F: Contig ordering

At the Wellcome Trust Sanger Institute we developed a tool called ABACAS (Assefa et al., 2009) to order contigs against a reference. Spaces between the contigs (gaps) will be *N* characters. The result is called a pseudo-molecule. It can be loaded into ACT (a bit like a sandwich of two Artemis views) and then be analysed.

In order to start ABACAS you need a reference sequence (Pf3D7_05.fasta) and the contigs (we assume k.assembly.49/contigs.fa - but you can use another assembly). Next you decide if you want to do a comparison of nucleotides (nucmer) or amino acids (promer).

```
$ abacas.1.3.1.pl -r Pf3D7_05.fasta -q k.columbus.55/
contigs.fa -p promer -b -d -a -o IT.ordered
```

Abacas has many options. We use
-b to generate a bin of contigs that don't map. This is very important
-a will append the bin onto the pseudo molecule
-d uses the standard comparison parameter, in this case faster
-o IT.ordered is the prefix for the output file.

The command

$ abacas.1.3.1.pl -h

will give you a complete list of all options.

```
-s  int     minimum length of exact matching word (nucmer default = 12,
            promer default = 4)
```

Higher values decrease the runtime for the price of sensitivity.

```
-e          Escape contig ordering i.e. go to primer design
```

If you just would like to generate primers over gaps regions.

```
-c          Reference sequence is circular
```

Once abacas is done, it indicates which data it generated and how to load them into Act:

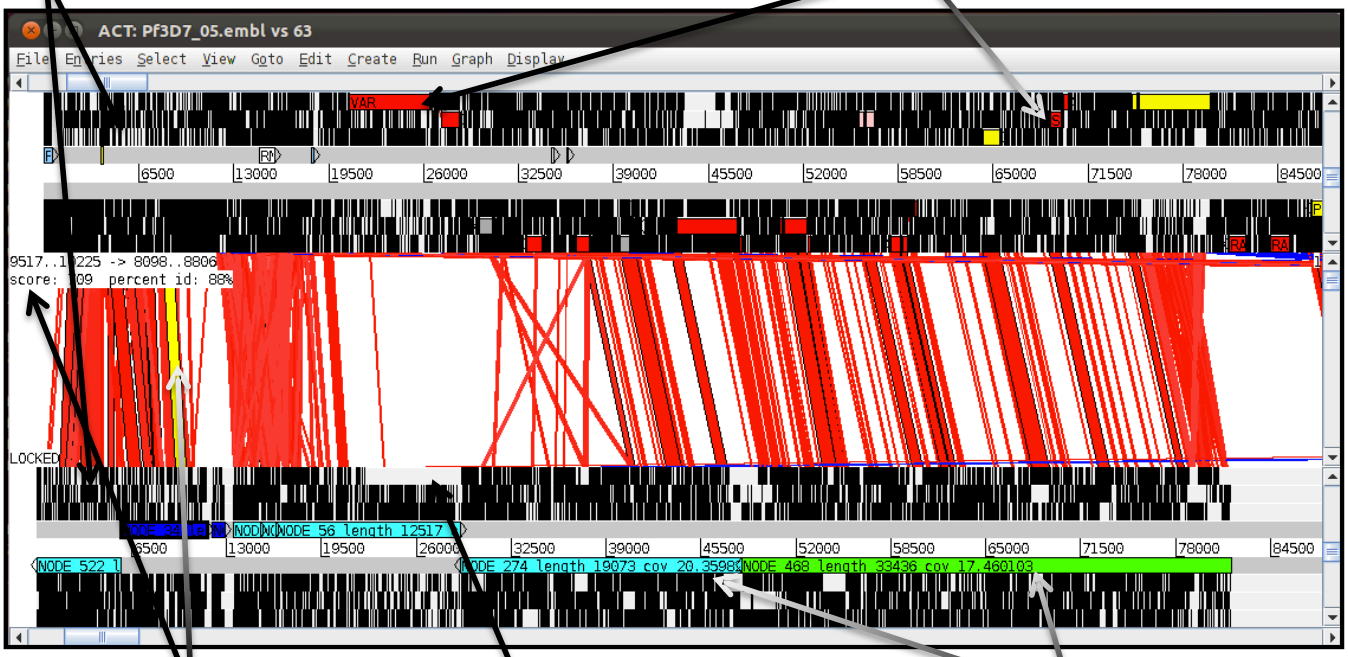
```
To view your results in ACT
                  Sequence file 1: Pf3D7_05.fasta
                  Comparison file 1: IT.ordered.crunch
                  Sequence file 2: IT.ordered.fasta

                 ACT feature file is: IT.ordered.tab

                 Contigs bin file is: IT.ordered.bin

                 Gaps in pseudomolecule are in: IT.ordered.gaps
```

Before opening the file in Act, we generate a BLAST comparison file:

```
$  formatdb -p F -i Pf3D7_05.fasta
$  blastall -p blastn -m 8 -e 1e-20 -d Pf3D7_05.fasta -i
IT.ordered.fasta -o comparison.blast
$ act Pf3D7_05.embl comparison.blast <(cat IT.ordered.tab
IT.ordered.fasta  ) &
```

