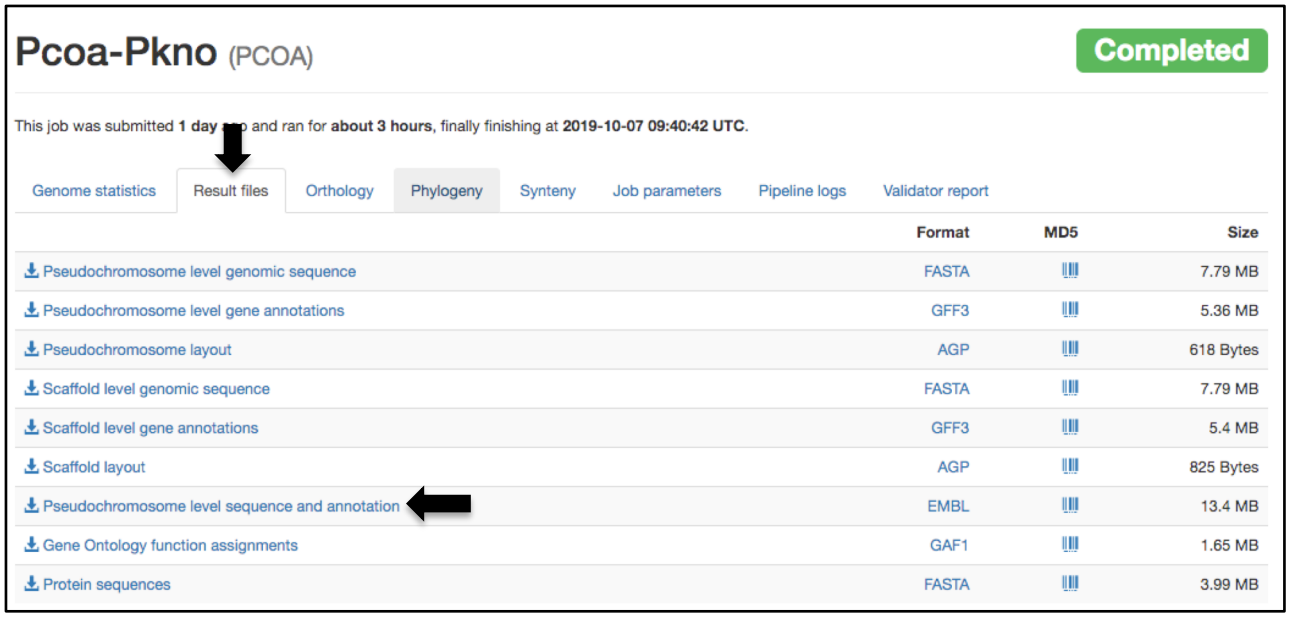
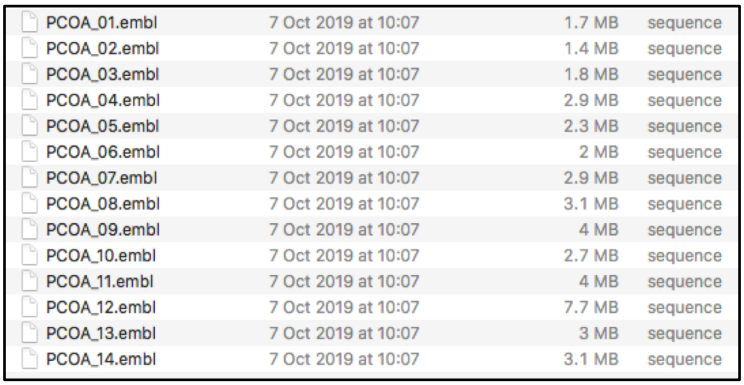
Viewing the Companion output in Artemis

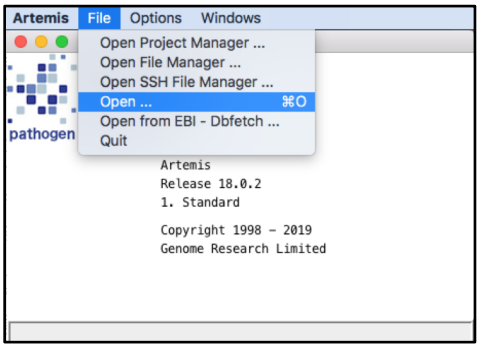
In the next exercise we will examine the Companion output in more detail. We will use Artemis to have a closer look. First, we need to download the files. Go to the tab 'Result files' and download the embl file 'Pseudochromosome level sequence and annotation'. The file is called embl.tar.gz.

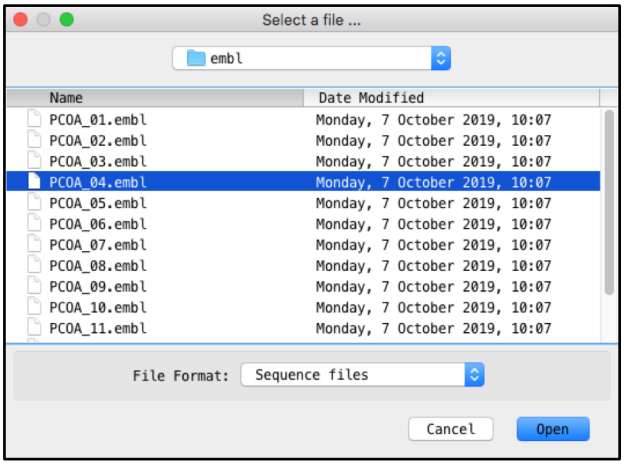


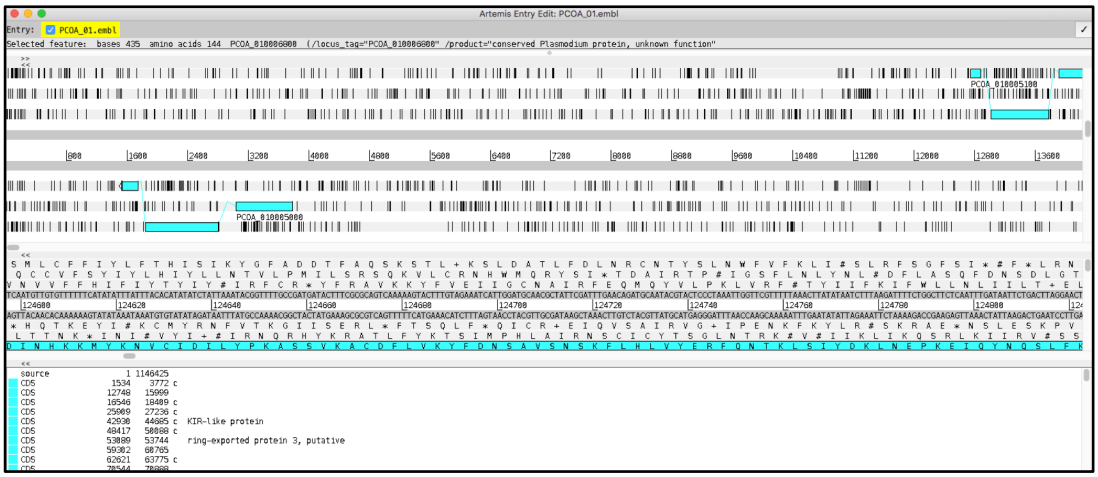
Locate the file called embl.tar.gz in your download folder. Double click on the file. A new folder called 'embl' will be created that contains all embl files. Here is the output for *P. coatneyi*. The file contains all chromosomes that were assembled and annotated by Companion. Contigs that could not be placed on one of the chromosomes are in the file *00.embl



Artemis is a great tool to visualise your Companion output. Choose one of the chromosomes you've just downloaded and open it in Artemis. As an example chromosome 4 of *P. coatneyi* (PCOA_04.embl) is shown.







Once you have your Artemis window open, scroll along the chromosome. Do you find any problems in the annotation? Can you see any missing genes? How many pseudogenes can you find? (Hint: search for the qualifier: pseudo). Do you think all of the pseudogenes are real or are some misannotated? You can answer this question quite easily by using a tool called ACT, Artemis Comparison Tool. We will show you how to use it in the next step!

