

Targeted de novo assembly

In Session 8 we learned that contig "deg7180000006464" is mtDNA. There was something weird with that:

```
cd ~/session_7

samtools tview ../session_3/data/sample-1_realn.bam reference/ninespine.fa \
-p deg7180000006464
```

We can look at it more closely:

```
samtools faidx reference/ninespine.fa deg7180000006464 | wc -c
samtools faidx http://wasabiapp.org/vbox/data/session_7/threespine.fa MT | wc -c

mkdir mtDNA

samtools faidx reference/ninespine.fa deg7180000006464 > mtDNA/ninespine_mt.fa
samtools faidx http://wasabiapp.org/vbox/data/session_7/threespine.fa MT \
> mtDNA/threespine_mt.fa

exonerate -t mtDNA/ninespine_mt.fa -q mtDNA/threespine_mt.fa |less -S
exonerate -t mtDNA/ninespine_mt.fa -q mtDNA/threespine_mt.fa --showalignment no

dottup -sreverse1 mtDNA/threespine_mt.fa mtDNA/ninespine_mt.fa
```

What do we learn?

We can extract reads that aligned to a specific region from a bam file, and then reassemble those reads separately:

```
samtools view -h -Obam ../session_3/data/sample-1_realn.bam deg7180000006464 \
| samtools sort -n -T temp - \
| bedtools bamtofastq -i - -fq mtDNA/sample-1_1.fastq -fq2 mtDNA/sample-1_2.fastq

less mtDNA/sample-1_1.fastq
```

Note that we used "samtools sort" in a new way. What did we do?

We then use "SOAPdenovo-Trans" (despite it being targeted for transcriptome!) to assemble these reads. Edit and save the config file below as "mtDNA/soap.config":

```
#maximal read length
max_rd_len=150
[LIB]
#maximal read length in this lib
rd_len_cutoff=155
#average insert size
avg_ins=200
#if sequence needs to be reversed
```

```
reverse_seq=0
#in which part(s) the reads are used
asm_flags=3
#minimum aligned length to contigs for a reliable read location (at least 32 for
short insert size)
map_len=32
#fastq file for read 1
q1=/home/evolutionary/session_7/mtDNA/sample-1_1.fastq
#fastq file for read 2 always follows fastq file for read 1
q2=/home/evolutionary/session_7/mtDNA/sample-1_2.fastq
```

We can then run SOAPdenovo-Trans (this requires lots of RAM; on my system it failed when the VM had 1Gb of RAM and worked when it had 3GB of RAM):

```
SOAPdenovo-Trans-127mer all -s mtDNA/soap.config -o mtDNA/sample-1
```

We should see improvement:

```
grep ">" mtDNA/sample-1.scafSeq
less mtDNA/sample-1.scafSeq
```

```
dottup mtDNA/threespine_mt.fa mtDNA/sample-1.scafSeq
```