Act has two Artemis windows

Gene models of reference



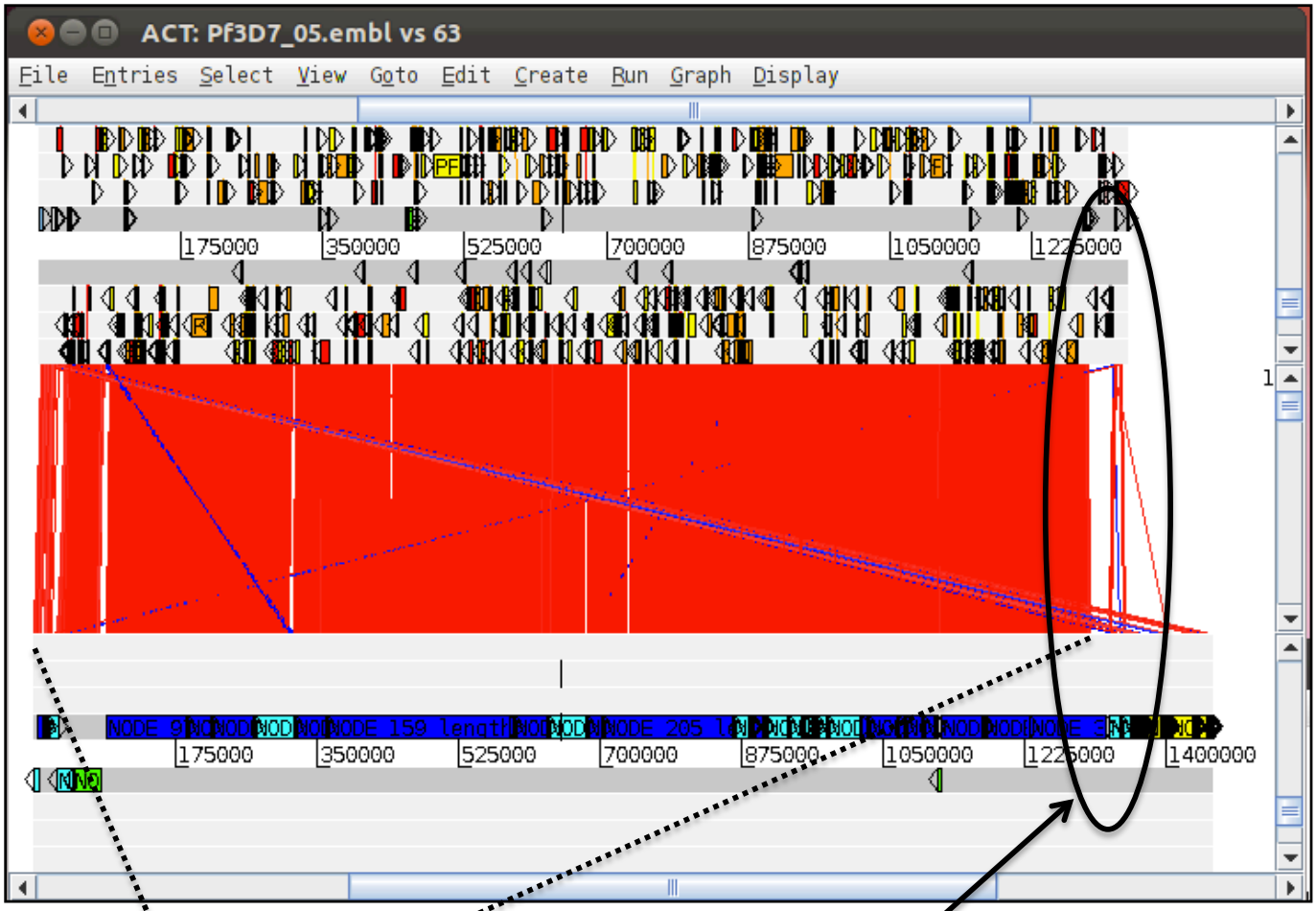
Similarity between the selected items

Interestingly, this is a VAR gene , which was not covered by mapped reads, see page 4!

Contigs ordered against the reference.

Scroll though the assembly. Maybe zoom in and out. How does it look? Are there any assembly errors?

What happened with the gene PF3D7_0532500?