Comparative Genomics
Visualising the Companion output in ACT

Introduction
In the next part of the exercise we will explore the Companion output in more detail with a tool called Artemis Comparison Tool (ACT). ACT was written by Kim Rutherford and was designed to extract 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 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 within 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. You can also use the online BLAST web server from NIH-NCBI to produce an alignment file that can be loaded into ACT. This option can only be used with BLASTN. We will cover this option in this Module.

Aims

The aim of this Module is for you to become familiar with the basic functions of ACT.

In the first part of this exercise you will learn the basic functions of ACT by looking at the companion output of *P. coatneyi* compared to *P. knowlesi*. By comparing two chromosomes you will be able to study the degree of conservation of gene order and identify small and large synteny breaks. You can also look for incorrectly annotated genes.

Once you are familiar with ACT we will show you how to create your own comparison file and explore your Companion output in ACT.

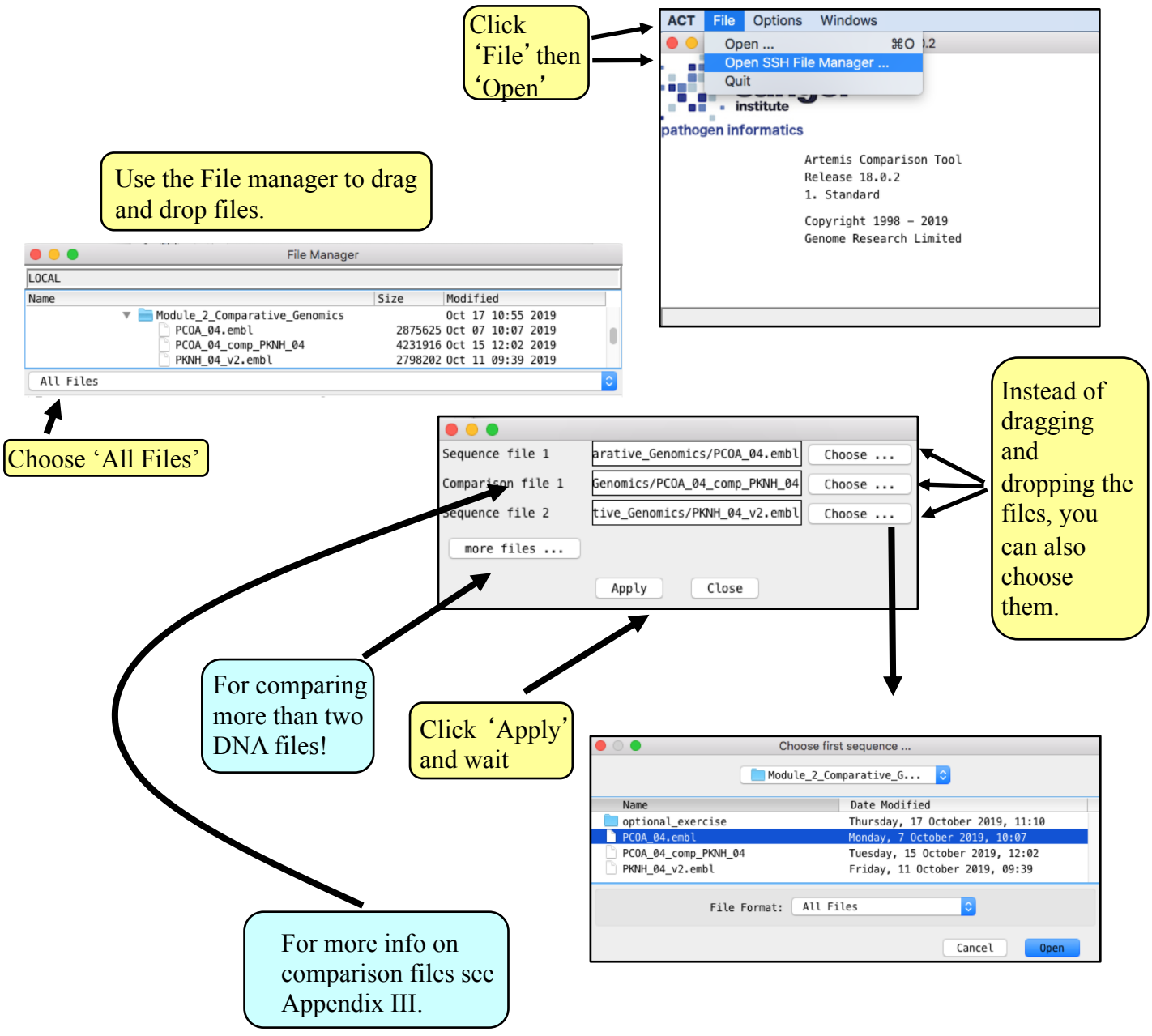
Part 1: Starting up the ACT software

In the first part of this exercise we will all use the same files. Make sure you're in the **Module_2_Comparative_Genomics** 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:

PCOA_04.embl	- embl file created by Companion
PCOA_04_comp_PKNH_04	- tblastx comparison file
PKNH_04_v2.embl	- *P. knowlesi* chr4 (reference used in Companion run)



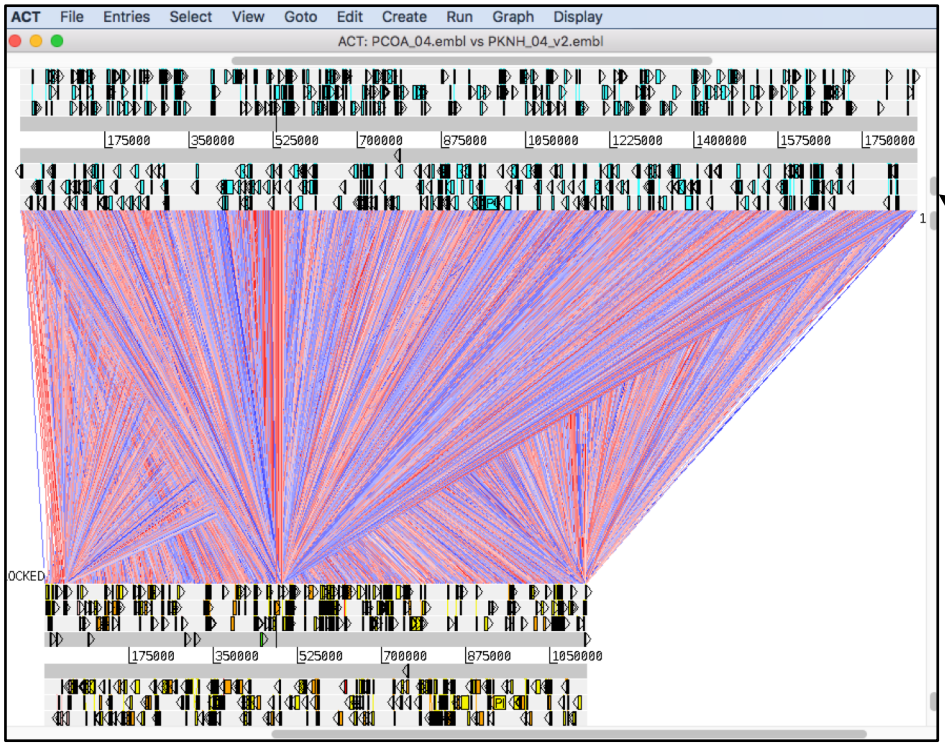
Once you have opened the files you will see a picture like this:

You can see the the name of the genomes displayed in ACT on the top of the window.

