

n	n:500	L50	min	N75	N50	N25	E-size	max	sum	name
7833	1218	260	504	2911	5169	8128	6149	33861	4229145	abyss-unitigs.fa
6724	375	64	504	10943	20576	33205	24093	68455	4219112	abyss-contigs.fa
5685	129	19	509	41283	82083	120219	81777	187944	4247698	abyss-scaffolds.fasta

2)

Install ABySS

```
conda install -c bioconda -c conda-forge abyss
```

Installs the conda packages

Assemble a small synthetic data set

```
wget http://www.bcgsc.ca/platform/bioinfo/software/abyss/releases/1.3.4/test-data.tar.gz
```

```
tar xzvf test-data.tar.gz
```

```
abyss-pe k=25 name=test B=1G \
```

```
in='test-data/reads1.fastq test-data/reads2.fastq'
```

the abyss folder used to assemble the data with k-mer values set to 25

The higher the k-mer value the higher the quality of data

Assembling a paired-end library

```
abyss-pe name=ecoli k=96 B=2G in='reads1.fa reads2.fa'
```

Assembles to make contigs from file to the named file.

In= 'with files that have fasta or fastq'

Name= what the file named will be. It will end with the -scaffolds.fa

Assembling with multiple libraries

```
abyss-pe k=96 B=2G name=ecoli lib='pea peb' \
```

```
pea='pea_1.fa pea_2.fa' peb='peb_1.fa peb_2.fa' \
```

```
se='se1.fa se2.fa'
```

Assembles fragment size of each library calculated empirically to the contigs

Two files will be created to view plots with calculated and expected.

Scaffoldings

```
abyss-pe k=96 B=2G name=ecoli lib='pea peb' mp='mpc mpd' \
```

```
pea='pea_1.fa pea_2.fa' peb='peb_1.fa peb_2.fa' \
```

```
mpc='mpc_1.fa mpc_2.fa' mpd='mpd_1.fa mpd_2.fa'
```

Larger assemblies can be used to create scaffolds that will have a final file name of -scaffolds.fa

Scaffolding with linked reads

```
abyss-pe k=96 B=2G name=hsapiens \
```

```
pe='pea' pea='lra.fastq.gz' \
```

```
mp='mpa' mpa='lra.fastq.gz' \
```

```
lr='lra' lra='lra.fastq.gz'
```

Can create a scaffold with linked reads, make-pairs and paired-ends

Rescaffolding with long sequences

```
abyss-pe k=96 B=2G name=ecoli lib='pe1 pe2' mp='mp1 mp2' long='longa' \
```

```
pe1='pe1_1.fa pe1_2.fa' pe2='pe2_1.fa pe2_2.fa' \
```

```
mp1='mp1_1.fa mp1_2.fa' mp2='mp2_1.fa mp2_2.fa' \
```

```
longa='longa.fa'
```

Assembles long contigs to make scaffolds

File will be final with -long-scaffs.fa

Assembling using a paired de Bruijn graph

```
abyss-pe name=ecoli K=16 k=96 in='reads1.fa reads2.fa'
```

the k-mer pairs are similar to long k-mers except it will be used twice and have been spaced out. Difference is that the spaces will not be shown.

Assembling a strand-specific RNA-Seq library

```
abyss-pe name=SS-RNA B=2G k=96 in='reads1.fa reads2.fa' SS=--SS
```

can be assembled with respect to the original transcript and the orientation of the unitigs, contigs and scaffolds.

3) Using hybrid assembly is for combining both long and short reads together to form a higher quality assembly. It corrects errors in third generation sequencing and anything that's unsure in the second generation sequencing.

Add the codes --trusted-contigs or --untrusted-contigs for the program to interact with files that have been assembled using other programs to allow it to make hybrids.

4)

SPAdes results

[All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs ≥ 0 bp)" and "Total length ≥ 0 bp)" include all contigs].

Assembly	scaffolds
# contigs ≥ 0 bp)	1186
# contigs ≥ 1000 bp)	132
# contigs ≥ 5000 bp)	108
# contigs ≥ 10000 bp)	88
# contigs ≥ 25000 bp)	49
# contigs ≥ 50000 bp)	31
Total length ≥ 0 bp)	4384077
Total length ≥ 1000 bp)	4223758
Total length ≥ 5000 bp)	4164770
Total length ≥ 10000 bp)	4022767
Total length ≥ 25000 bp)	3354060
Total length ≥ 50000 bp)	2684365
# contigs	142
Largest contig	210769
Total length	4230630
GC (%)	34.87
N50	62320
N90	16282
auN	74034.3
L50	22
L90	72
# N's per 100 kbp	2.36

ABYSS results

All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

Assembly	scaffolds
# contigs (≥ 0 bp)	5685
# contigs (≥ 1000 bp)	120
# contigs (≥ 5000 bp)	98
# contigs (≥ 10000 bp)	78
# contigs (≥ 25000 bp)	47
# contigs (≥ 50000 bp)	31
Total length (≥ 0 bp)	4967305
Total length (≥ 1000 bp)	4247711
Total length (≥ 5000 bp)	4196856
Total length (≥ 10000 bp)	4060598
Total length (≥ 25000 bp)	3517537
Total length (≥ 50000 bp)	2935056
# contigs	130
Largest contig	188244
Total length	4254156
GC (%)	34.87
N50	82133
N90	18606
auN	81861.6
L50	19
L90	62
# N's per 100 kbp	140.36

7)

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Select columns

Show

100

?

