

# **Computational Practical 13**

## **Genomics Surveillance of AMR Analysis of resistance in genomes**

### **13.1 Using genomics to investigate antibiotics resistance**

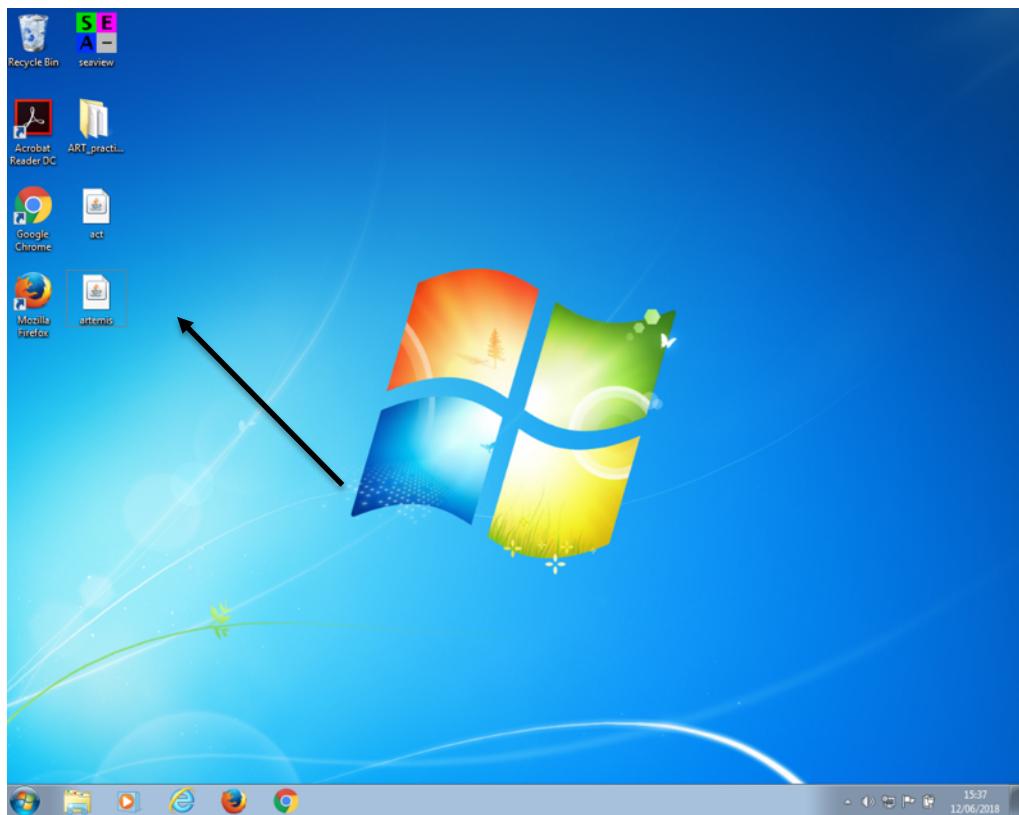
### **13.2 Overview**

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.

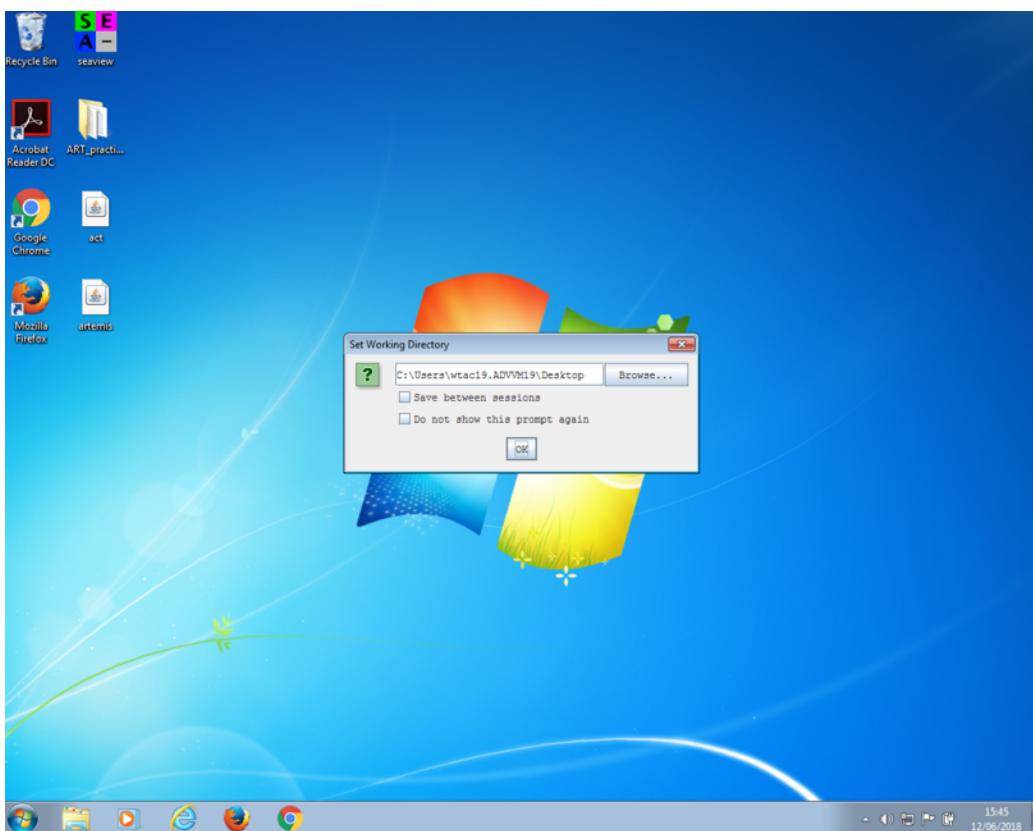
In the previous practical you investigate resistance using online tools that identify the presence of genes and resistance conferring mutations, we are now going to zoom down to individual bacterial genome which can help us further understand antibiotic resistance and extract data for further analysis.

### **13.3 Part I: Getting going with Artemis**

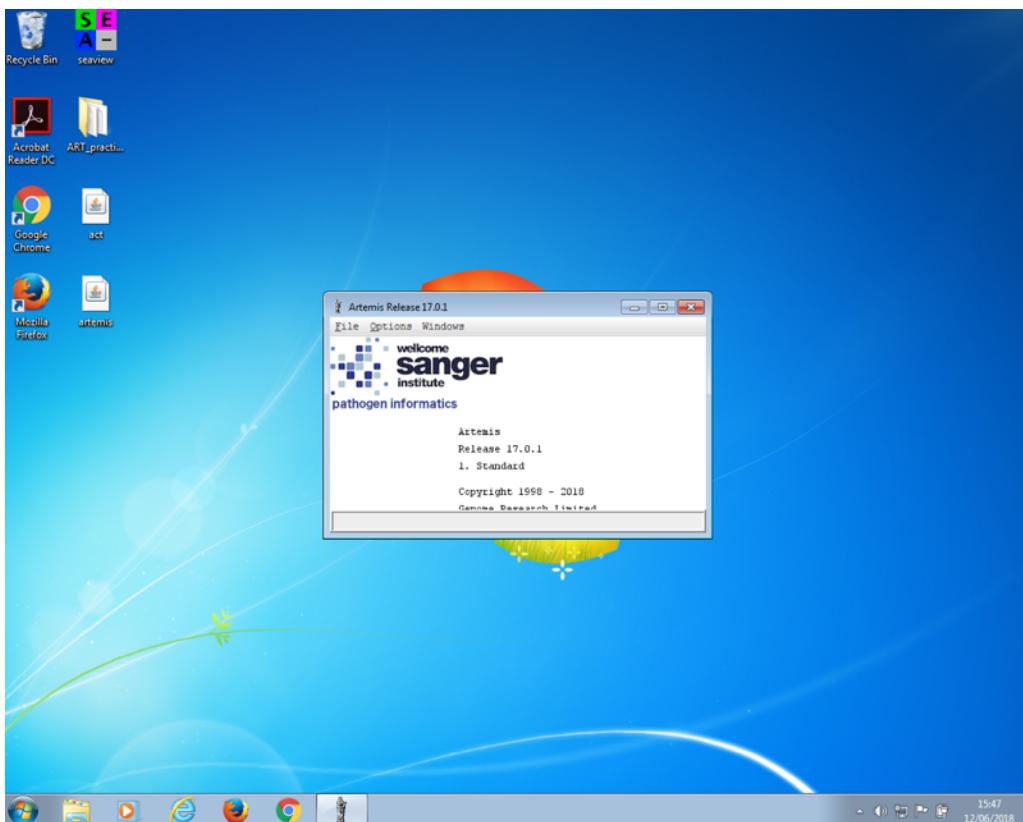
- Artemis is a genome browser and annotation software developed at the Wellcome Sanger Institute (<http://www.sanger.ac.uk/science/tools/artemis>). It is freely available to download for PCs and Mac.
- The practical is designed to give you a basic understanding of Artemis software and to get a better 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!



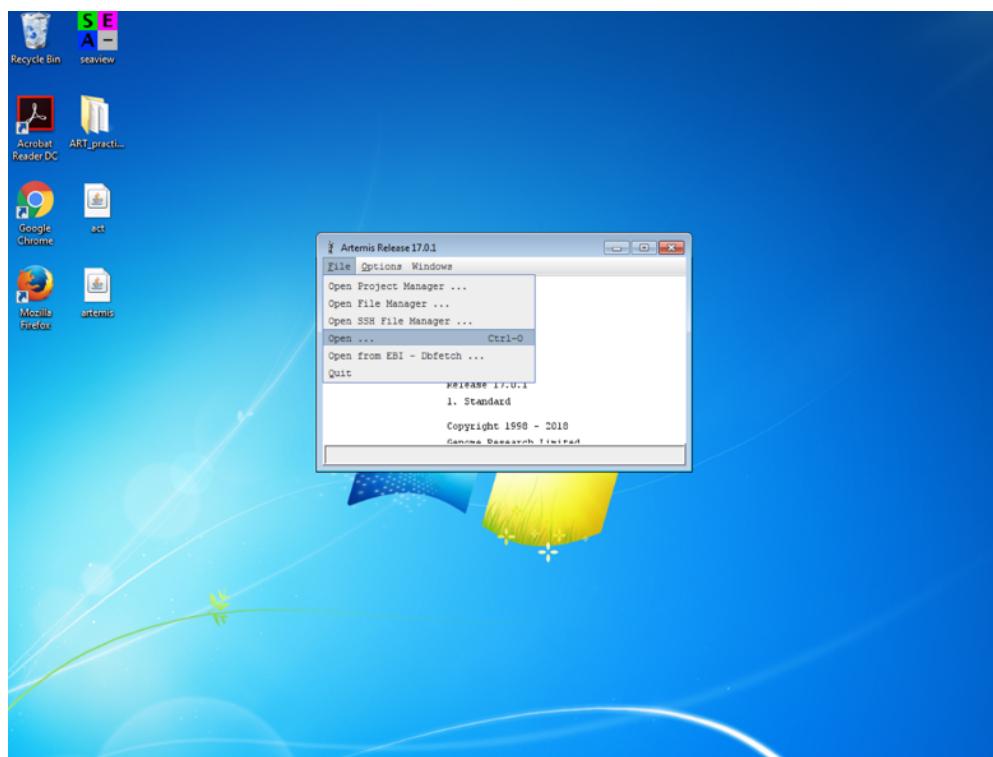
**Step 1:** Your screen should look like this. Now click on the 'artemis' icon as indicated. **NOTE: if you cannot locate the 'artemis' icon on your desktop, open a new terminal window and type in 'art'. This will also open Artemis tool.**



