

Artemis and analysis of genomes

13.1 Using genomics to investigate antibiotic resistance and genomic features

13.2 Overview

As has been outlined in previous lectures and practical's genomics is an incredibly powerful tool to understand the molecular epidemiology, transmission and evolution of bacterial pathogens.

In the previous practical you investigate resistance using online tools that identify the presence of genes and resistance conferring mutations, we are now going to zoom down to individual bacterial genome which can help us further understand antibiotic resistance and extract data for further analysis.

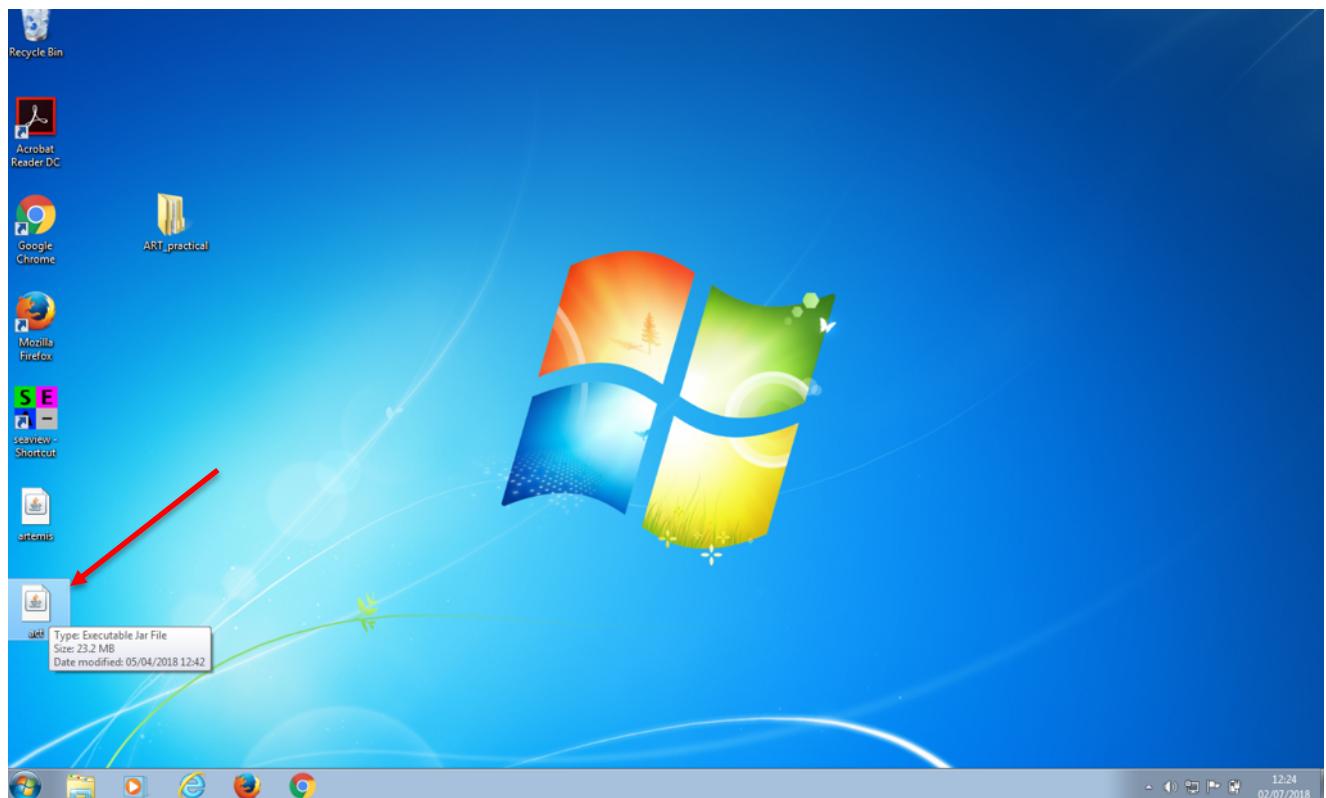
- Artemis is a genome browser and annotation software developed at the Wellcome Sanger Institute (<http://www.sanger.ac.uk/science/tools/artemis>). It is freely available to download for PCs and Mac.
- The practical is designed to give you a basic understanding of Artemis software and to get a better idea of genome structure and content. It is not expected for you to master this software in a single session.
- **Important note – Artemis has a huge number of features for many different tasks – we are just concentrating on the basics – so don't worry about most of the what is there.**
- If you have any questions during the practical please ask!

Analysis of resistance in genomes

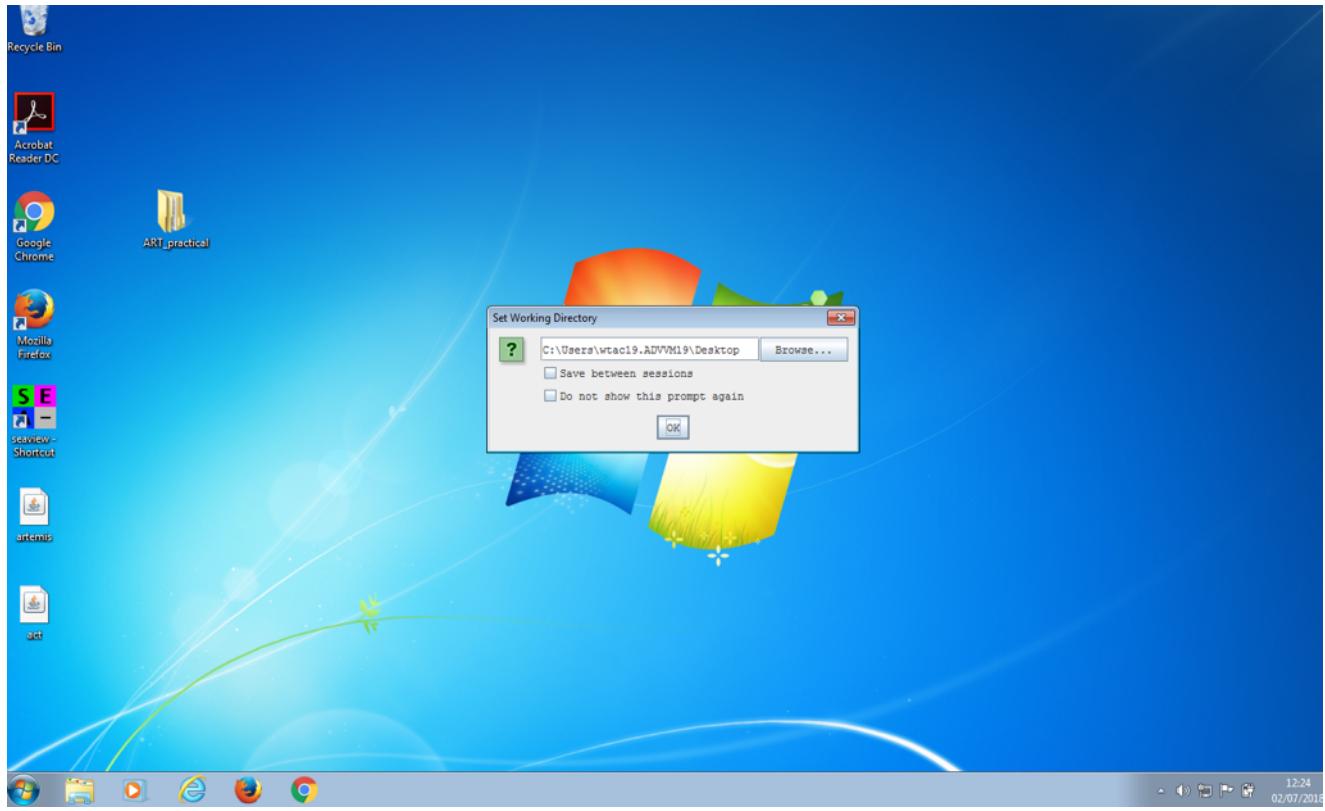
14.1 Comparative genomics

We are now going to investigate the *mecA* region in more detail to understand the origins and evolution of methicillin resistance in *S. aureus*. We are going to use a variation of Artemis called Artemis Comparison Tool (ACT):

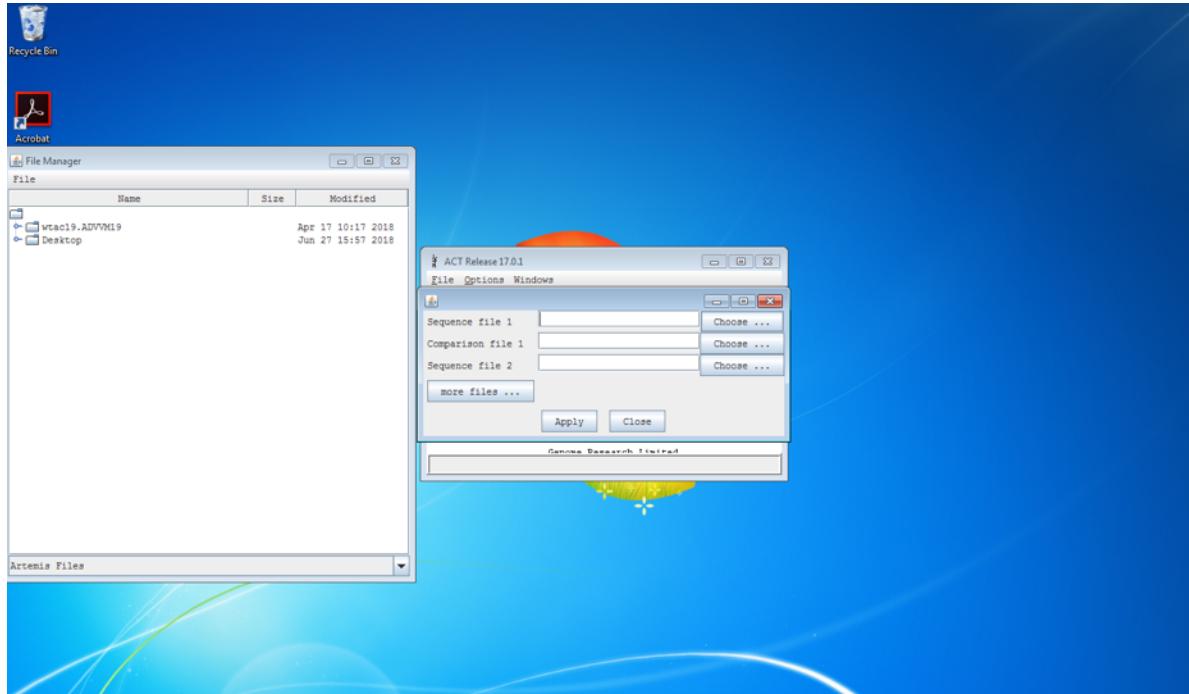
- Artemis Comparison Tool is a genome browser that lets you view comparisons of genomes. It was developed at the Wellcome Sanger Institute (<http://www.sanger.ac.uk/science/tools/artemis>). It is freely available to download for PCs and Mac.
- The practical is designed to give you a basic understanding of the ACT software and to get a better idea of genome structure and content. It is not an expected for you to master using the software in a single session.
- **Important note – ACT like Artemis has a huge number of features for many different tasks – we are just concentrating on the basics – so don't worry about most of what is there.**
- If you have any questions during the practical please ask!



Step 1: Double click on the ACT symbol to open ACT. **NOTE:** if you cannot locate ‘act’ icon on your desktop, open up a new terminal window and type in ‘act’. This will also open ACT tool too.

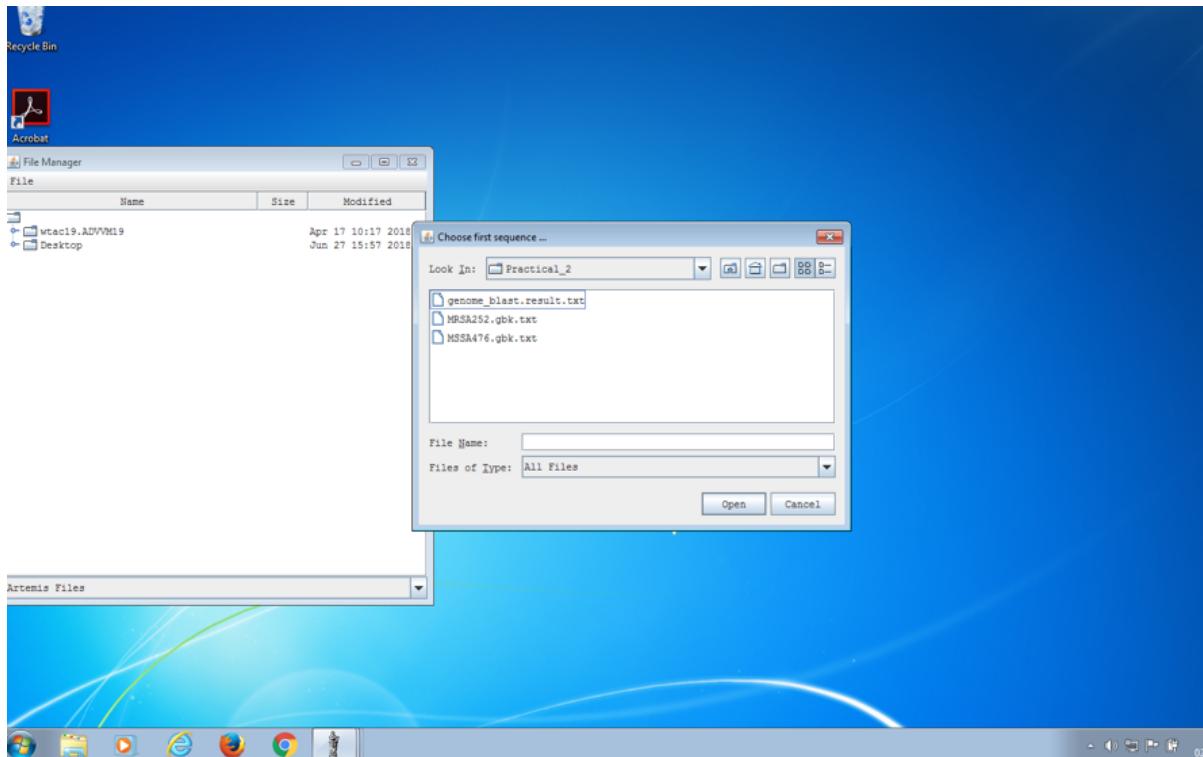


Step 2: A dialog box will appear. Click ‘OK’.

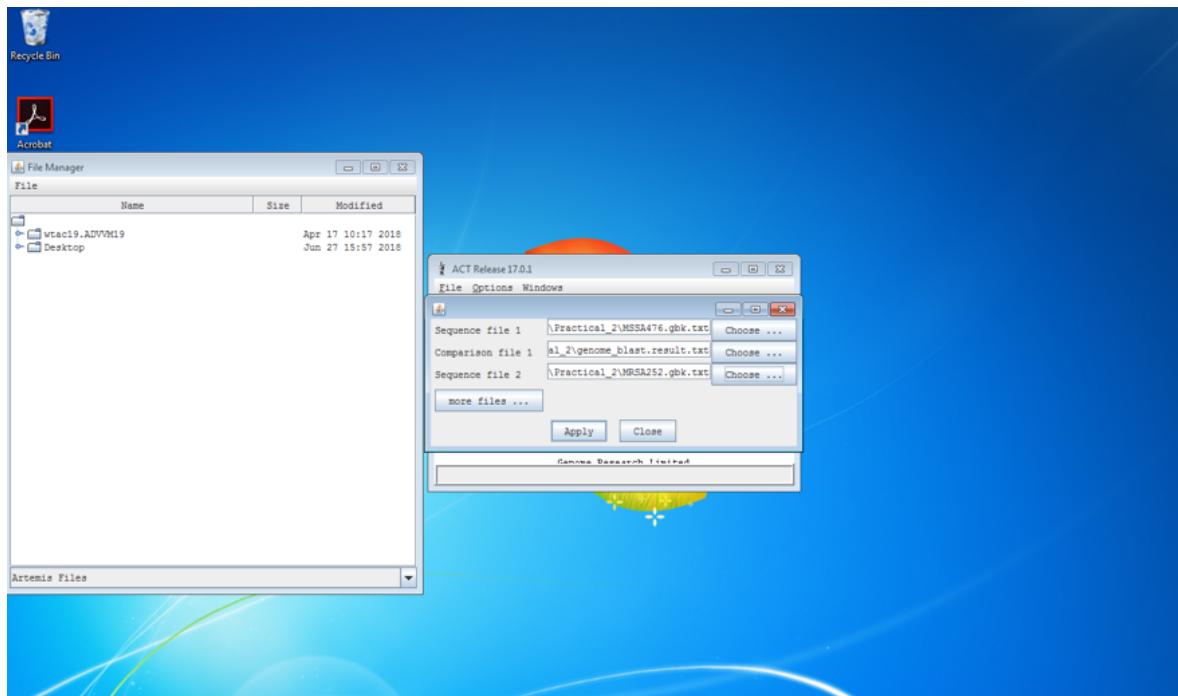


Step 3: We are now going to look at two genomes of *Staphylococcus aureus*; MSSA476 - a methicillin-sensitive strain and MRSA252 – methicillin-resistant strain. Click ‘File’ and then select ‘Open ...’. A box will open like the one shown above. For

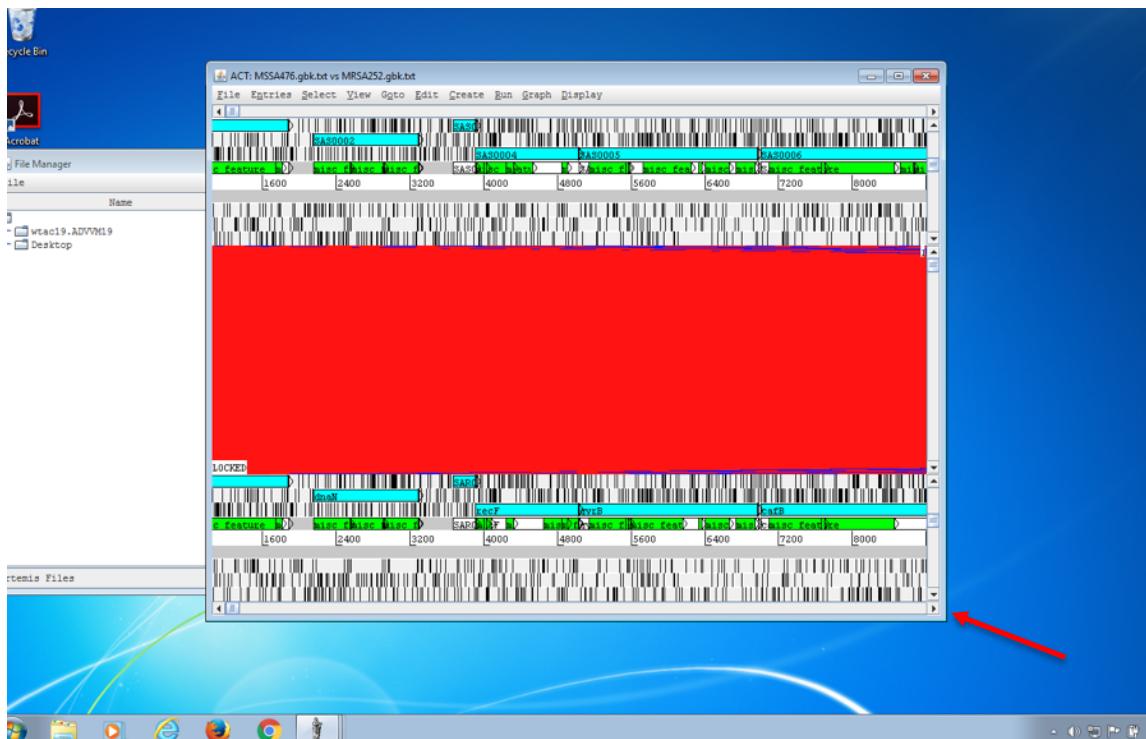
Sequence file 1 – click the ‘Choose …’ box. Navigate to the directory **on your desk top** and then click the ‘ACT_practical’ folder, and then open the Practical 2 folder.