P. coatneyi

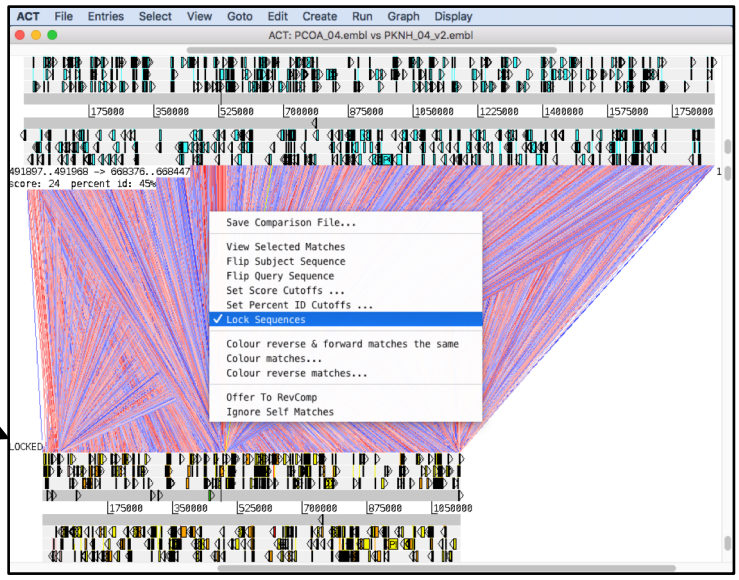
tblastx comparison

P. knowlesi chr4



1. 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

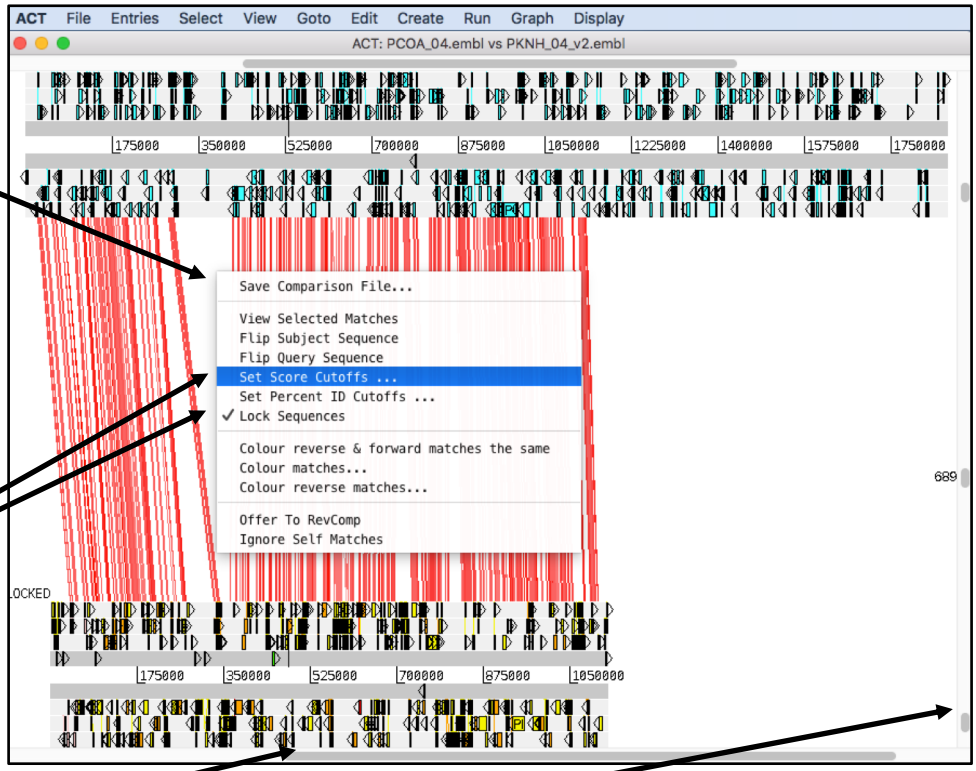
You can optimise your image by either removing 'low scoring' (or percentage ID) hits from view, as shown below **1-3** or by using the slider on the comparison view panel (**4**). The slider allows you to filter the regions of similarity based on the length of sequence over which the similarity occurs,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

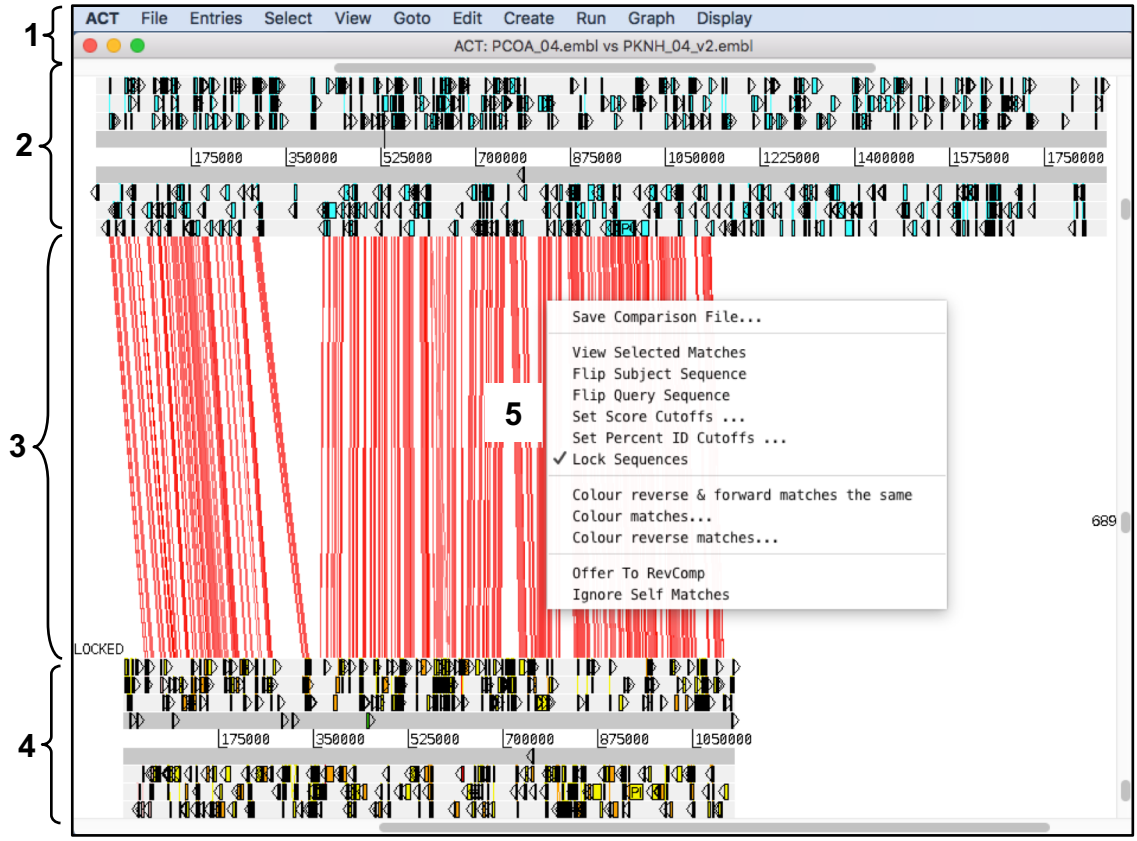


4

3

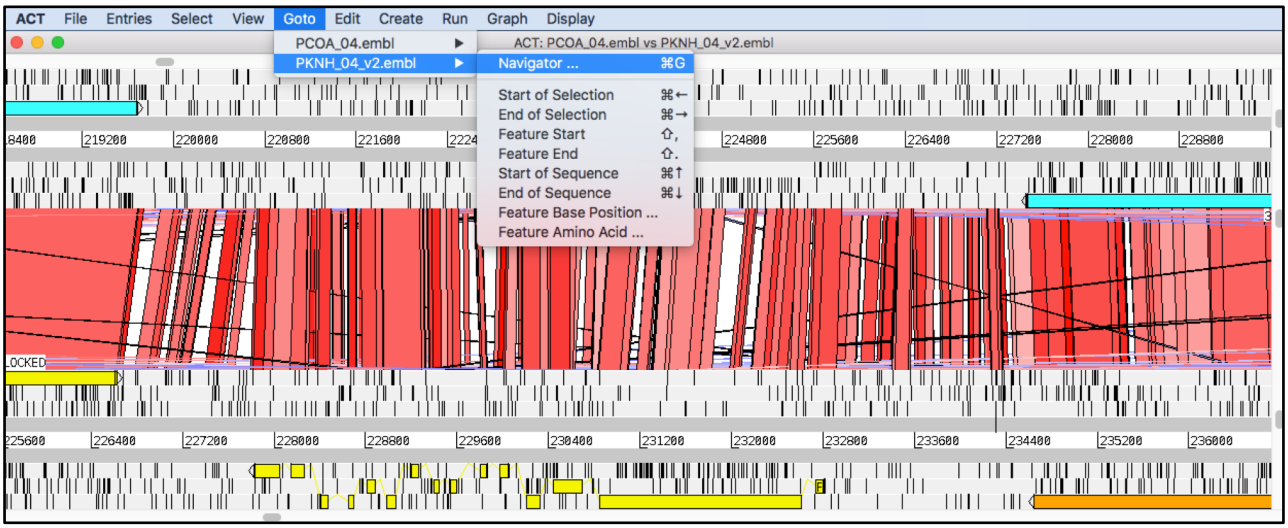
Move the sliders to manipulate the comparison view image.