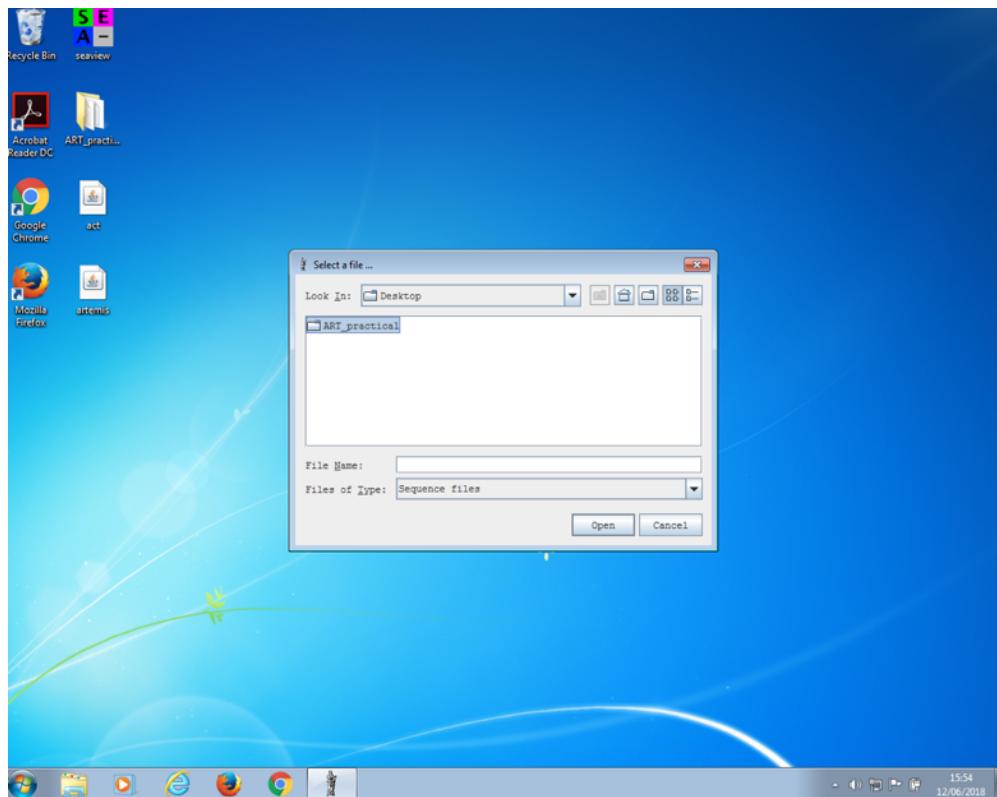
**Step 2:** The ‘Set Working directory box should appear as shown above. Click ‘OK’.



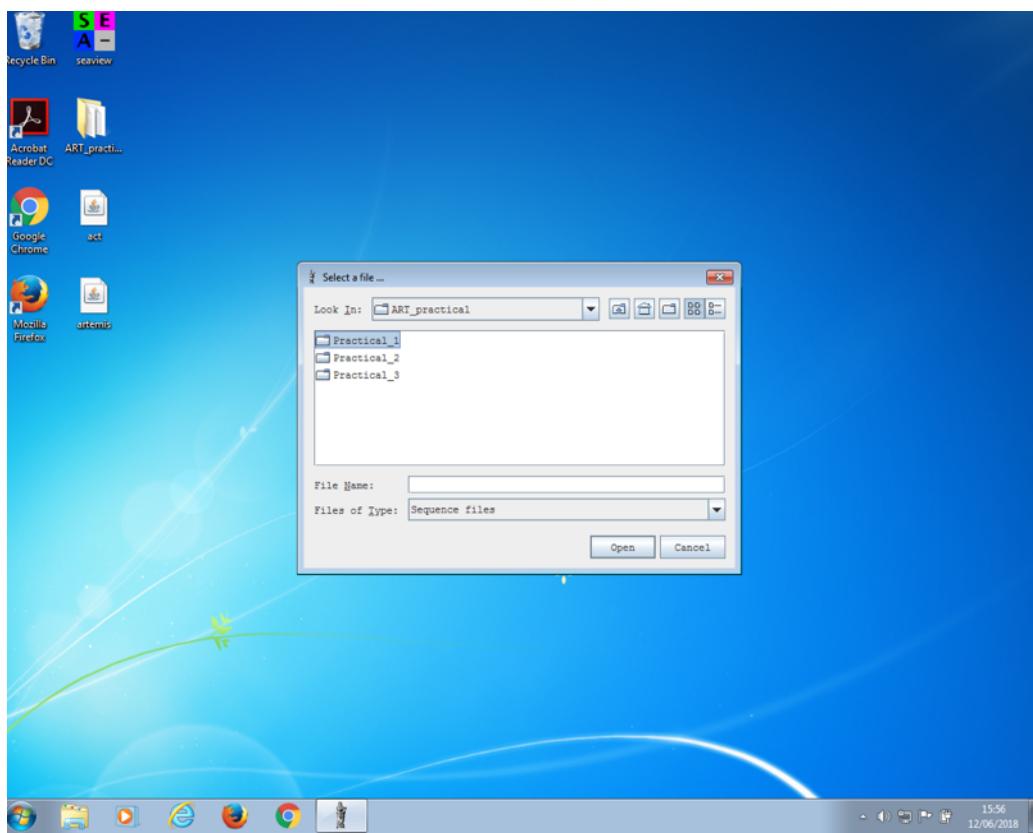
**Step 3:** The Artemis window should appear like this. Now click ‘File’.



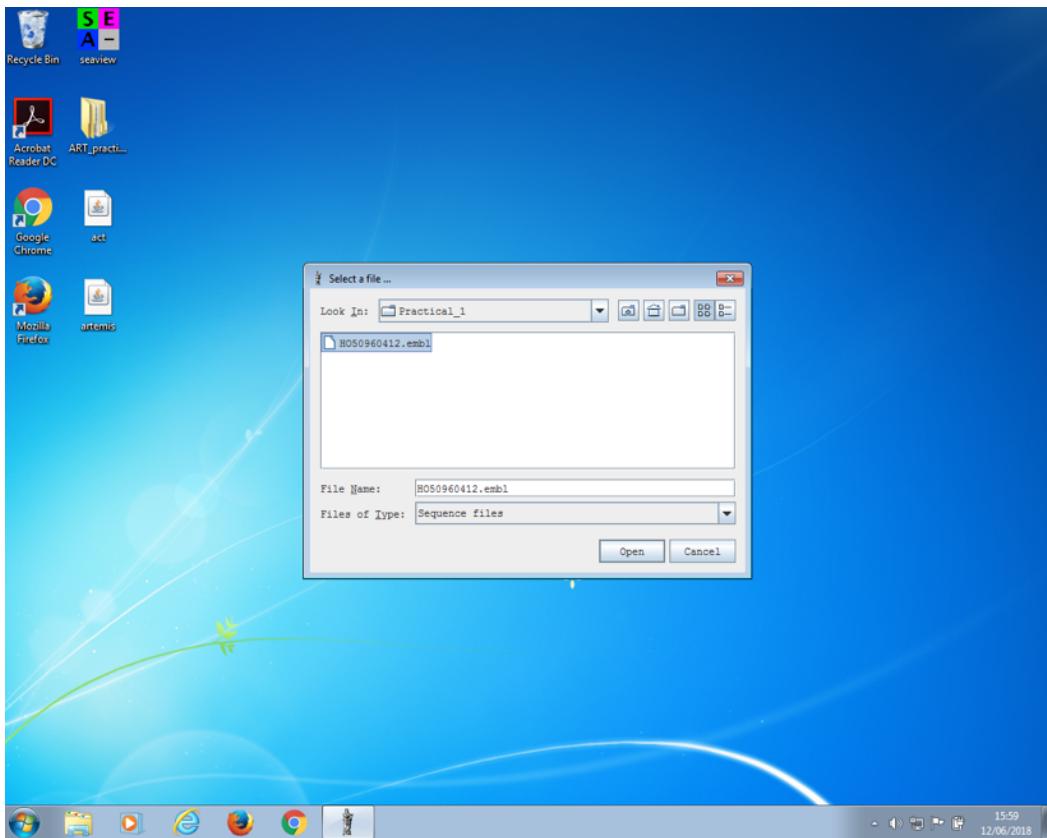
**Step 4:** Now select ‘Open ...’.