Step 4: The folder contains three files, these are the two genome sequences (MSSA476.gbk.txt, MRSA252.gbk.txt) and a blast results file of the two genomes (MSSA476_vs_MRSA252_blast_results.txt). Click MSSA476.gbk.txt and click ‘open’.



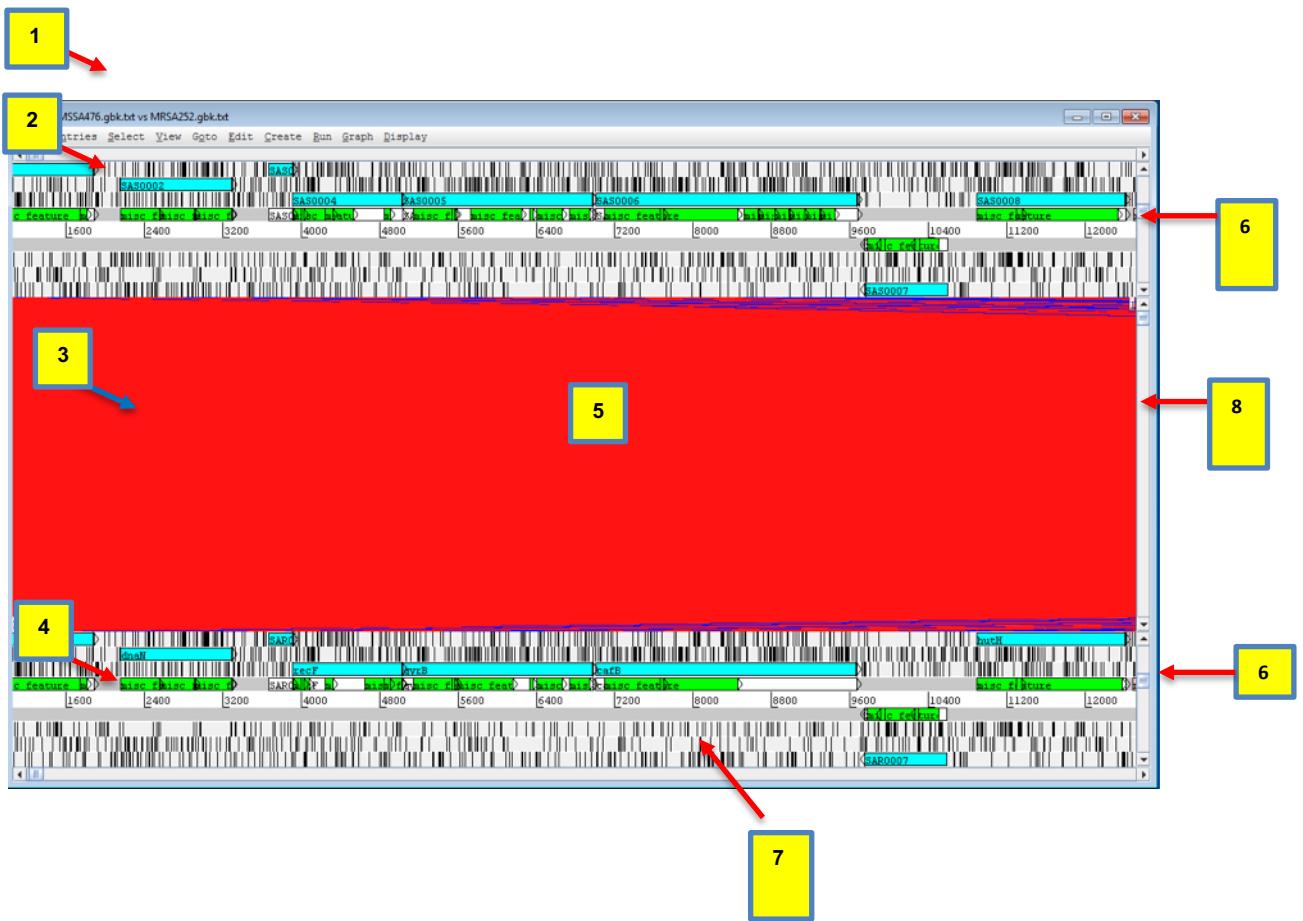
Step 5: Now click the ‘Choose …’ button for ‘Comparison file 1’ and then select MSSA476_vs_MRSA252_blast_results.txt and click ‘open’. Now do the same for Sequence file 1 selecting the file MRSA252.gbk.txt. Now press the ‘Apply’ button. A message will appear asking if you want view errors. Just click yes.



Step 6: A window like this will appear. Drag the corner to make the window bigger.



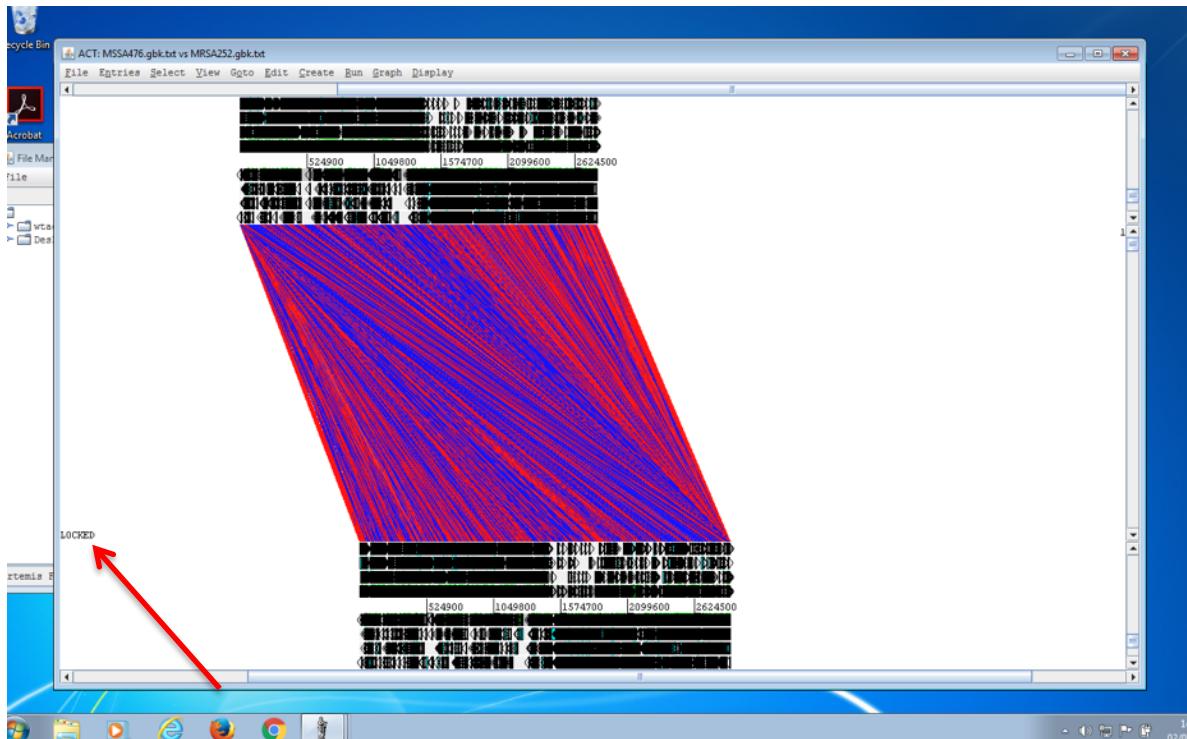
Step 7: You should now have a view like the one above. As you can see it is very similar to the view from Artemis. But you have two sequences on top of each other – the top one being MSSA476 and the bottom one MRSA252. The red colour in the middle indicates that these sequences are the same (conserved) in the two genomes. You can see that the genes and orientations are the same. If this is in blue it means they are the same but their orientation is flipped (don't worry about this now we will see this again later on).



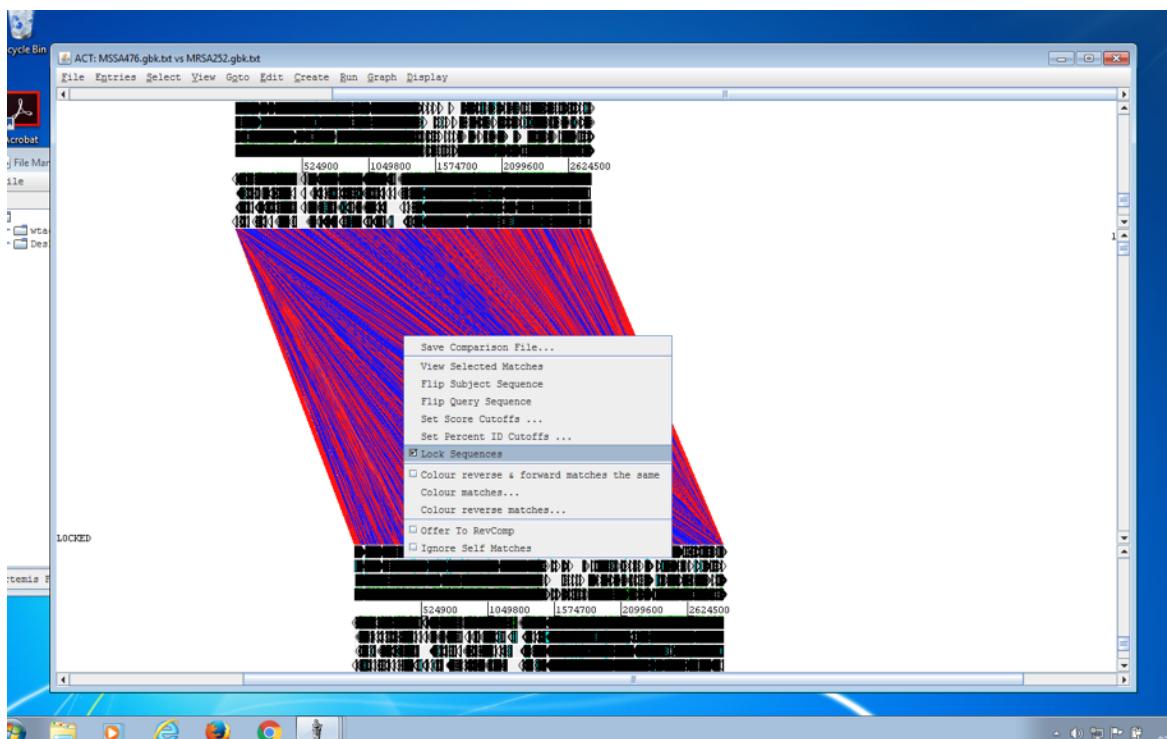
Step 8: The ACT window is very similar to the Artemis window we used earlier but gives you a comparative view.

1. Drop-down menus: These are basically the same as you used 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 (in this case MSSA476 or MRSA252) before going to the full drop-down menu.
2. This is the Sequence view panel for 'Sequence file 1' (Subject Sequence – MSSA476) 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 – MRSA252).
5. Right button click in the Comparison View panel brings up this important ACT-specific menu which we will use later.
6. Sliders for zooming view panels.
7. Sliders for scrolling along the DNA.
8. Slider that allows you to filter the regions of similarity based on the length of sequence over which the similarity occurs.

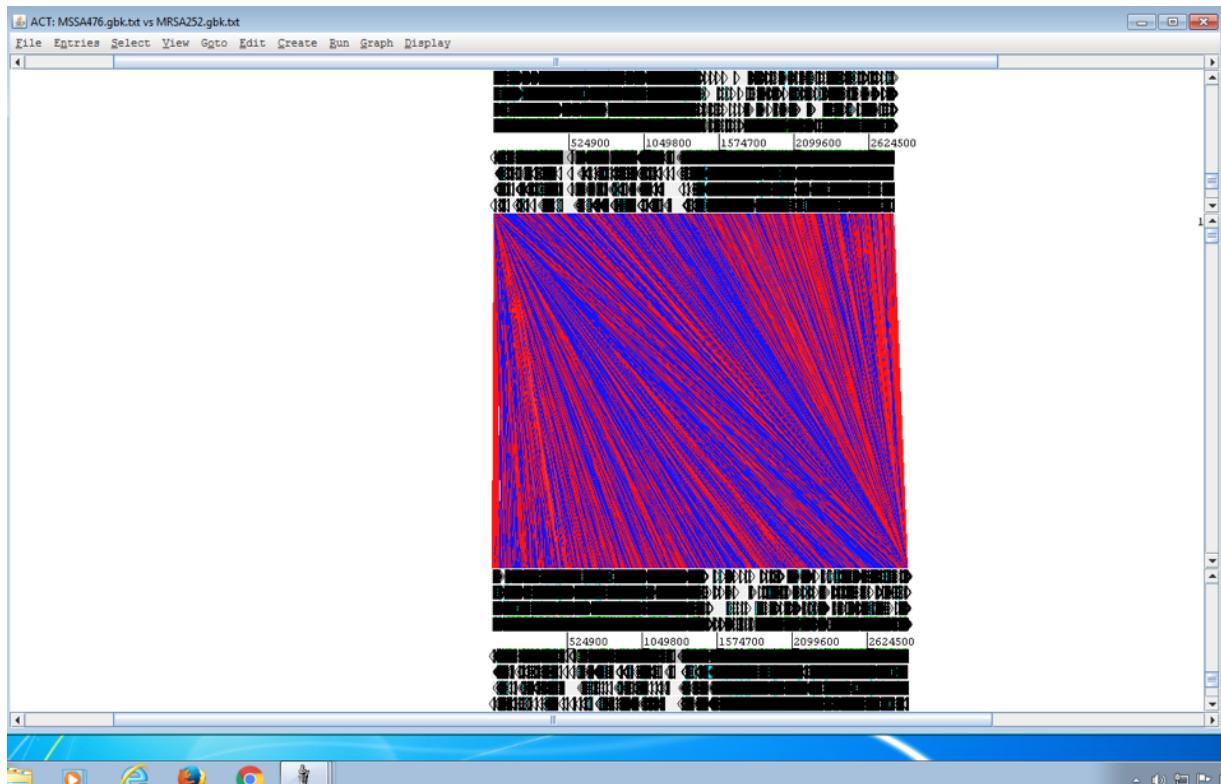
Step 9: Use one of the zooming sliders (marked 6) and zoom the view all the way out to the maximum.



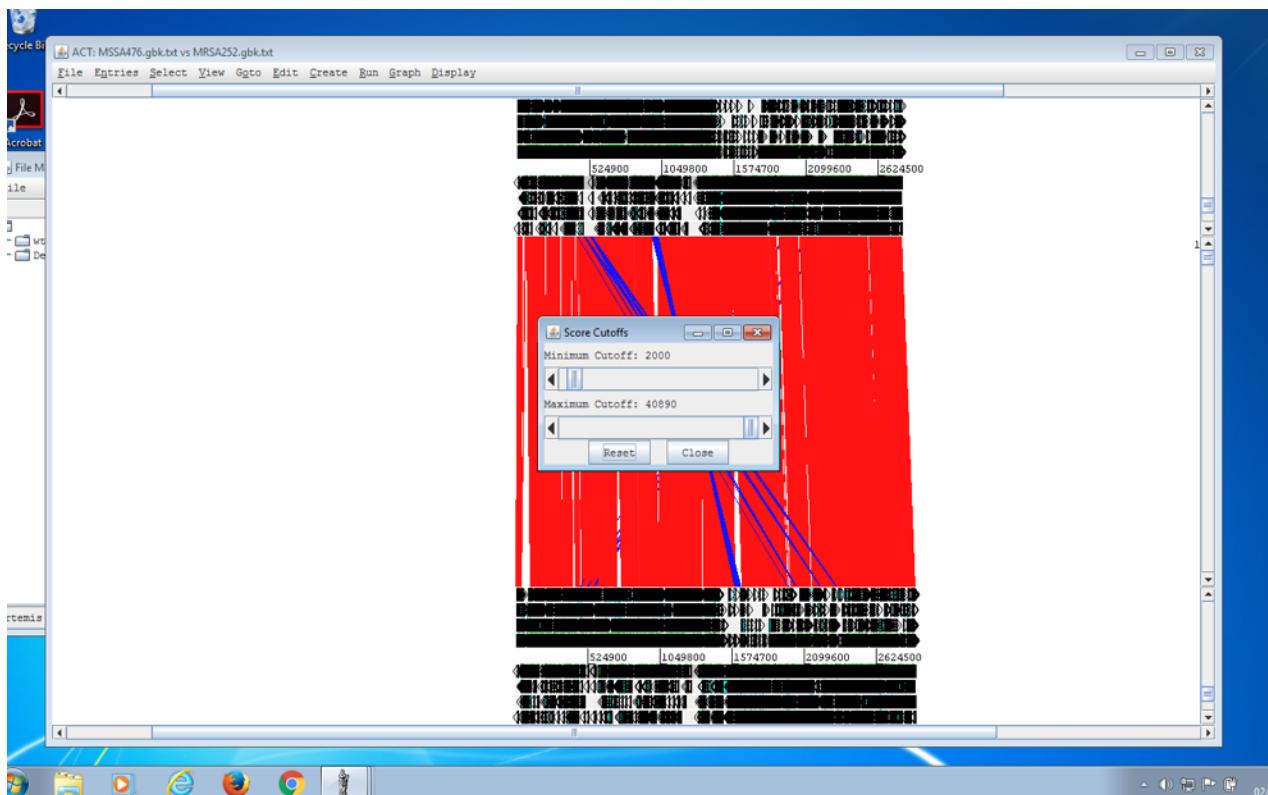
Step 10: You will notice here that it says 'LOCKED'. If you use one of the sliders to move down the genome – both genomes will move together. We are going to turn this off so the genomes move independently.



