***Recipes for genome assemblies***

The following commands were used to generate the assemblies:

**ABySS**

```
abyss-pe k=K l=1 n=5 s=100 name=asm lib='reads' reads=reads.fastq
aligner=bowtie
```

with the following values for the kmer parameter `K`:

K=31 for *R. sphaeroides* HiSeq data

K=49 for *B. cereus* MiSeq data, *R. sphaeroides* MiSeq data *X. axonipodis* HiSeq data

K=51 for *V. cholerae* HiSeq data

K=53 for *M. abscessus* HiSeq data

K=58 for *M. abscessus* MiSeq data

K=65 for *A. hydrophila* HiSeq data*, B. cereus* HiSeq data, *V. cholerae* MiSeq data, *S. aureus* HiSeq data

K=83 for *B. fragilis* HiSeq data

**CABOG**

```
fastqToCA  -insertsize M S -libraryname reads -mates
reads1.fastq,reads2.fastq > reads.frg
runCA –d . –p asm –s config reads.frg>&runCA.log
```

with config file specifying

```
unitiger = bog
```

and the corresponding mean `M` and standard deviation `S` of the library's insert size:

M=180 and S=20 for *B. cereus* HiSeq data, *B. fragilis* HiSeq data, *A. hydrophila* HiSeq data, *S. aureus* HiSeq data

M=220 and S=25 for *R. sphaeroides* HiSeq data

M=335 and S=35 for *M. abscessus* HiSeq data, *M. abscessus* MiSeq data, *V. cholerae* HiSeq data, *V. cholerae* MiSeq data

M=400 and S=40 for *X. axonopodis* HiSeq data

M=540 and S=60 for *R. sphaeroides* MiSeq data

M=600 and S=60 for *B. cereus* MiSeq data

**MIRA**

```
srrname=SRRxxxxxx
numreads=xxxxxxx
strainname="NA"
numlines=$((4*${numreads}))
```

```
cat reads1.fastq | head  -${numlines} | sed  -e ‘s/SRR[0-9.]*/&\/1/’
>${strainname}-${numreads} in.solexa.fastsq
cat reads2.fastq | head  -${numlines} | sed  -e ‘s/SRR[0-9.]*/&\/1/’
>${strainname}-${numreads}_in.solexa.fastsq
grep “@SRR” ${strainname}-${numreads}_in.solexa.fastq | cut –f 1 –d ‘
‘ | sed  -e ‘s/@//’  -e “s/$/  ${strainname}/” >> ${strainname}-
${numreads}_straindata_in.txt
ln –s NA-numreads_in.solexa.fastq mira_in.solexa.fastq
ln –s NA-numreads_straindata_in.txt mira_straindata_in.txt
mira  –fastq  –job=denovo,genome,accurate,solexa SOLEXA_SETTINGS –
GE:tismin=MIN:tismax=MAX –LR:file_type=fastq –
AS:mrpc=5>&log_assembly.txt
```

with `srrname` and `numreads` containing the correct values for each run, and `MIN` and `MAX` having the following values:

`MIN`=90 and `MAX`=270 for *A. hydrophila HiSeq* data, *B. cereus* HiSeq data, *B. fragilis* HiSeq data, *S. aureus* HiSeq data

`MIN`=110 and `MAX`=330 for *R. sphaeroides* HiSeq data

`MIN`=167 and `MAX`=502 for *M. abscessus* HiSeq data, *M. abscessus* MiSeq data, *V. cholerae* HiSeq data, *V. cholerae* MiSeq data

`MIN`=200 and `MAX`=600 for *X. axonopodis* HiSeq data

`MIN`=270 and `MAX`=810 for *R. sphaeroides* MiSeq data

`MIN`=300 and `MAX`=900 for *B. cereus* MiSeq data

**MSRCA**

```
runSRCA.pl config
./assemble
```

where `config` file contains the following information:

```
PATHS
JELLYFISH_PATH=/full/path/to/MSR-CA/bin
SR_PATH=/full/path/to/MSR-CA/bin
CA_PATH=/full/path/to/Cabog_installation/bin
END

DATA
PE= p1 M S reads1.fastq reads2.fastq
END

PARAMETERS
GRAPH_KMER_SIZE=K
NUM_THREADS=t
```

```
JF_SIZE=2000000000

END
```

with M and S set to correct mean and standard deviation values for a particular data set (see the values for M and S in the description of Cabog assembler), and the following values of kmer K were used:
K=49 for *B. cereus* HiSeq data
K=55 for *R. sphaeroides* HiSeq data
K=63 for *R. sphaeroides* MiSeq data
K=79 for *S. aureus* HiSeq data
K=89 for *A. hydrophila* HiSeq data, *B. fragilis* HiSeq data, *M. abscessus* HiSeq data, *V. cholerae* HiSeq data, *X. axonopodis* HiSeq data
K=99 for *M. abscessus* MiSeq data, *V. cholerae* HiSeq data
K=101 for *B. cereus* MiSeq data

