ABUNDANT_SEQUENCE_READS_PERCENTAGE,0.0954594
FORWARD_READS_SIMPSON,0.961111834025
FORWARD_READS_TOTAL_READS,1077945
FORWARD_READS_UNIQUE_READS,353086
FORWARD_READS_UNIQUE_READS_PERCENTAGE,32.7555
OVERLAP_DEREPEP_CHIM_TOTAL_READS,
OVERLAP_DEREPEP_NONCHIM_MATCHED_TOTAL_READS,41
OVERLAP_DEREPEP_NONCHIM_NOTMATCHED_TOTAL_READS,47
OVERLAP_DEREPEP_NONCHIM_READS_ENTROPY,3.7135720667
OVERLAP_DEREPEP_NONCHIM_READS_INVERSE_SIMPSON,41.0
OVERLAP_DEREPEP_NONCHIM_READS_MEAN_IDENT,98.9829268293
OVERLAP_DEREPEP_NONCHIM_READS_SIMPSON,0.975609756098
OVERLAP_DEREPEP_NONCHIM_TOTAL_READS,88
OVERLAP_MATCHED_TOTAL_READS,45
OVERLAP_NOTMATCHED_TOTAL_READS,45
OVERLAP_READS_ENTROPY,3.65262978298
OVERLAP_READS_INVERSE_SIMPSON,36.8181818182
OVERLAP_READS_MEAN_IDENT,98.6022222222
OVERLAP_READS_SIMPSON,0.972839506173
REVERSE_MATCHED_TOTAL_READS,602332
REVERSE_NOTMATCHED_TOTAL_READS,602332
REVERSE_READS_ENTROPY,3.46928575527
REVERSE_READS_INVERSE_SIMPSON,25.9454082853
REVERSE_READS_MEAN_IDENT,99.7560904951

```
REVERSE_READS_MOST_ABUNDANT_SEQUENCE,TCCTTAGAGTGCCCGAAGTGCCTGGCAACTAAGGATGTGGGTGCGCTCGTTGCCGGACTTAACCGAACATCTCAC
GACACGAGCTGACGACAGCCATGCAGCACCTGTCACTGCGTCACCGAAGTGAACGCCCAGATCTCTCGGGTTAGCACAGGATGTCAAGGGTTGGTAAGGTTCTGCGCGTTGCTTCGA
ATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTT
REVERSE_READS_MOST_ABUNDANT_SEQUENCE_READS,1280
REVERSE_READS_MOST_ABUNDANT_SEQUENCE_READS_PERCENTAGE,0.118744
REVERSE_READS_SIMPSON,0.961457534643
REVERSE_READS_TOTAL_READS,1077945
REVERSE_READS_UNIQUE_READS,338070
REVERSE_READS_UNIQUE_READS_PERCENTAGE,31.3625
[uzi@quince-srv2 ~/AMPLImock/test_AMPLImock]$
```

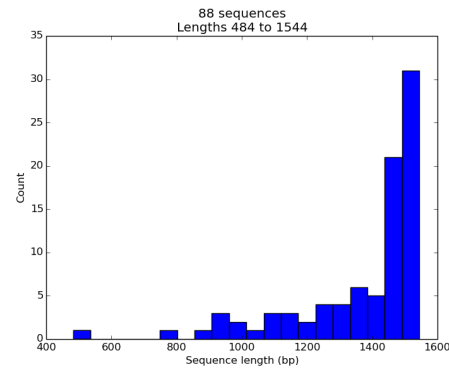
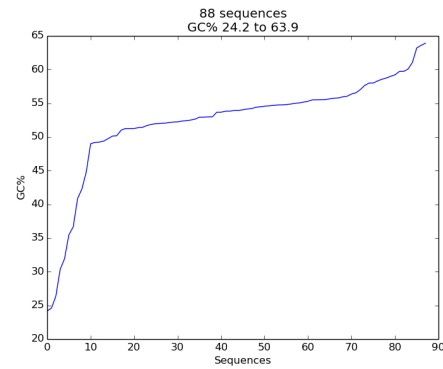
R1:



R2:



Assembled long amplicons:



[454 single-end reads](#)

```
[uzi@quince-srv2 ~/AMPLImock/test_AMPLImock]$ ../AMPLImock -q 454/454AllReads.fna -d References/16S.fa -e  
References/16S_R.fa -l ~/TSBAmplicons/References/IDs.txt -u
```