Step 11: Right click the mouse anywhere in the comparison area (the red / blue). A menu should appear. On the menu un-tick the box that says 'Lock sequences'.

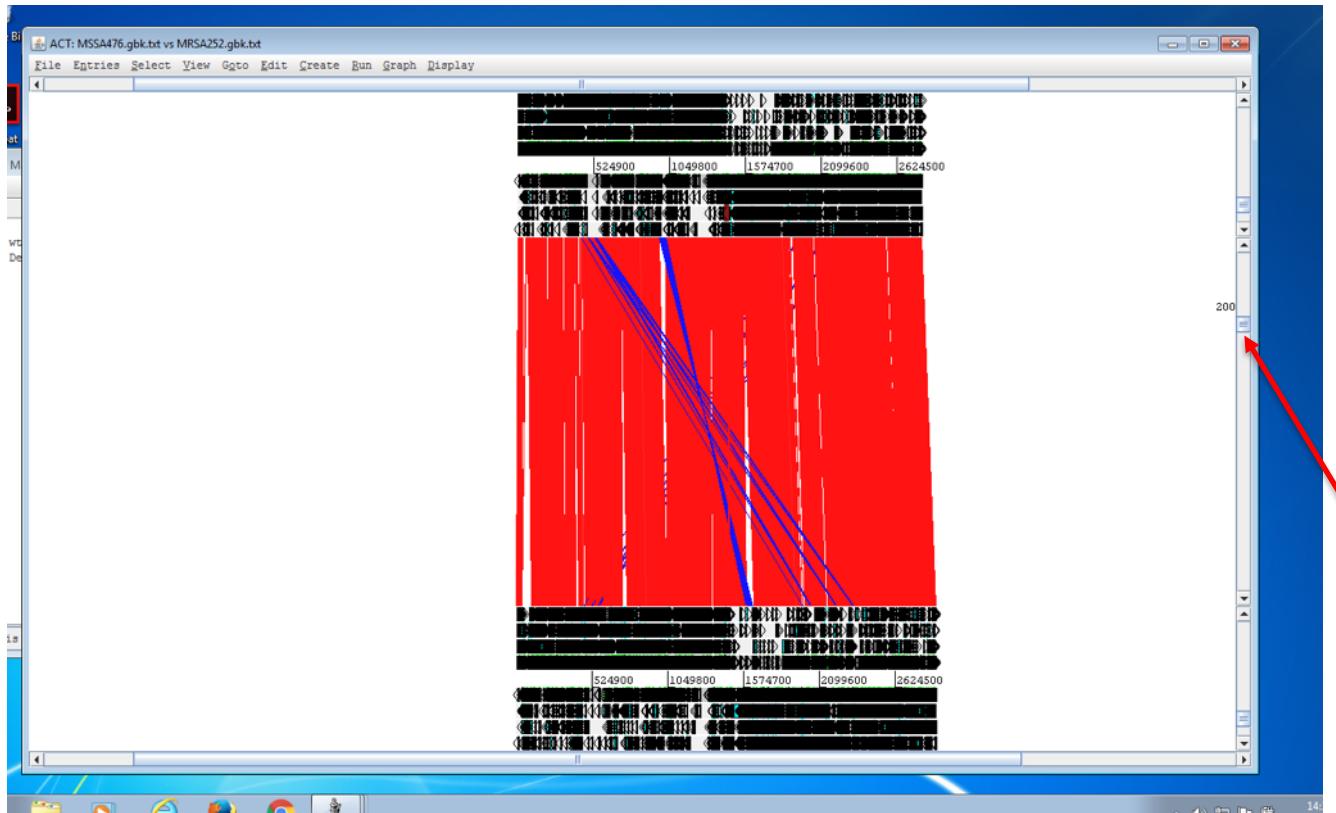


Step 12: The sequences will now move independently. Try moving them around yourself. Once you have tried, line them up again so it looks like the picture above.

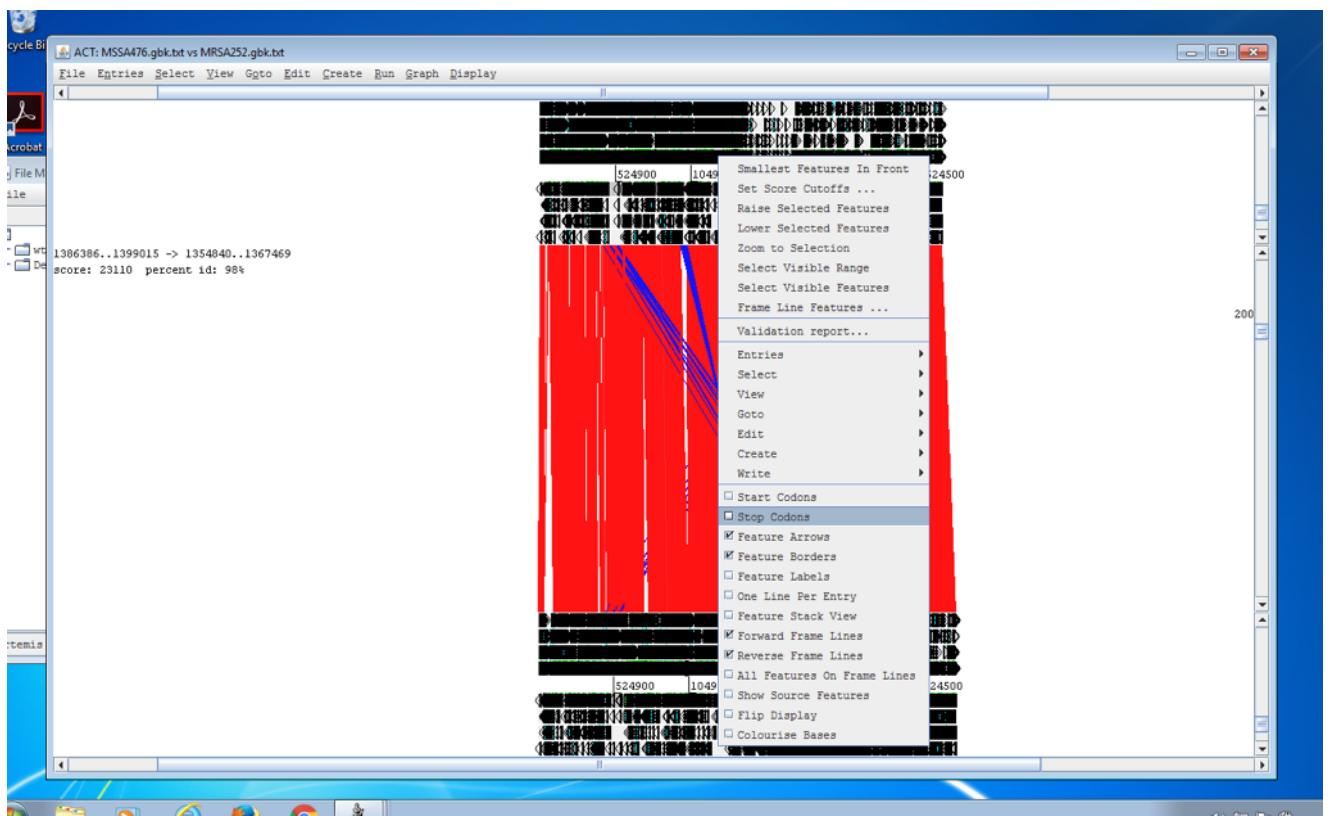


Step 13: Now, right click the mouse anywhere in the comparison area (the red / blue). Select the 'Set score cutoff'. Move the upper slider in the Score Cutoff of menu to 2000. This will remove any BLAST hits from the view with a score of less than 2000. Try moving the cutoff up and down to see what happens. Then leave it set at 2000

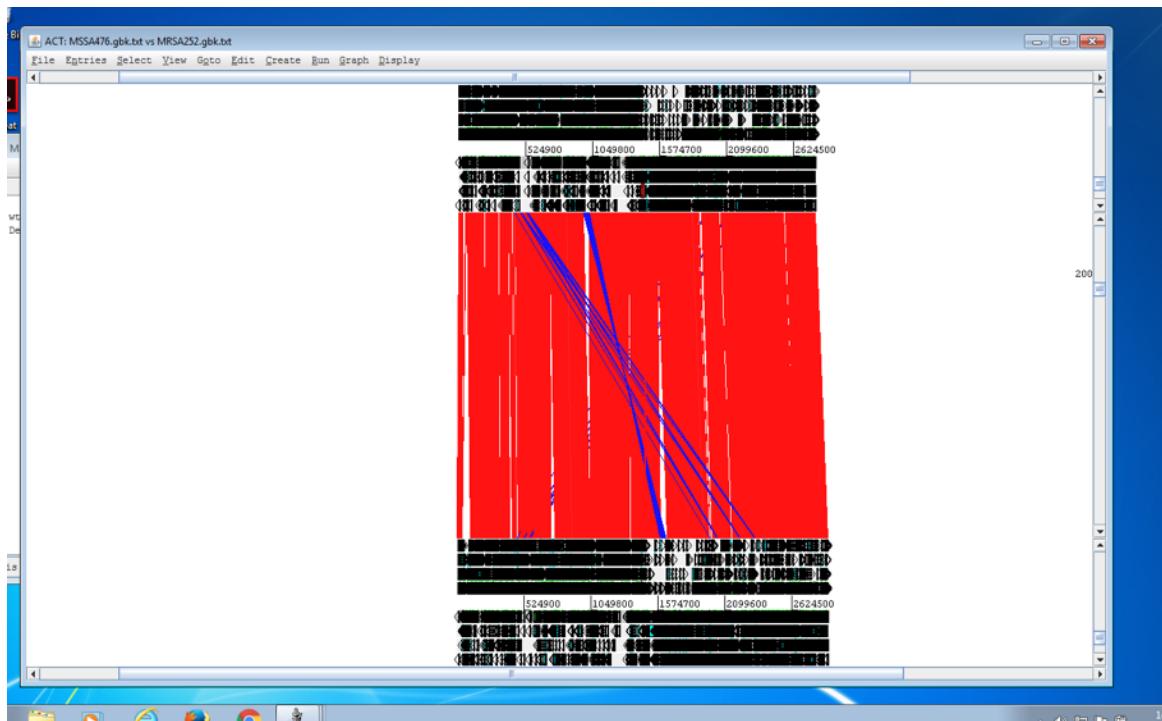
and click anywhere outside in the white to hide the menu or just move it out of the main window.



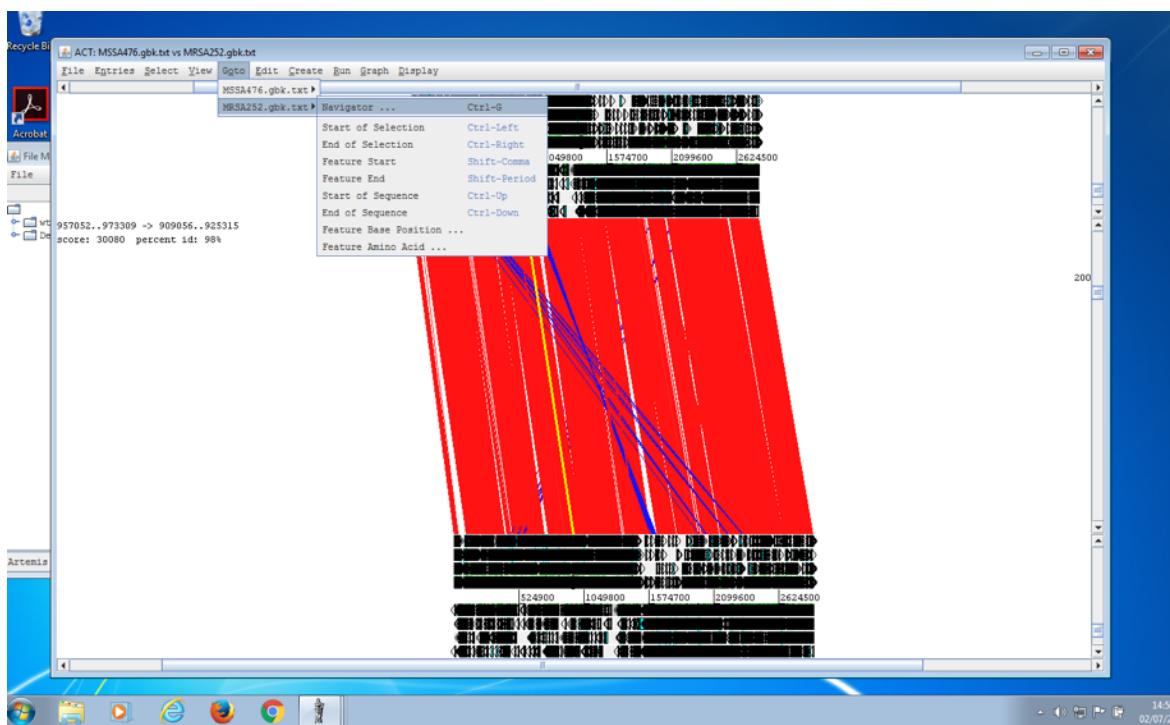
Step 14: Now slide the length of the sequence slider to 200. The window will now only show BLASTN hits longer than 200bp.



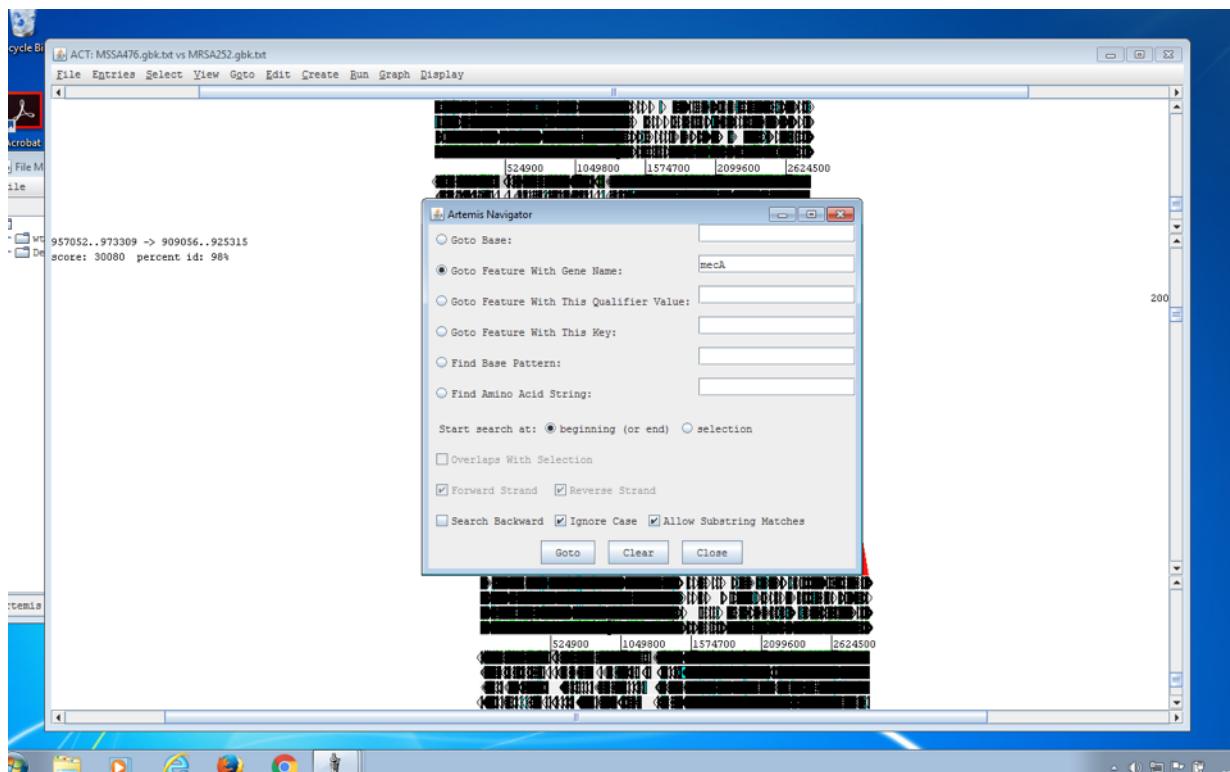
Step 15: Next, right click on the sequence viewer for the top sequence (MSSA476) and un-tick the ‘Stop codons’ box. Now do the same for the bottom sequence (MRSA252). We are now ready to explore the sequences.



