Module 3
Comparative Genomics

Introduction

The Artemis Comparison Tool (ACT) was designed to better understand the additional information that can only be gained by comparing the growing number of sequences from closely related organisms (Carver *et al*. 2005). ACT is based on Artemis, so you will already be familiar with many of its core functions. It is is essentially composed of three layers or windows. The top and bottom layers are mini Artemis windows (with their inherited functionality), showing the linear representations of the DNA sequences with their associated features. The middle window shows red and blue blocks, which span this middle layer and link conserved regions between the two sequences, in the forward and reverse orientation respectively. Consequently, if you were comparing two identical sequences in the same orientation you would see a solid red block extending over the length of the two sequences in this middle layer. If one of the sequences was reversed, and therefore present in the opposite orientation, there would be a blue 'hour glass' shape linking the two sequences. Unique regions in either of the sequences, such as insertions or deletions, would show up as breaks (white spaces) between the solid red or blue blocks.

In order to use ACT to investigate your own sequences of interest you will have to generate your own pairwise comparison files. Data used to draw the red or blue blocks that link conserved regions is generated by running pairwise BLASTN or TBLASTX comparisons of the sequences. ACT is written so that it will read the output of several different comparison file formats; these are outlined in Appendix III. Two of the formats can be generated using BLAST software freely downloadable from the NCBI, which can be loaded and run on a PC or Mac. Another way of generating comparison files for ACT is to use the WebACT web resource (http://www.webact.org/). This site allows you to cut and paste or upload your own sequences, and generate ACT readable BLASTN or TBLASTX comparison files.

Aims

The aim of this Module is for you to become familiar with the basic functions of ACT. In the first exercise you will be looking at a comparison between chromosome 11 of *P. falciparum* 3D7 (a human malaria parasite) and chromosome 9 of *P. chabaudi* AS (a rodent malaria parasite). By comparing the two chromosomes you will be able to study the degree of conservation of gene order and identify small and large synteny breaks. You will explore how the genomes of malaria parasites have changed over time as they have evolved to infect different hosts.

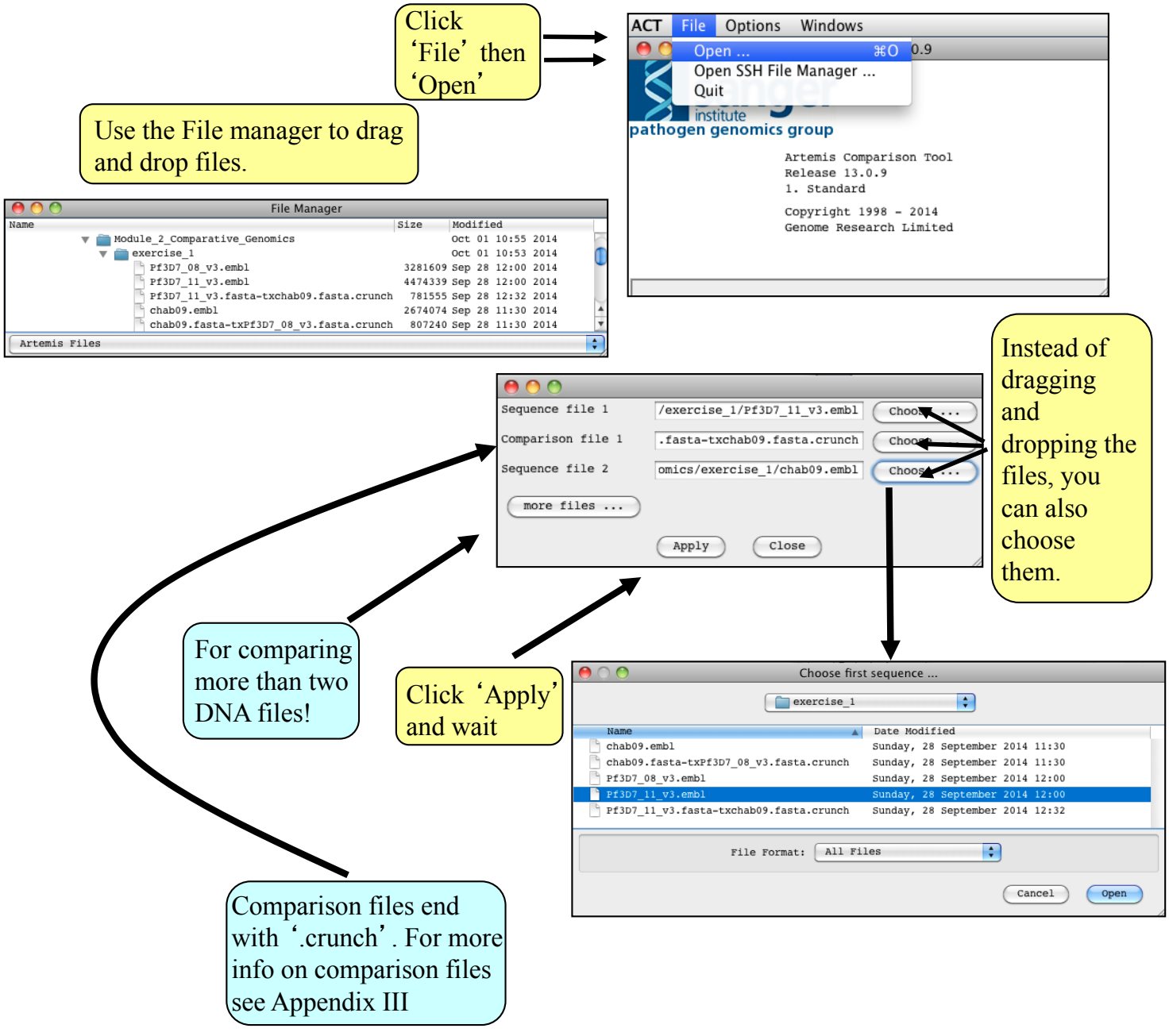
Exercise 1
Starting up the ACT software

Make sure you're in the **Module_3_Comparative, exercise_1** directory.
Then type
act & [return]
A small start up window will appear.

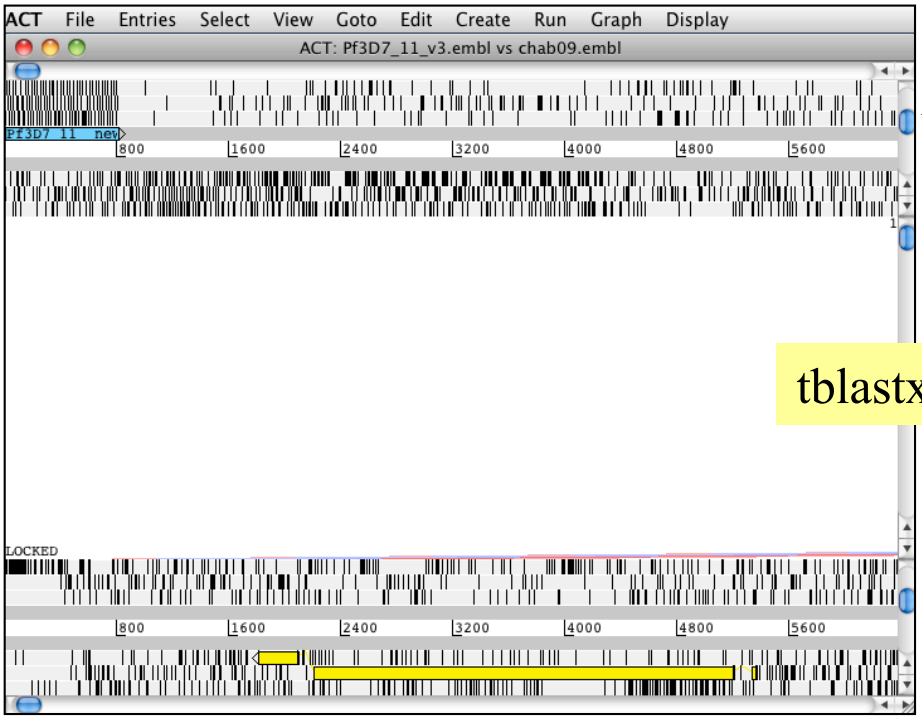
To open ACT you can also double click the ACT icon on your Desktop.