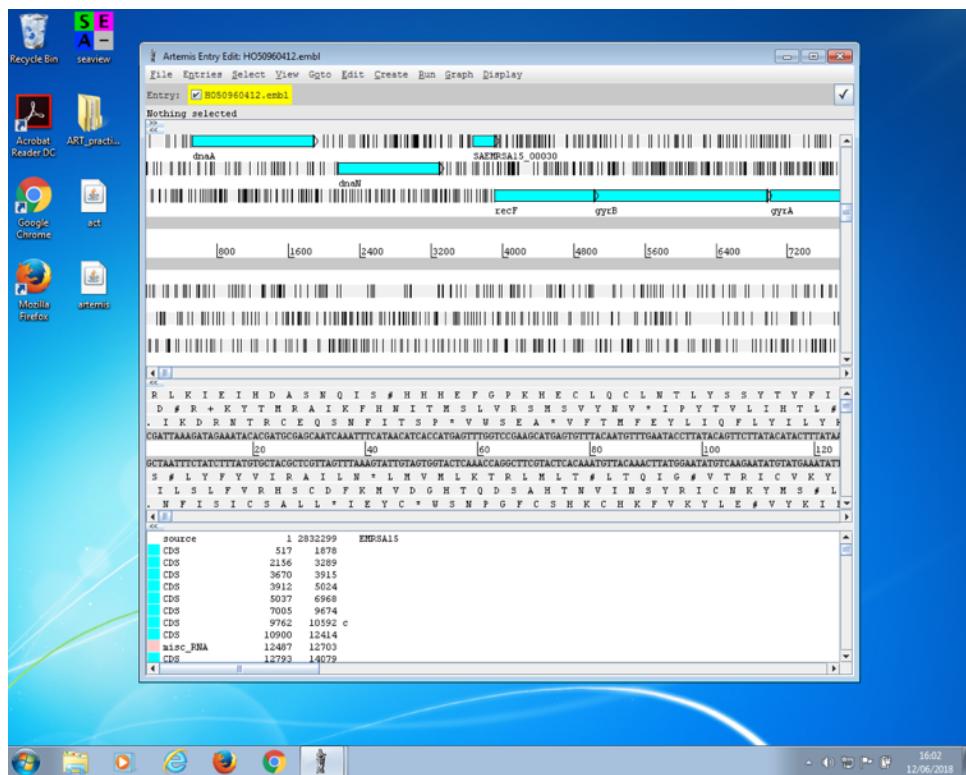
**Step 5:** Your window should now look like this – navigate to ~/course/ directory



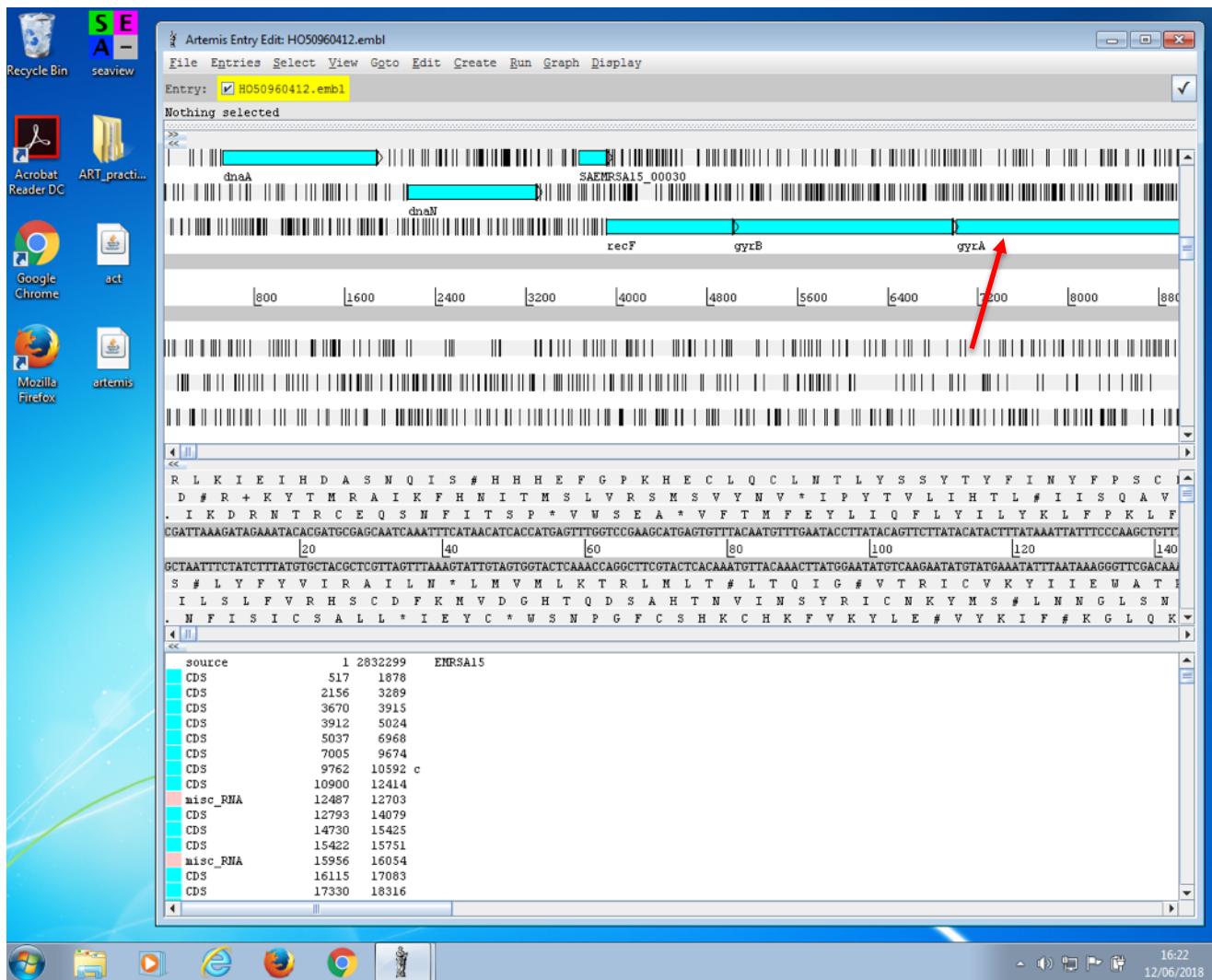
**Step 6:** Now select the 'cp13' folder and click 'Open'.



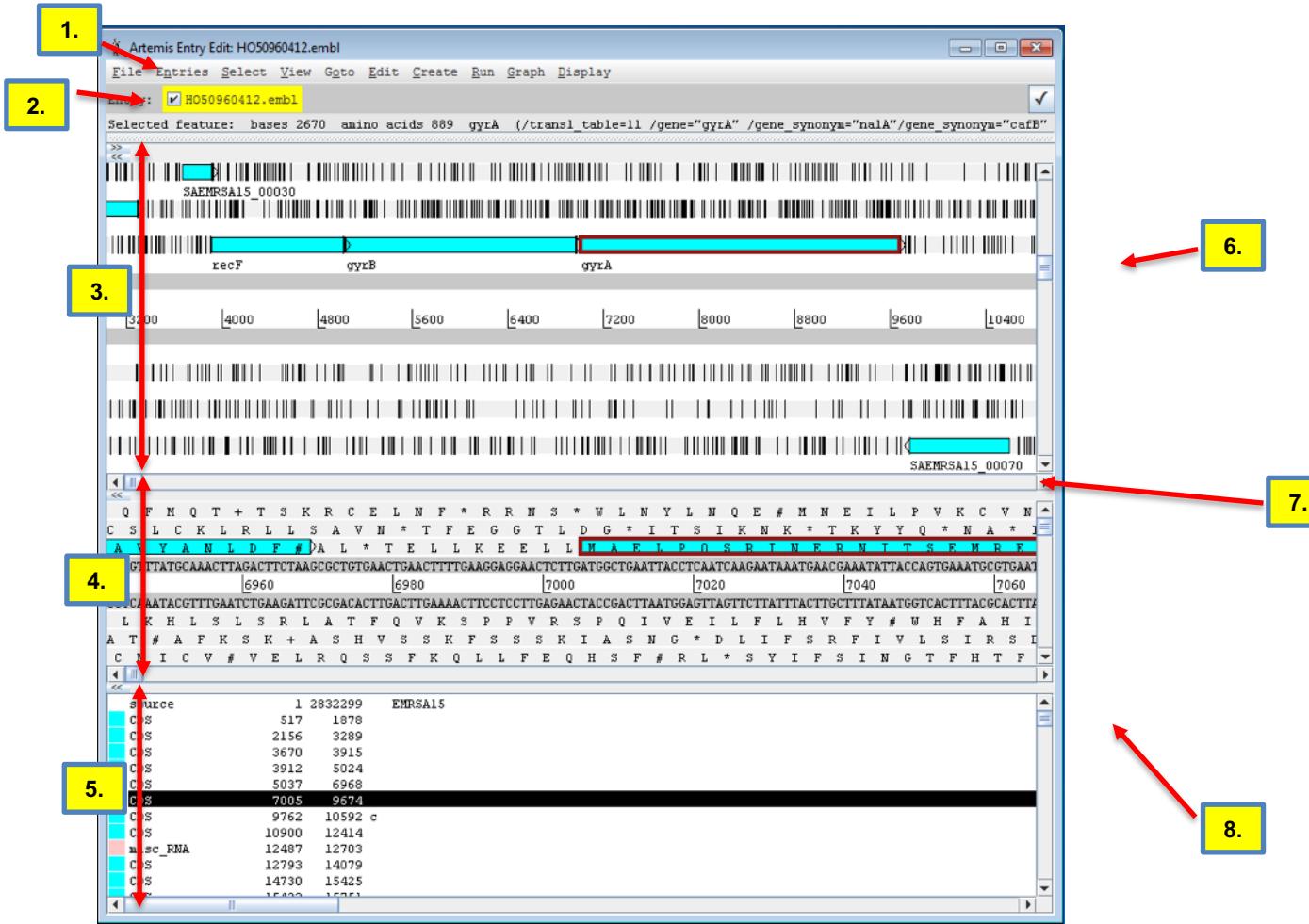
**Step 7:** Now select the file named: 'HO50960412.embl' and click 'open'.



**Step 8:** Your window should now look like this.



**Step 9:** Now double click on the gene labelled 'gyrA'. The window should move and the *gyrA* gene should be highlighted in red. The next page explains what all the different parts of the Artemis window are for.



1. **Drop-down menus:** There's lots in there so don't worry about all the details right now.

2. **Entry (top line):** shows which entries are currently loaded with the default entry highlighted in yellow (this is the entry into which newly created features are created). Selected feature: the details of a selected feature are shown here; in this case gene *gyrA* (blue box surrounded by thick red line).

3. This is the main **sequence view panel**. The central 2 grey lines represent the forward (top) and reverse (bottom) DNA strands. Above and below those are the 3 forward and 3 reverse reading frames. Stop codons are marked on the reading frames as black vertical bars. Genes and other annotated features (eg. Pfam and Prosite matches) are displayed as coloured boxes. We often refer to predicted genes as coding sequences or CDSs.

4. This panel has a similar layout to the main panel but is zoomed in to show nucleotides and amino acids. Double click on a CDS in the main view to see the zoomed view of the start of that CDS. Note that both this and the main panel can be scrolled left and right (7, below) zoomed in and out (6, below).

5. **Feature panel:** This panel contains details of the various features, listed in the order that they occur on the DNA. Any selected features are highlighted. The list can be scrolled (8, below).