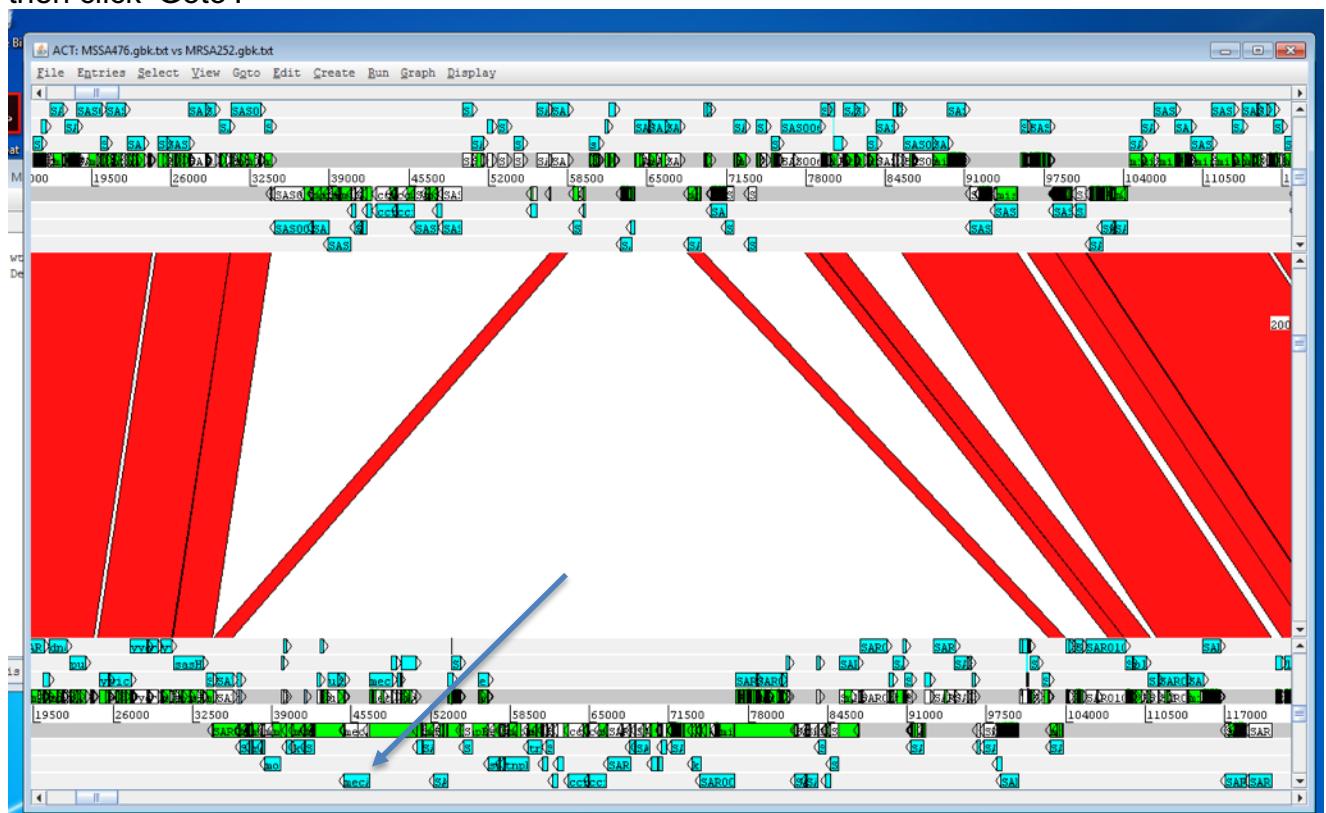
Step 16: We can see that the two genomes share a great deal of conservation and synteny (genes in the same order). You can also see a number of regions of difference between the two areas that are white in one genome but present in the other. If you use the slide bar we can zoom in to take a closer look at these differences. Try this and take a look around the genome at some of the differences.



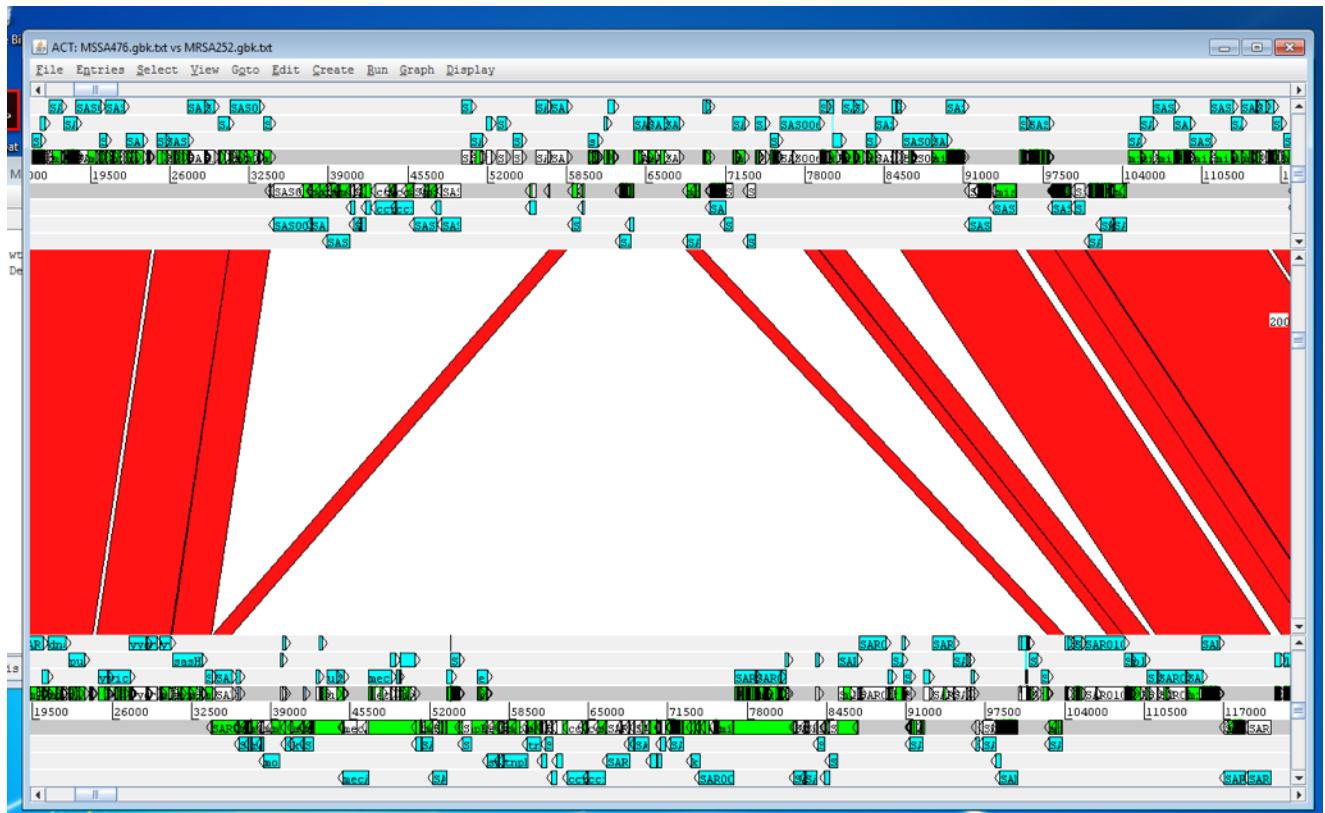
Step 17: We are now going to take a look at some of these differences that contain antibiotic resistance genes. Click on the 'Goto' menu at the top and select the lower sequence 'MRSA252.gbk.txt' and then click 'Navigator'.



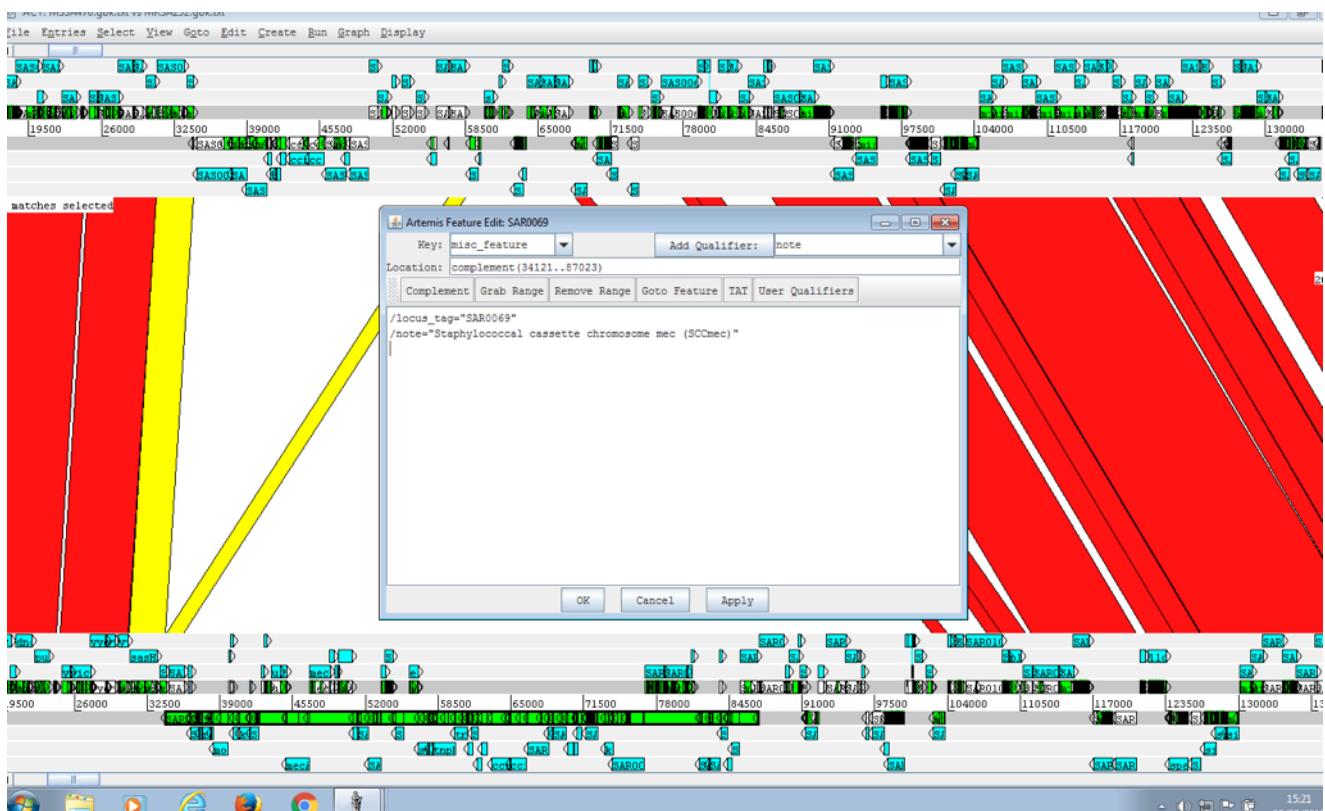
Step 18: Click the 'Goto Feature With Gene Name' and type 'mecA' in the box and then click 'Goto'.



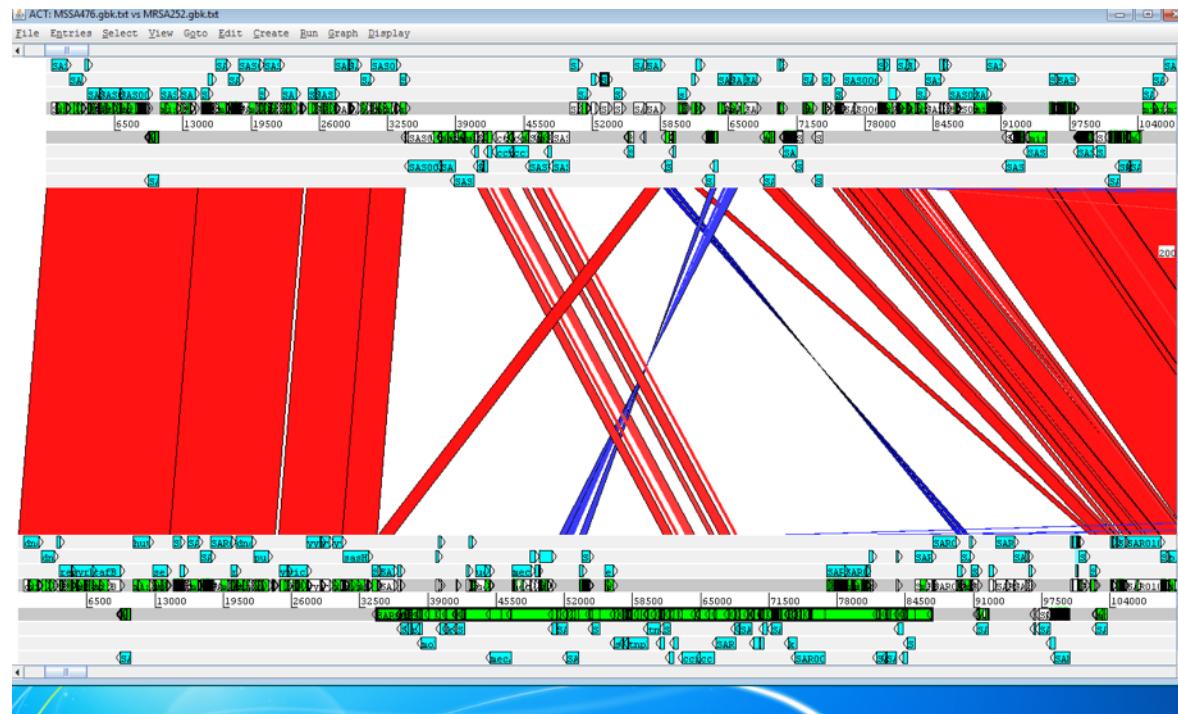
Step 19: Using the scrolling bars and zoom move the viewer until the window looks something like this. *mecA* should be highlighted (shown here with the arrows).



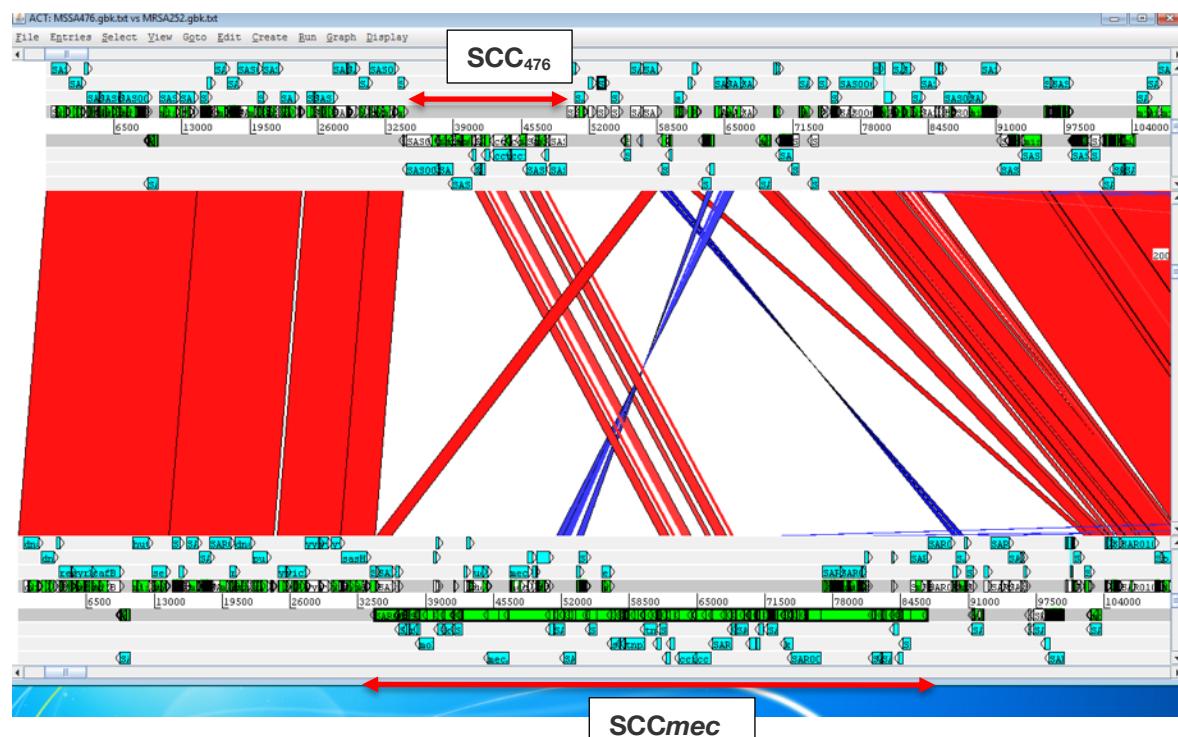
Step 20: If you click on the green region in the bottom genome, and then right click and select 'View' and 'Selected feature'.



Step 21: A menu will appear showing you that this region is a ‘Staphylococcal cassette chromosome *mec* also known as a SCC*mec*. This is a mobile element that brought the *mecA* gene into the strain.