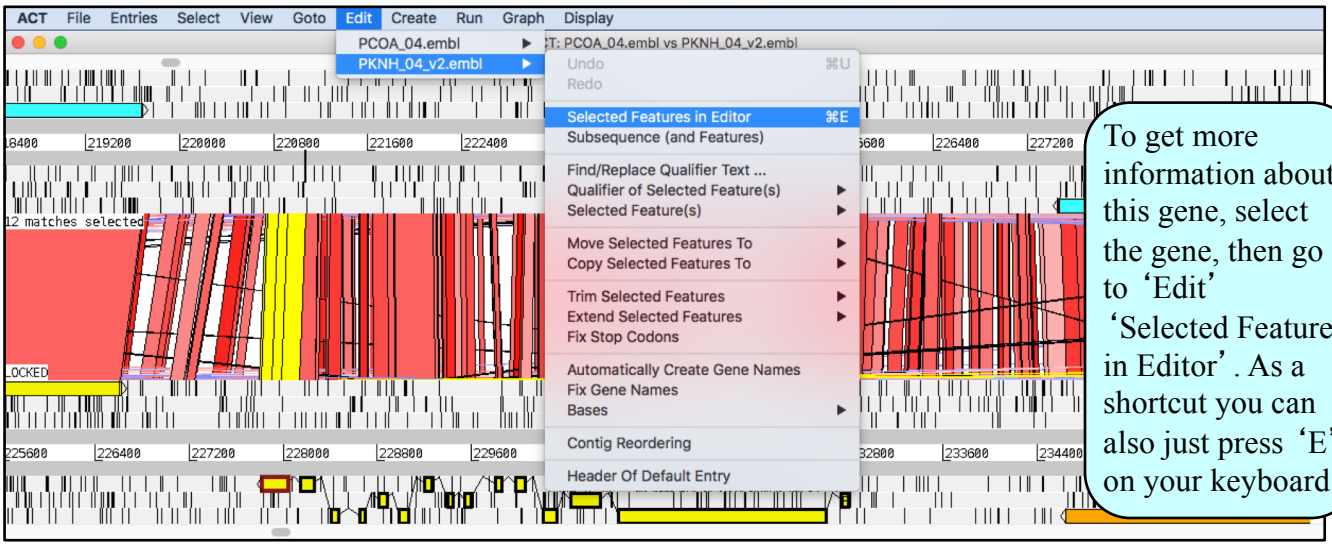
Now that you have an ACT window open let's look what is in there.



1. Drop-down menus. These are mostly the same as in Artemis. The major difference you will 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 ACT-specific menu which we will use later.

ACT is a great tool to spot any problems in the automatic Companion annotation. Scroll along the genome to find genes that were missed by Companion. Go to the *P. knowlesi* gene PKNH_0405900 by using the Navigator. Compare it to the *P. coatneyi* annotation on the top. Can you see that this gene has been missed by Companion?

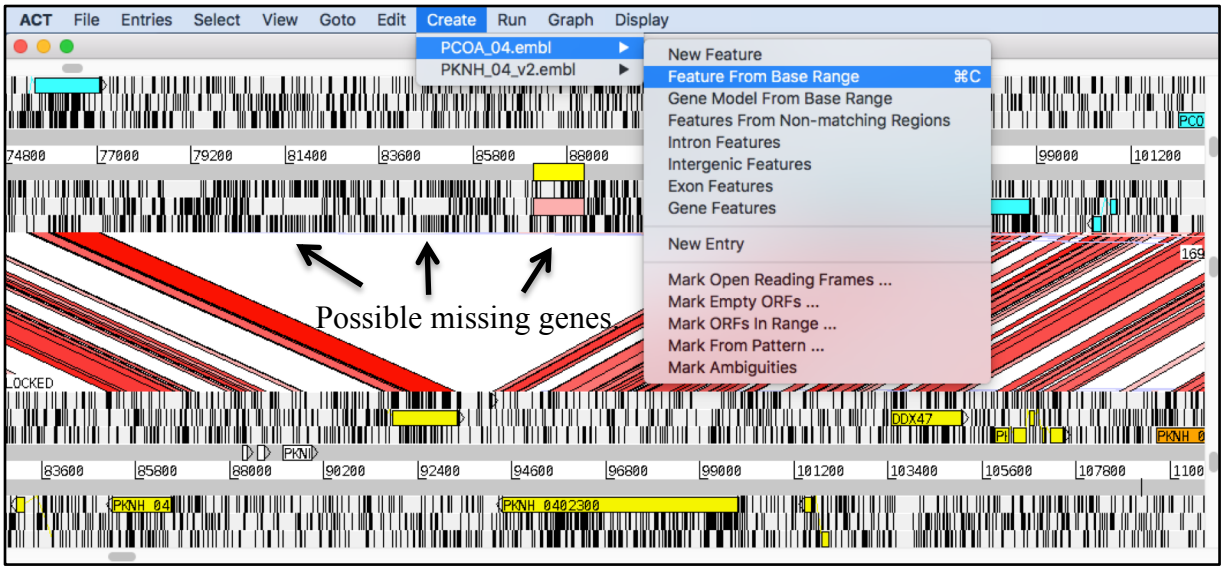




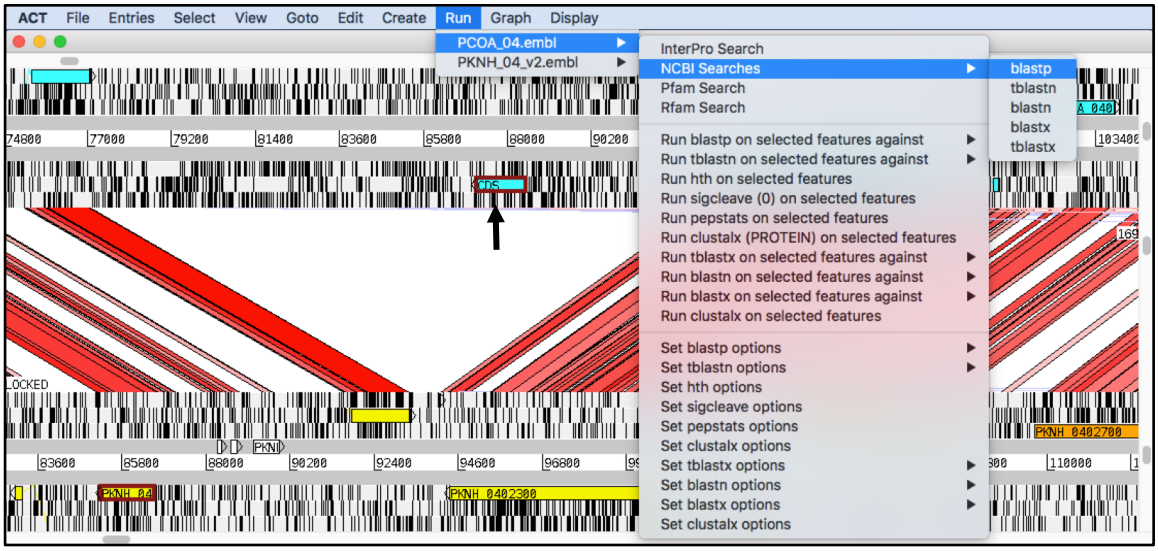
To get more information about this gene, select the gene, then go to 'Edit' 'Selected Features in Editor'. As a shortcut you can also just press 'E' on your keyboard.

Scroll along the chromosome and try to get an estimate on the number of synteny breaks. Do you think there are any missing genes in the synteny breaks?