The files that you are going to need are:

Pf3D7_11_v3.embl	- *P. falciparum* chr11
Pf3D7_11_v3.fasta-txchab09.fasta.crunch	- tblastx comparison file
chab09.embl	- *P. chabaudi* chr9



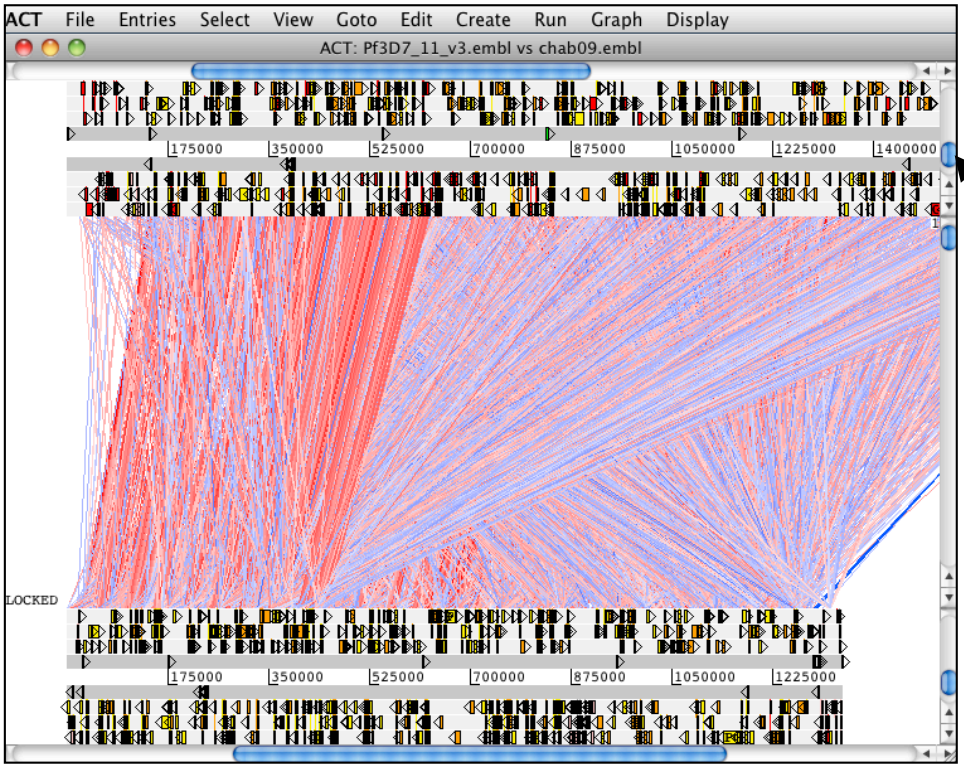
Once you have opened the files you will see a picture like this:



P. falciparum chr11

tblastx comparison

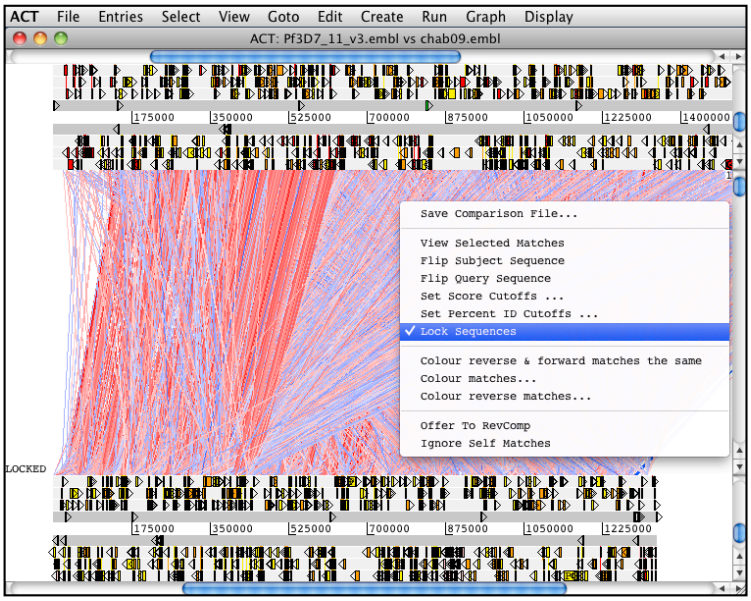
P. chabaudi chr9



Use the vertical sliders to zoom out. Drag or click the slider downwards from one of the genomes. The other genome will stay in synch.

When you scroll along with either slider both genomes move together. This is because they are 'locked' together. Right click over the middle comparison view panel. A small menu will appear, select Unlock sequences and then scroll one of the horizontal sliders. Notice that 'LOCKED' has disappeared from the comparison view panel and the genomes will now move independently.

LOCKED



You can optimise your image by either removing 'low scoring' (or percentage ID) hits from view, as shown below **1-3** or by using the slider on the comparison view panel (**4).** The slider allows you to filter the regions of similarity based on the length of sequence over which the similarity occurs, sometimes described as the "footprint".