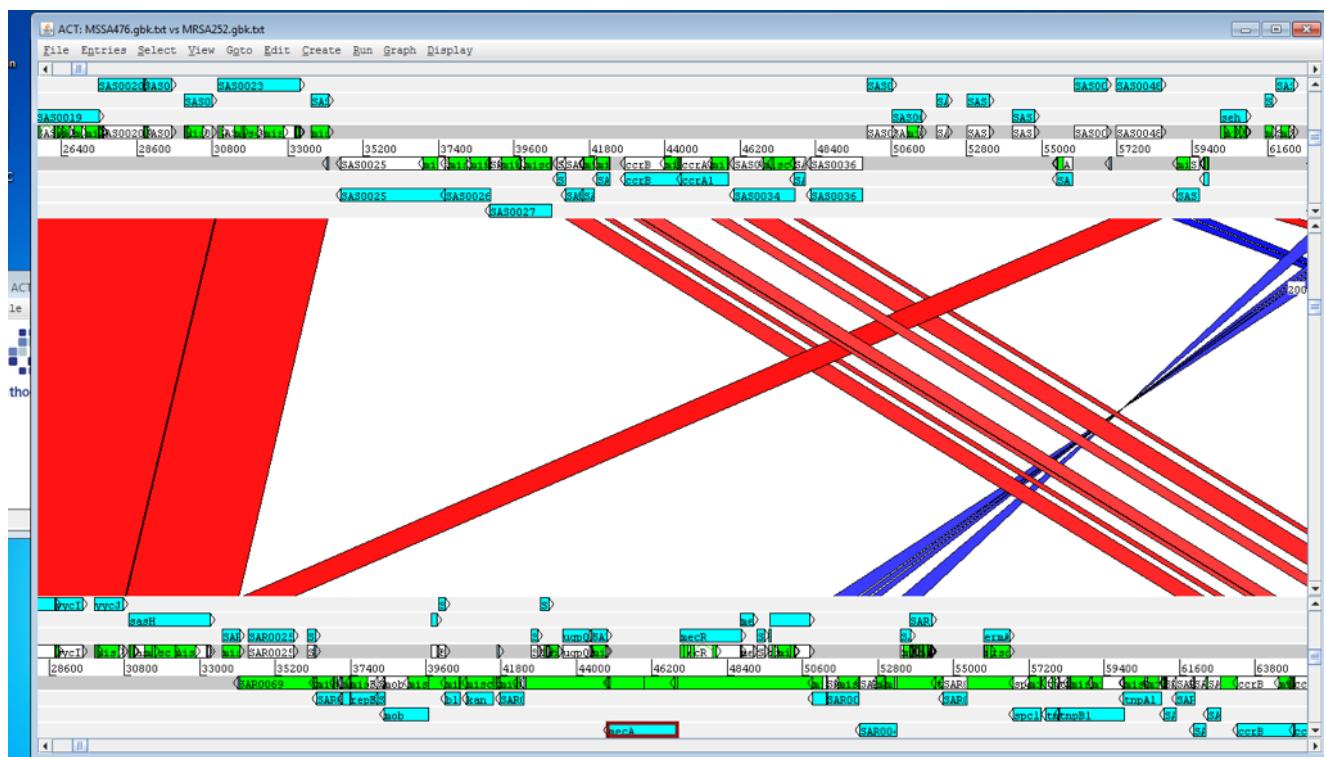


Step 22: If right click the mouse anywhere in the comparison area (the red / blue). Select the ‘Set score cutoff’. Move the slider down in the Score Cutoff of menu to back to 0. You should then see the above view. You can now see that in both genomes something has inserted in the chromosome and is flanked either side by conserved regions of the genome. This is why a comparative view is useful – it allows you to see the differences in context.

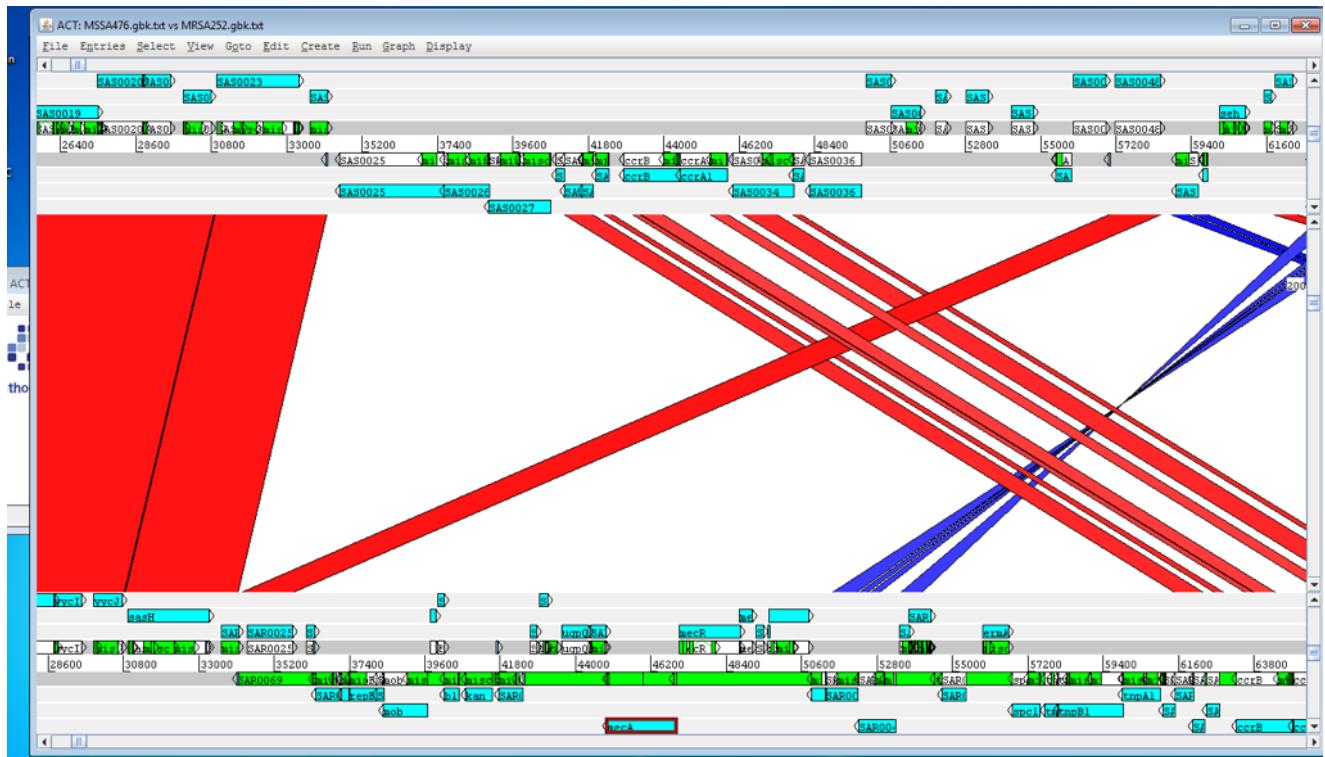


Step 23: You can see that there is some degree of similarity between the SCCmec element in MRSA252 (bottom) genome and a region in the MSSA476 (top) genome. This is because MSSA476 has another mobile element of the same family called SCC₄₇₆ (see highlighted region above). Click some of the genes regions of conservation and see what they are (use ctrl-v or right click – View – View selected features. This region is known as the ‘orfX region’ – as SCC elements use a sequence which is part of the *orfX* gene to insert into the genome and is regarded as a hotspot for the acquisition of horizontally transferred DNA.

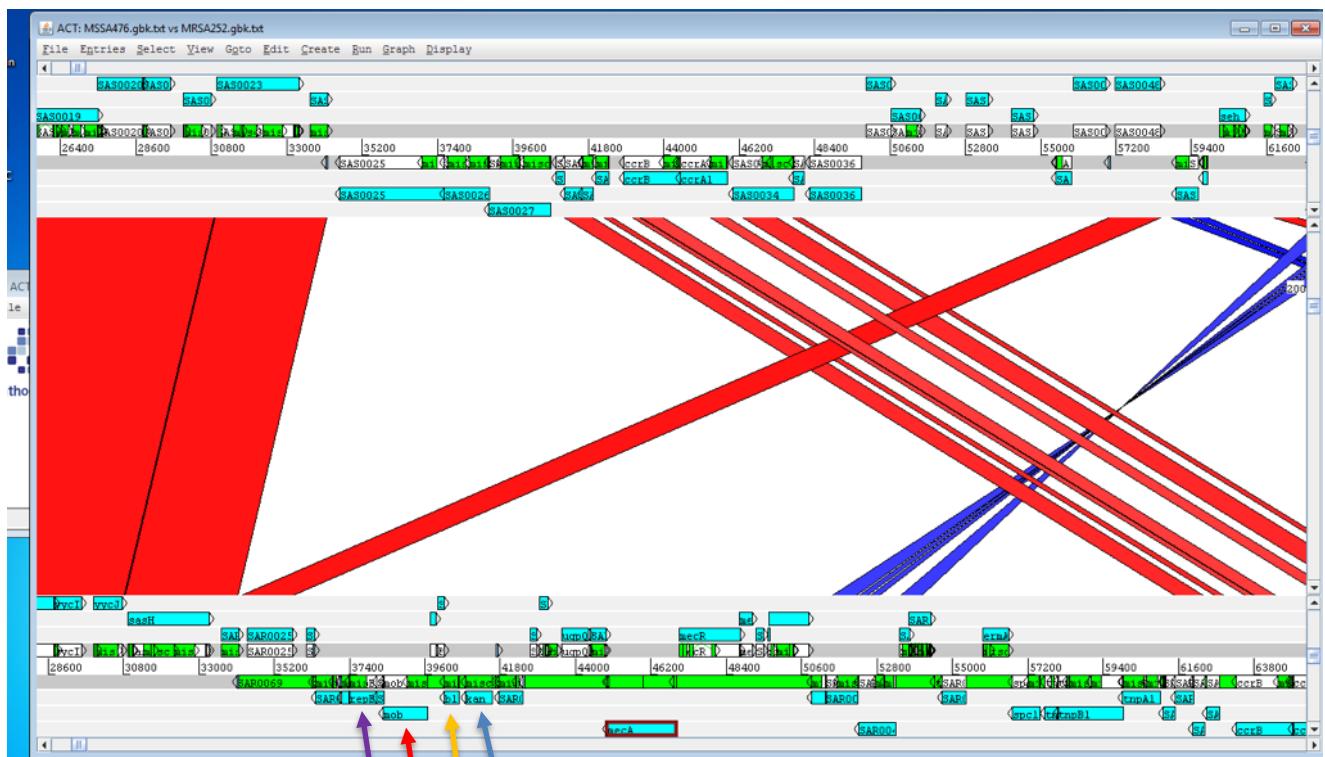
Step 24: You should have found that in both genomes – there are some genes annotated as *ccrA* and *ccrB*. These are the site-specific recombinase genes – that mediated the site-specific integration and excision from the genome, enabling horizontal transfer. Finding these genes alongside resistance genes is common and a useful indicator of how the genes got into the genome.



Step 25: Next zoom into the region around the *mecA* gene a bit closer. (If you need to find *mecA* again - Click on the ‘Goto’ menu at the top and select the lower sequence ‘MRSA252.gbk.txt’ and then click ‘Navigator’. Click the ‘Goto Feature With Gene Name’ and type ‘*mecA*’ in the box and then click ‘Goto’). You should then have a view like above (or something close it).



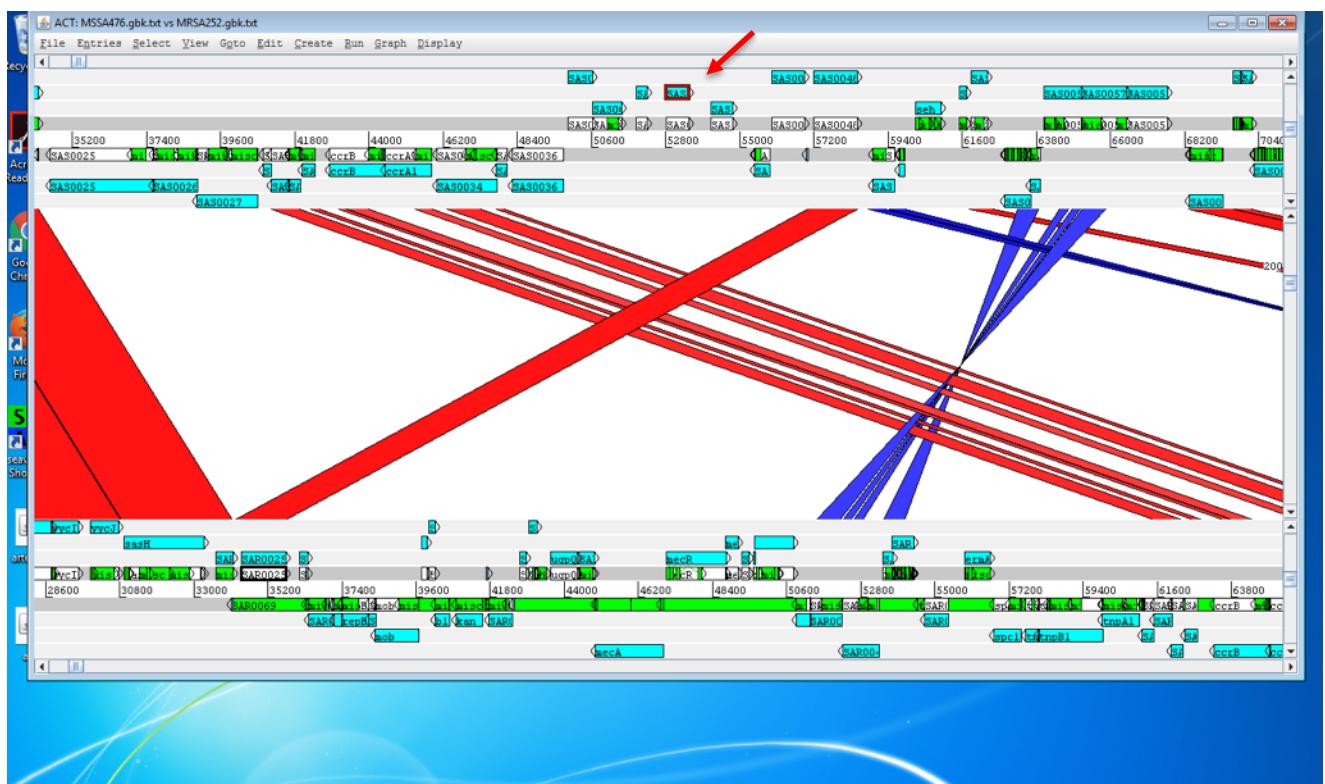
Step 26: Next zoom into the region around the *mecA* gene a bit closer. (If you need to find *mecA* again - Click on the 'Goto' menu at the top and select the lower sequence 'MRSA252.gbk.txt' and then click 'Navigator'. Click the 'Goto Feature With Gene Name' and type '*mecA*' in the box and then click 'Goto'). You should then have a view like above (or something close it).



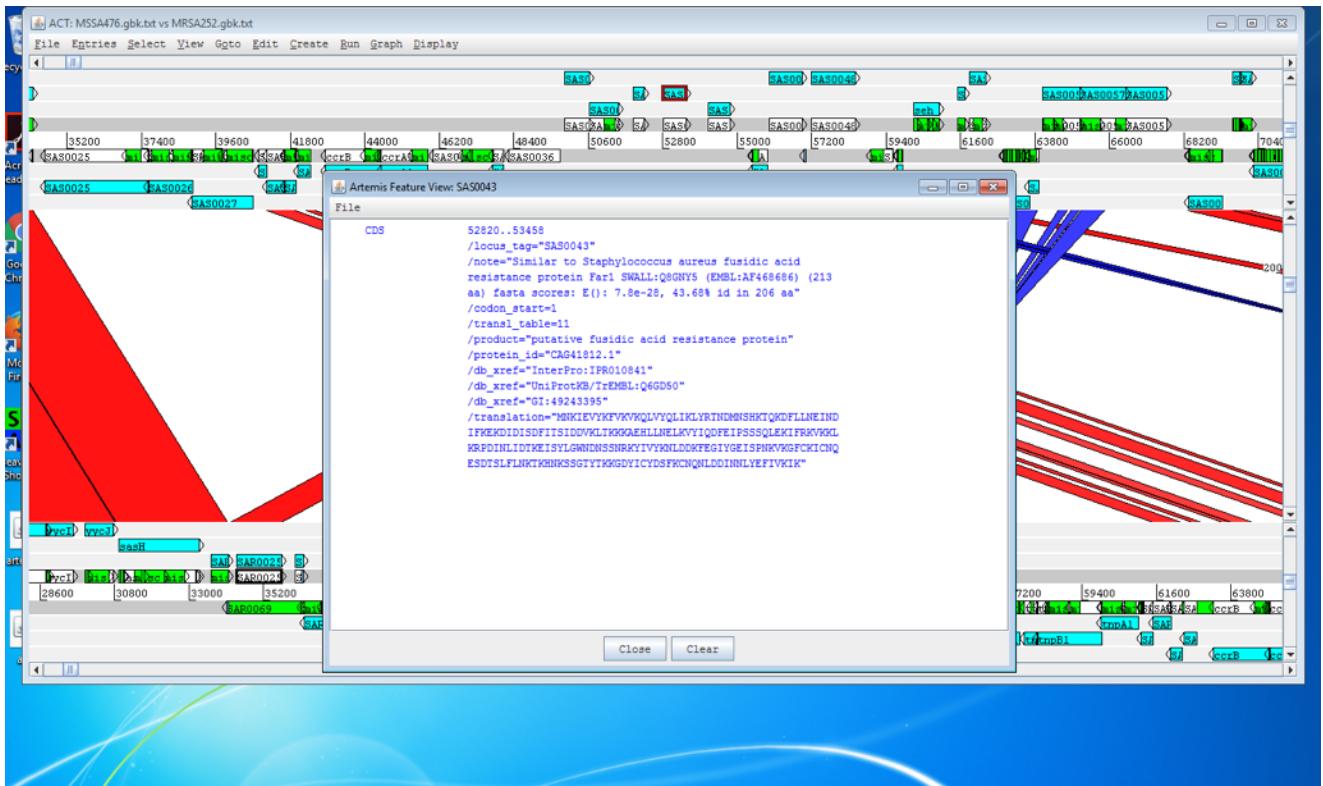
Step 27: Try clicking and viewing a few of the genes (use **ctrl-v** or right click – 'View' – 'View selected features') downstream (indicated by arrows – above) of the *mecA* gene. What are they annotated as?

Step 28: You should have found that the blue arrow was highlighting a gene called ‘kan’ – this is now known as *aadD* (this genome was annotated a long time ago – more about that in a bit). and mediates resistance to kanamycin. The gene indicated with an orange arrow is annotated as ‘ble’ – and mediates resistance to bleomycin. The other two genes indicated with red and purple arrows are ‘mob’ and ‘repB’ – these are both genes involved in plasmid mobilisation and replication – in fact what you are looking at here is an integrated copy of a plasmid called pUB110. This mosaic of mobile elements, making up a single mobile element is quite common and often includes other multiple resistance mechanisms to antibiotics, heavy metals and disinfectants.

