Module 4
Contig Ordering

Aims

The previous module will have demonstrated how you can use ACT to compare related sequences, which makes ACT a useful tool for ordering and finishing of genomes. With ACT you can identify specific regions, such as areas where conserved gene order between otherwise co-linear genomes break down, or regions where repeats are expanded in one genome versus its comparator, or alternatively, that may be problem areas (e.g. partial gene models) for further work. ACT thus enables you to hone in on and prioritise certain regions.

As detailed in Module 3, you will need 3 files to load into ACT:

- sequence of the reference contig/chromosome/genome
- blast comparison file
- sequence of your comparator set of contigs

There are a number of ways that these files, in particular the comparison file, can be generated. Some of the options are:

- Command line BLAST (detailed in Appendix II)
- WebACT (see Module 2)
- A script called ABACAS (detailed on page 9 of Module 4)

Exercise 1: Contig ordering against a reference genome with ACT (*Chlamydia*)

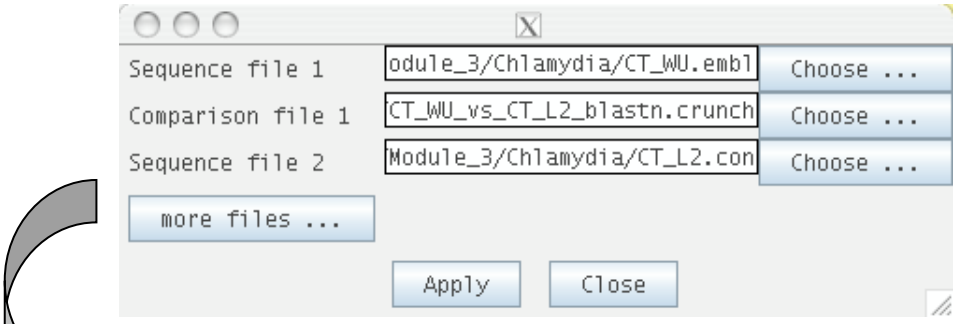
***Chlamydia trachomatis* L2**

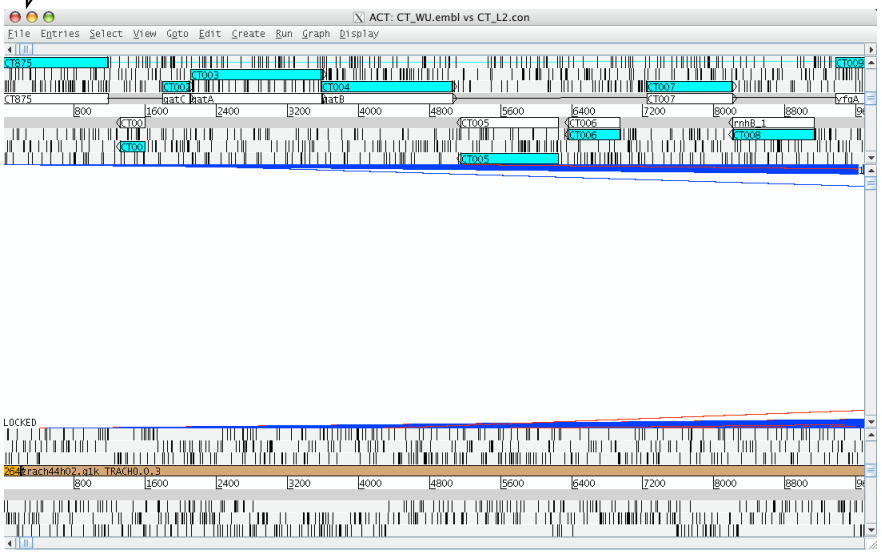
The two genomes you are going to look at are both *Chlamydia trachomatis* strains. The first is the published sequence *Chlamydia trachomatis* strain WU, the other is a consensus file representing the latest assembly of the *Chlamydia trachomatis* strain L2.