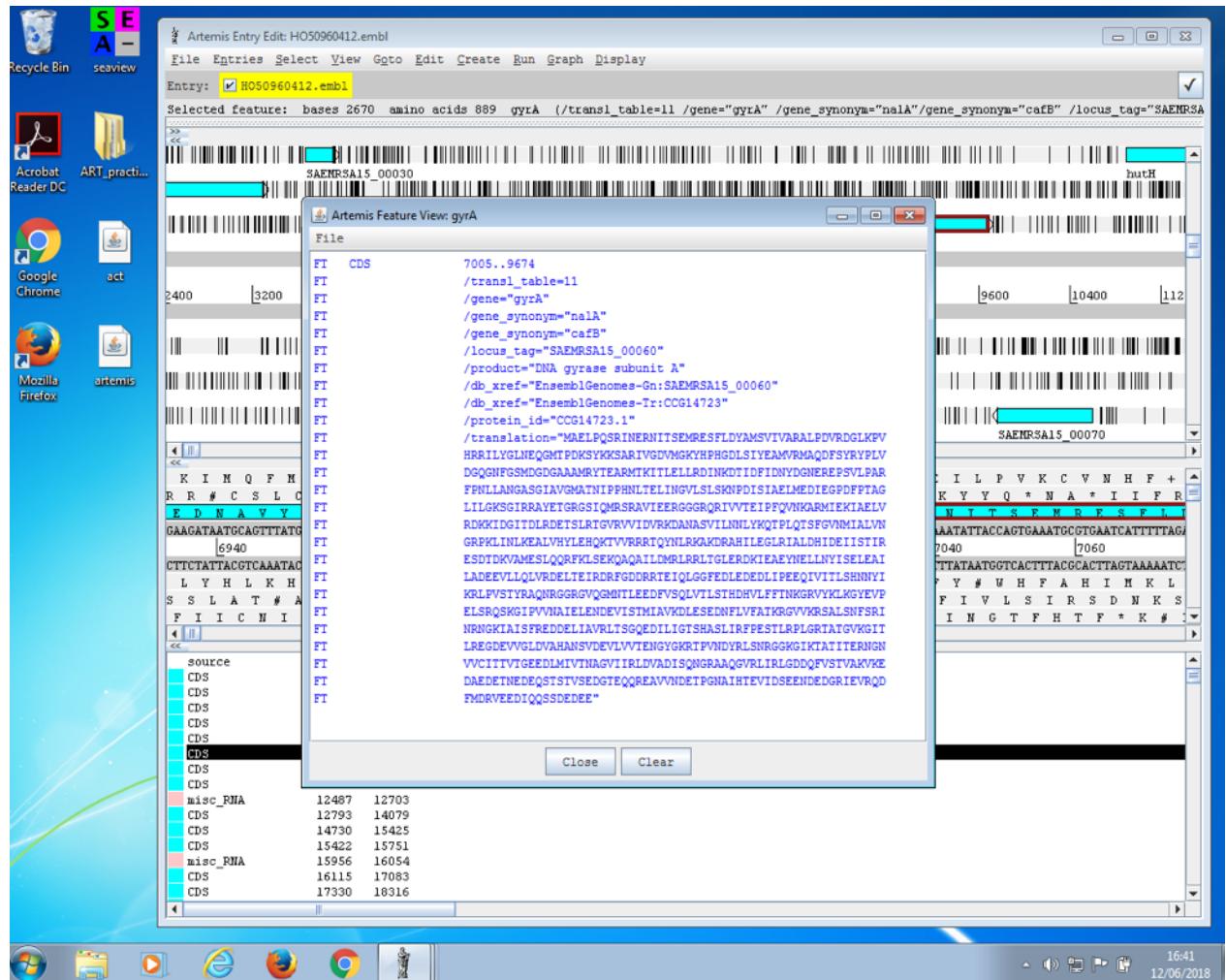
6. **Sliders** for zooming view panels.

7. **Sliders** for scrolling along the DNA.

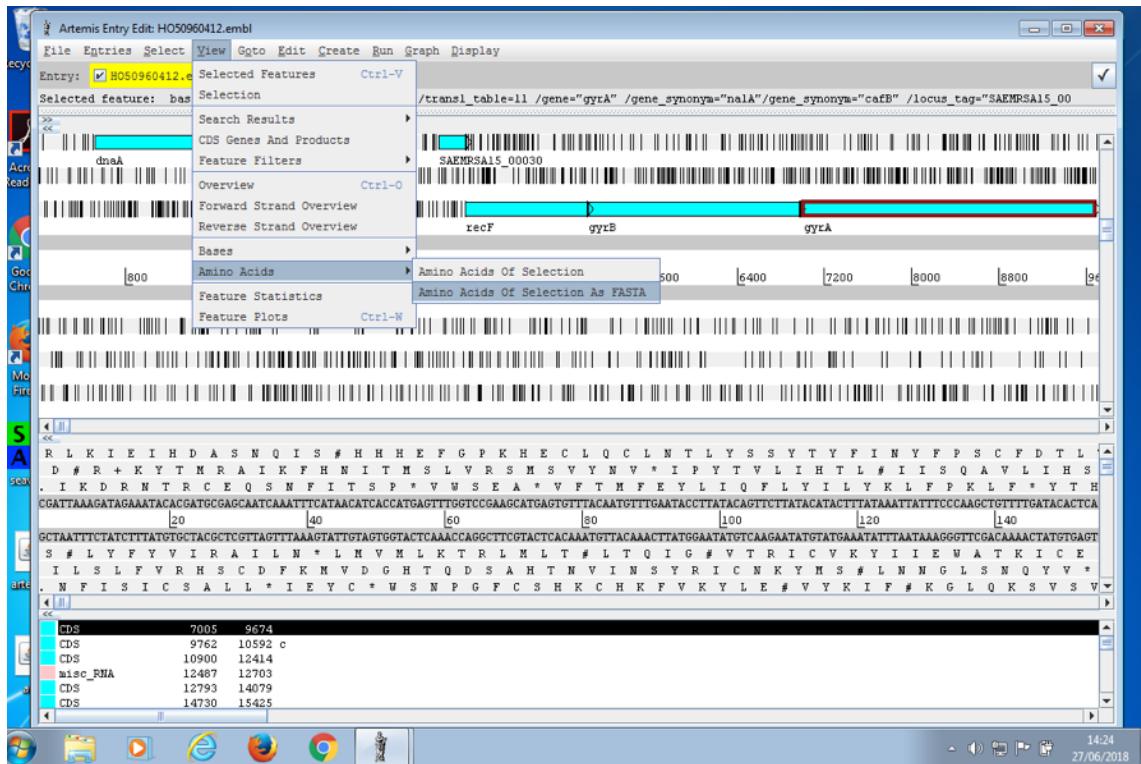
## 9. Slider for scrolling feature list.

**Step 10:** Try zooming the view in and out (no. 6 in figure above) and moving genome location using the slider for scrolling the DNA (no. 7 in figure above).

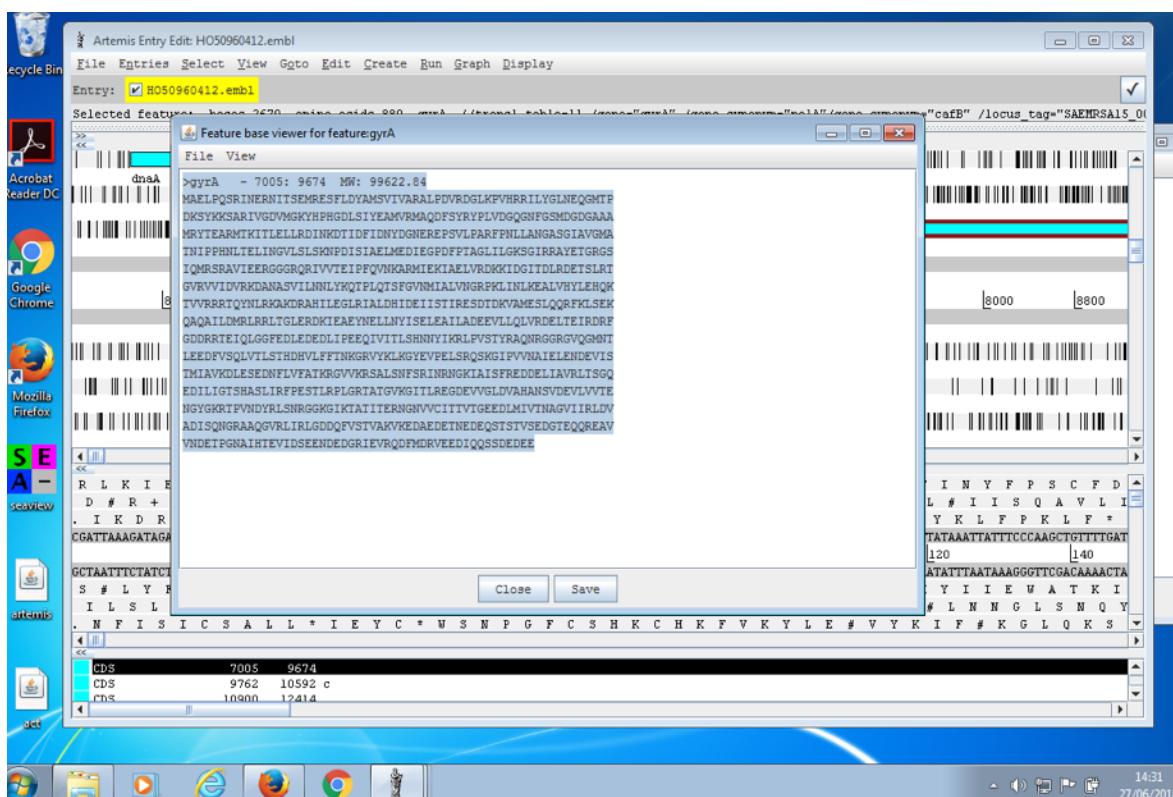
**Step 11:** Now scroll back to the start of the genome and click again on the *gyrA* gene. Now on the drop-down menu (No.1 in the figure above) and select 'View' and 'Selected features' (a short cut for this is ctrl – v). This brings up all the information that is stored in the entry about this gene or feature. In this case you can see that the product of *gyrA* is 'DNA gyrase subunit A'.