☒ select all
 100 sequences selected

[GenBank](#)
[Graphics](#)
[Distance tree of results](#)
[MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_AL20115_chromosome_complete_genome	Cyanobacterium...	2739	8219	100%	0.0	100.00%	4267295	CP138348.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_PCC_10605_chromosome_complete_genome	Cyanobacterium...	2739	8180	100%	0.0	100.00%	4114099	CP003947.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_UTEX_3221_chromosome_complete_genome	Cyanobacterium...	2739	8208	100%	0.0	100.00%	4280321	CP141254.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_AL20118_chromosome_complete_genome	Cyanobacterium...	2739	8213	100%	0.0	100.00%	4266140	CP149437.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_UTEX_3222_chromosome_complete_genome	Cyanobacterium...	2734	8202	100%	0.0	99.93%	4417119	CP141255.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_strain_PCC_10605_16S_ribosomal_RNA_partial_sequence	Cyanobacterium...	2723	2723	99%	0.0	100.00%	1474	NR_102443.1
<input checked="" type="checkbox"/>	Cyanobacterium_sp._Dongsha4_chromosome_complete_genome	Cyanobacterium...	2695	8086	100%	0.0	99.46%	4250085	CP084098.1
<input checked="" type="checkbox"/>	Cyanobacterium_IHB-410_16S_ribosomal_RNA_gene_partial_sequence	cyanobacterium...	2643	2643	97%	0.0	99.79%	1440	JF966679.1
<input checked="" type="checkbox"/>	Cyanobacterium_sp._MCCB_115_16S_ribosomal_RNA_gene_partial_sequence	Cyanobacterium...	2625	2625	98%	0.0	99.38%	1449	KM275589.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_IkSCC30_16S_ribosomal_RNA_gene_and_16S-23S_ribosomal_RNA_intergenic_spac...	Cyanobacterium...	2625	2625	96%	0.0	100.00%	1728	KM438201.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_ETS-03_16S_rRNA_gene_(partial)_ITS1_tRNA-Ile_gene_tRNA-Ala_gene_and_23S_r...	Cyanobacterium...	2608	2608	97%	0.0	99.44%	1786	AM238427.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_ThrSCSp4_16S_ribosomal_RNA_gene_and_16S-23S_ribosomal_RNA_intergenic_sp...	Cyanobacterium...	2567	2567	94%	0.0	99.79%	1710	KM438200.1
<input checked="" type="checkbox"/>	Cyanobacterium_sp._MCCB_238_16S_ribosomal_RNA_gene_partial_sequence	Cyanobacterium...	2523	2523	94%	0.0	99.28%	1397	KM275590.1
<input checked="" type="checkbox"/>	Cyanobacterium_sp._MCCB_114_16S_ribosomal_RNA_gene_partial_sequence	Cyanobacterium...	2488	2488	93%	0.0	99.13%	1384	KM275588.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_Kh.A_16S_ribosomal_RNA_gene_partial_sequence	Cyanobacterium...	2486	2486	92%	0.0	99.63%	1360	MH179053.1
<input checked="" type="checkbox"/>	Cyanobacterium_aponinum_KSU-WH-5_16S_ribosomal_RNA_gene_partial_sequence	Cyanobacterium...	2477	2477	91%	0.0	99.78%	1350	KT807478.1
<input checked="" type="checkbox"/>	Uncultured_bacterium_clone_LVB1_16S_ribosomal_RNA_gene_partial_sequence	uncultured bacte...	2473	2473	97%	0.0	97.58%	1449	KJ465919.1
<input checked="" type="checkbox"/>	Uncultured_bacterium_clone_VDB19_16S_ribosomal_RNA_gene_partial_sequence	uncultured bacte...	2468	2468	97%	0.0	97.51%	1450	KJ465958.1
<input checked="" type="checkbox"/>	Synechocystis_aquatilis_SAG_90.79_16S_ribosomal_RNA_gene_partial_sequence	Synechocystis a...	2399	2399	98%	0.0	96.37%	1459	KM020011.1
<input checked="" type="checkbox"/>	Uncultured_bacterium_clone_PHS5_16S_ribosomal_RNA_gene_partial_sequence	uncultured bacte...	2385	2385	88%	0.0	99.77%	1300	KX062019.1
<input checked="" type="checkbox"/>	Geminocystis_sp._NIES-3709_DNA_complete_genome	Geminocystis sp...	2383	7149	100%	0.0	95.69%	4150181	AP014821.1
<input checked="" type="checkbox"/>	Synechococcus_elongatus_str._Ramsar_16S_ribosomal_RNA_gene_partial_sequence	Synechococcus ...	2374	2374	90%	0.0	98.73%	1340	JQ771323.1

10)

spadesoutgalaxy/scaffolds.fasta	neighbors/Dongsha4.fasta	92.9712	1094	1342
spadesoutgalaxy/scaffolds.fasta	neighbors/stanieri.fasta	77.5466	244	1342