Go to the Module_4, exercise1 directory. The files you will need are in the 'Chlamydia' directory:

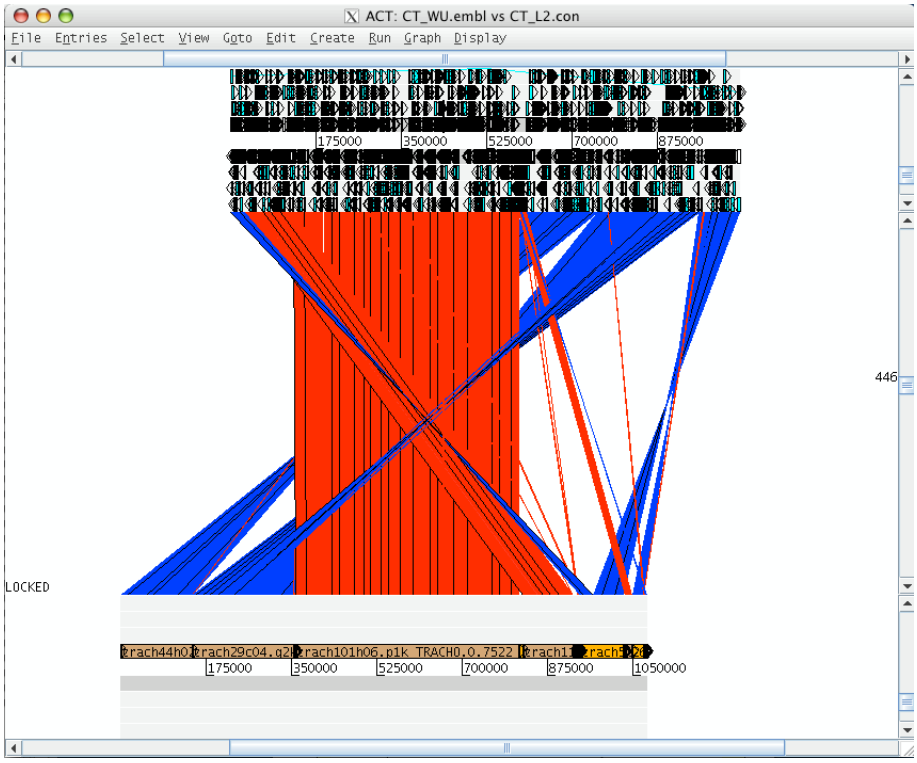
CT_WU.embl	- sequence file in EMBL format
CT_WU_vs_CT_L2_blastn.crunch	- blastn comparison file
CT_L2.con	- Consensus file from Gap4 in fasta format

Open up the sequence files as previously. Load in the sequence and comparison files in the order as they appear below.





- Use the sliders at the side of the genome sequences (1) to zoom out and the sliders (2) on the Comparison View panel to remove the smaller “footprint” hits.
- If you do the above and align the sequences you should see a view similar to that shown below. Note that the inverted matches are shown in blue.

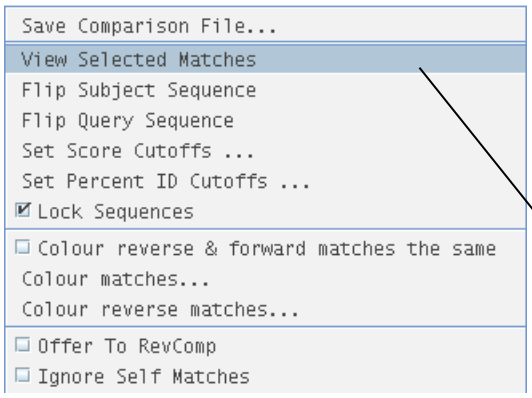


1.

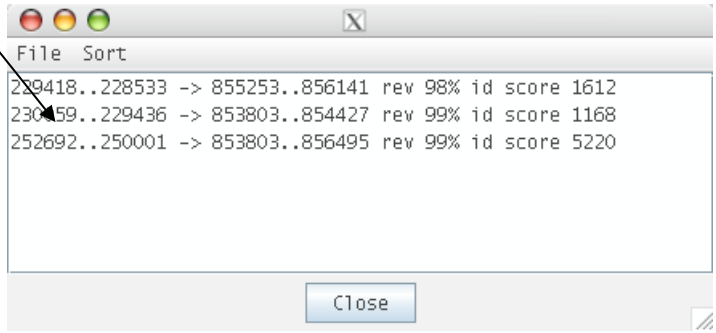
2.

1.