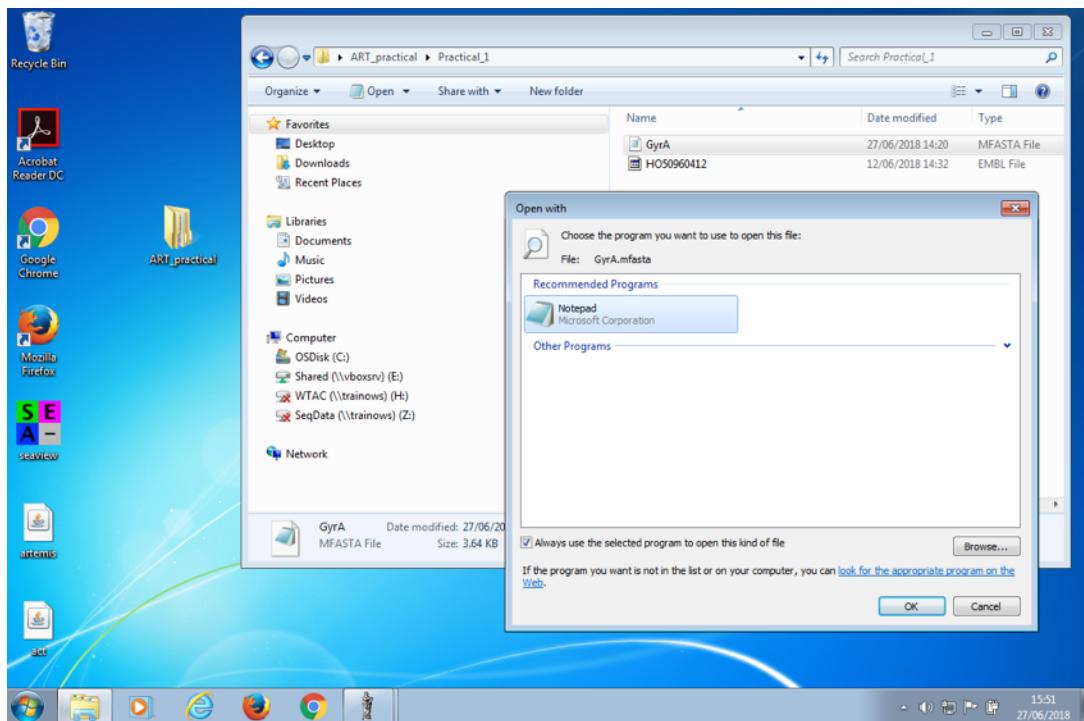
**Step 12:** You might have noticed this is the one of the genes that we looked at when we were tracking the origin of the EMRSA-15 (ST22) MRSA clone. If you remember a point mutation in *gyrA* that generates the Ser84Leu substitution can mediate resistance to fluoroquinolone antibiotics such as ciprofloxacin. We are now going to investigate if this isolate is likely resistance by seeing if this gene contains the substitution necessary for resistance.



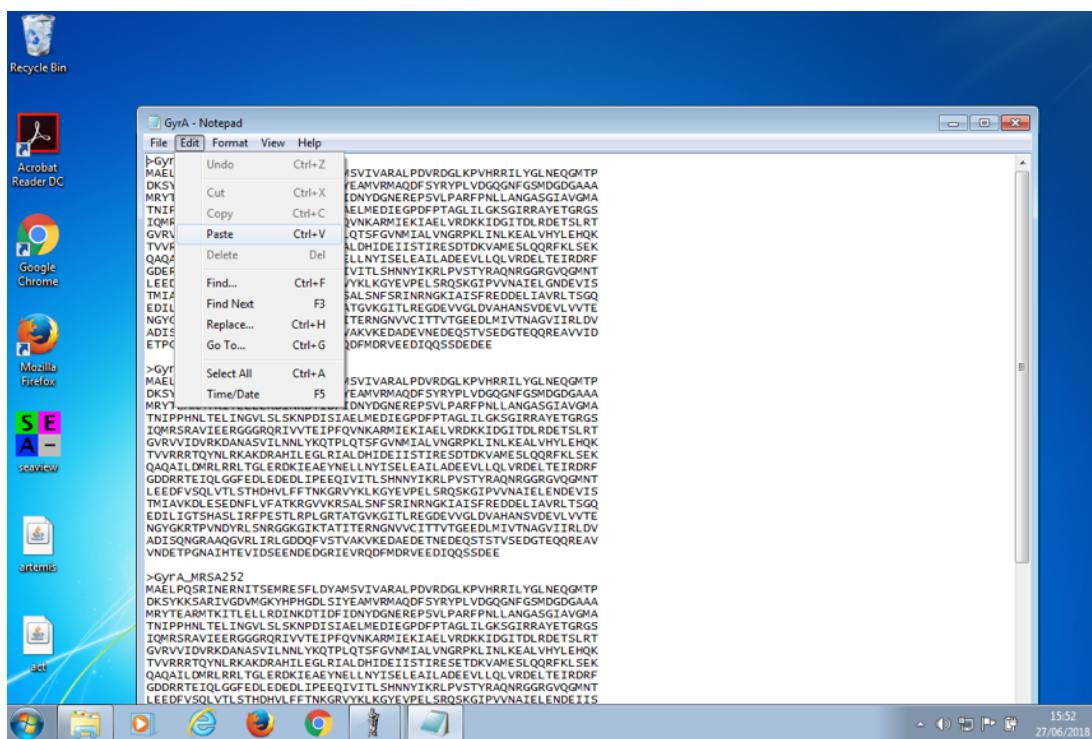
**Step 13:** Now making sure that *gyrA* is still selected. Click on the ‘View’ window and select ‘Amino Acids’ and then select ‘Amino Acids of Selection As Fasta’



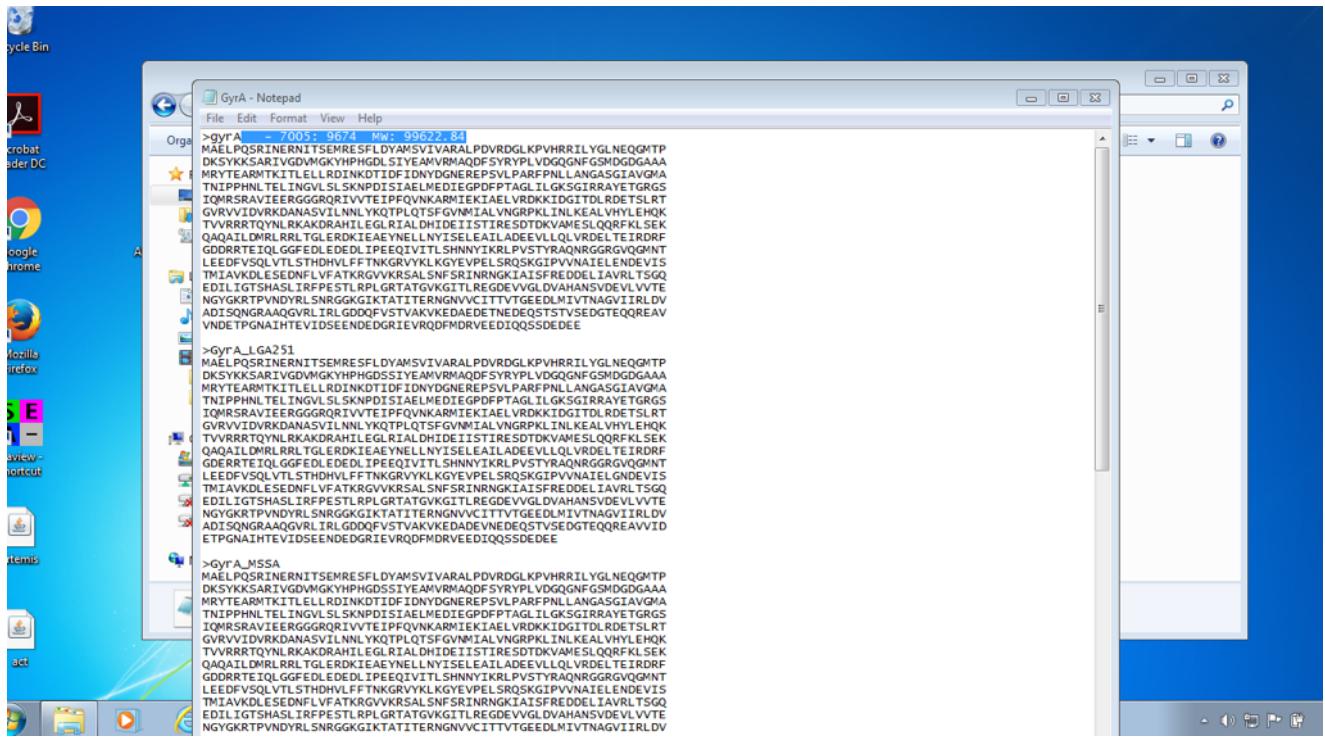
**Step 14:** A window will now open with the GyrA amino acid sequence. Now press ctrl + A to select all the whole sequence and then press ctrl + c to copy the amino acid sequence.



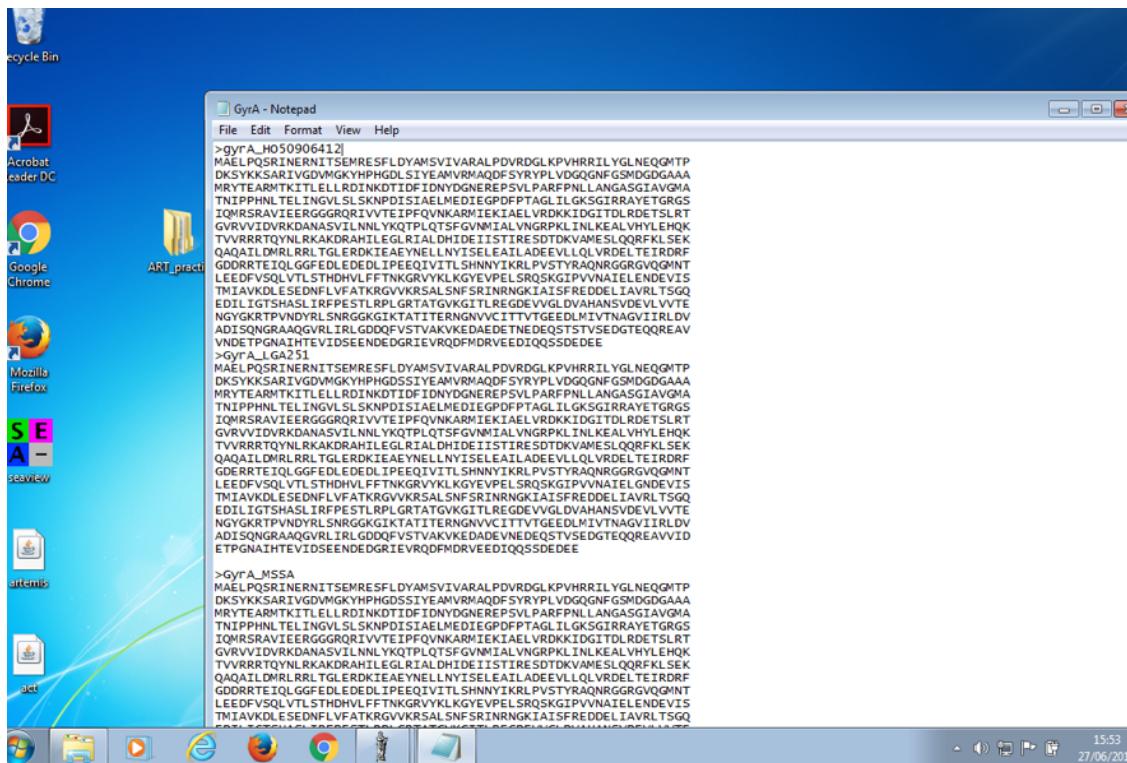
**Step 15:** Minimise the Artemis window and navigate to `~/course/` directory. Now open the 'cp13' folder. Inside this file you will find a file called 'GyrA.mfasta' – right click on this and select 'Open with Text Editor'.