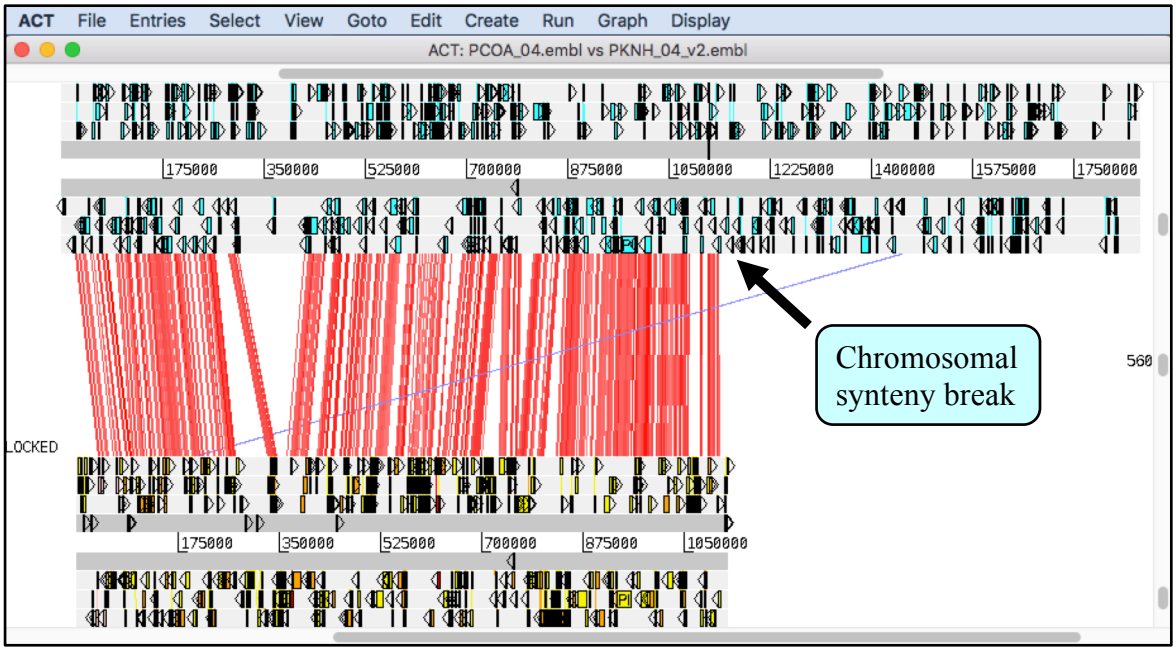
If you think there is a missing gene, you can just mark that area with your mouse. Then go to “Create” and choose “Feature from Base Range”. There is also a shortcut. Just press “C” on your keyboard.



Once you have created a feature, run blast to find out more about the possible missing gene. Can you assign a product?



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



We will show you in the next part how to open a three-way comparison in ACT and explore the synteny break. This is an optional exercise. You can skip it and proceed to part 2, exploring your own Companion output.

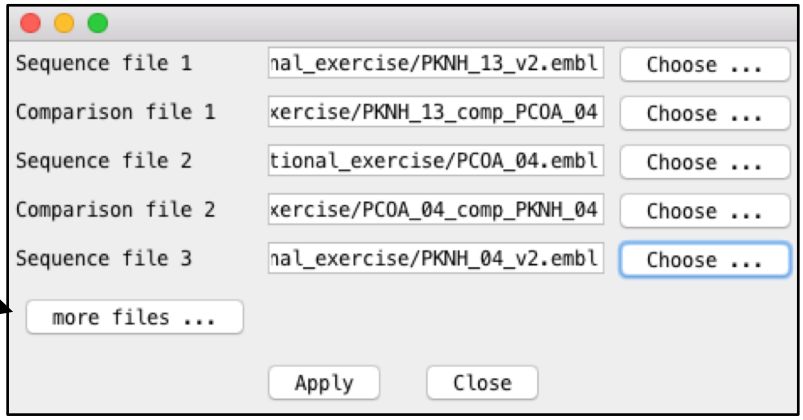
Optional exercise

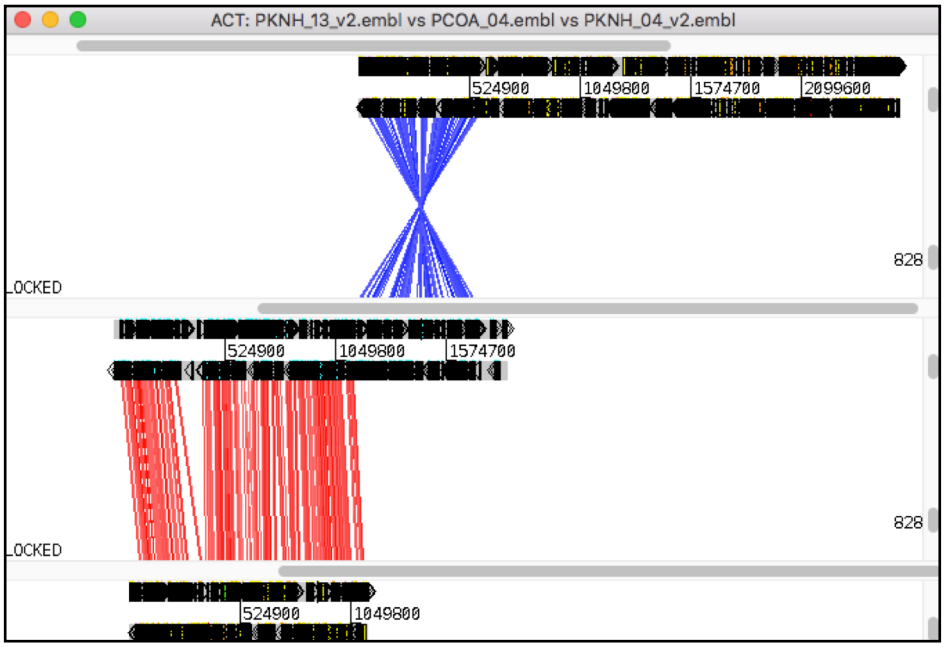
In this optional exercise we will show you how to open a three-way comparison in ACT and explore the chromosomal synteny break in *P. coatneyi*.

The files you are going to need are:

PKNH_13_v2.embl	- *P. knowlesi* chr13
PKNH_13 _comp_ PCOA_04	- tblastx comparison file
PCOA_04.embl	- *P. coatneyi* chr04
PCOA_04_comp_PKNH_04	- tblastx comparison file
PKNH_04_v2.embl	- *P. knowlesi* chr4

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





P. knowlesi chr13

P. coatneyi chr4

P. knowlesi chr04

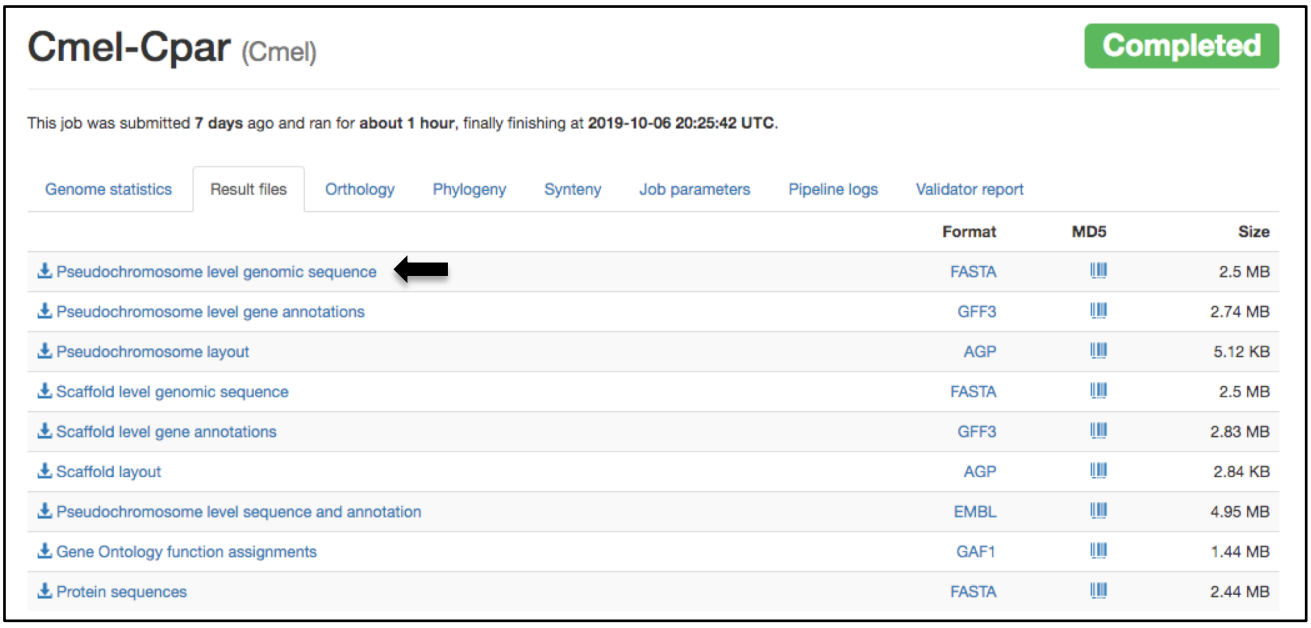
The blue 'hour glass' shape indicates that one of the chromosomes is reversed. With a right click in the middle area you can get an additional menu. Select 'Flip Subject Sequence' to flip one of the sequences.