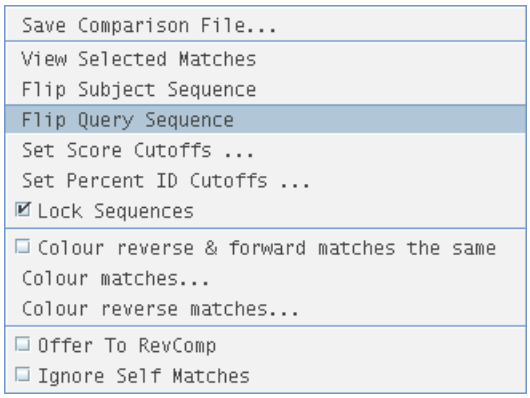
- Using the feature selector we used to navigate in Artemis earlier find the rRNA genes in *C. trachomatis* WU. How many rRNA genes are there in WU and how are they arranged? Hint: they appear under the key rRNA.
- Can you find the locations of the rRNA genes in L2. To determine this you will have to find the complementary regions in L2. Select one of the rRNA genes in *C. trachomatis* WU by clicking on it. In the central window click with the right mouse button to get the ACT specific menu and select ‘view selected matches’. A window will appear showing all the matches for that gene between the two genomes double click on these to align the sequences.
- Hint you may need to first zoom in and also to invert the lower genome in order to make the view more comprehensible.



To view selected matches right click in the central ACT window and select 'view selected matches'.

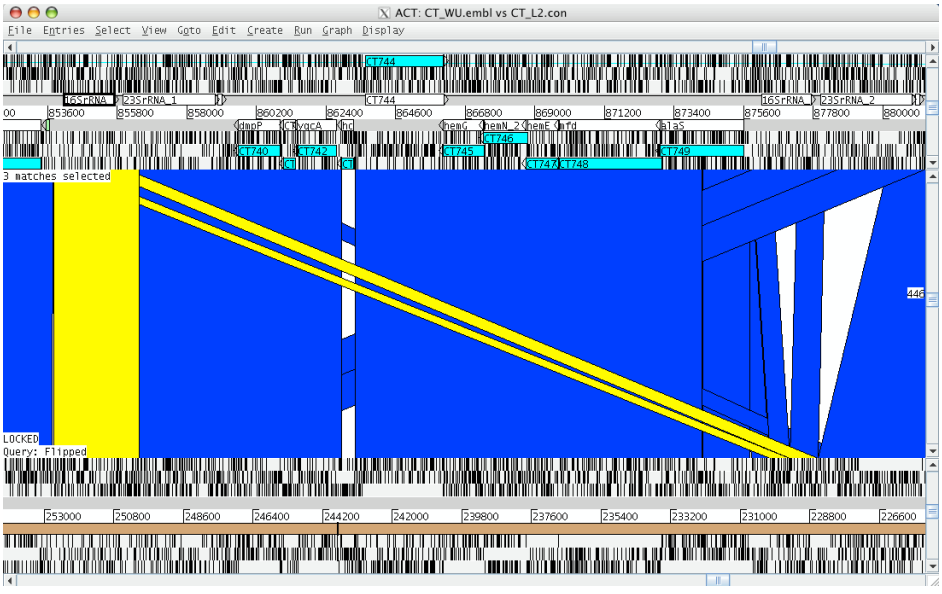


Double click on these to align sequences



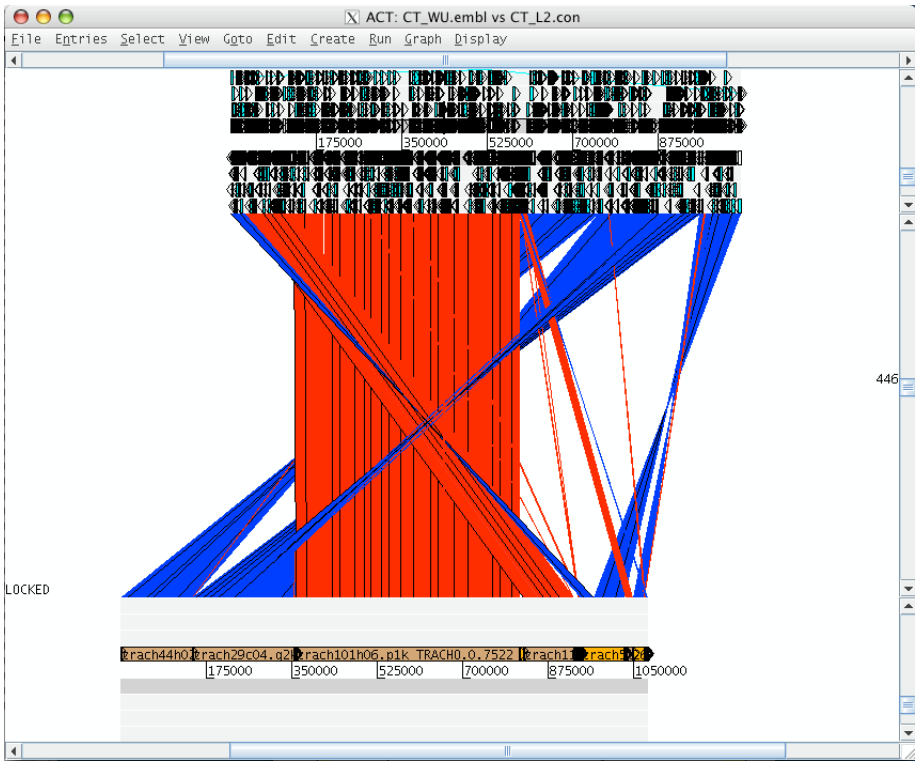
To flip one genome or the other use the central menu and select appropriately