1

Right button click in the Comparison View panel

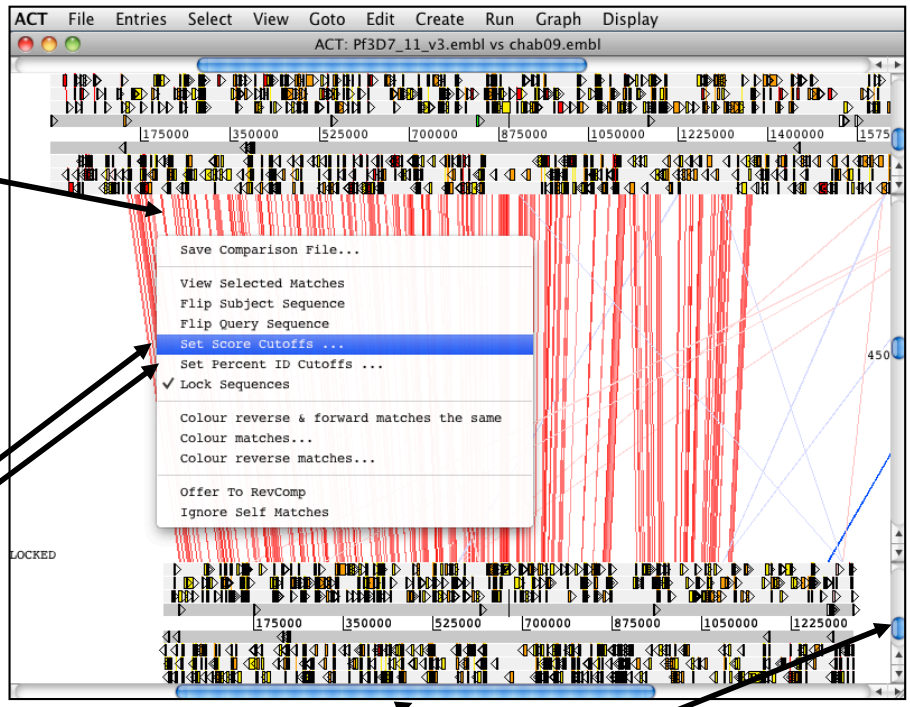
2

Select either Set Score Cutoffs or Set Percent ID Cutoffs

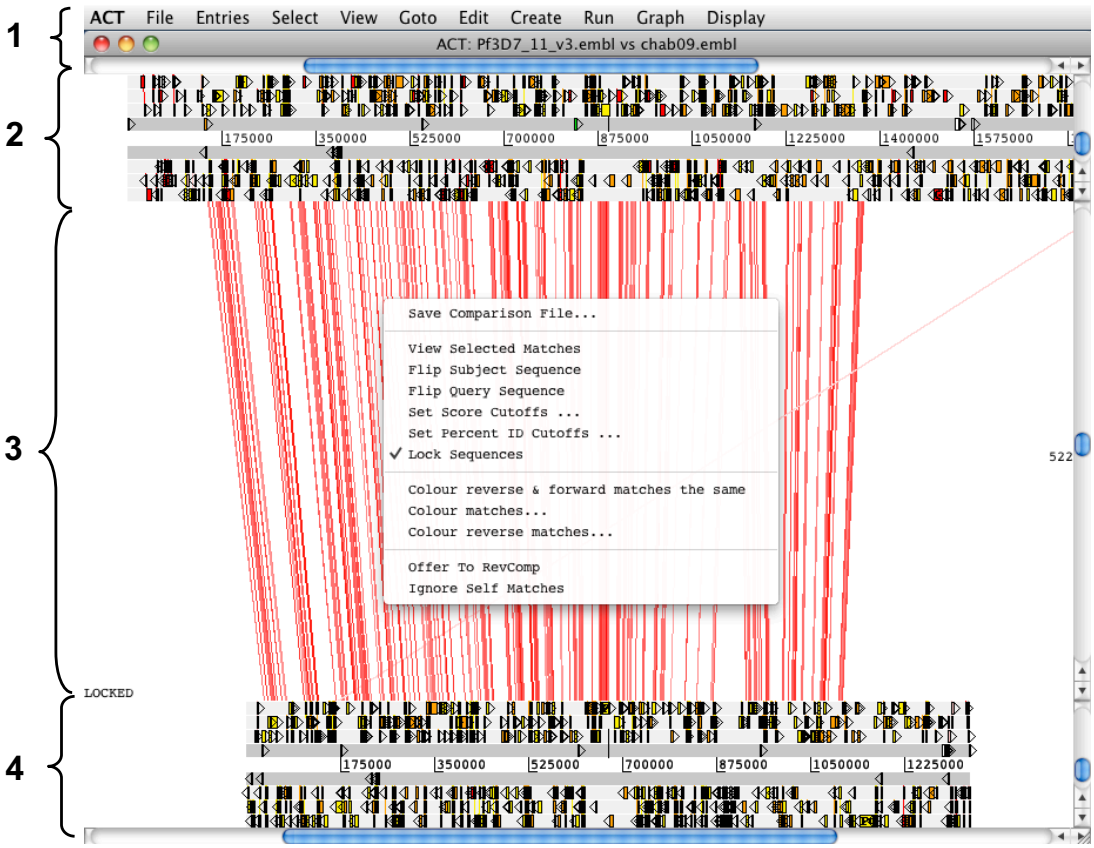
3

Move the sliders to manipulate the comparison view image

4

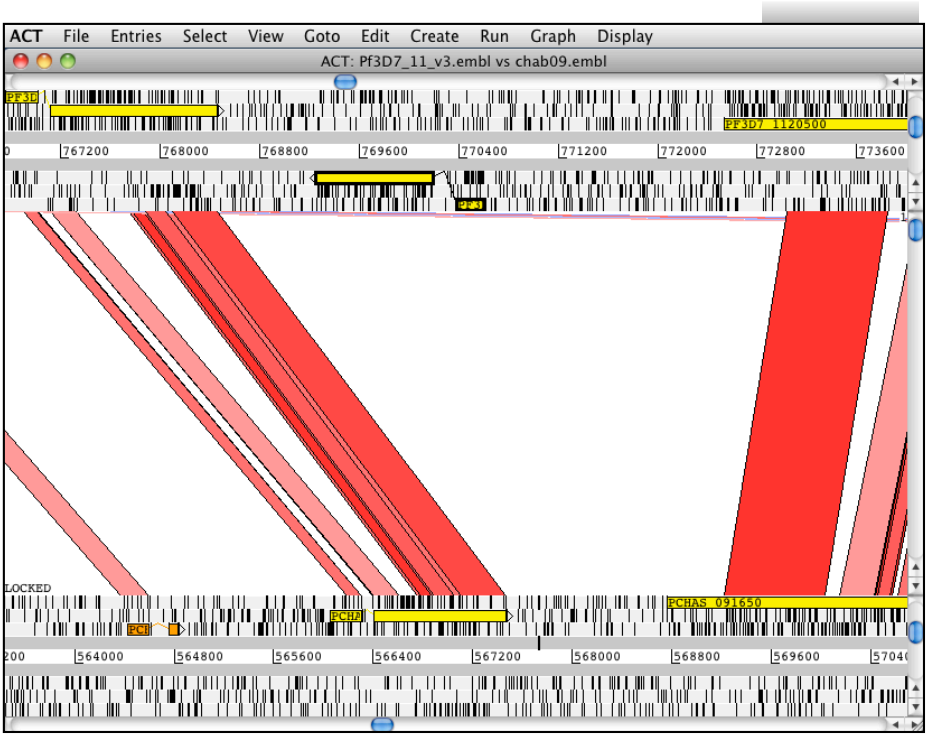


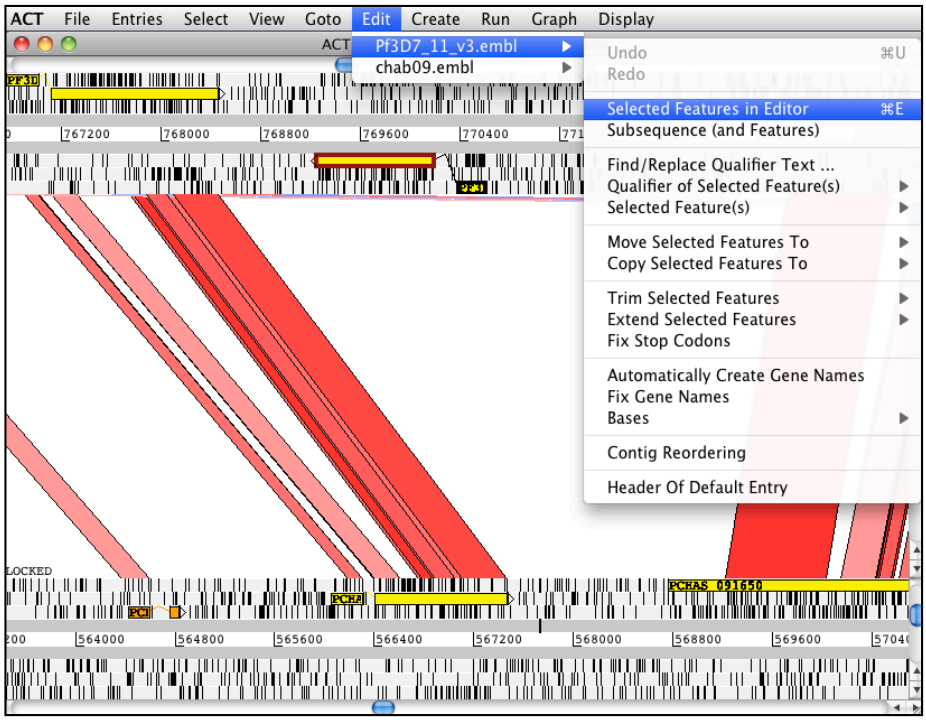
Now that you have an ACT window open let's look what's in there.



1. Drop-down menus. These are mostly the same as in Artemis. The major difference you'll find is that after clicking on a menu header you will then need to select a DNA sequence before going to the full drop-down menu.
2. This is the Sequence view panel for 'Sequence file 1' (Subject Sequence) you selected earlier. It's a slightly compressed version of the Artemis main view panel. The panel retains the sliders for scrolling along the genome and for zooming in and out.
3. The Comparison View. This panel displays the regions of similarity between two sequences. Red blocks link similar regions of DNA with the intensity of red colour directly proportional to the level of similarity. Double clicking on a red block will centralise it. Blue blocks link regions that are inverted with respect to each other.
4. Artemis-style Sequence View panel for 'Sequence file 2' (Query Sequence).
5. Right button click in the Comparison View panel brings up this important ACT-specific menu which we will use later.

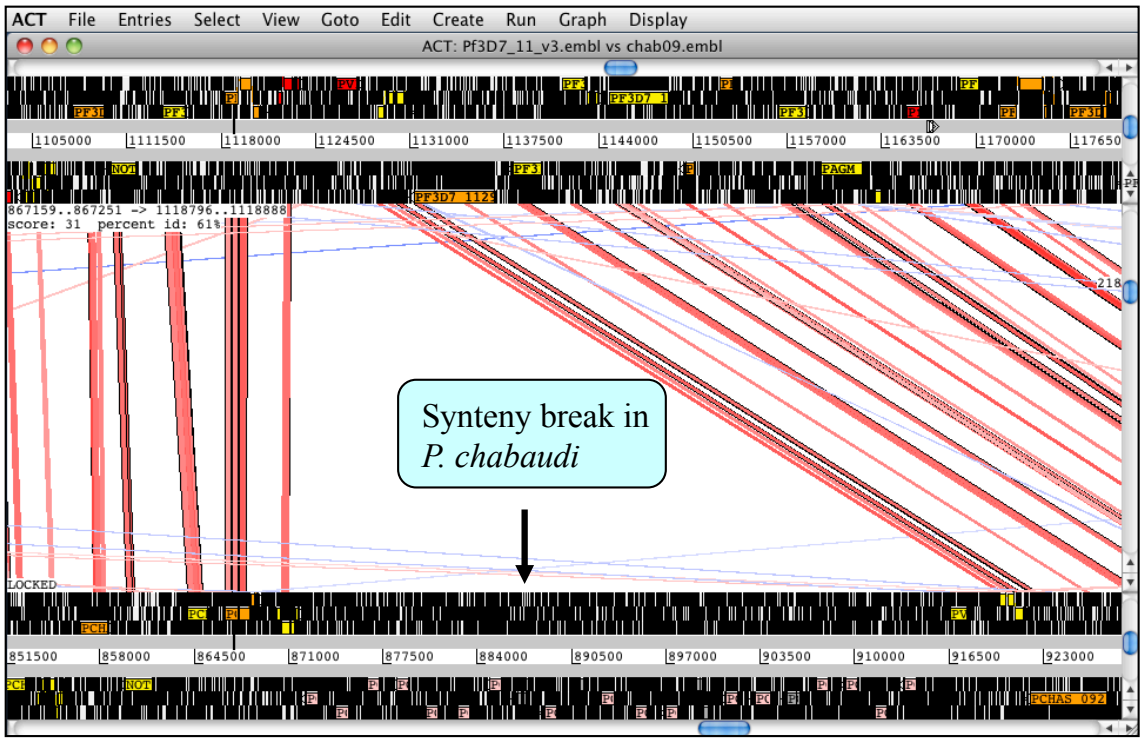
Scroll along the chromosome and look for small synteny breaks between *P. falciparum* and *P. chabaudi*. One example is shown here. Use the 'Goto' option to go to this region. You can either use the option 'Goto base' or 'Goto Feature With Gene Name' (e.g. PF3D7_1120400).