This is AMPLImock v0.1. Copyright (c) 2013 Computational Microbial Genomics Group, University of Glasgow, UK

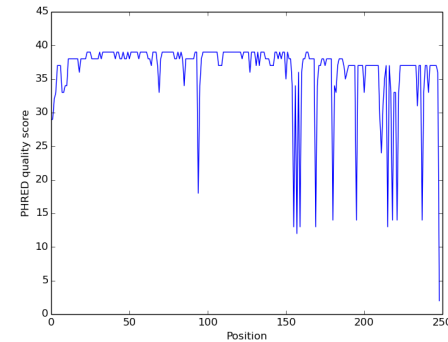
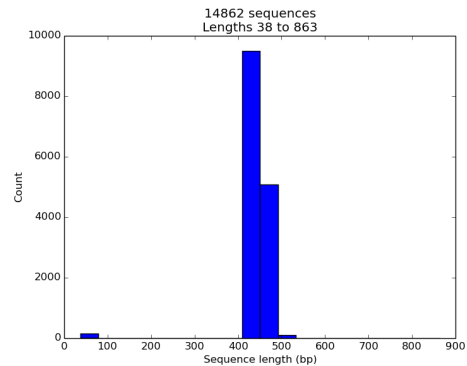
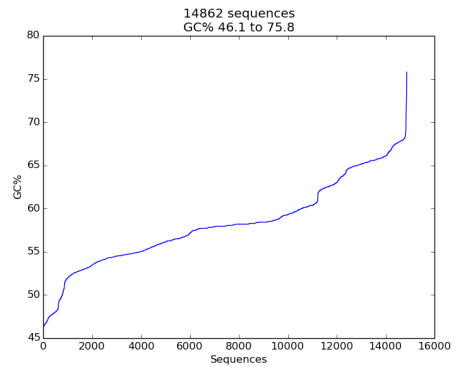
Checking for 'sickle': found /home/opt/sickle/sickle

Checking for 'usearch': found /home/opt/usearch_v6.0.307/usearch

Checking for 'pandaseq': found /home/opt/pandaseq/pandaseq

Forward reference database:References/16S.fa


```
Reverse reference database:References/16S_R.fa
List file containing mapping from multiple 16S genes:/home/uzi/TSBAmplicons/References/IDs.txt
Single reads:454/454AllReads.fna
Query file is in fasta format
Query file will not be trimmed
Processing the data in /home/uzi/AMPLImock/test_AMPLImock
[2013-08-20 10:04:18] 454/454AllReads.fna has 14862 reads
[2013-08-20 10:04:29] Plotting sequences lengths histogram for 454/454AllReads.fna
[2013-08-20 10:04:30] Generated 454AllReads_length-histogram.png
[2013-08-20 10:04:30] Plotting GC percentage plot for 454/454AllReads.fna
[2013-08-20 10:04:31] Generated 454AllReads_GC-percentage.png
[2013-08-20 10:04:31] Searching 454/454AllReads.fna against References/16S.fa using usearch on 10 threads and
matching identity of 0.95
[2013-08-20 10:04:32] Generated 454AllReads.aln, 454AllReads.ualn, 454AllReads_matched.fasta, and
454AllReads_notmatched.fasta
[2013-08-20 10:04:32] 454AllReads_matched.fasta has 14428 reads
[2013-08-20 10:04:32] 454AllReads_notmatched.fasta has 434 reads
[2013-08-20 10:04:32] Computing usearch statistics for 454AllReads.ualn
[2013-08-20 10:04:32] Generated 454AllReads_freq.csv. Mean identity is 98.8918353202, simpson index is
0.929426580512, inverse simpson is 14.1696407409, entropy is 3.31984340836
[2013-08-20 10:04:32] Saving pipeline statistics in AMPLImock_stats.csv. Here is the summary of results:
SINGLE_MATCHED_TOTAL_READS,14428
SINGLE_NOTMATCHED_TOTAL_READS,434
SINGLE_READS_ENTROPY,3.31984340836
SINGLE_READS_INVERSE_SIMPSON,14.1696407409
SINGLE_READS_MEAN_IDENT,98.8918353202
SINGLE_READS_SIMPSON,0.929426580512
SINGLE_READS_TOTAL_READS,14862
```



