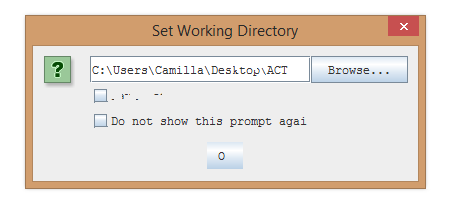
How to use Artemis Comparison Tool (ACT) to align two genomes.

# Downloading and installing ACT (on Windows)

1. Make an ACT directory (on desktop is probably easier) – ie a folder labelled ‘ACT’ or something similar, in which you will put all your files associated with this program.
2. The program can be downloaded from sanger.ac.uk

<https://www.sanger.ac.uk/resources/software/act/#downloads>

1. Choose ‘ACT for Windows’ and click to download the file. If your computer tried to ‘unzip’ the file, don’t let it. This is a java file and doesn’t need unzipping. It does need you have to the most up to date version of java so check that before proceeding to the next step.
2. Once you have downloaded the file, find it in your downloads – cut and paste it into your ACT directory. It should be called ‘act.jar’
3. Double Click on this, and it should just run, bringing up a window like this ->

Asking you to set working directory. Click ‘OK’ as it has the correct directory listed already.

1. Before going any further, you will need to download the fasta files of the genomes you would like to compare, as well as the ‘comparison file’ of these two sequences.

# Getting started: Finding the fasta sequence of the two genomes

1. Go to

**ftp://ftp.ncbi.nih.gov/genomes/**

1. Scroll down to end and double click ‘genbank’ folder
2. Double click on ‘BACTERIA’ folder, then search for the bacterium whose genome sequence you would like.
3. Double click this and will open a page listing all the files associated with that genome. The useful ones (in general, and for ACT) are:

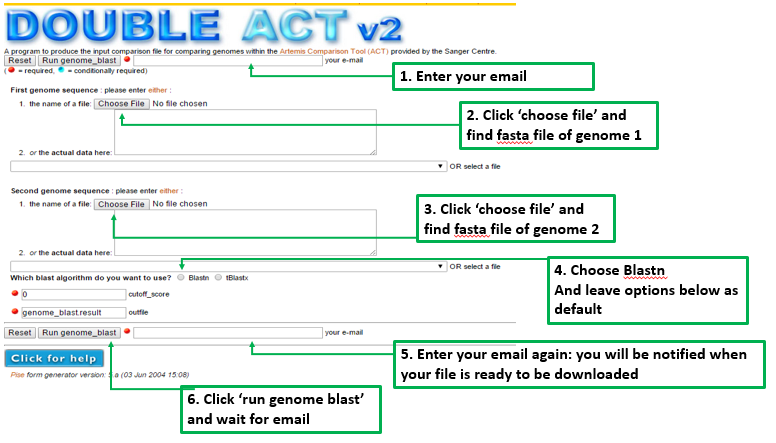
* fna : FASTA file of the chromosomal sequence (think "n" = nucleotide)
* gbk : Genbank file containing meta-data, sequence, and annotations
* gff : GFF3 file containing annotations only (coordinates relative to the .fna file)
* faa : FASTA file of the translated coding regions (proteins) annotated in the .gbk/.gff ("aa" = amino acids)

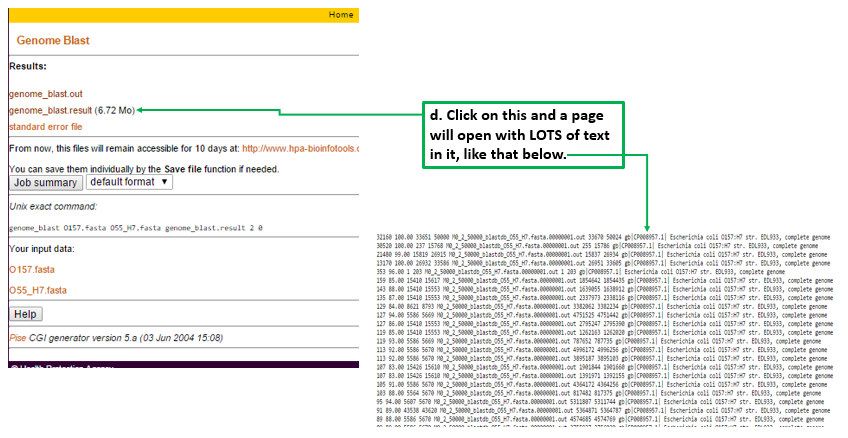
1. Click the fasta file (ie .fna file) for your bacterium of interest -> this will open a new tab with the fasta sequence in. Click CTRL+A to highlight all the text in the page, then click CTRL+C to copy all this text.
2. Open a text editor, like ‘Notebook’ -> NOT MS Word!
3. Paste (CTRL+V) and then save the file as ‘nameofbacteria.fasta’ for example EcoliO157H7.fasta -> make sure to save it in your ACT directory.
4. Do exactly the same to get the genome sequence of the second bacteria you would like to compare. Name it appropriately.