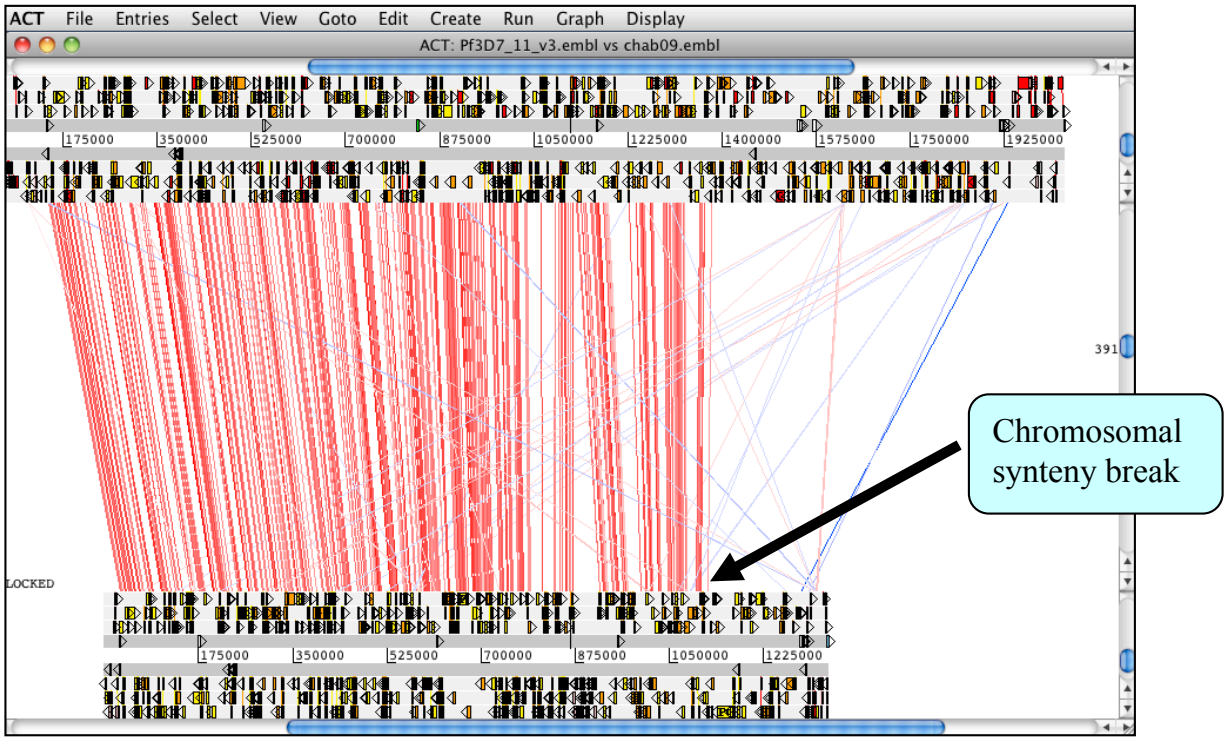


To get more information about this gene, select the gene, then go to 'Edit' 'Selected Features in Editor'. As a shortcut you can also just press 'E' on your keyboard.

Scroll along the chromosome and try to get an estimate on the number of synteny breaks. Can you find the largest synteny break? Identify the genes that are located in this area.



Can you locate the region of a chromosomal synteny break point?



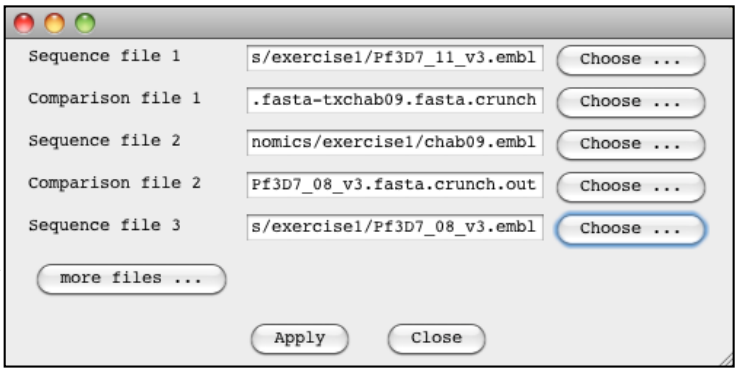
In this part of the exercise we will look at a three-way comparison and explore the chromosomal synteny break in *P. chabaudi* further.

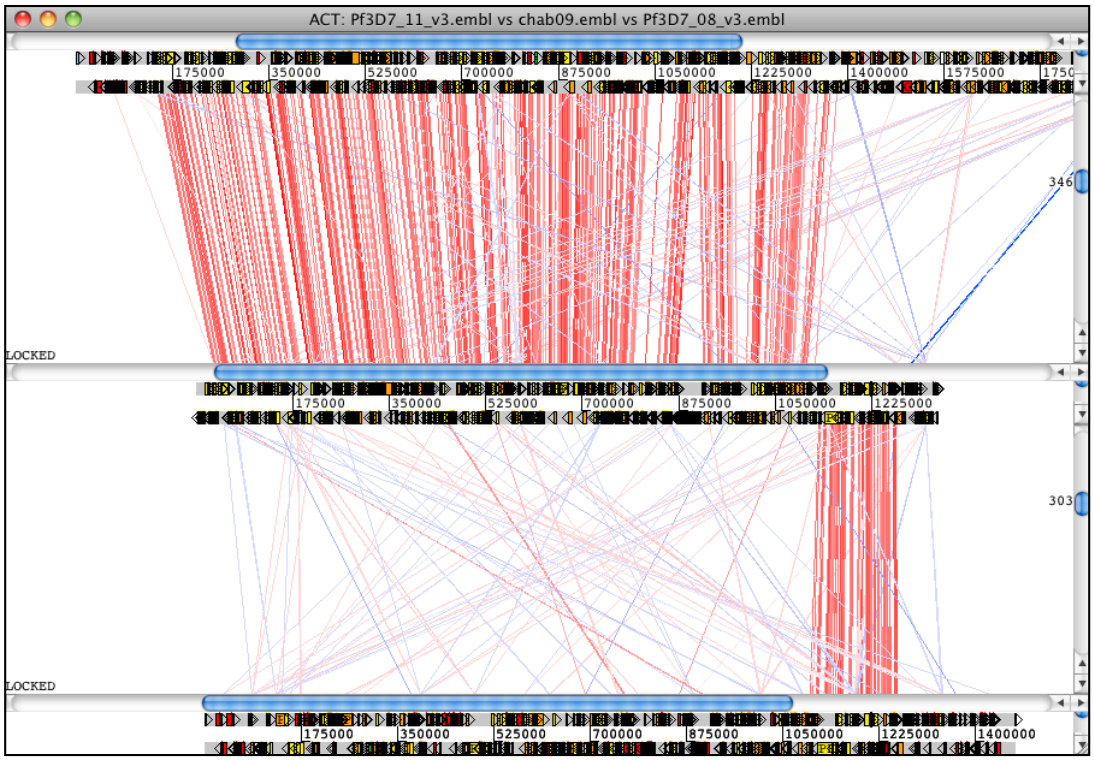
The files you are going to need are:

Pf3D7_11_v3.embl	- *P. falciparum* chr11
Pf3D7_11_v3.fasta-txchab09.fasta.crunch	- tblastx comparison file
chab09.embl	- *P. chabaudi* chr09
chab09.fasta-txPf3D7_08_v3.fasta.crunch	- tblastx comparison file
Pf3D7_08_v3.embl	- *P. falciparum* chr8

Close the current window, then open these five files.

Click on 'more files' to compare more than 2 files.





P. falciparum chr11

P. chabaudi chr09

P. falciparum chr08

What has happened to these chromosomes in evolution?

Once you have finished this exercise remember to close this ACT session down completely before starting the next exercise.

Exercise 2

Genomes are often highly similar and one has to look quite closely to find small differences. In this exercise we will be comparing the genomes of the human malaria parasite *P. falciparum* and the chimpanzee malaria parasite, *P. reichenowi* and try to identify differences. Can we find any differences that might relate to host specificity?