Part 2: Explore your own Companion output in ACT

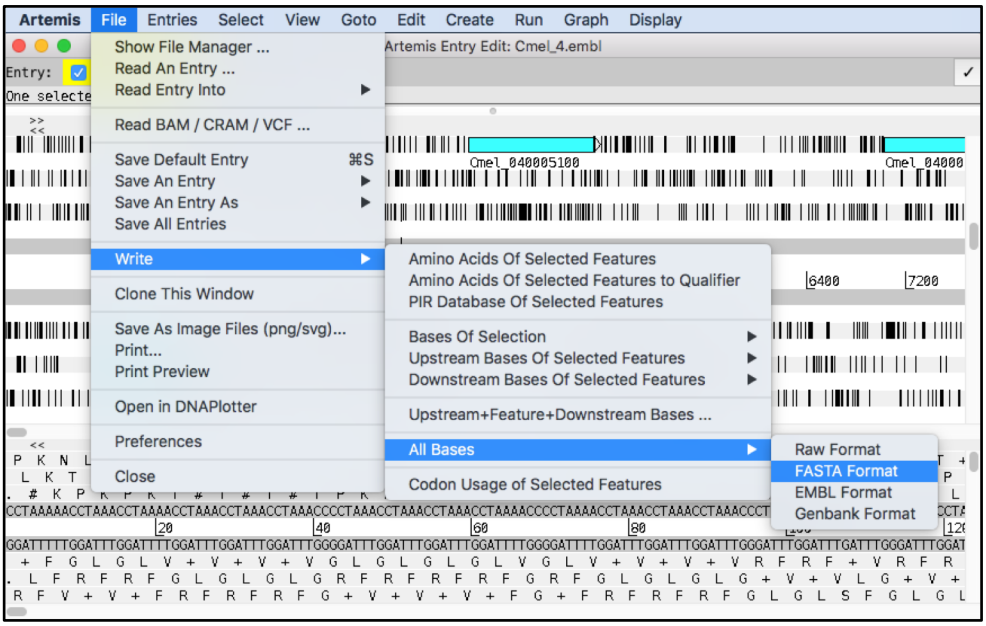
Now that you are familiar with the basic functions of ACT, let's explore your own Companion output. To do so, you need the Companion output (sequence and annotation), a comparison file and the reference (sequence and annotation). Have a look at your Companion output and choose one of the chromosomes you want to explore in more detail.

1. Download sequence and annotation of your Companion run

You've already downloaded the embl files in the first part of this exercise. To create the comparison file, you also need to download the sequence without annotation. Go to the 'Result files' and download the 'Pseudochromosome level genomic sequence'. This is a sequence file that contains all the chromosomes. Extract by copying and pasting the sequence of the chromosome you are interested in.



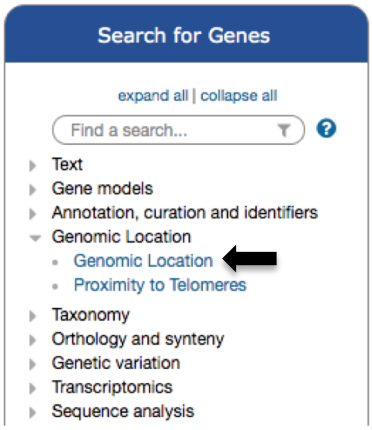
Alternatively, open the embl file you've downloaded from Companion in Artemis and save the sequence as shown below.

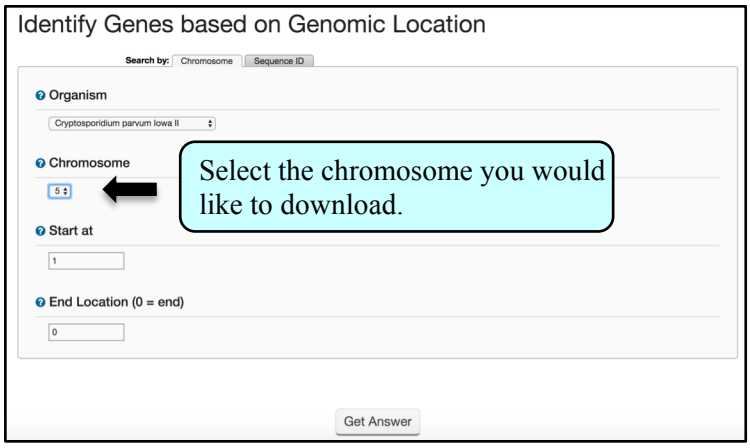


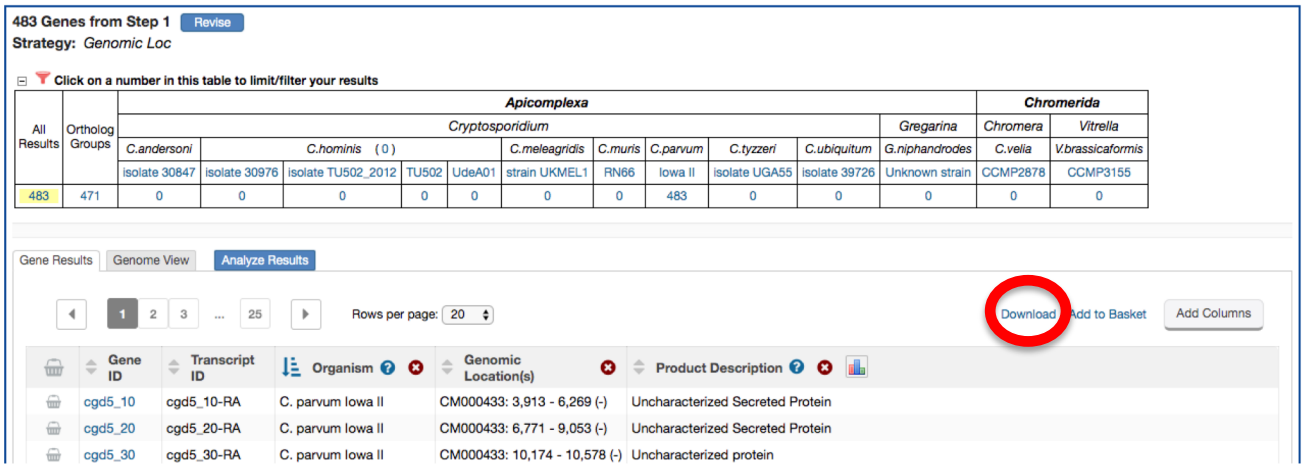
2. Download sequence and annotation of your reference genome

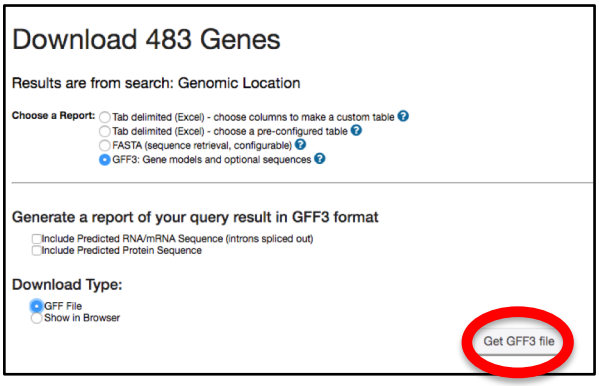
Downloading the sequence and annotation of your reference, can either be done on EuPathDB or on one of the main repositories like GenBank or ENA. How to download your sequence in GenBank is shown in the Appendix.

