

PRACTICE EXERCISE

SEQUENCE ALIGNMENT

Objective: Align closure reads to BAC contigs to determine which gaps might be closed and to determine which POMP PCR products (TBCNZ01, TBCNZ02 and TBCNZ03) span which physical ends.

From your home directory, move to the closure data directory:

```
cd AssemblyData/BAC/closure/
```

This directory contains sequence data from closure sequencing for the sample BAC.

There are two directories contained in the closure directory:

`chromat_dir`- contains all the chromatograms for reads generated from a POMP PCR experiment and reads done for the sequencing gaps of the sample BAC.

`vector_dir`- contains all the vectors that the sequences need to be trimmed against.

Step 1: Basecalling/trimming

Before they can be aligned, the closure reads need base-calls assigned and need to be trimmed of vector and low quality.

The program *preTA* runs the base-caller and trimming software as required for TIGR Assembler. *preTA* converts the chromatograms into .phd files, trims all the low quality areas at the end of the sequences as well as all traces of vector, then generates a new set of .phd files.

preTA has already been run for the closure reads as part of the Assembly exercises.

Step 2: Generate fasta files for each closure read

The program *phd2fasta* converts the phd files in the `phd_dir` directory to fasta sequence and quality outputs suitable for alignments and TIGR Assembler input.

```
phd2fasta -id phd_dir -os closure.seq -oq closure.qual
```

Step 3: Separate the file `closure.seq` into individual files

```
seqretsplit closure.seq  
ls
```

There should now be a fasta file for each closure read.

Step 4: Align sequence gap reads against BAC contigs

Goal: Determine which sequencing gaps may close using these closure reads. Determine also the size of each sequence gap.

First, copy the BAC contigs fasta file to your current location:

```
cp ../scratch/edit_dir/bac.fasta .
```

Now, let's align some sequence gap walks. For example:

```
fasta34 TBCNC26_S6BD.i1.fasta bac.fasta | more
```

Try aligning the following sequence gap reads to the BAC contigs:

```
TBCNC26_S6BD.i1.fasta  
TBCNG17_S6BD.i1.fasta  
TBCNH10_S2EC.i1.fasta  
TBCNJ56_S1BC.i1.fasta  
TBCNJ56_S2EC.i1.fasta
```

Which gaps do these reads appear to close?

What is the size of gap 10? Gap 11?

Step 5: Align POMP PCR reads against BAC contigs

Goal: Determine which physical contig ends are adjacent and may close due to POMP reads. Determine physical gap sizes.

We have reads from a POMP PCR experiment in which 3 PCR products were obtained:

TBCNZ01 from primer POOLs 4 and 5.

TBCNZ02 from primer POOLs 4 and 6.

TBCNZ03 from primer POOLs 5 and 6.

Using the following sequences, run alignments to the BAC contigs to determine which contigs are adjacent and whether the physical gaps will close:

```
TBCNZ01_POOL4B.i1.fasta  
TBCNZ01_POOL4C.i1.fasta  
TBCNZ01_POOL5B.i1.fasta  
TBCNZ01_POOL5C.i1.fasta
```

```
TBCNZ02_POOL4A.i1.fasta  
TBCNZ02_POOL4C.i1.fasta  
TBCNZ02_POOL6B.i1.fasta  
TBCNZ02_POOL6C.i1.fasta
```

```
TBCNZ03_POOL5A.i1.fasta
```

TBCNZ03_POOL5C.i1.fasta
TBCNZ03_POOL6A.i1.fasta
TBCNZ03_POOL6C.i1.fasta

POMP PCR Experiment Conclusions:
(Fill in the blanks and select either beginning/end)

1. The PCR product TBCNZ01 links the beginning/end of contig _____ and the beginning/end of contig _____. The gap size is: _____.
2. The PCR product TBCNZ02 links the beginning/end of contig _____ and the beginning/end of contig _____. The gap size is: _____.
3. The PCR product TBCNZ03 links the beginning/end of contig _____ and the beginning/end of contig _____. The gap size is: _____.