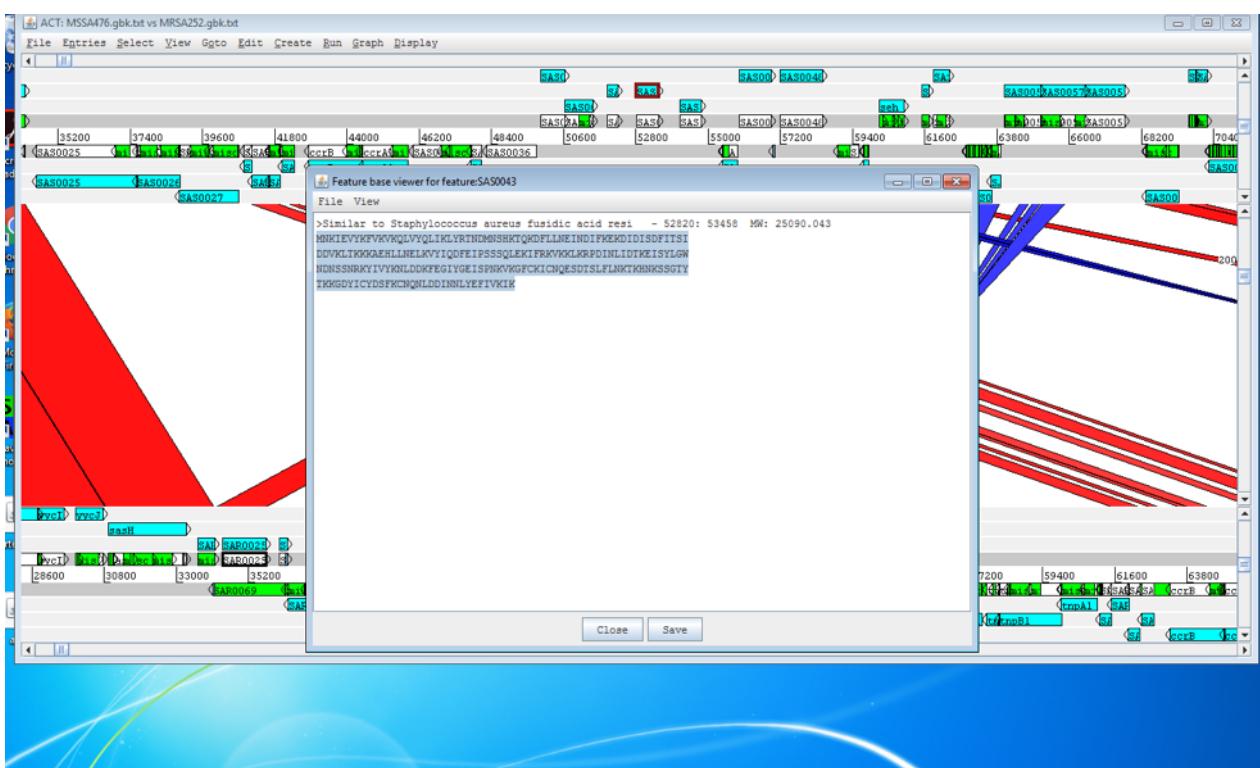
Step 29: The discrepancy between what a gene is annotated as and its actual function, is a problem that can occur for various reasons including genomes that were annotated a long time ago, or if a new sequence was annotated by an automated method. In this latter case, we would want to check if the automated annotation makes sense by checking our sequence against a reference database. One way (there are many different ways of doing this we don’t have time to go into) to do this is to use a BLAST search of a DNA or amino acid sequence. We are going to do this now to check the annotation of another gene.



Step 30: If you click on the top sequence (MSSA476) and then click on the ‘Goto’ menu at the top and select the upper sequence ‘MSSA476.gbk.txt’ and then click ‘Navigator’. Click the ‘Goto Feature With Gene Name’ and type ‘SAS0043’ in the box and then click the ‘Goto’ **TWICE**) – You should then have a view like above (or something close it). As you can see this gene is not present in the other genome (MRSA252). But as it is in the *orfX* region – it is likely that this gene was also horizontally transferred – but that the element that carried it in has degraded over time.

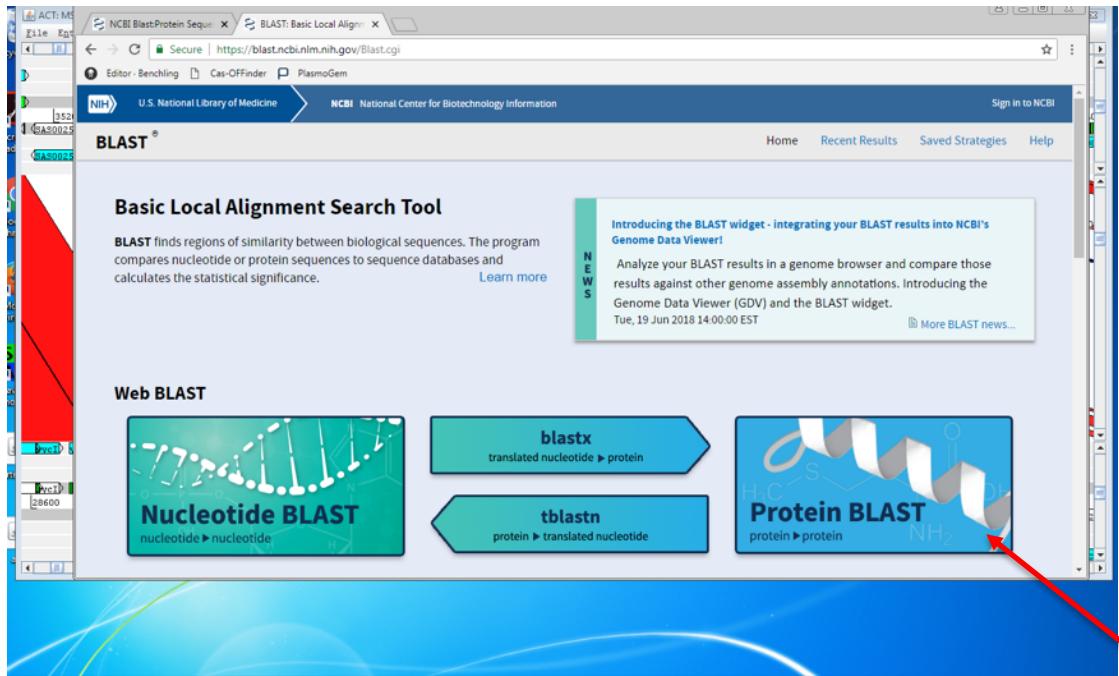


Step 31: Use ctrl-v or right click – ‘View’ – Select the upper sequence ‘MSSA476.gbk.txt’ then ‘View selected features’ – As you can see the gene is annotated as ‘Similar to *Staphylococcus aureus* fusidic acid resistance protein’. We will now check this annotation.

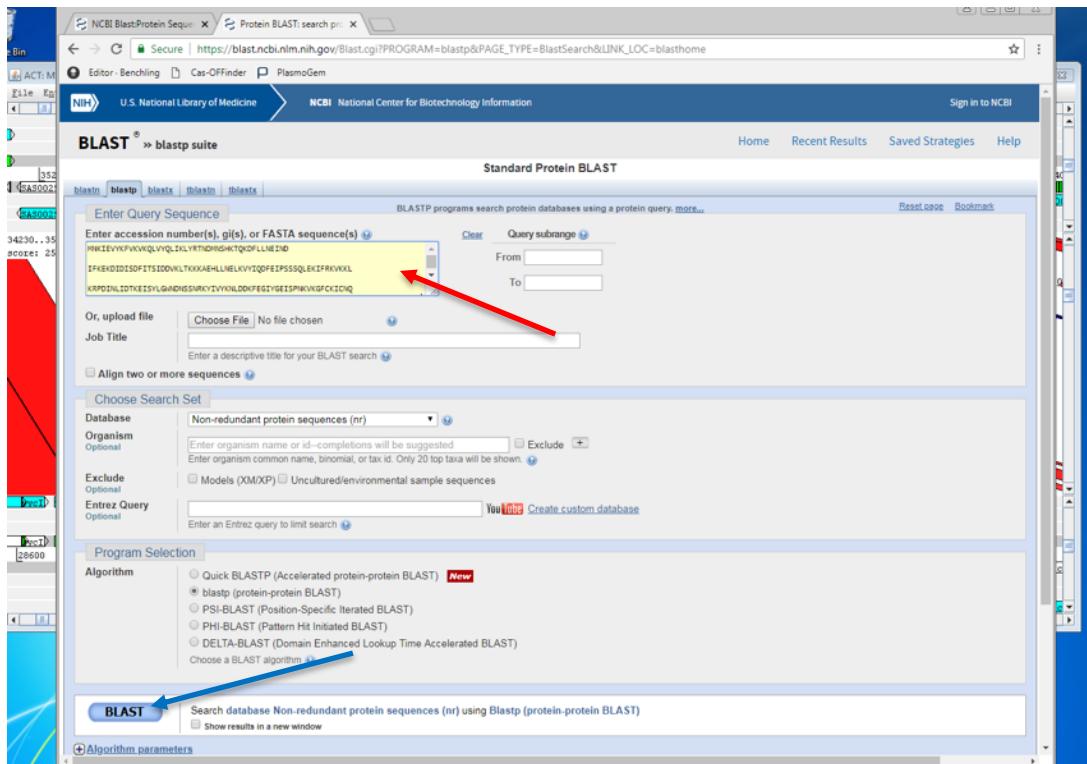


Step 32: Now Click on the ‘View’ window and select upper sequence ‘MSSA476.gbk.txt’ ‘Amino Acids’ and then select ‘Amino Acids of Selection As Fasta’

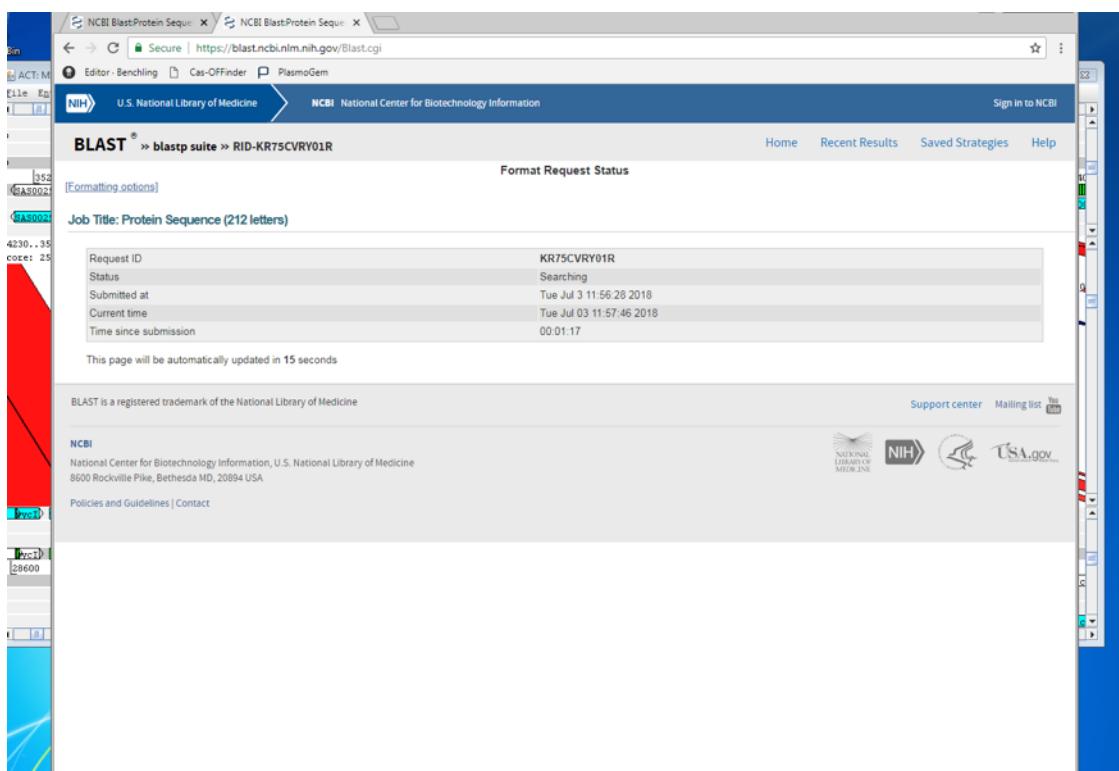
and then highlight and copy the sequence (like shown above) and press **ctrl + C** to copy the amino acid sequence.



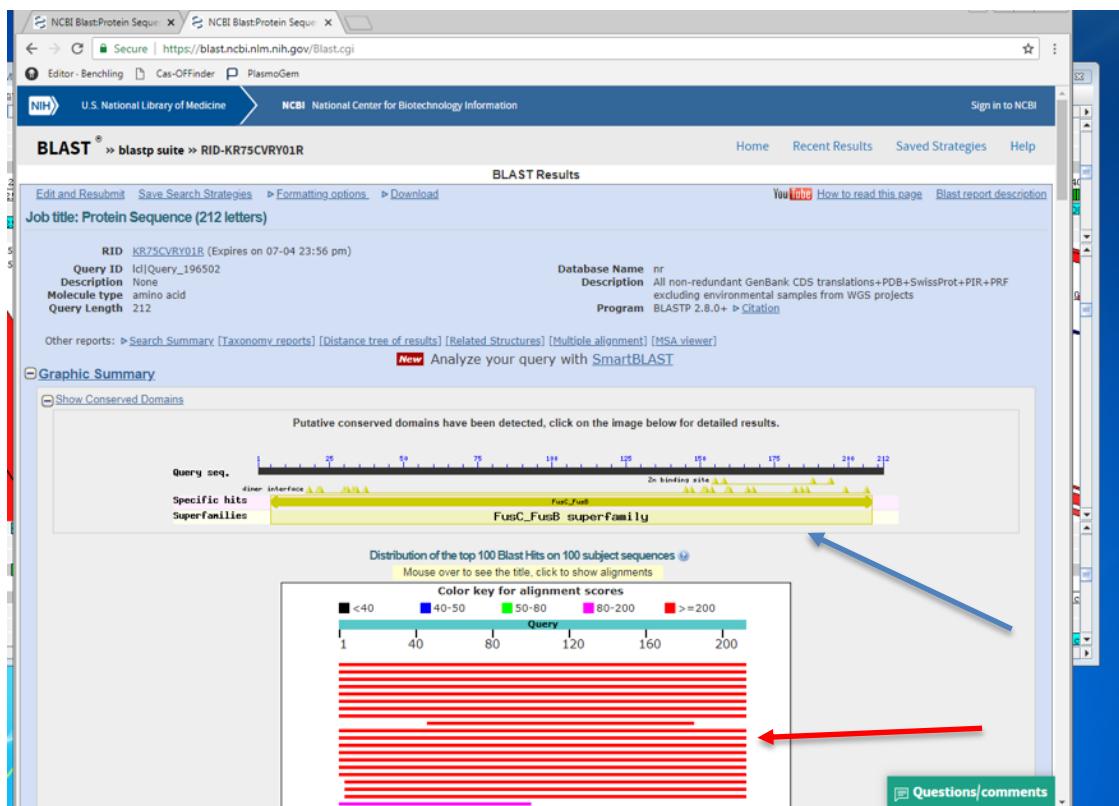
Step 33: Now minimise the ACT window and click on the ‘Google chrome’ web browser icon on the desktop. And then type: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> in the web address bar. Then click on ‘Protein blast’ (see red arrow).



Step 34: Now paste the sequence into the ‘Query sequence’ box (see red arrow) and click the BLAST button (blue arrow). Don’t worry about any of the other settings will just run things on default.



Step 35: You will then get a window like this while the BLAST runs.



Step 36: Once the BLAST is complete you will get a view like this. The Blue arrow is indicating the section showing that BLAST has found that your amino acid query sequence contains hits to conserved protein domain families. The red arrow is

showing you the summary of the alignments and the blast hit scores (the higher the number the better the score and therefore how closely related the BLAST hit score).

The screenshot shows a web browser window with the URL <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. The main content is a table titled 'Descriptions' showing 'Sequences producing significant alignments'. The table has columns for 'Description', 'Max score', 'Total score', 'Query cover', 'E value', 'Ident', and 'Accession'. The first row, highlighted with a green background, corresponds to the hit mentioned in the text. The 'Description' column for this row contains the text 'Chain A. Structure Of The Fusidic Acid Resistance Protein Fusc'.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Chain A. Structure Of The Fusidic Acid Resistance Protein Fusc	419	419	100%	4e-148	100%	ADC40001_1
MULTISPECIES: fusidic acid resistance EF-G-binding protein FusC [Staphylococcus]	419	419	100%	5e-148	100%	WP_001033157_1
FusC family: fusidic acid resistance EF-G-binding protein [Staphylococcus cohnii]	419	419	100%	5e-148	100%	ZYB5_A
FusC family: fusidic acid resistance EF-G-binding protein [Staphylococcus haemolyticus]	417	417	100%	3e-147	99%	WP_107386114_1
unnamed protein product [Staphylococcus aureus subsp. aureus]	416	416	100%	7e-147	99%	WP_059747714_1
FusC family: fusidic acid resistance EF-G-binding protein [Staphylococcus aureus]	412	412	100%	3e-145	98%	CCP89052_1
MULTISPECIES: FusC family fusidic acid resistance EF-G-binding protein [Macrococcus]	410	410	100%	8e-145	98%	WP_025176262_1
fusidic acid resistance protein [Staphylococcus haemolyticus]	404	404	100%	2e-142	97%	WP_096077718_1
hypothetical protein [Staphylococcus agnetis]	277	277	65%	6e-93	100%	AY30227_1
hypothetical protein [Staphylococcus agnetis]	265	265	100%	5e-87	60%	WP_095622151_1
elongation factor G-binding protein [Staphylococcus agnetis]	263	263	100%	2e-86	60%	WP_060552383_1
elongation factor G-binding protein [Staphylococcus agnetis]	263	263	100%	2e-86	60%	WP_037566393_1
hypothetical protein [Staphylococcus agnetis]	262	262	100%	4e-86	60%	WP_103346372_1
hypothetical protein [Staphylococcus agnetis]	259	259	100%	6e-85	59%	WP_107391064_1
hypothetical protein [Staphylococcus agnetis]	258	258	100%	2e-84	59%	WP_105994924_1
hypothetical protein [Leptospirillum halophilum]	250	250	100%	2e-81	55%	WP_092594841_1
hypothetical protein [Macrococcus sp. IME1552]	238	238	98%	2e-76	55%	WP_096076399_1
hypothetical protein [Macrococcus goetzii]	225	225	98%	2e-71	57%	WP_099578357_1
hypothetical protein [Macrococcus caseolyticus]	220	220	98%	2e-69	56%	WP_101035271_1
fibronectin-binding domain protein [Staphylococcus aureus subsp. aureus 21304]	194	194	47%	1e-60	100%	EZ190619_1
hypothetical protein [Streptococcus carabensis]	189	189	91%	2e-57	51%	WP_103346991_1
Fibronectin-binding protein (FBP) [Streptococcus carabensis]	189	189	91%	2e-57	51%	PNY19232_1
elongation factor G-binding protein [Enterococcus faecium]	183	183	98%	7e-55	46%	WP_002343616_1
elongation factor G-binding protein [Enterococcus faecium]	183	183	98%	9e-55	46%	WP_002318946_1
elongation factor G-binding protein [Enterococcus faecium]	183	183	98%	9e-55	46%	WP_053542686_1
elongation factor G-binding protein [Enterococcus faecium]	182	182	98%	9e-55		

Step 37: If you scroll down you will get a view like the one above, which lists all the hits. You can see that most of the top hits label the protein to be involved in Fusidic acid resistance and – and one is labelled ‘Chain A structure of the **Fusidic acid Resistance Protein Fusc**’ (See arrow).