In EuPathDB you need to download the annotation and sequence for the chromosome you are interested in separately. In the first part of the Companion exercise we showed you how to download a FASTA file. Go to 'Genomic Sequences' and select 'Organism'. To download the annotation, use the search option 'Identify Genes based on Genomic Location'.



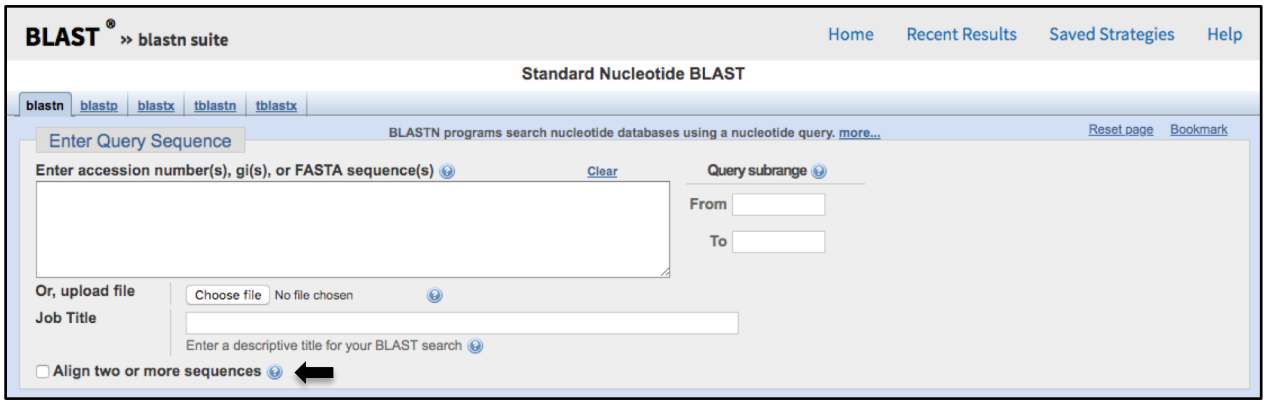




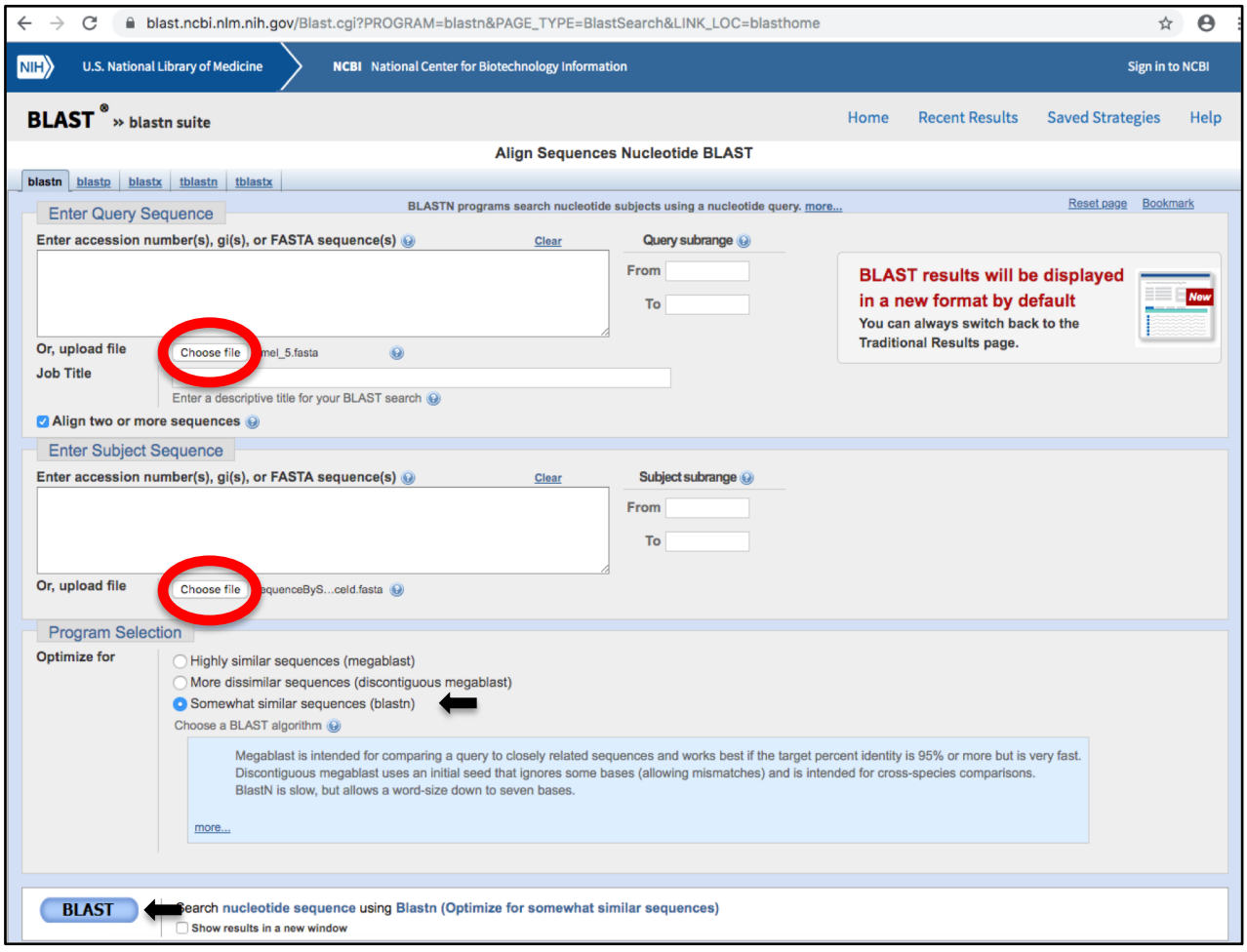


2. Create the ACT comparison file

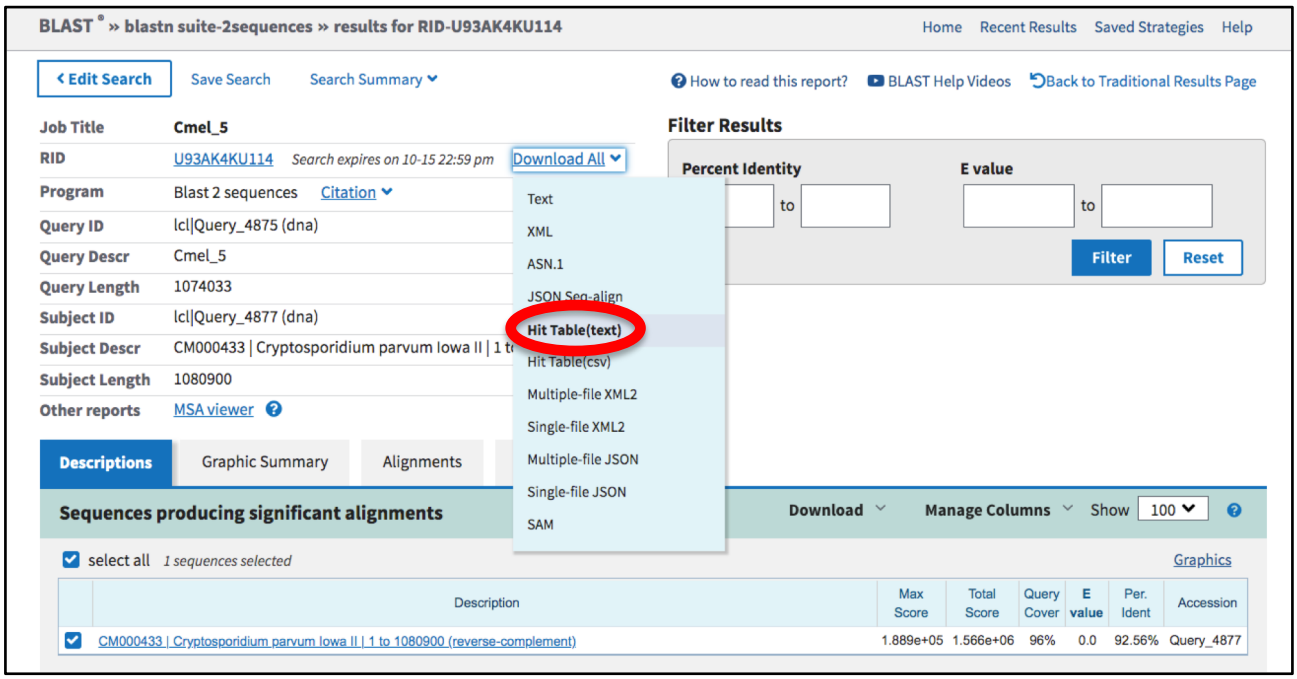
To create the comparison file, go to the following website: **https://blast.ncbi.nlm.nih.gov/Blast.cgi** and select 'Nucleotide Blast', 'Align two or more sequences'.