**SGA**

```
ln –s reads1.fastq frag1
ln –s reads2.fastq frag2
#!/bin/bash
K=kmer_value
CPU=8
MIN_OVERLAP=min_overlap
ASSEMBLE_OVERLAP=assemble_overlap
MIN_PAIRS=5
sga preprocess --pe-mode 1 –o reads.pp.fastq frag1 frag2
sga index --algorithm=ropebwt –t $CPU reads.pp.fastq
sga correct –k $K –t $CPU –o reads.ec.fastq reads.pp.fastq
sga index --algorithm=ropebwt –t $CPU reads.ec.fastq
sga filter –t $CPU reads.ec.fastq
sga overlap –m $MIN_OVERLAP –t $CPU reads.ec.filter.pass.fa
sga assemble –o primary reads.ec.filter.pass.asqg.gz
ln –s primary-contigs.fa.ctg.fasta
bwa index ctg.fasta
bwa aln –t $CPU ctg.fasta frag1 > frag1.sai
bwa aln –t $CPU ctg.fasta frag2 > frag2.sai
bwa sampe ctg.fasta frag1.sai frag2.sai frag1 frag2 > frag.sam
samtools view –Sb frag.sam > libPE.bam
sga-bam2de.pl –n $MIN_PAIRS --prefix libPE libPE.bam
sga-astat.py libPE.bam > libPE.astat
sga scaffold –m 200 –a libPE.astat –o scf --pe libPE.de ctg.fasta
sga scaffold2fasta –a primary-graph.asqg.gz –o scf.fasta scf
```

with the following values used for `kmer_value (K)`,`min_overlap (M)`, and `assemble_overlap (A)`:

K=23, M=85, A= 111 for *R. sphaeroides* MiSeq data
K=41, M=45, A= 45 for *R. sphaeroides* HiSeq data
K=55, M=45, A= 45 for *A. hydrophila* HiSeq data
K=65, M=45, A= 45 for *B.cereus* HiSeq data, *B. fragilis* HiSeq data, *M. abscessus* HiSeq data, *V. cholerae* HiSeq data, *X. axonopodis* HiSeq data
K=73, M=45, A= 45 for *S. aureus* HiSeq data
K=65, M=85, A= 111 for *B.cereus* MiSeq data, *M. abscessus* MiSeq data, *V. cholerae* MiSeq data

**SOAPdenovo2**

```
SOAPdenovo2 all -K kmer_value -F -R -E -w -u -s config -o asm -p 8>>
SOAPdenovo.log
GapCloser -b config -a asm.scafSeq -o asm.new.scafSeq -t 8 >>
SOAPdenovo.log
```

with config file containing the following information:

```
[LIB]
avg_ins=mean
reverse_seq=0
asm_flags=3
rank=1
q1=reads1.fastq
q2=reads2.fastq
```

with corresponding `mean` value for insert size (see cabog values for `M`), and with `kmer_value`:

K=47 *M. abscessus* MiSeq data
K=49 *M. abscessus* HiSeq data, *V. cholerae* MiSeq data
K=51 *V. cholerae* HiSeq data
K=55 for *B. cereus* MiSeq data *,* *R. sphaeroides* HiSeq data
K=65 for *B. cereus* HiSeq data
K=71 for *S. aureus* HiSeq data
K=79 for *A. hydrophila* HiSeq data, *B. fragilis* HiSeq data, *R. sphaeroides* MiSeq data, *X. axonopodis* HiSeq data

**SPAdes**

```
spades.py -t 2 -k K1,K2,K3 -1 reads1.fastq -s reads2.fastq -o output
```

with the following values for kmer values `K1,K2,K3`:

21,33,55 for *R. sphaeroides* HiSeq data

31,43,65 for *R. sphaeroides* MiSeq data
41,53,75 for *B. cereus* HiSeq data
51,63,85 for *B. cereus* MiSeq data, *X. axonopodis* HiSeq data
61,73,95 for *A. hydrophila* HiSeq data, *B. fragilis* HiSeq data, *S. aureus* HiSeq data
33,55,65,75,85,99 for *M. abscessus* HiSeq data, *M. abscessus* MiSeq data, *V. cholerae* HiSeq data, *V. cholerae* MiSeq data

**Velvet**

```
shuffleSequences_fastq.pl reads1.fastq reads2.fastq inputReads.fastq
velveth . K –fastq –shortPaired inputReads.fastq
velvetg . –exp_cov auto –ins_length M –ins_length_sd S –scaffolding
yes
```

with the corresponding mean `M` and standard deviation `S` (see values for M and S in cabog description), and the following values for kmer `K`:
K=31 for *R. sphaeroides* MiSeq data
K=49 for *R. sphaeroides* HiSeq data, *M. abscessus* HiSeq data, *V. cholerae* HiSeq data
K=63 for *A. hydrophila* HiSeq data, *B. cereus* HiSeq data, *B. cereus* MiSeq data, *X. axonopodis* HiSeq data
K=73 for *B. fragilis* Hiseq data, *S. aureus* HiSeq data
K=97 for *M. abscessus* MiSeq data, *V. cholerae* MiSeq data