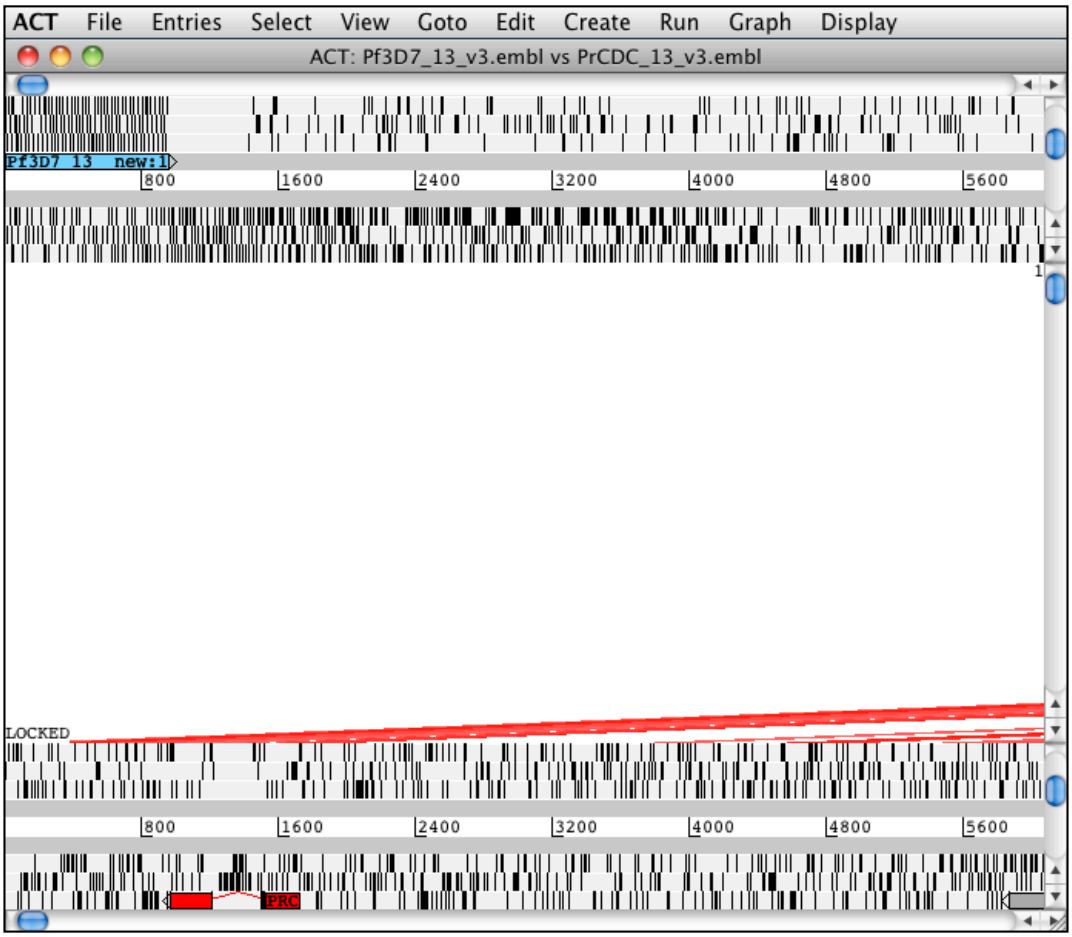
Make sure you're in the **Module_3_Comparative, exercise_2** directory.
Then type
act & [return]
A small start up window will appear.

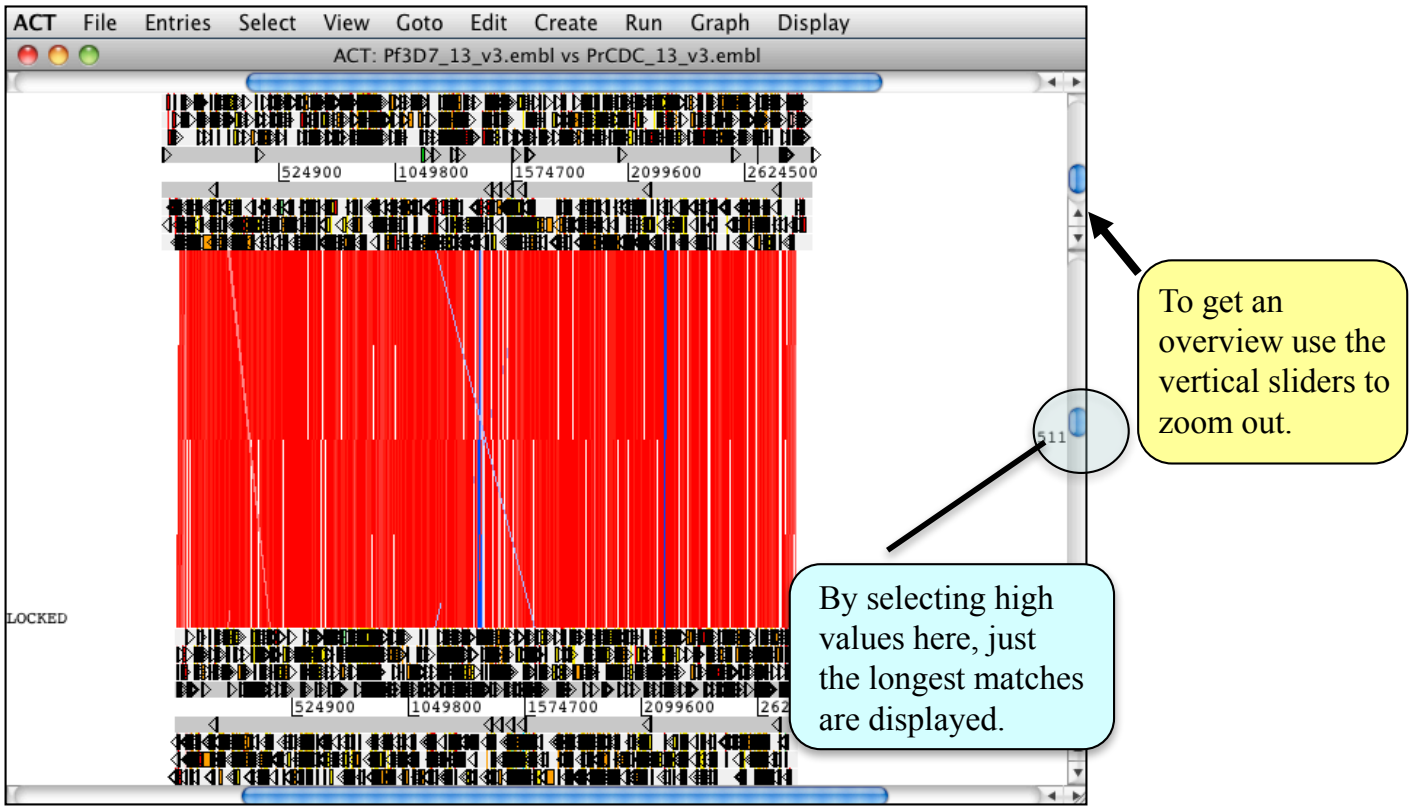
Or, to open ACT you can double click the ACT icon on your Desktop.

The files that you are going to need are:

Pf3D7_13_v3.embl	- *P. falciparum* chr13
Pf3D7_13_v3.fasta-txPrCDC_13.fasta.crunch	- tblastx comparison file
PrCDC_13_v3.embl	- *P. reichenowi* chr13

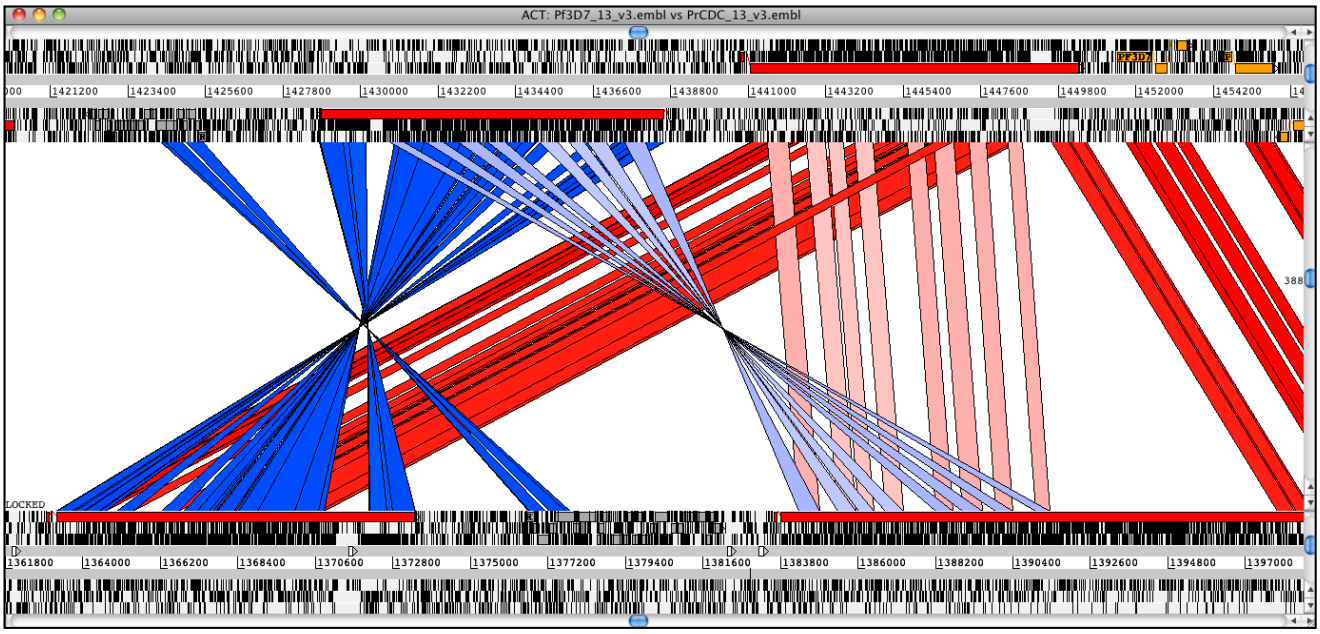
Once you have opened the files you will see a picture like this:



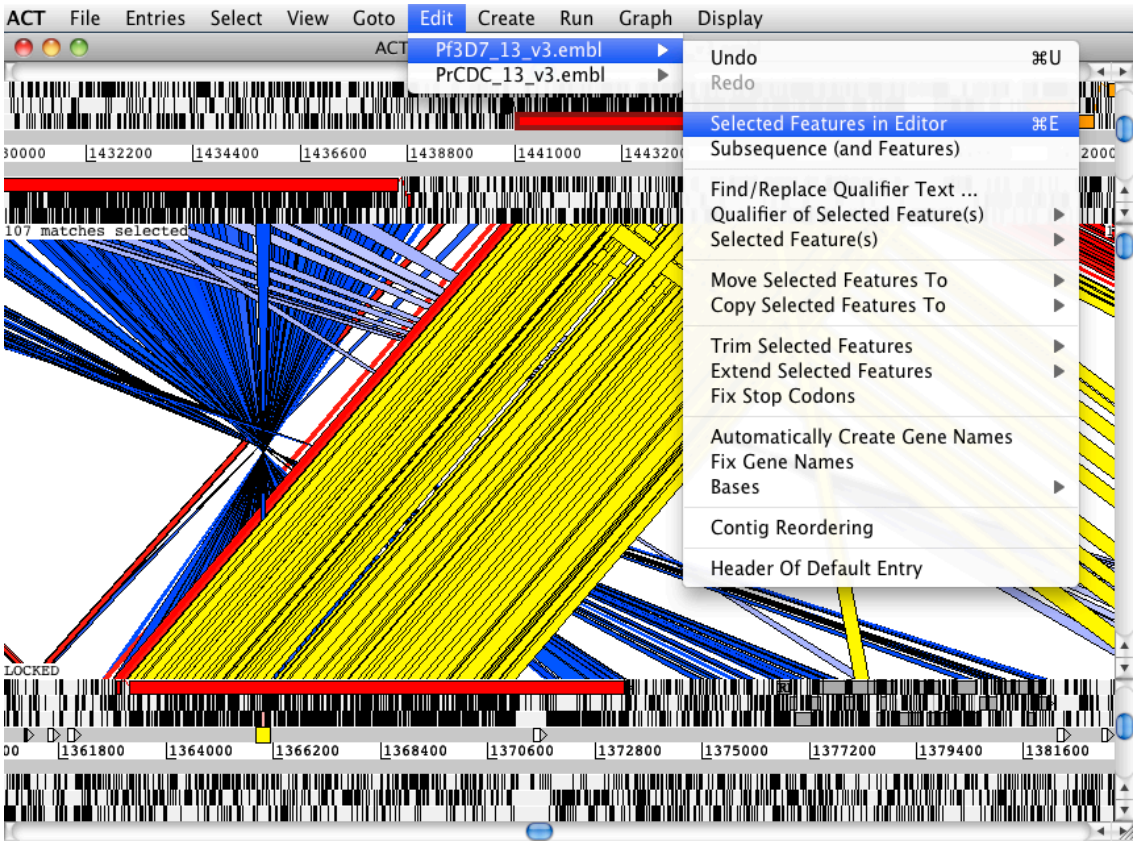


How many differences can you identify? Zoom in to see the genomes in more detail.

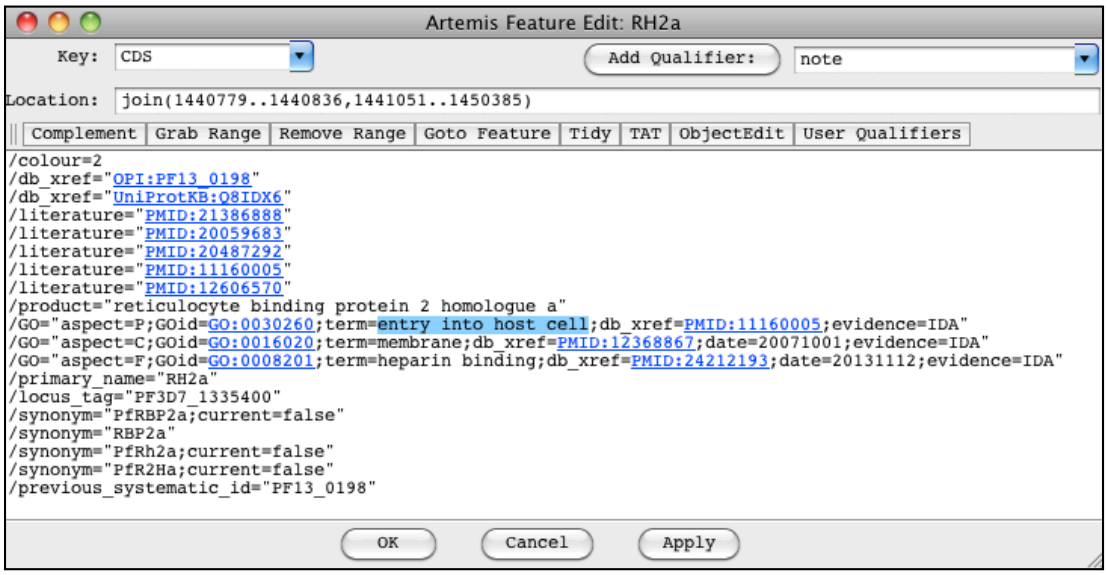
Have you already come across this area? Here are the coordinates: 1422300 - 1450390. Let's have a more detailed look at the genes.



To look at the annotation, mark one of the genes in this area (e.g. PF3D7_1335400), go to 'Edit' and select 'Selected Features in Editor'. As a shortcut you can just press 'e'.



What are the gene products? Are they important for host specificity? Does the Gene Ontology annotation give any additional information?



More information about Gene Ontology can be found here: http://geneontology.org/

Optional exercise: Comparison of *Leishmania* spp genomes

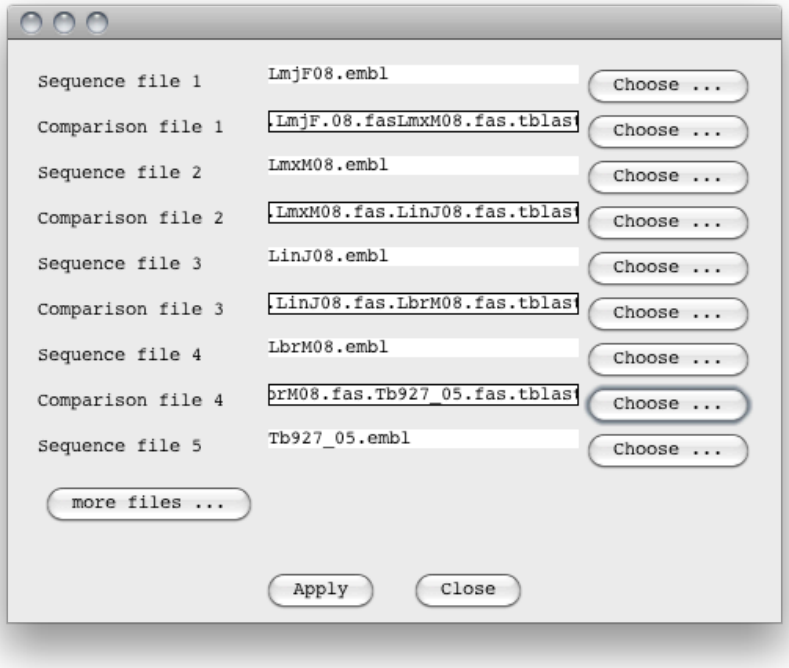
Leishmania are protozoan parasites that, depending on the species, cause a range of disease phenotypes from self-curing cutaneous lesions to destruction of facial tissue (mucocutaenous) to potentially deadly visceral disease. In this exercise you will compare the genomes of 4 species:

1. *Leishmania major* (cutaneous, Old World) –the original reference genome species; high quality, manually improved sequence
2. *L. infantum* (visceral, Old World) – A human-improved draft, assembled *de novo* and aligned against *L. major*
3. *L. braziliensis* (mucocutaneous, New World) - A human-improved draft, assembled *de novo* and aligned against *L. major*
4. *L. mexicana* (cutaneous, New World) - A human-improved draft, assembled *de novo and* aligned against *L. major*

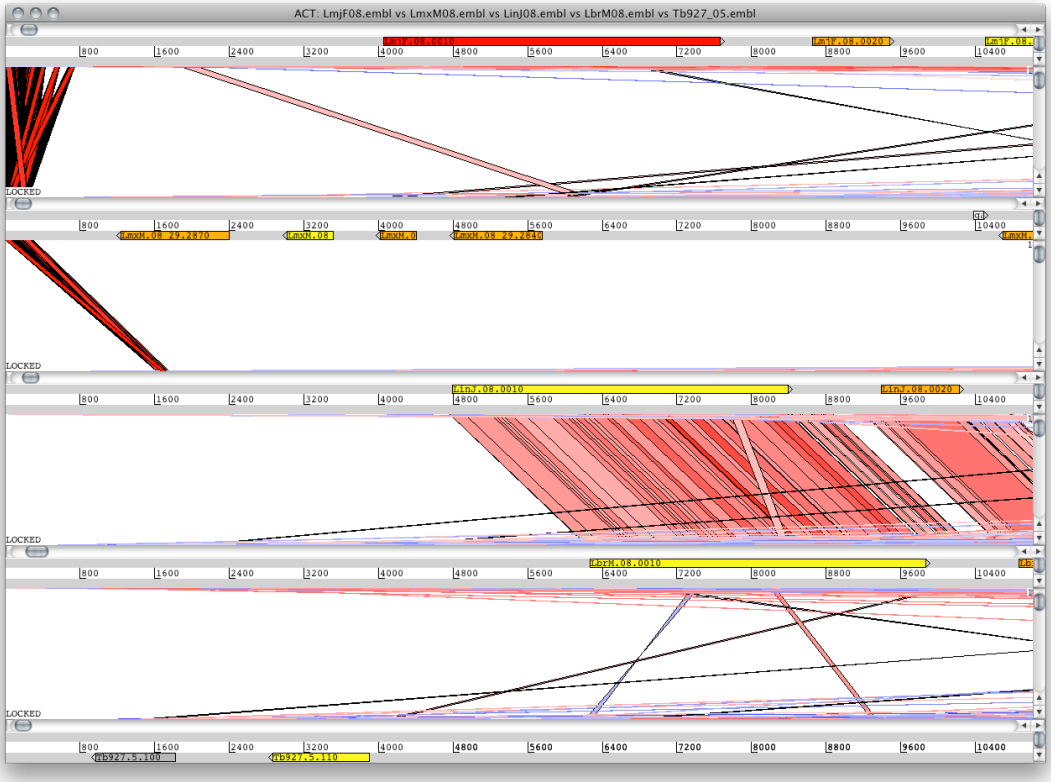
Close the previous ACT session.