Step 38: This shows you that the sequence that we BLASTed is 100% identical to this entry in the database. If you click the link indicated by the arrow it will take you to the protein database entry for this hit.

Chain A, Structure Of The Fusidic Acid Resistance Protein Fusc

PDB: 2YB5_A
Identical Proteins FASTA Graphics

Go to: ▾

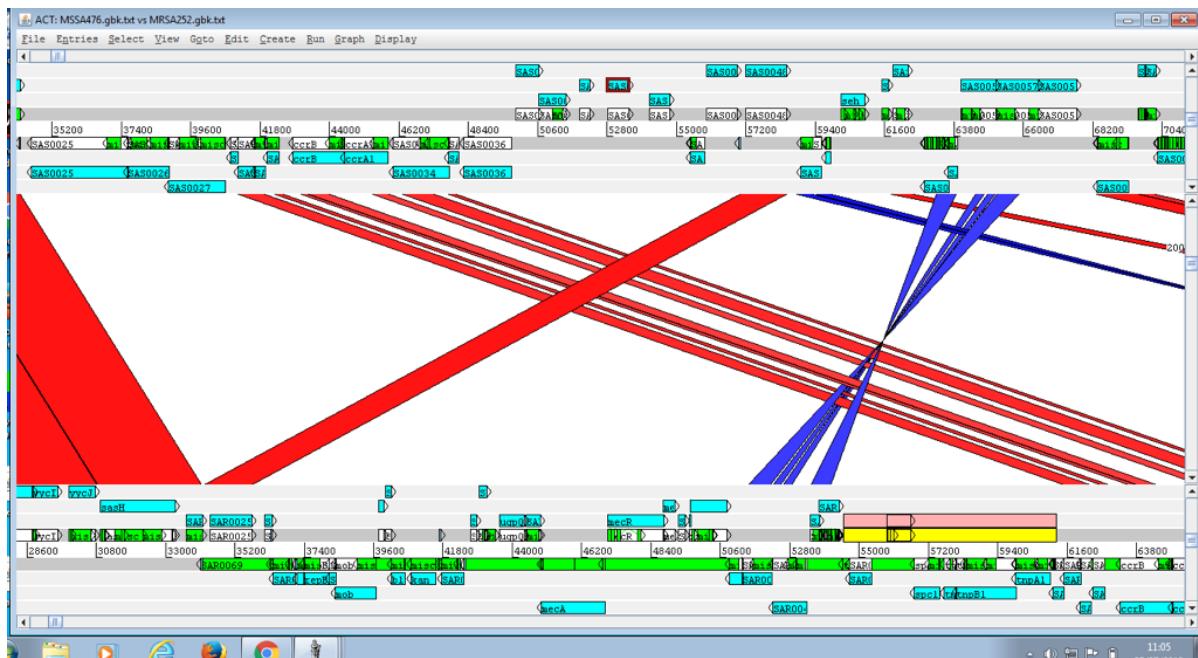
Locus: 2YB5_A 215 aa linear BCT 23-SEP-2015
Definition: Chain A, Putative Fusidic Acid Resistance Protein.
Accession: 2YB5_A
Version: 2YB5_A
DBSOURCE: pdb: molecule 2YB5, chain 65, release Sep 23, 2015; deposition: Mar 1, 2011; class: Translation; source: Mmdb_id: 96621, Pdb_id 1: 2YB5; Exp. method: X-Ray Diffraction.
Keywords: Staphylococcus aureus
Source: Staphylococcus aureus
Organism: Bacteria; Firmicutes; Bacilli; Bacillales; Staphylococcaceae; Staphylococcus
Reference: 1 (residues 1 to 215)
Authors: Cox,G., Thompson,G.S., Jenkins,H.T., Peske,F., Savelbergh,A., Rodnina,M.V., Wintermeyer,N., Homans,S.W., Edwards,T.A. and O'Neill,A.J.
Title: Ribosome clearance by FusB-type proteins mediates resistance to the antibiotic fusidic acid
Journal: Proc. Natl. Acad. Sci. U.S.A. 109 (6), 2102-2107 (2012)
PubMed: 22308410

Analyze this sequence
Run BLAST
Identify Conserved Domains
Highlight Sequence Features
Find in this Sequence

Protein 3D Structure
Structure Of The Fusidic Acid Resistance Protein Fusc
PDB: 2YB5
Source: Staphylococcus aureus
Method: X-Ray Diffraction
Resolution: 2.1 Å

Related information
Similar protein sequences using SmartBlast

Step 39: This entry shows that there is experimental evidence that this protein is involved in resistance to Fusidic acid – so we can be quite confident that this protein is involved in resistance to Fusidic acid. If you now go back and have a look at some of the other hits, you will see that the information available – can be considerable vaguer – with proteins annotated as ‘hypothetical protein’ or ‘unnamed protein product’.



Step 40: Now return to the ACT view. The SCCmec in MRSA252 contains some more resistance genes – that are highlighted above – using what you have just learned look up the annotation of these genes and check the annotations using BLAST. Which resistance genes are present in this region? What other genes are present? Do you have any idea of how these genes might have got in to the SCCmec element?

14.2 Part II: Investigating multidrug resistance plasmids

In the previous two parts of this practical you have been looking at completed genomes – that is genomes that have been completely sequenced, aligned into a single chromosome and manually annotated (that is each gene annotation checked manually). But if you generate your own data, you will probably end up with a less ‘polished’ genome to work with.

In this practical you are going to look at two of the plasmids from two of the outbreak (these are in the folder Practical 3):

1. CP004 – *Klebsiella pneumoniae*
2. CP0069 – *E. coli*

Both the plasmids from these strains have the same *blaNDM-1* gene, we are going to use ACT to compare the plasmid sequences to see how much homology the two plasmids share and to determine any structural features that might be of importance.

Step 1: Firstly, we need to generate the BLAST comparison file that ACT takes to show how related the two sequences you are comparing. Open: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> in a web browser (Firefox).

You should get a window like this:

The screenshot shows the BLAST homepage. At the top, there's a blue header bar with the NIH logo, the text "National Library of Medicine", and "National Center for Biotechnology Information". On the right side of the header is a "Log in" button. Below the header, the word "BLAST®" is prominently displayed. To the right of "BLAST®" are links for "Home", "Recent Results", "Saved Strategies", and "Help". A green "NEWS" button is located on the left side of the main content area. The main content area has a light gray background. It features a section titled "Basic Local Alignment Search Tool" with a brief description of what BLAST does. Below this, there's a "Web BLAST" section with three buttons: "Nucleotide BLAST" (nucleotide → nucleotide), "blastx" (translated nucleotide → protein), and "tblastn" (protein → translated nucleotide). To the right of these buttons is a "Protein BLAST" button (protein → protein). At the bottom of the page, there's a "BLAST Genomes" section with a search bar and dropdown menus for "Human", "Mouse", "Rat", and "Microbes".

Step 2: Now click on 'Nucleotide BLAST'

You should get a window like this

BLAST® » blastn suite

Align Sequences Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide subjects using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [?](#)

From _____ To _____

Or, upload file Choose File [no file selected](#) [?](#)

Job Title _____

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Subject subrange [?](#)

From _____ To _____

Or, upload file Choose File [no file selected](#) [?](#)

Program Selection

Optimize for Highly similar sequences (megablast) More dissimilar sequences (discontiguous megablast) Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

Step 3: Click the tick box where it says 'Align two or more sequences' (see red arrow above). This will allow a comparison of two sequences.

BLASTN programs search nucleotide sub...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [?](#)

From _____ To _____

Or, upload file Choose File [cpe001_plasmid_HI1B.fasta](#) [?](#)

Job Title _____

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Subject subrange [?](#)

From _____ To _____

Or, upload file Choose File [cpe069_plasmid_IncAC.fasta](#) [?](#)

Program Selection

Optimize for Highly similar sequences (megablast) More dissimilar sequences (discontiguous megablast) Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences) Show results in a new window

Step 4: Now click the 'choose file' on the upper section (green arrow) and select the file – 'cpe004_plasmid_IncFIB-HI1B.fasta' (you will need to navigate to the file location). Do the same for the lower sequence (blue arrow) and select – 'cpe069_plasmid_IncAC.fasta'. Then press BLAST (blue button)

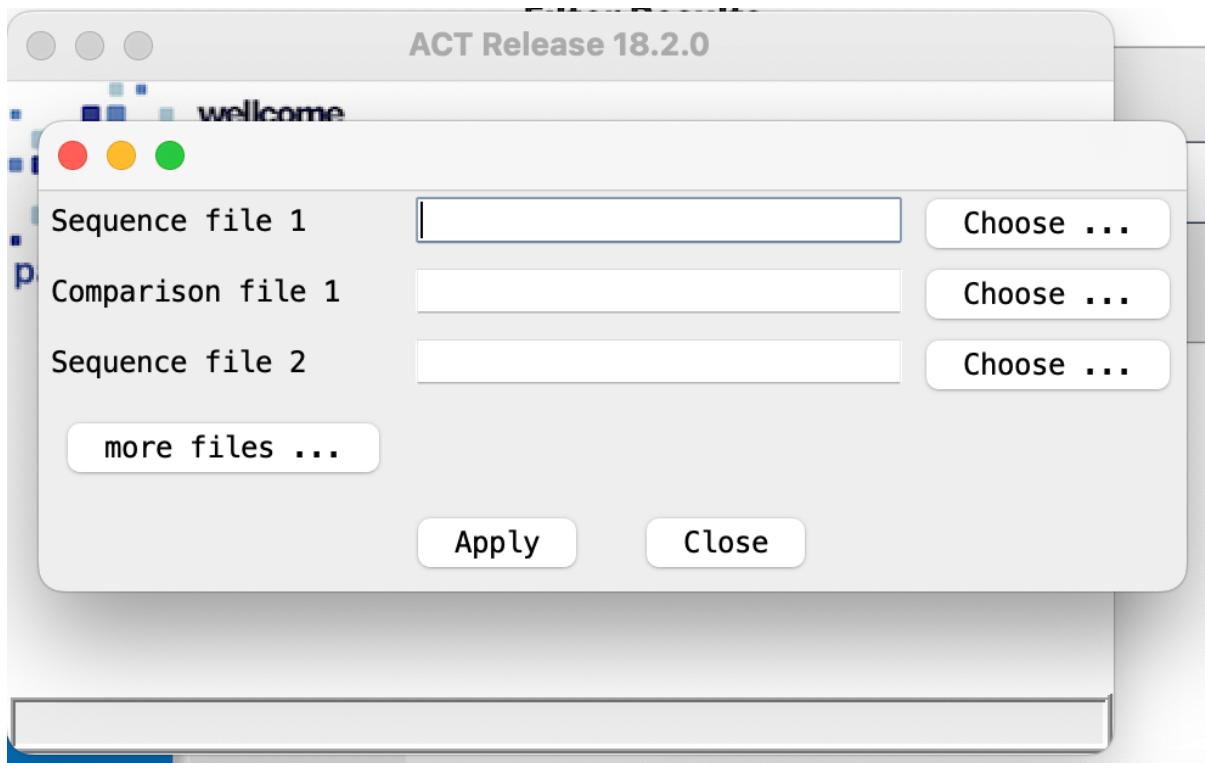
Step 5: You should now get a view like this. You need to click on the alignment tab (see the red arrow below)

The screenshot shows the National Library of Medicine BLAST search results for RID-U65S56RT114. The 'Alignments' tab is highlighted with a red arrow pointing to it. The results table displays one significant alignment entry:

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
cpe069_hybrid_plasmid-IncAC Escherichia coli strain strain		11705	1.093e+05	17%	0.0	99.97%	145843	Query_2194687

Step 6: You should now see a view like this – select ‘Download’ from the righthand side (red arrow) and select ‘Hit Table (text)’. This will generate and download a file that you can read into ACT. Note – where this file has downloaded as you are going to need it. It will be called ‘xxxxxxxxxx--Alignment.txt’. (if you have problems there is a folder called (backup – which contains one we have pre-generated).

The screenshot shows the same BLAST search results page as before, but the 'Download' dropdown menu is open, revealing options including 'Hit Table (text)'. The 'Alignments' tab is still selected.

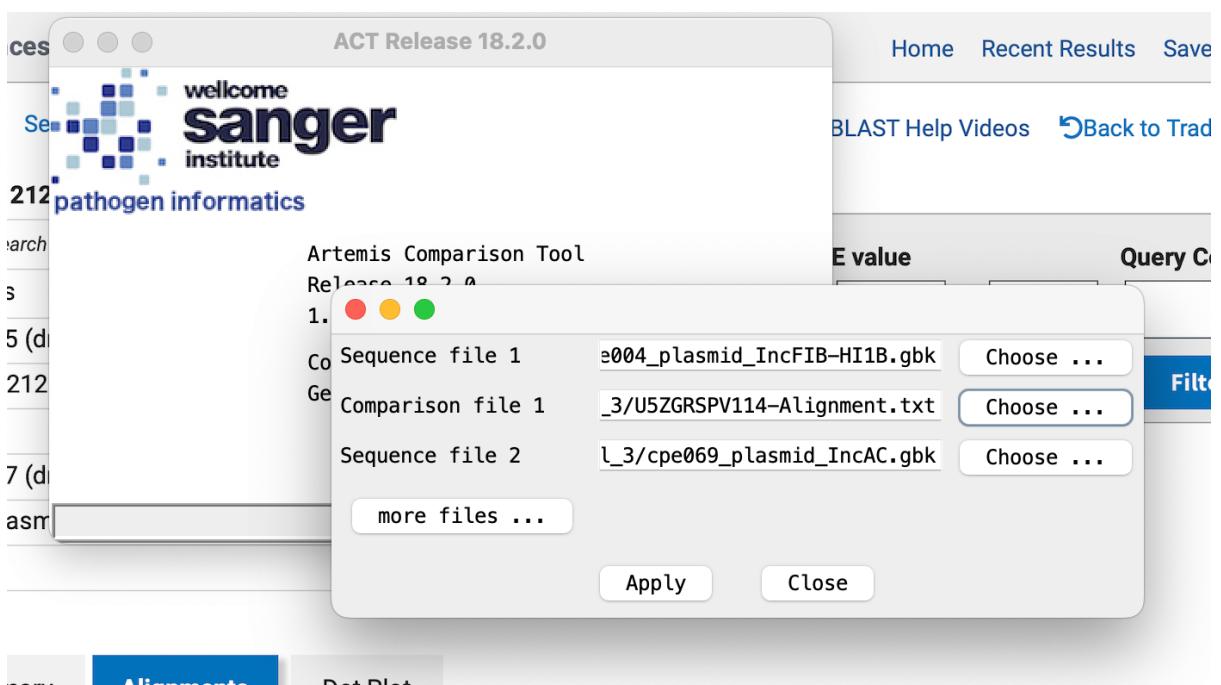


Step 7: Open ACT and load the following into:

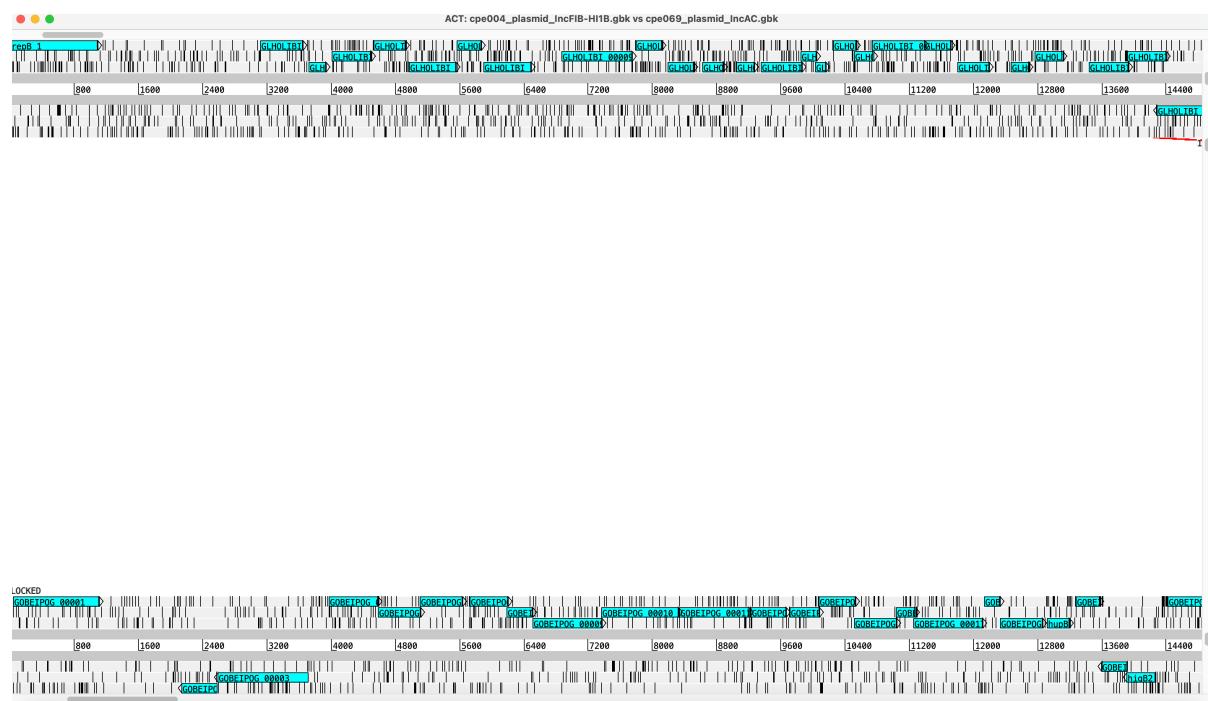
- (a) Sequence file 1: 'cpe004_plasmid_IncFIB-HI1B.gbk'
- (b) Comparison file 1: 'xxxxxxxx--Alignment.txt' (this is file you just downloaded).
- (c) Sequence file 2: 'cpe069_plasmid_IncAC.gbk'

Note – we are now using the .gbk files as they contain the annotation information.