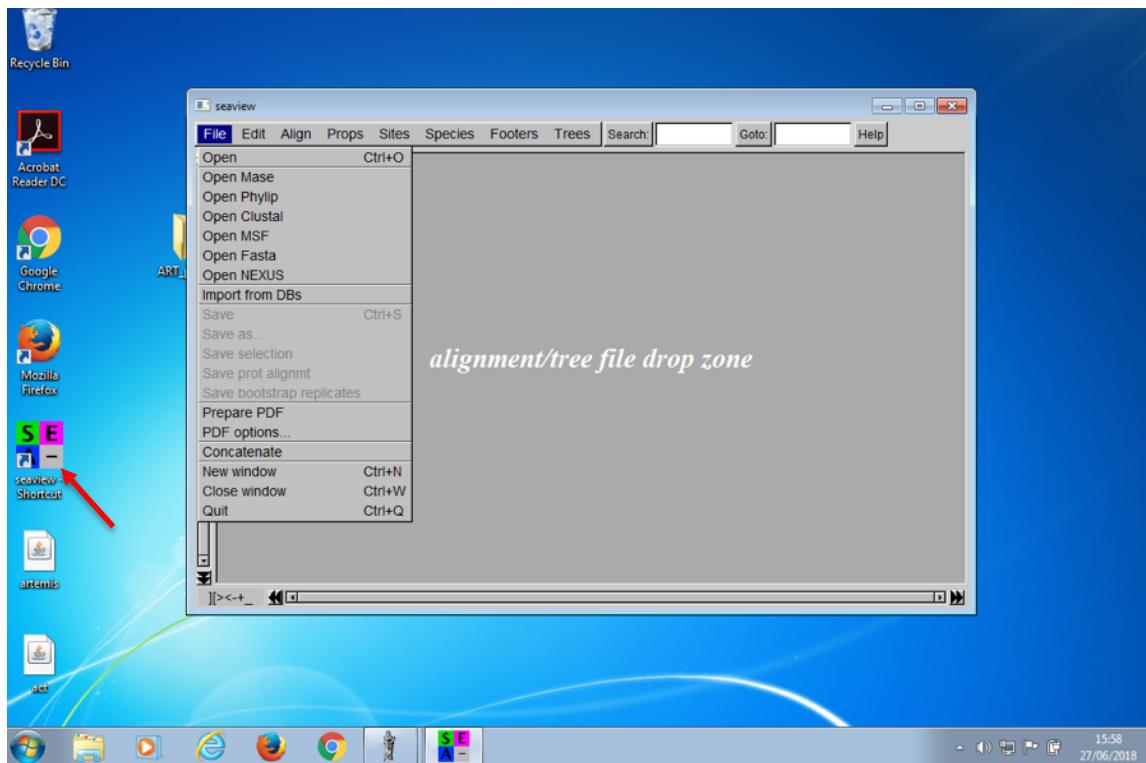
**Step 16:** The file should now open. The file contains the amino acid sequences of GyrA from three other *S. aureus* isolates. Making sure the cursor is at the top, select the 'edit' menu and then click 'paste'. This will now copy the GyrA sequence you copied from the HO50960412 sequence into the file.



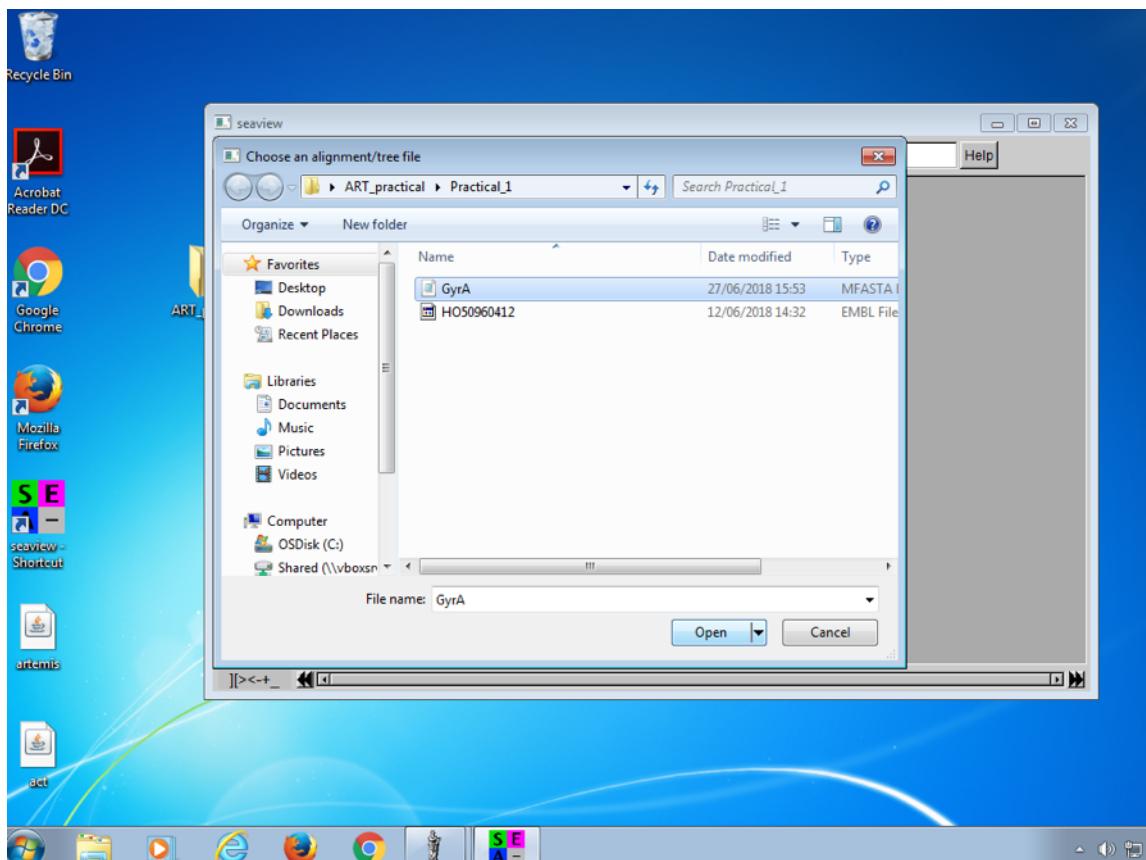
**Step 17:** Now move the cursor to highlight the text shown at the top of the sequence you just pasted into the file. Once you have highlighted this text – select the ‘edit’ menu and then click ‘delete’ to delete this text.



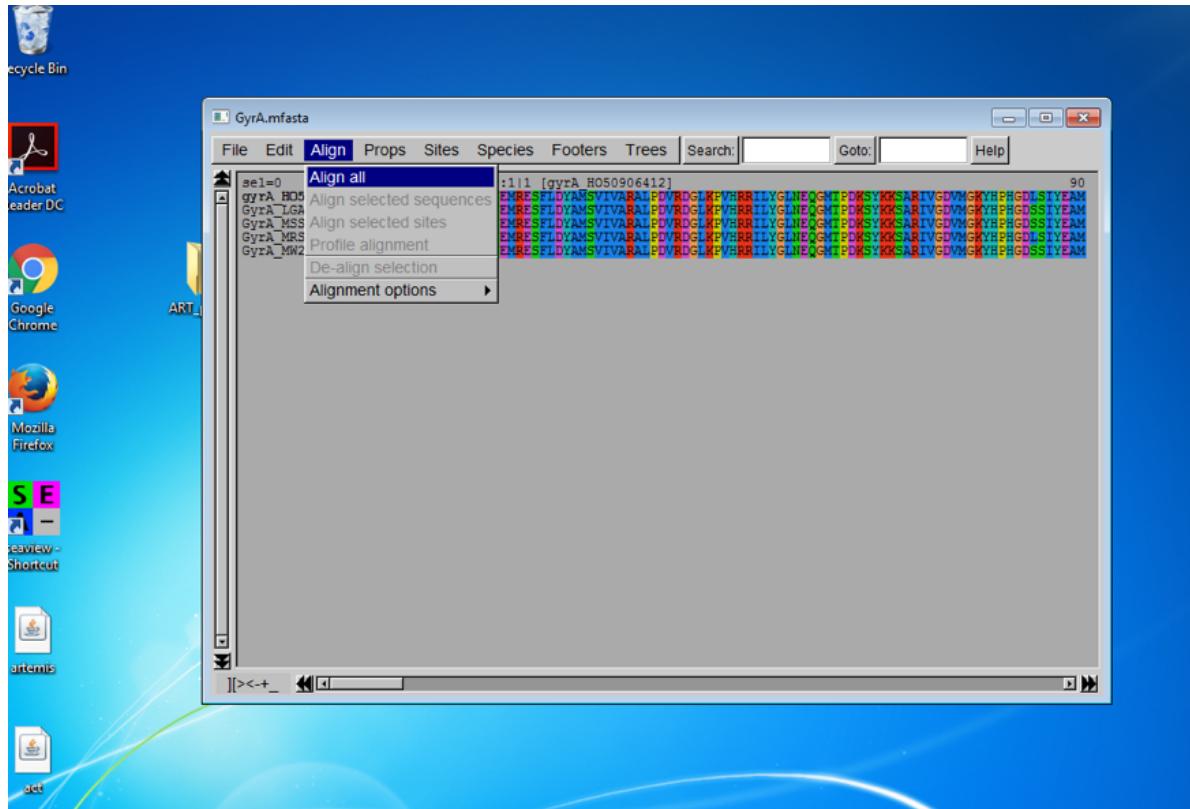
**Step 18:** Now type ‘H050960412’ – where you deleted the text. Then select ‘File’ from the menu and click ‘Save’.