# Create a comparison file

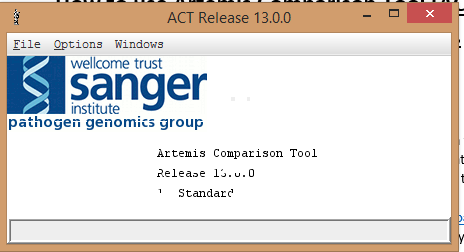
As the ACT tool you downloaded is only for *visualising* the comparison of two genomes, the ‘alignment’ program which actually compares the bases is online.

1. Go to: **http://www.hpa-bioinfotools.org.uk/pise/double\_act.html**
2. The page will look like this (follow instructions in picture)

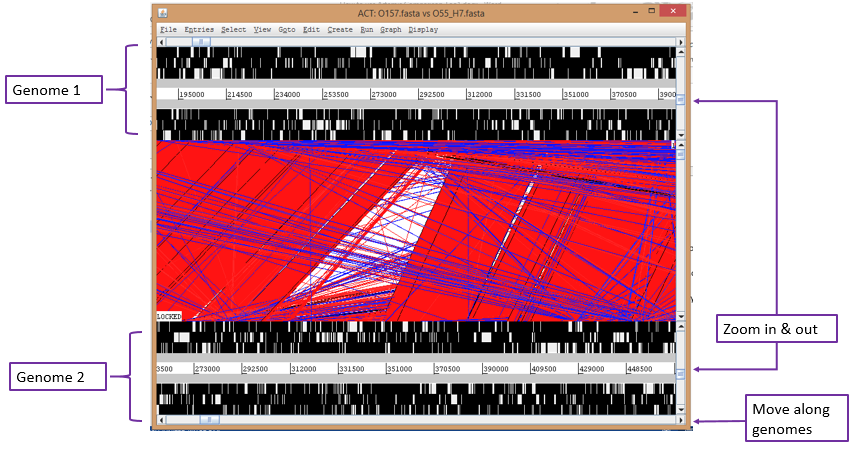


1. You will receive a link to the page containing your result file by email. Once clicking on the link it will look something like this:
2. Click CTRL+A to highlight all the text in the page, then click CTRL+C to copy all this text.
3. Open a text editor, like ‘Notebook’ -> NOT MS Word!
4. Paste (CTRL+V) and then save the file as ‘comparison.fasta’ in your ACT directory again.

# Running ACT



1. Go to the pop up window that opens when you run ACT (as in part 1e)
2. Click on ‘File’ and then ‘Open’
3. A window will open asking to enter Sequence 1, comparison file and sequence 2. (If you would like to compare more than 2 genomes click more.)
4. Click ‘choose’ for Sequence one, & choose the fasta file for your first genome of interest.
5. Do the same for comparison file and sequence 2, then click apply and wait for the display window to load.
6. A result window should come up, see image below (you may need to zoom out to see it better):
   1. Red connections = represent matches that run in the *same direction* on the two genomes being compared,
   2. blue connections = alignments that run in *opposite directions* on each genome
   3. White areas = signify where there is *no alignment* in one of the two genomes. (eg a pathogenicity island).



# Interpreting results to find unmatched regions (ie unique to one genome)

1. In the results window, click on the ‘*Create’* option, then choose your genome of interest and a list of options will come up.
2. Click on ‘*Features From Non-matching Regions*’ and this will label genes/intergenic regions/etc that are not matched with a bright green colour.
3. To Blast this region in order to determine if it is specific to other bacteria aside from the one compared by ACT:
4. click on one of the bright green regions
5. it will become highlighted (thick black line around green region)
6. Click on the ‘*Run’* option -> Choose the same genome of interest -> *NCBI blast* -> *blastn*
7. Leave default settings in pop up window and
8. This will take a little time, and will open a new tab in your default web browser with your NCBI blastn results displayed as usual.