Change to optional_exercise directory, then Restart ACT by typing act & on the command line or double click the icon on the Desktop.

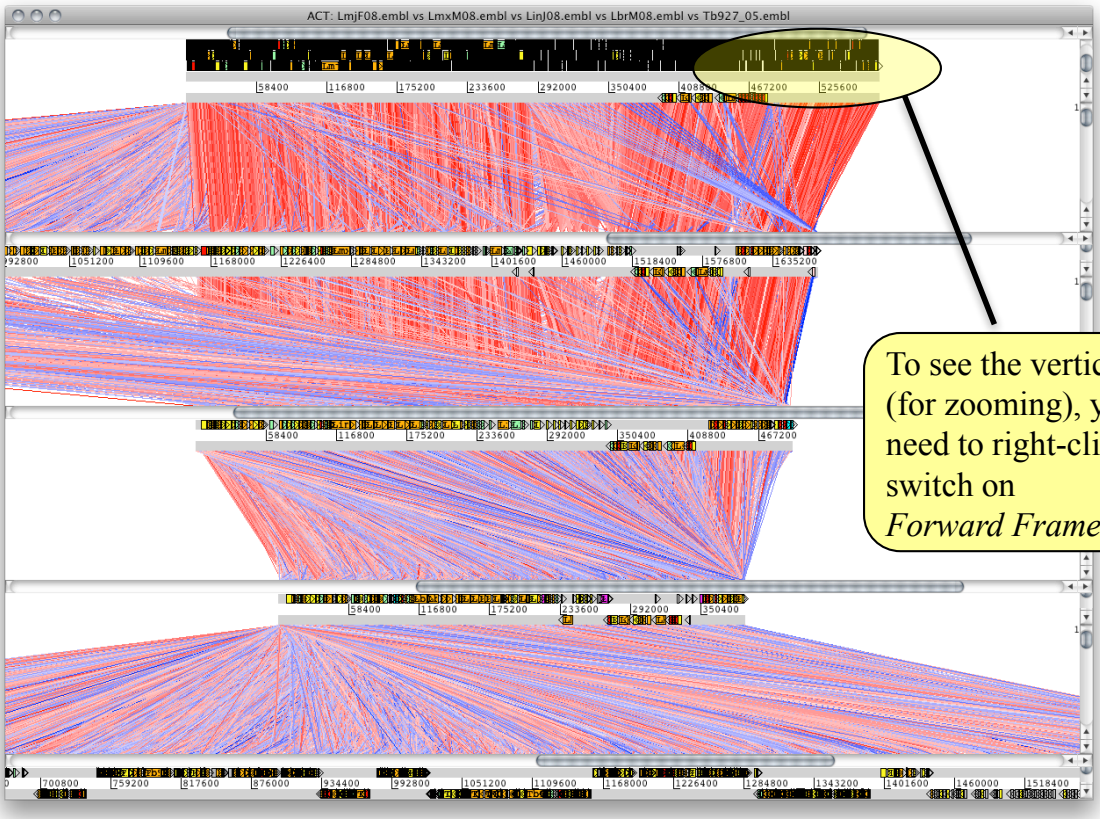
Load sequence and comparison files for 4 *Leishmania* species plus *Trypanosoma brucei* (outgroup)



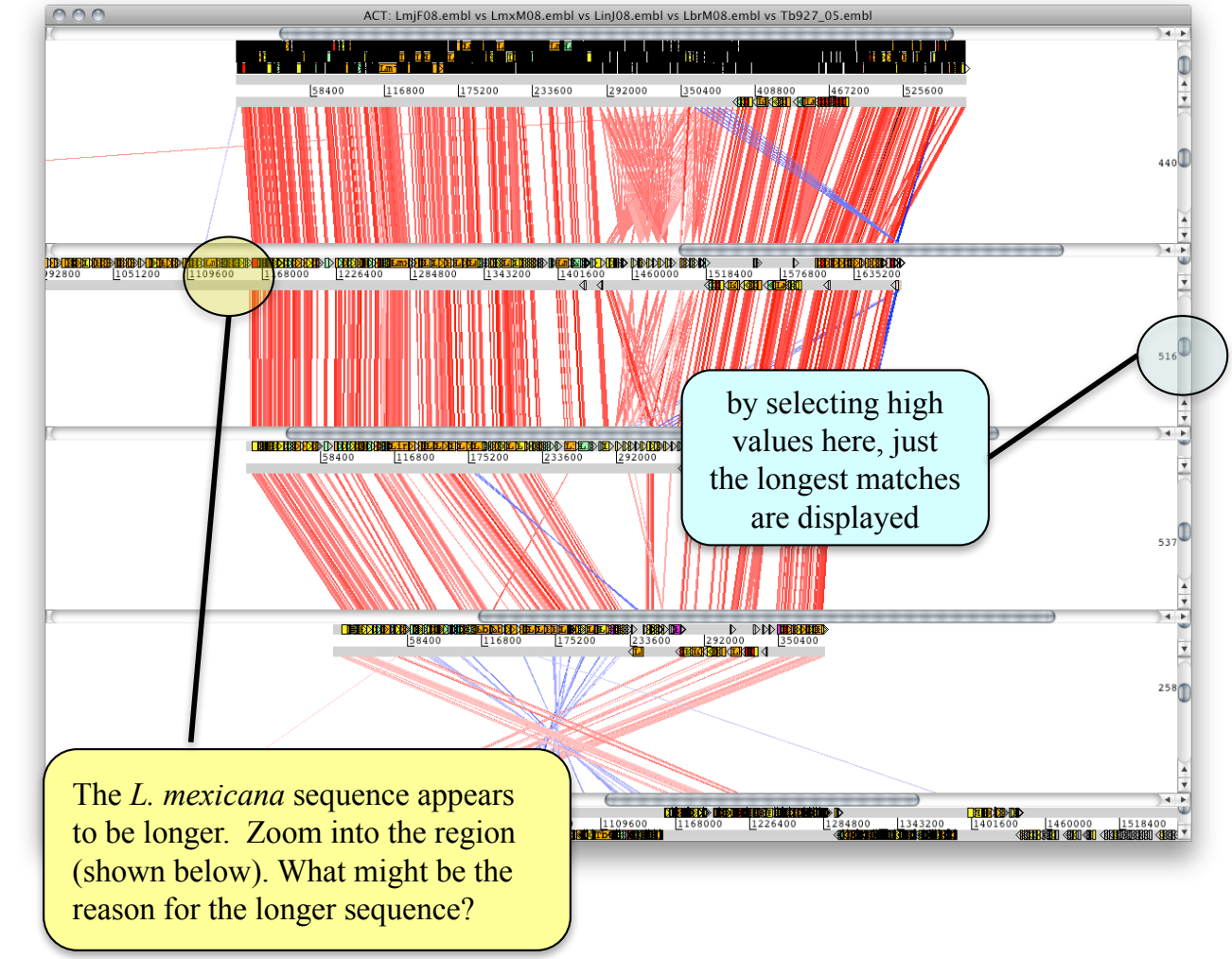
Your ACT window should now look something like this:

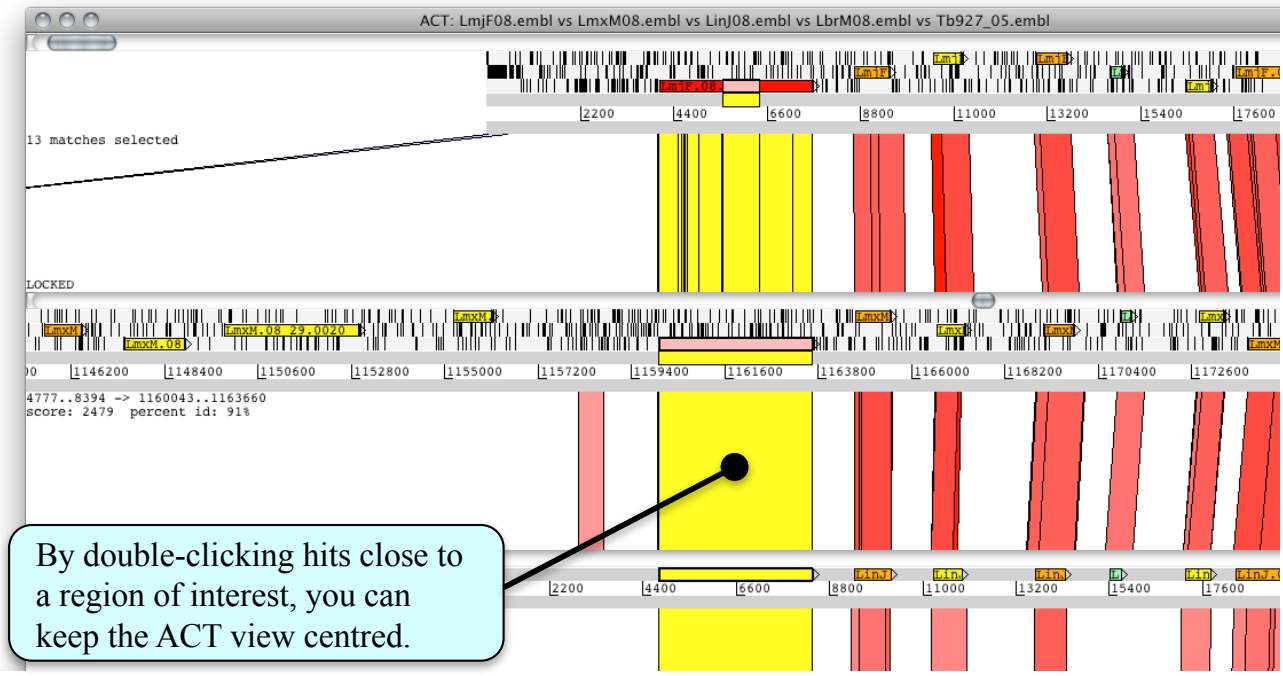


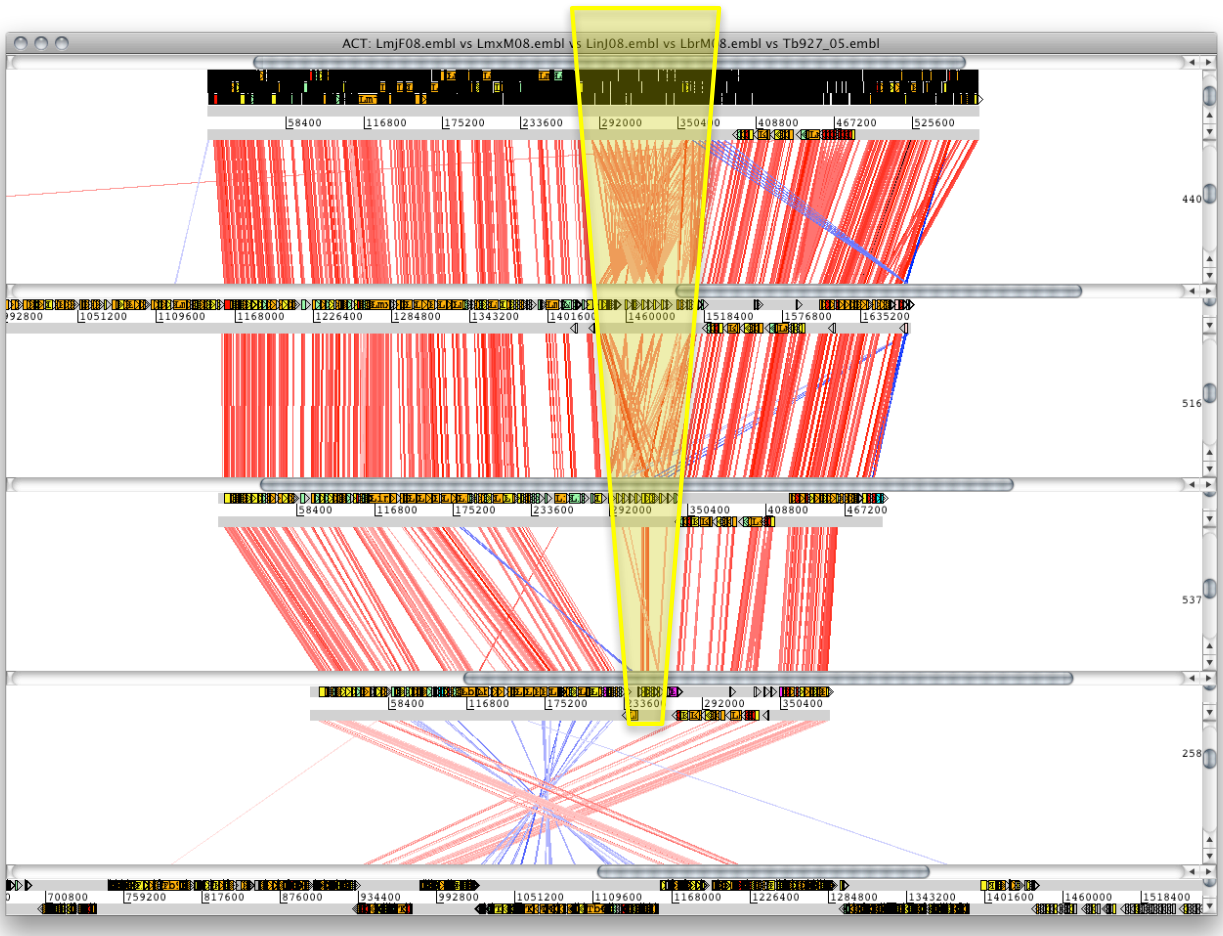
Zoom out (and scroll) as appropriate to identify the regions of conserved synteny.



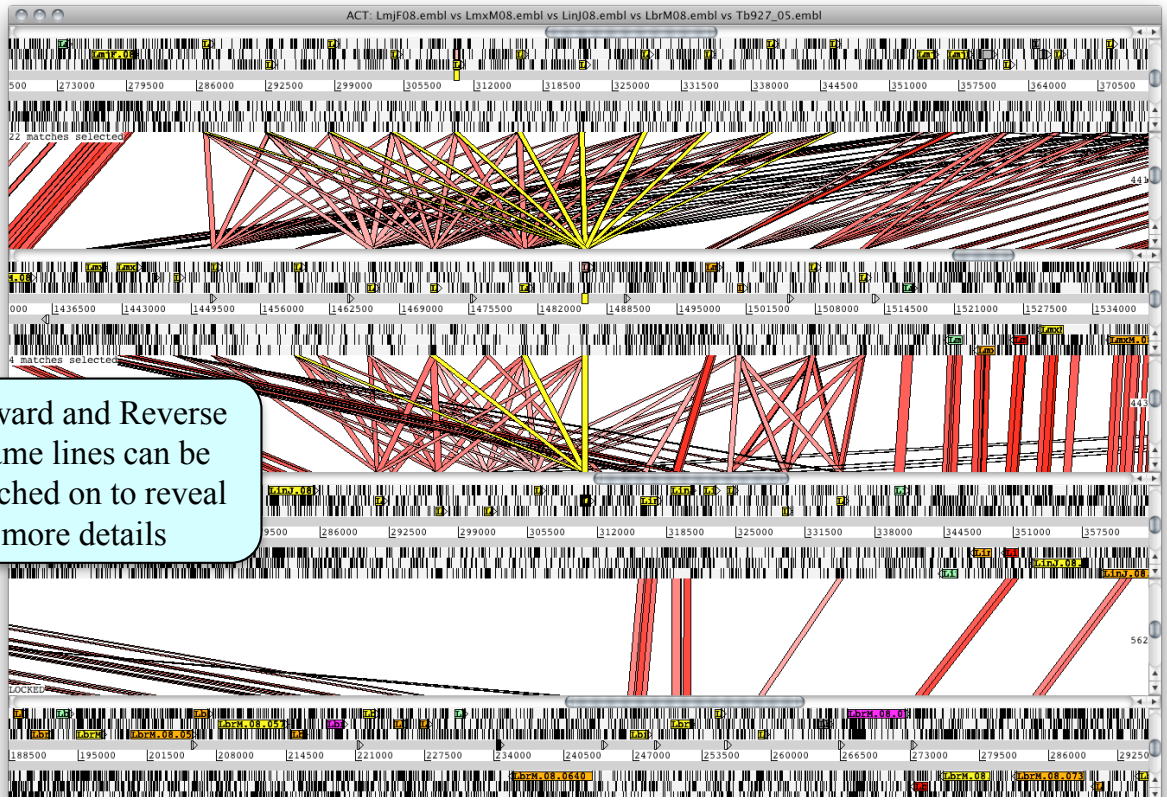
Use the filters (percentage identity, score or length of match) to adjust the signal:noise ratio







Several regions appear different sizes across different species. One is shown above and below. What is a likely reason?



Forward and Reverse Frame lines can be switched on to reveal more details

Try to find the species-specific difference below. What is this gene? What additional information can the *T. brucei* outgroup provide about this locus? Can you develop an evolutionary hypothesis about what has happened?