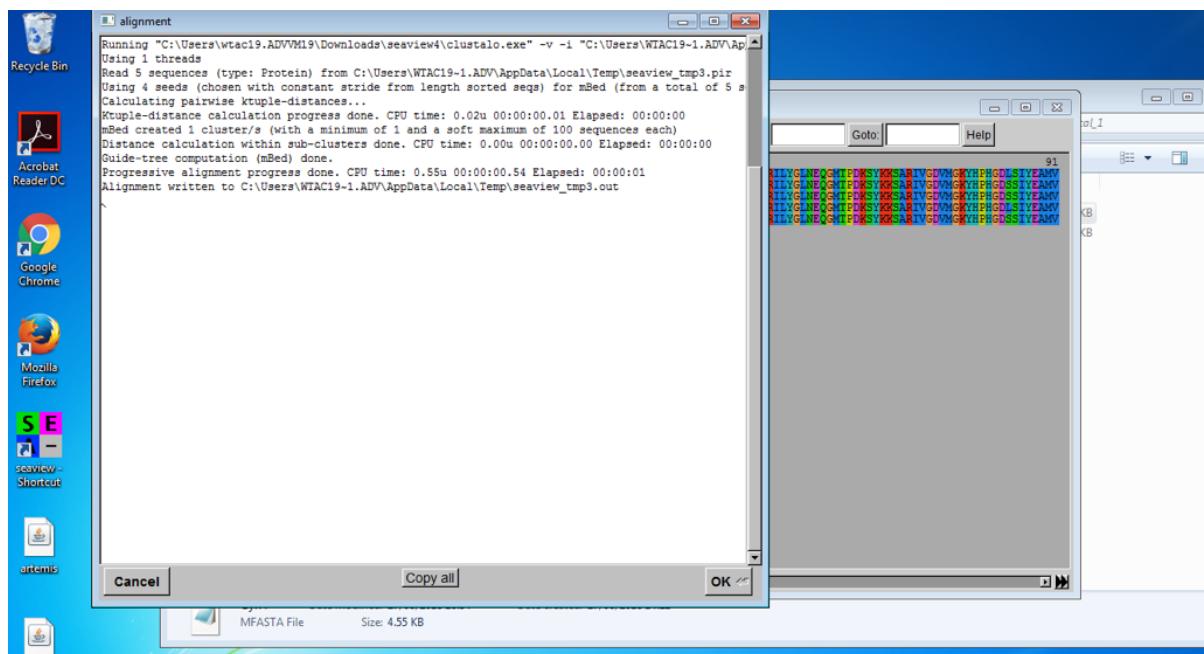
**Step 19:** Now got to the desktop and select the icon ‘Seaview’ and double click to open the program. Once the window opens select ‘File’ and then click ‘open’. **NOTE: if you cannot locate ‘Seaview’ icon on your desktop, open up a new terminal window and type in ‘seaview’. This will also open Seaview tool too.**



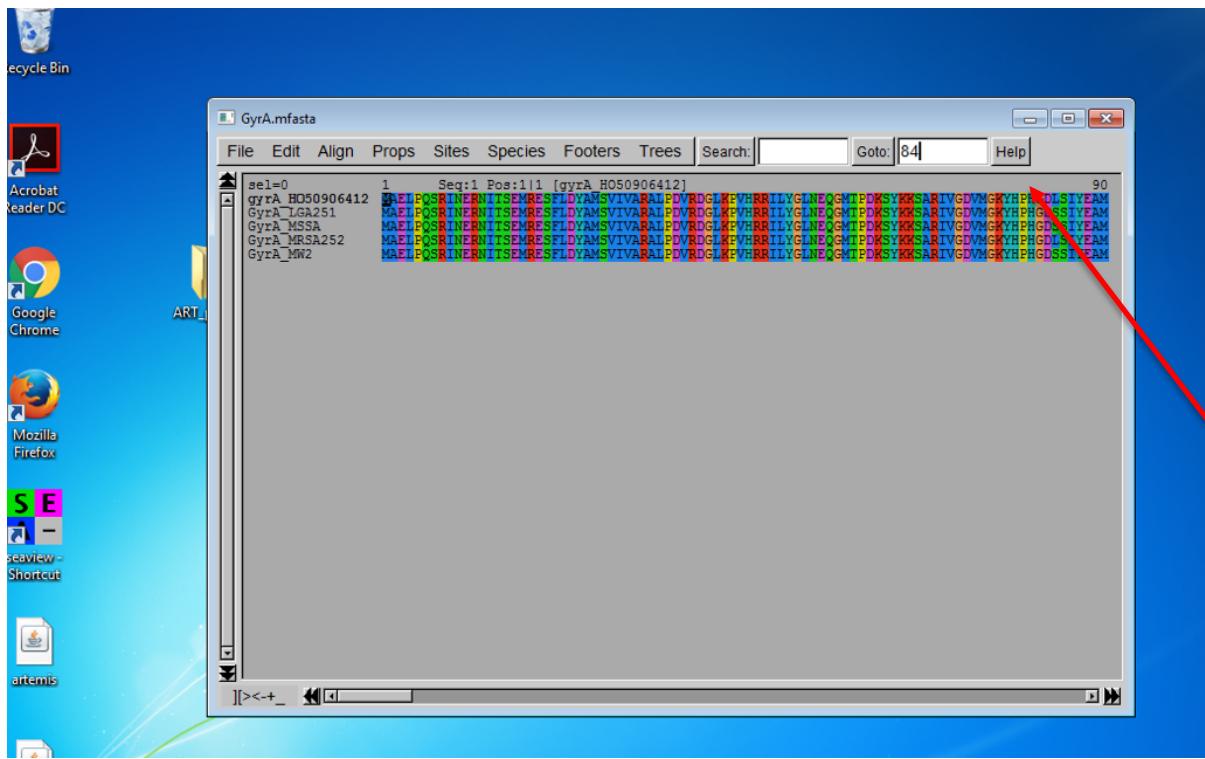
**Step 20:** Now navigate to `~/course/cp13` directory. Then click on the file 'GyrA.mfasta' you just saved and click 'open'.



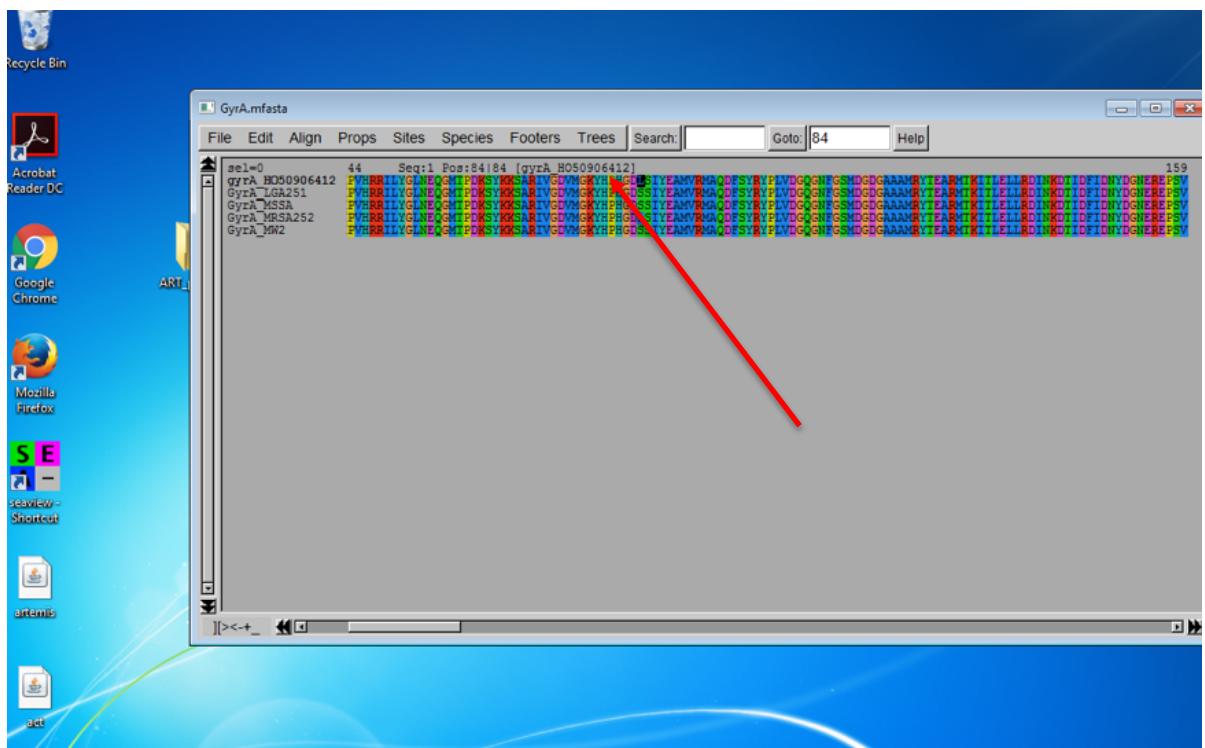
**Step 21:** Now select the 'Align' menu and click 'Align all' in the drop down menu. This will run a program that will align your sequences.



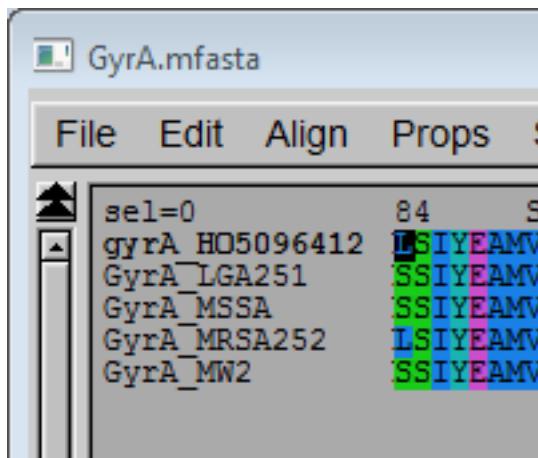
**Step 22:** The program will run and you will see a screen like this. Click 'okay'.