Upload the two sequences you would like to compare and then select a blast option.



The BLAST run will take a few minutes. Once this is done, select the Download option ‘Hit Table(text)’. This is the comparison file that you can open in ACT.

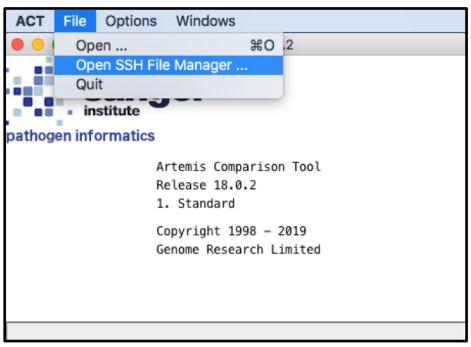


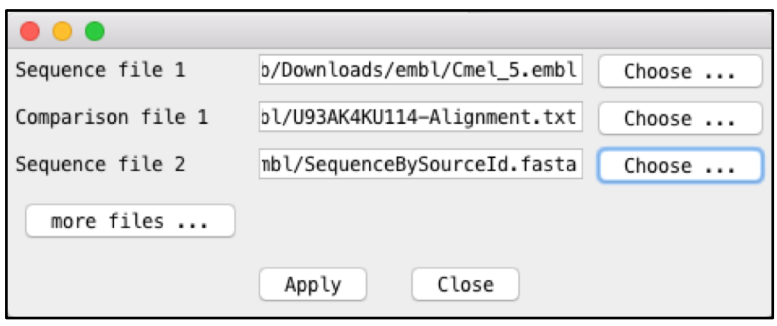
3. Open your file in ACT

Let’s open the two sequences in ACT.

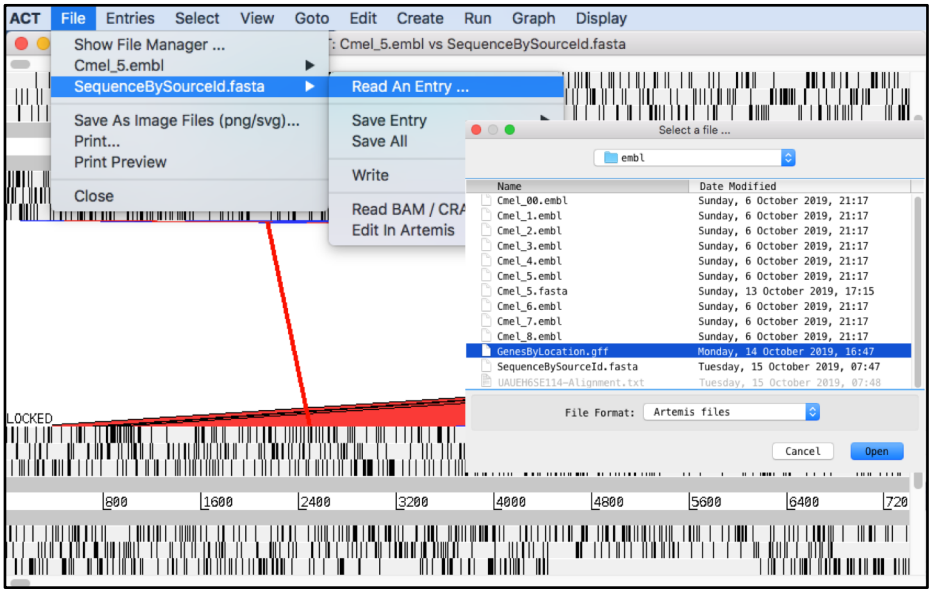
The files that you are going to need are:

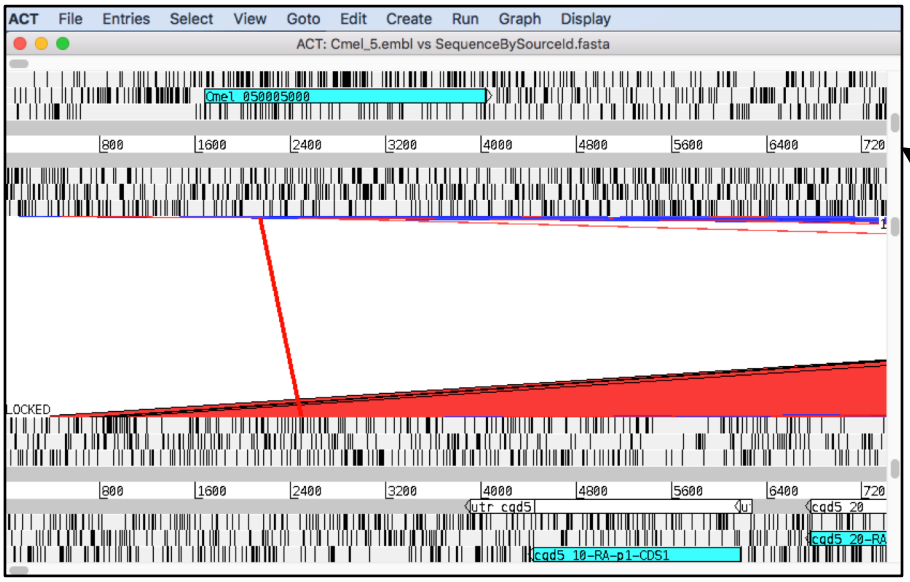
Companion embl file
Comparison file (Hit Table – downloaded from NCBI)
Reference (Fasta file and GFF file downloaded from PlasmoDB)



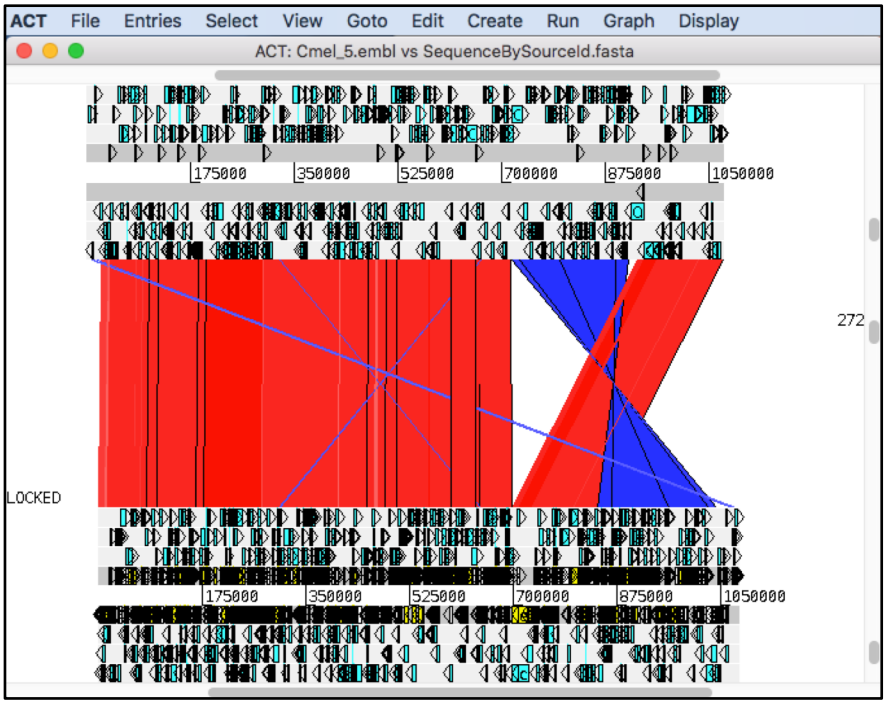


Once you've opened ACT load in the GFF file you've downloaded from EuPathDB.





Use the vertical sliders to zoom out to get an overview. Drag or click the slider downwards from one of the genomes.



Once you've got an overview, zoom in to explore interesting areas, i.e. synteny breaks.