[IonTorrent single-end reads](#)

```
[uzi@quince-srv2 ~/AMPLImock/test_AMPLImock]$ ../AMPLImock -q ../IonTorrent/TED65mircea_mock_V4.fastq -d
References/16S.fa -e References/16S_R.fa -l ~/TSBAmplicons/References/IDs.txt -u -s 20
This is AMPLImock v0.1. Copyright (c) 2013 Computational Microbial Genomics Group, University of Glasgow, UK
Checking for 'sickle': found /home/opt/sickle/sickle
Checking for 'usearch': found /home/opt/usearch_v6.0.307/usearch
Checking for 'pandaseq': found /home/opt/pandaseq/pandaseq
Forward reference database:References/16S.fa
Reverse reference database:References/16S_R.fa
List file containing mapping from multiple 16S genes:/home/uzi/TSBAmplicons/References/IDs.txt
Single reads:../IonTorrent/TED65mircea_mock_V4.fastq
Query file is in fastq format
Query file will be trimmed
Processing the data in /home/uzi/AMPLImock/test_AMPLImock
[2013-08-20 10:07:58] Trimming ../IonTorrent/TED65mircea_mock_V4.fastq using quality type as 'sanger', quality score
of 20, and minimum of 10 bp to keep
[2013-08-20 10:08:05] Generated TED65mircea_mock_V4_trim.fastq
[2013-08-20 10:08:18] Plotting PHRED quality scores for TED65mircea_mock_V4_trim.fastq
[2013-08-20 10:08:18] Generated TED65mircea_mock_V4_trim_phred-quality.png
[2013-08-20 10:08:22] TED65mircea_mock_V4_trim.fastq has following stats:
Total reads:1399727
Unique reads:318235
Unique reads percentage:22.7355
```

Most abundant sequence:TCAGACATACGCGTGTGGTCC
Most abundant sequence reads:121010
Most abundant sequence read percentage:8.64526
[2013-08-20 10:08:22] Converting TED65mircea_mock_V4_trim.fastq to TED65mircea_mock_V4_trim.fasta
[2013-08-20 10:08:34] Generated TED65mircea_mock_V4_trim.fasta
[2013-08-20 10:08:34] Plotting sequences lengths histogram for TED65mircea_mock_V4_trim.fasta
[2013-08-20 10:08:56] Generated TED65mircea_mock_V4_trim_length-histogram.png
[2013-08-20 10:08:56] Plotting GC percentage plot for TED65mircea_mock_V4_trim.fasta
[2013-08-20 10:09:56] Generated TED65mircea_mock_V4_trim_GC-percentage.png
[2013-08-20 10:09:56] Searching TED65mircea_mock_V4_trim.fasta against References/16S.fa using usearch on 20 threads and matching identity of 0.95
[2013-08-20 10:11:11] Generated TED65mircea_mock_V4_trim.aln, TED65mircea_mock_V4_trim.ualn, TED65mircea_mock_V4_trim_matched.fasta, and TED65mircea_mock_V4_trim_notmatched.fasta
[2013-08-20 10:11:12] TED65mircea_mock_V4_trim_matched.fasta has 361390 reads
[2013-08-20 10:11:12] TED65mircea_mock_V4_trim_notmatched.fasta has 1038337 reads
[2013-08-20 10:11:12] Computing usearch statistics for TED65mircea_mock_V4_trim.ualn
[2013-08-20 10:11:14] Generated TED65mircea_mock_V4_trim_freq.csv. Mean identity is 95.8572782867, simpson index is 0.956480548939, inverse simpson is 22.9782310124, entropy is 3.4093415412
[2013-08-20 10:11:14] Saving pipeline statistics in AMPLImock_stats.csv. Here is the summary of results:
SINGLE_MATCHED_TOTAL_READS,361390
SINGLE_NOTMATCHED_TOTAL_READS,1038337
SINGLE_READS_ENTROPY,3.4093415412
SINGLE_READS_INVERSE_SIMPSON,22.9782310124
SINGLE_READS_MEAN_IDENT,95.8572782867
SINGLE_READS_MOST_ABUNDANT_SEQUENCE,TCAGACATACGCGTGTGGTCC
SINGLE_READS_MOST_ABUNDANT_SEQUENCE_READS,121010
SINGLE_READS_MOST_ABUNDANT_SEQUENCE_READS_PERCENTAGE,8.64526
SINGLE_READS_SIMPSON,0.956480548939
SINGLE_READS_TOTAL_READS,1399727
SINGLE_READS_UNIQUE_READS,318235
SINGLE_READS_UNIQUE_READS_PERCENTAGE,22.7355