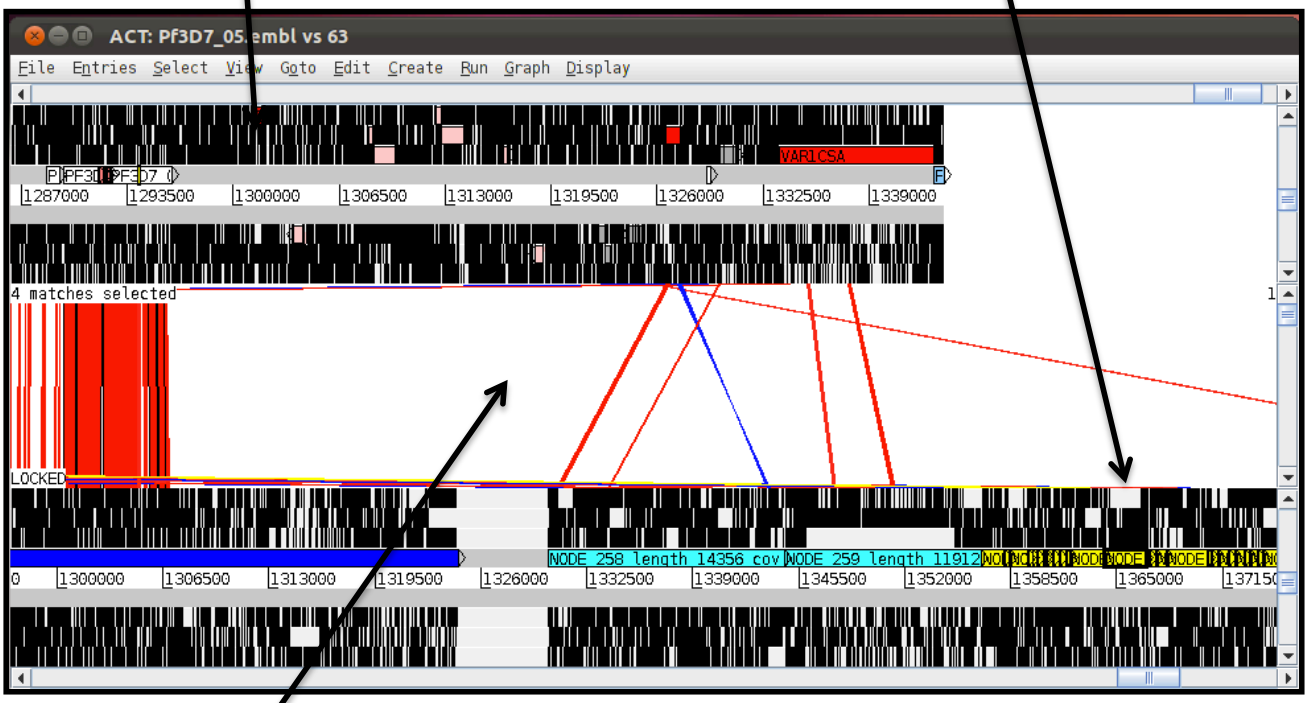
There are no big mis-assembly, as there are no synteny breaks.

Why is there no contig for the assembly here?

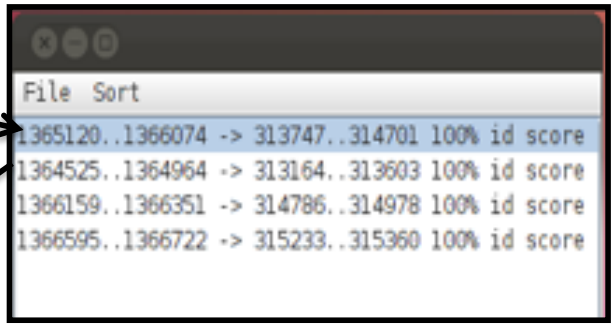
We ran abacas with the -a option. This means that contigs that didn't map against the reference are appended at the end. Scroll to the right hand site. Any idea what those contigs are? Could you order some into the core of the chromosome?

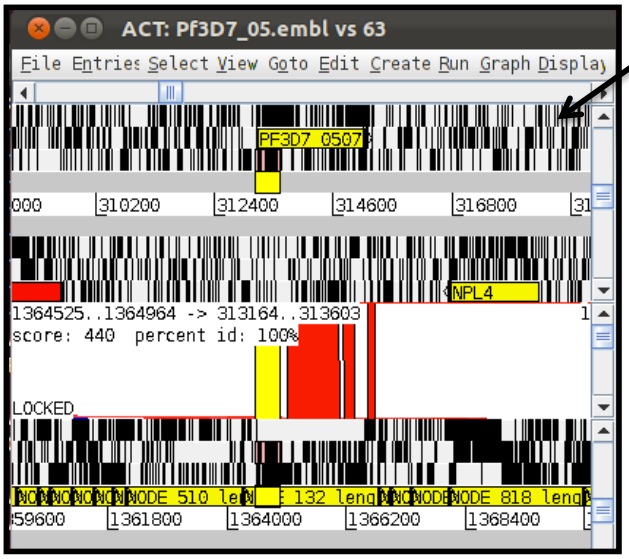
Has this gene not enough similarity or were the contigs not ordered well enough?

This looks like on open reading frame. Can you determine the function? It seems to have similarity with the reference.
Why is it not ordered against the reference?



1. Right click -> View selected matches. Double click on it.





This contig is indeed wrongly ordered. If you want you can do the following optional exercise to order the contig manually.

Don't close Act for the next exercise.