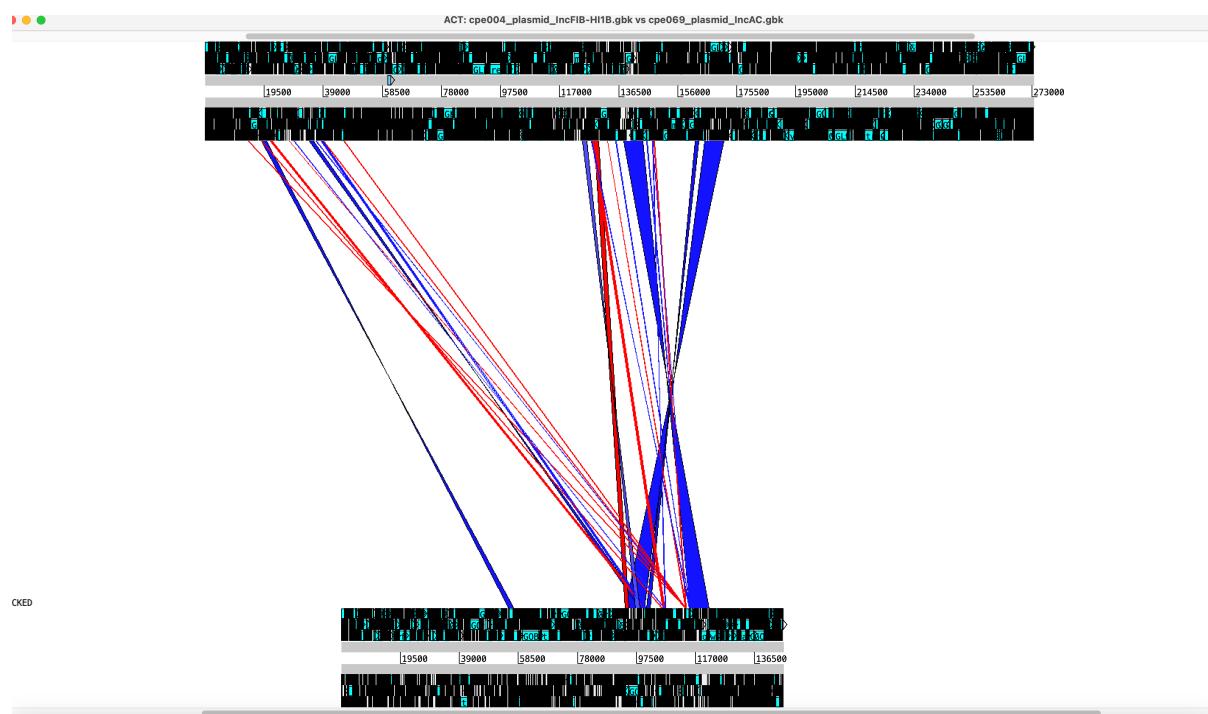
The image below shows what it should look like.



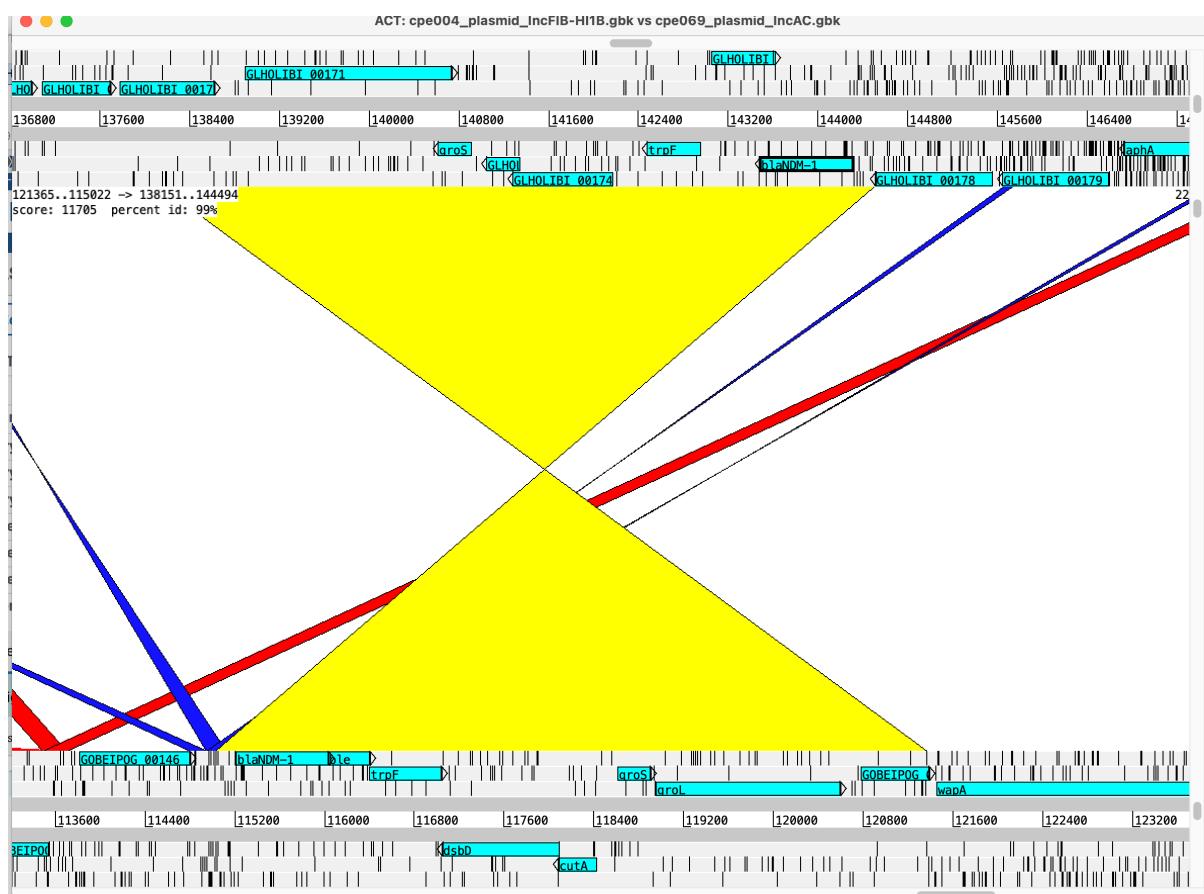
Step 8: You should now have a view like this. Remember to unlock the two sequences and then zoom out and take a look at the two sequence.

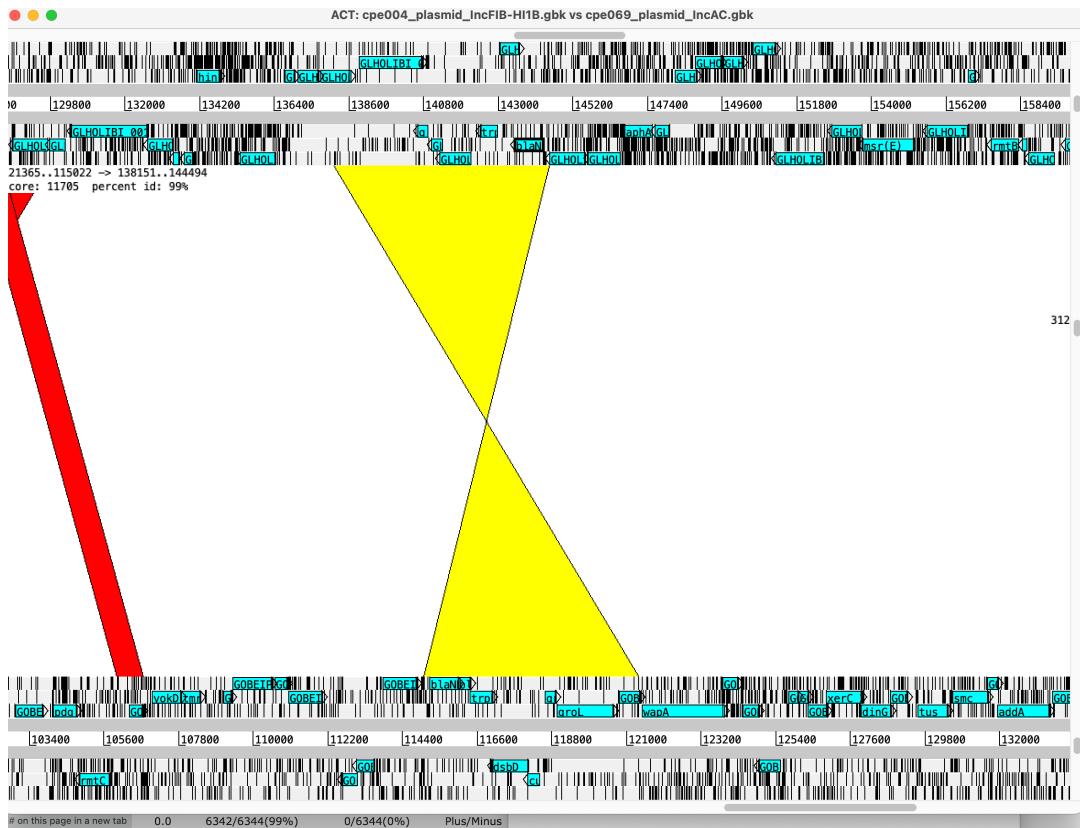


Step 8: You should have a view like this. You can clearly see although the two plasmids are generally distinct, they do share some blocks of homologous sequences.

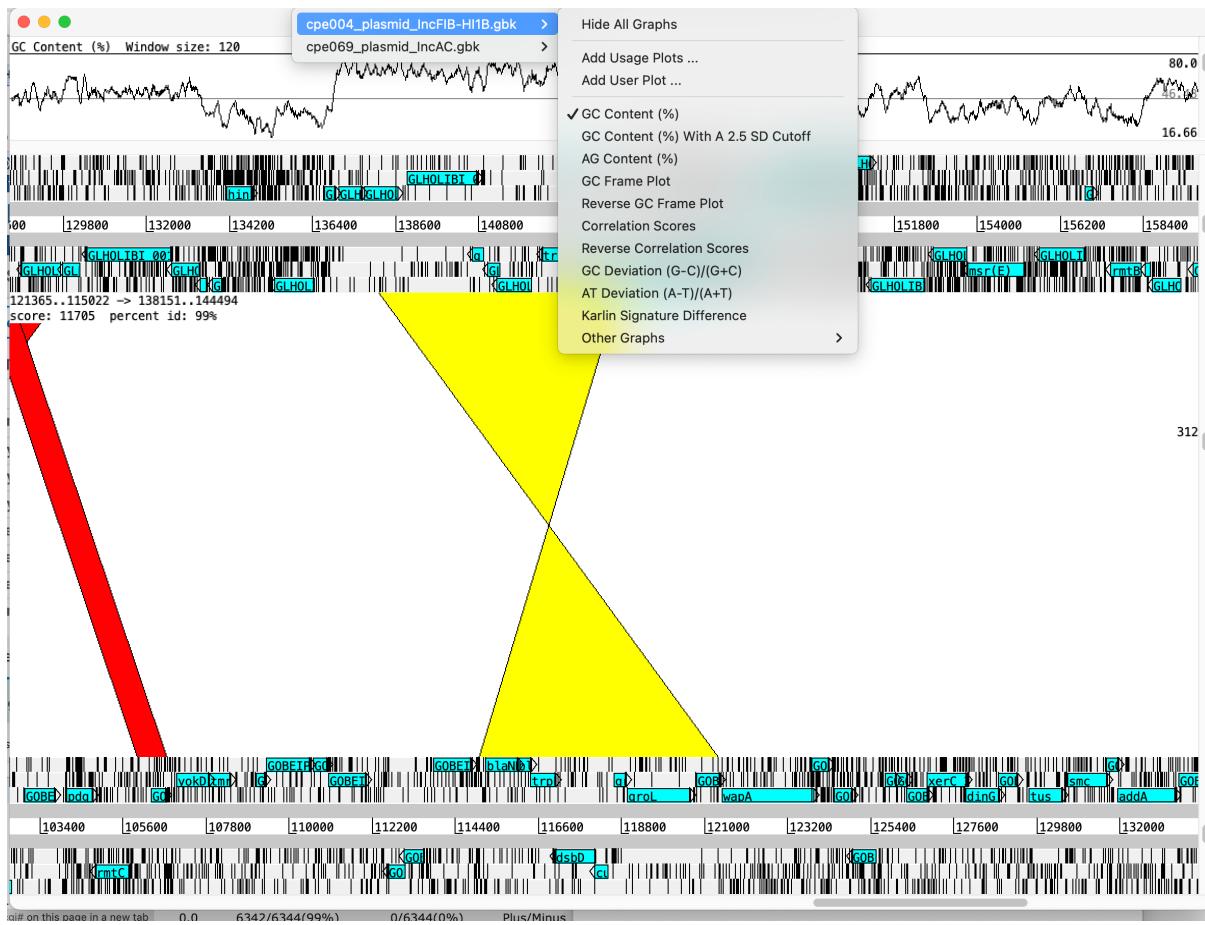


Step 9: If you use the navigator (control/command – G) or on the menu ‘Goto’ and select ‘cpe004_plasmid_IncFIB-HI1B.gbk’ – Navigator to use the ‘Goto Feature with Gene Name’ and type: ‘ndm’ (to take you to the *bla*_{NDM-1} gene. Then zoom in to this area (like shown below). You can clearly see the two plasmids both share the same block.

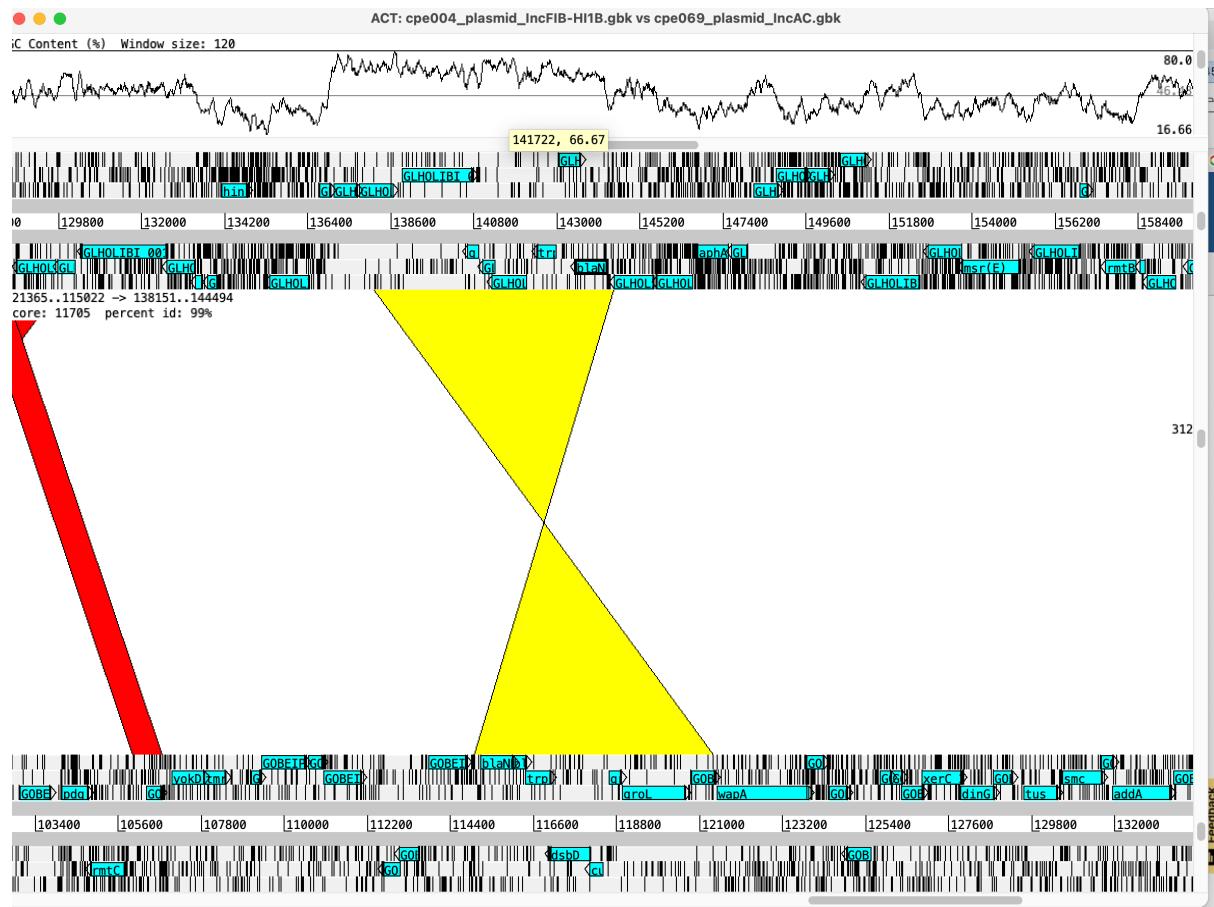




Step 10: What genes are upstream of the *bla*_{NDM-1} in both plasmids? How might they have contributed to the presence of *bla*_{NDM-1}?



Step 11: Go to the ‘Graph’ menu at the top for the cpe004_plasmid_IncFIB-HI1B.gbk’ and select ‘GC content (%)’. You should then see a graph appear at the top.



Step 12: Looking at the GC plot what is different for the region that has the *bla*_{NDM-1} gene?

Step 13: Add the G/C graph for the other sequence – is it the same?

Step 14: You can now explore the plasmids what are the other blocks of homology?

Step 15: Can you find other resistance genes and other genes or features of interest?