You should be able to get a view like the one shown below
Can you think why the second set of rRNA genes Should appear to be smaller and even incomplete in L2 assembly?



- Now we will attempt to predict the order of the contigs in this early assembly. First zoom back out again. You will see that the central blocks of colour generally finish at the end of each contig. Go and have a look at a few of the contigs.

- By using the published *C. trachomatis* WU sequence as a guide see if you can predict the order of the *C. trachomatis* L2 contigs and note them below.



- Predicted Contig order is:

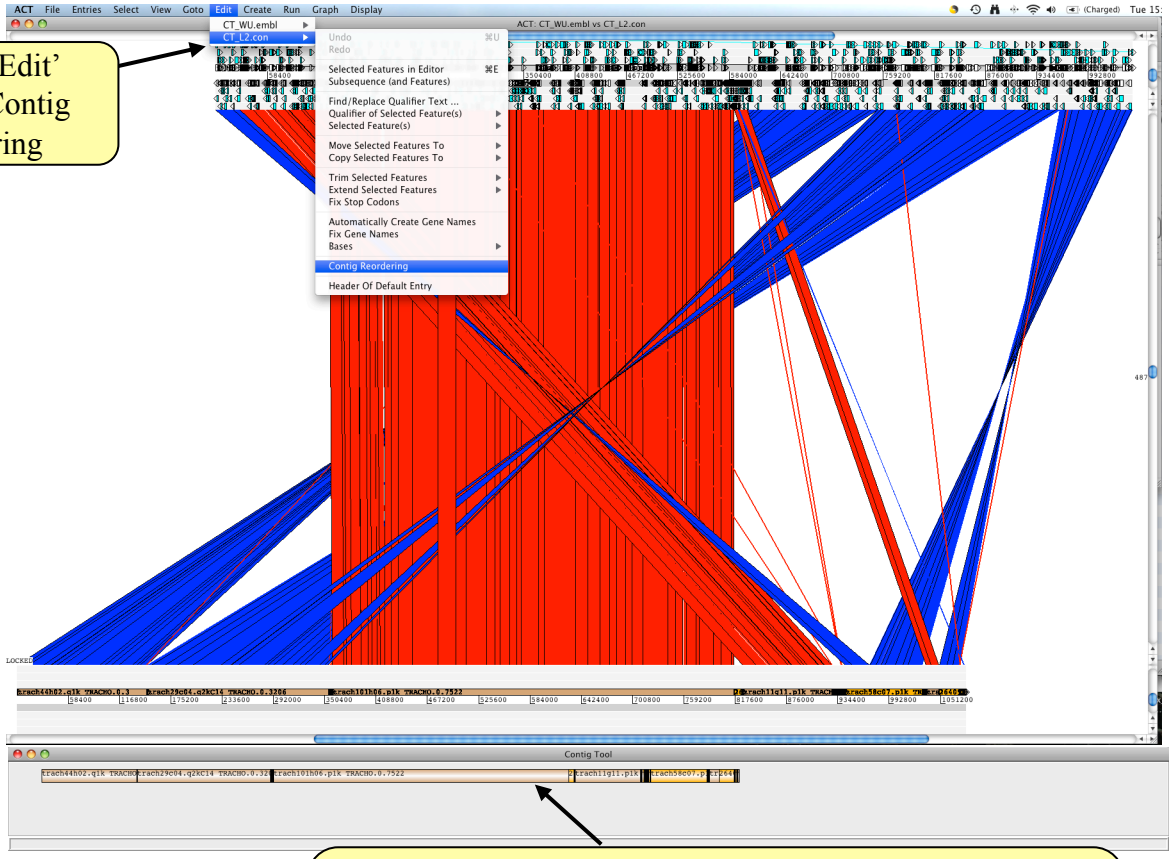
_____

_____

_____

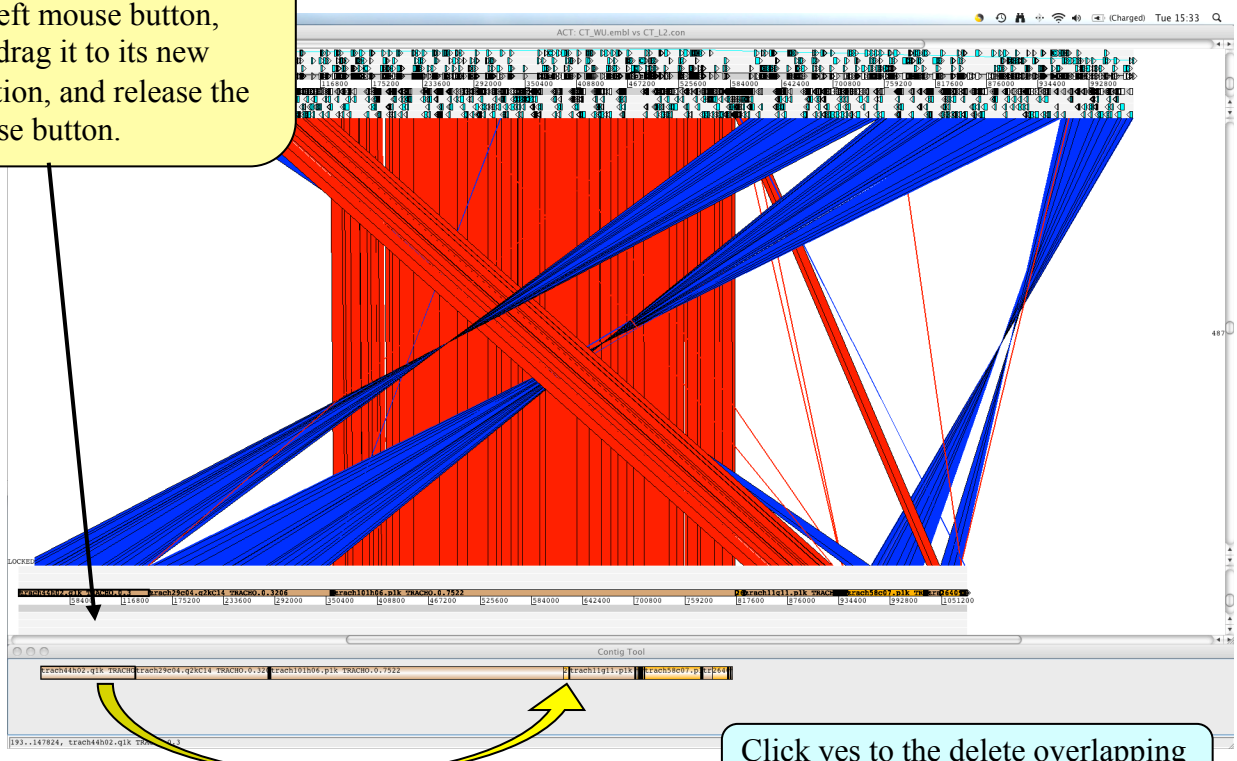
How to re-order contigs:

Click ‘Edit’ then ‘Contig reordering

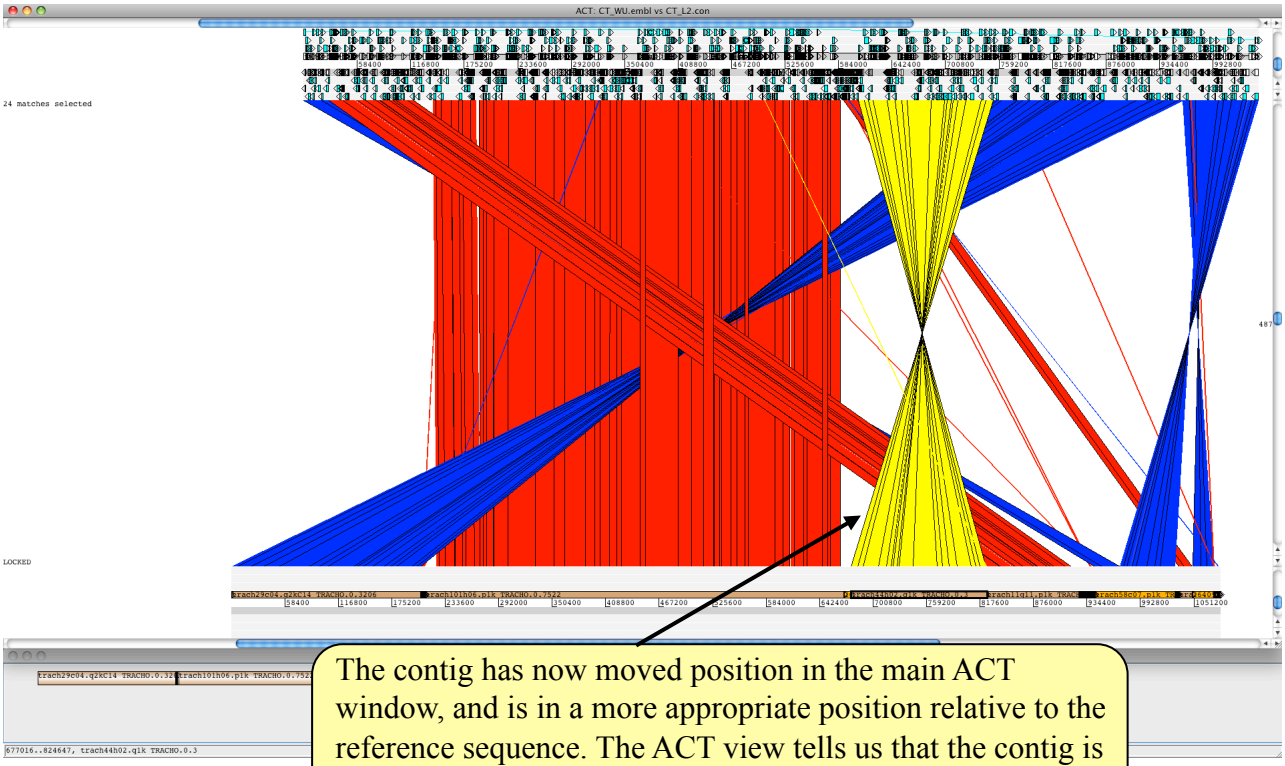


Contig tool window appears with a linear representation of the contigs as they appear in the consensus file. Contigs selected in this window are simultaneously selected in the main ACT window, and can be re-ordered.

Select the contig you want to move, hold down the left mouse button, and drag it to its new position, and release the mouse button.

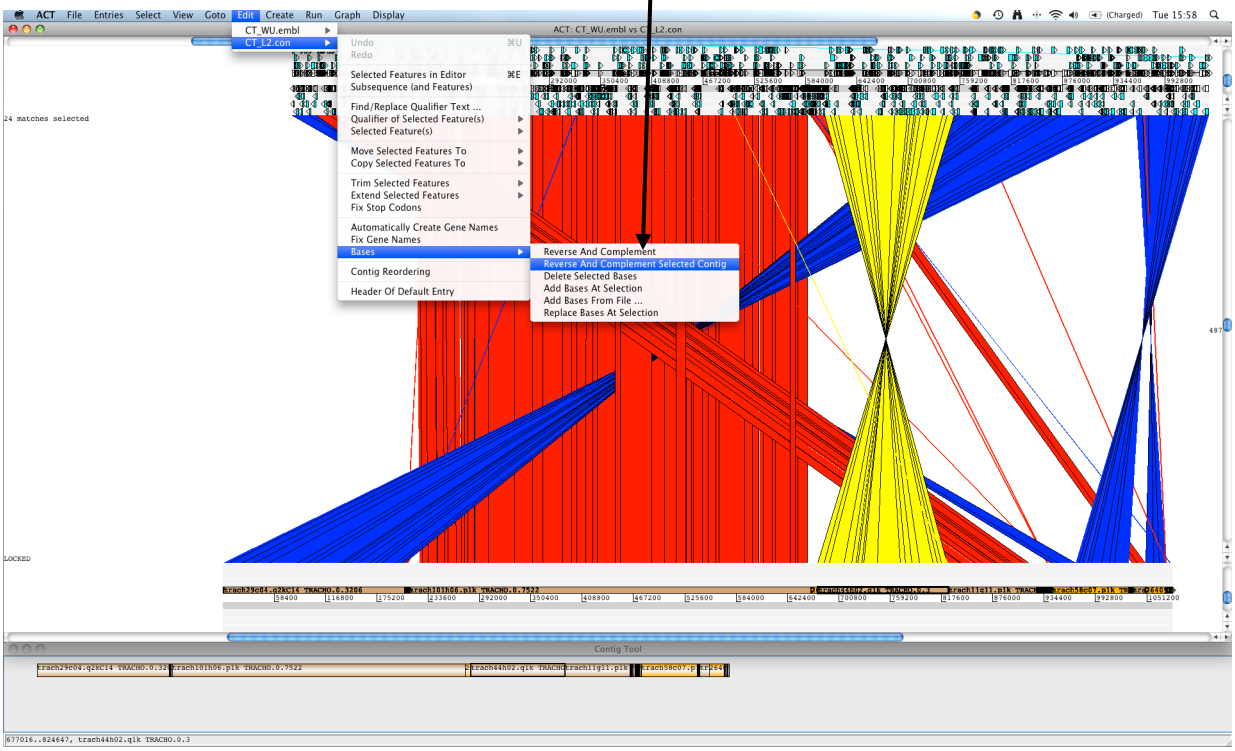


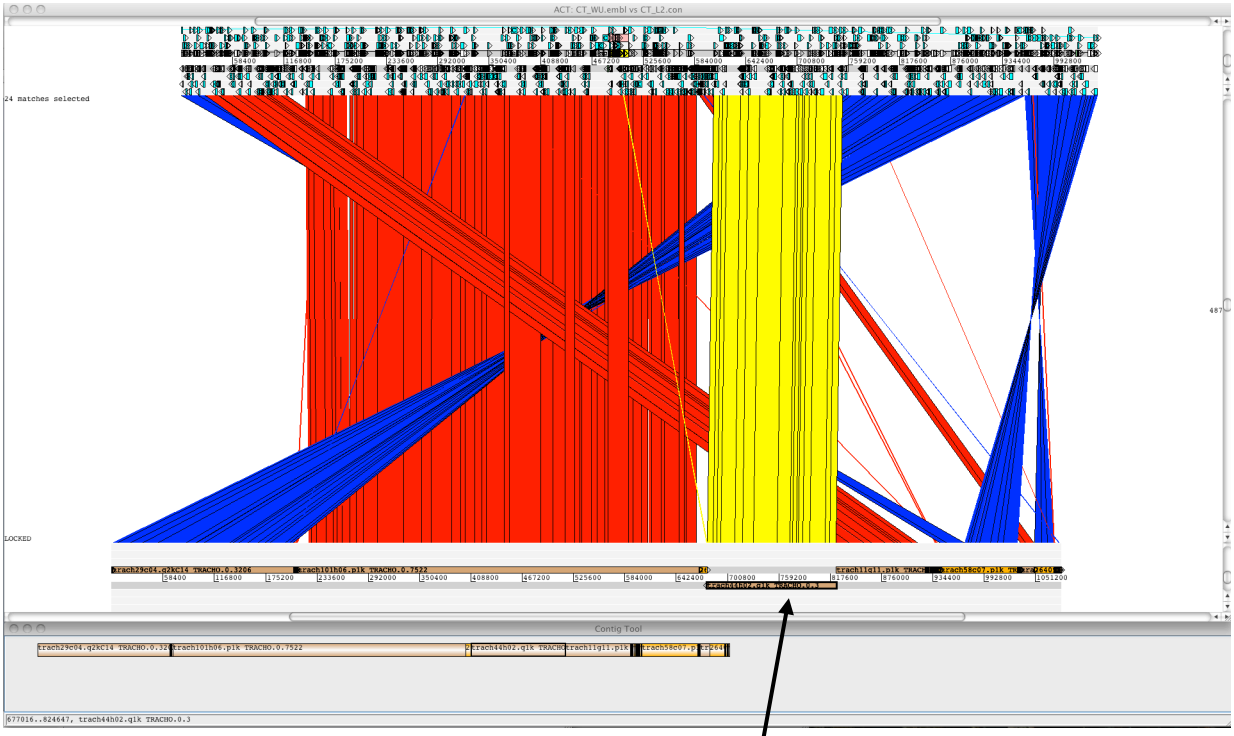
Click yes to the delete overlapping matched if a window appears



The contig has now moved position in the main ACT window, and is in a more appropriate position relative to the reference sequence. The ACT view tells us that the contig is now in the wrong orientation, so it needs to be reversed.

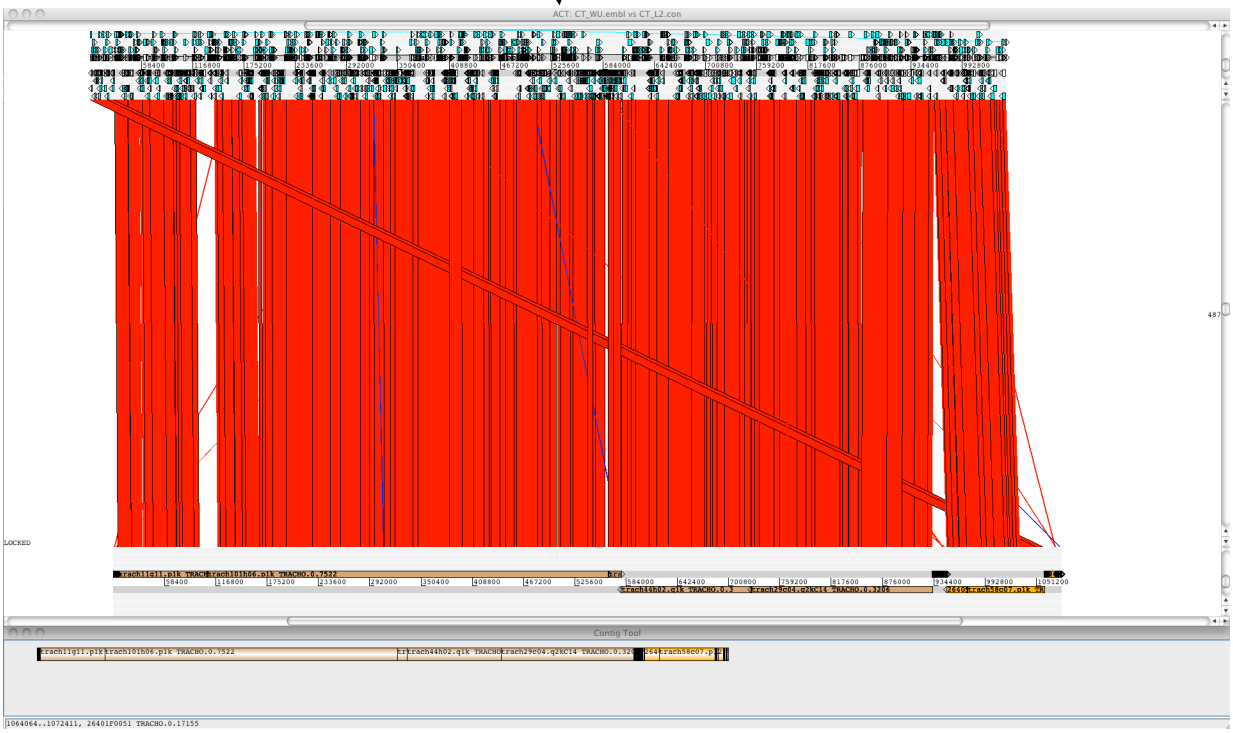
Select the contig, and from the CT_L2.con edit menu, select bases, and then Reverse And Compliment Selected Contig.





The contig is now in the opposite orientation

Re-order and re-orientate the remaining contigs to a new alignment that is colinear to the reference sequence



Remember that now you have re-ordered the contigs and the blast matches, if you want to keep this configuration, you will have to save the contig file and the blast match. To save the re-arranged blast matches, right click on the blast view, and Save comparison file.