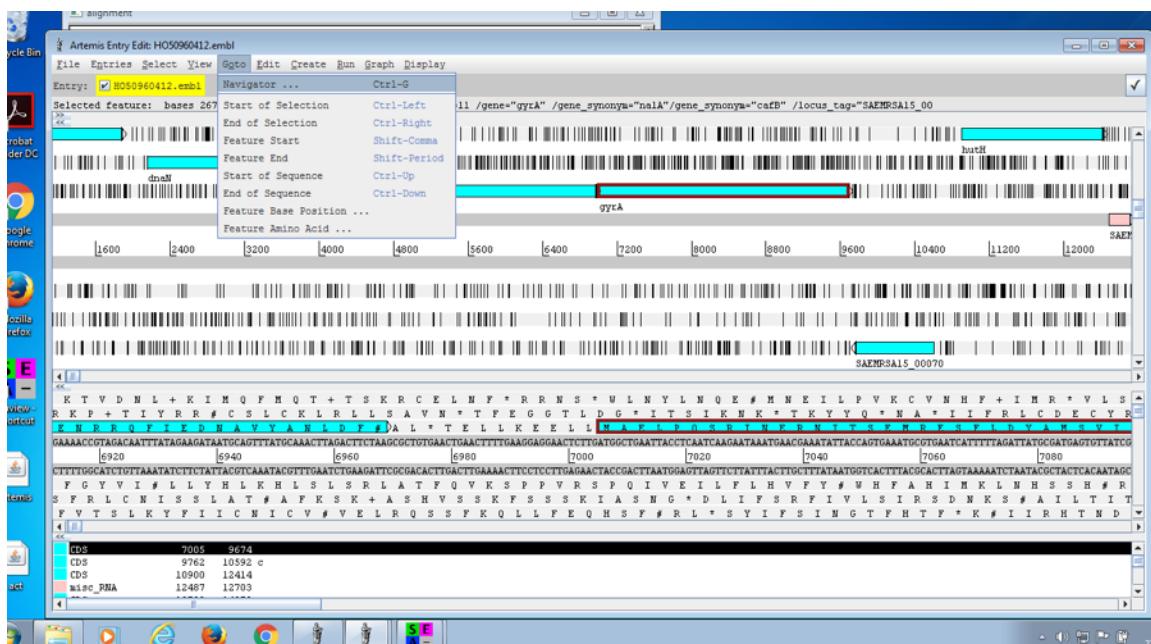
**Step 23:** You can now use the scroll bar to move up and down the alignment. This view allows you to compare differences between the sequences. Now click on the 'Goto' box at the top (see arrow) and type '84' and click the 'Goto' button. The program will now take you to amino acid position 84 in the sequence.



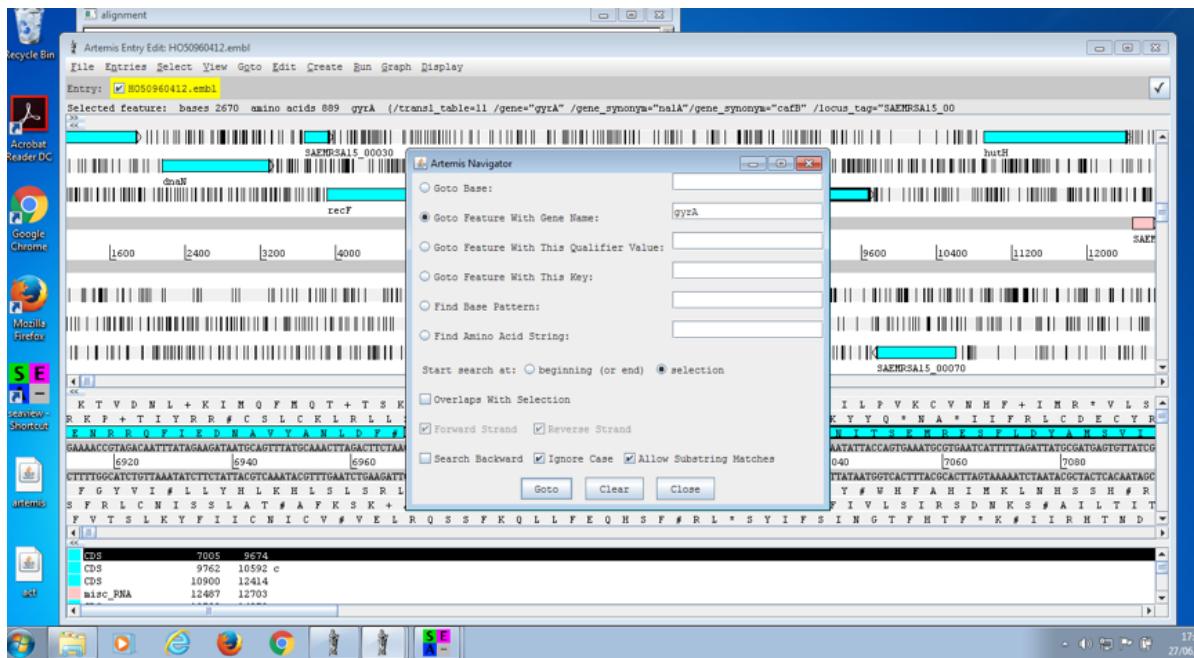
**Step 24:** You should now have a view like this, with the black cursor highlighting position 84.



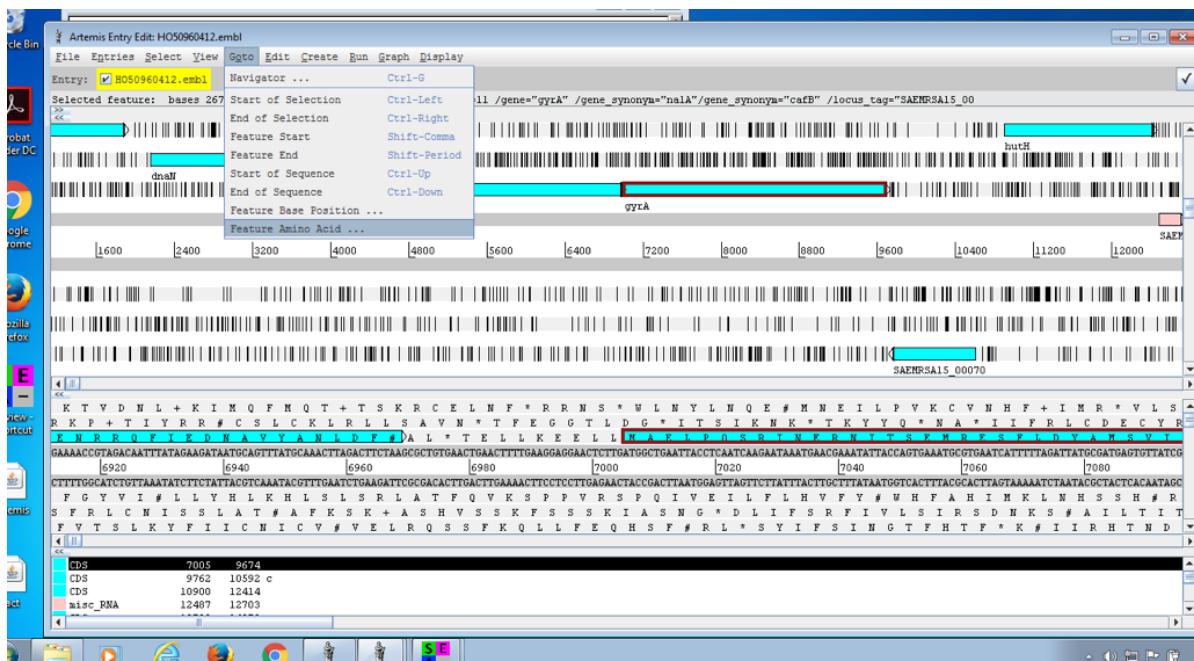
**Step 25:** Above is a zoomed in view. What you can see if that two of the isolates (MRSA252 and the one we have been working with HO5906412) both have a L (L = Leucine = Leu). While the other three isolates (MSSA, MW2 and LGA251) have an S (S = Serine = Ser). If you remember back to the previous practical – a Serine to Leucine (Ser84Leu) substitution at this position mediates resistance to ciprofloxacin (a fluoroquinolone). As you can see our isolate HO5906412 has the Ser84Leu substitution and so is resistant. This is what we would expect from an EMRSA-15 isolate. The other isolate with the substitution MRSA252 is from another hospital associated lineage of MRSA known as EMRSA-16, in which fluoroquinolone resistance is also common.



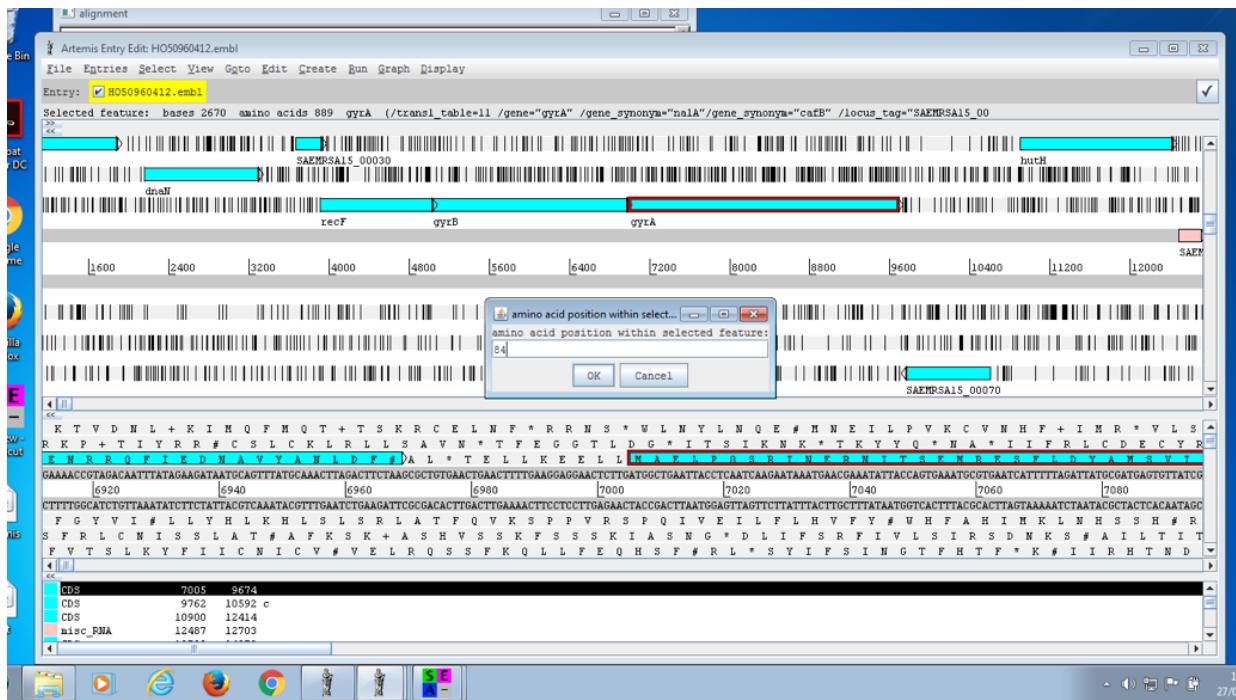
**Step 26:** If you go back to the Artemis window. There is another way to see if a gene or protein has a mutation or substitution at a particular position in the gene or amino acid sequence. Select the 'Goto' menu at the top of the Artemis window and then select and click on 'Navigate'.



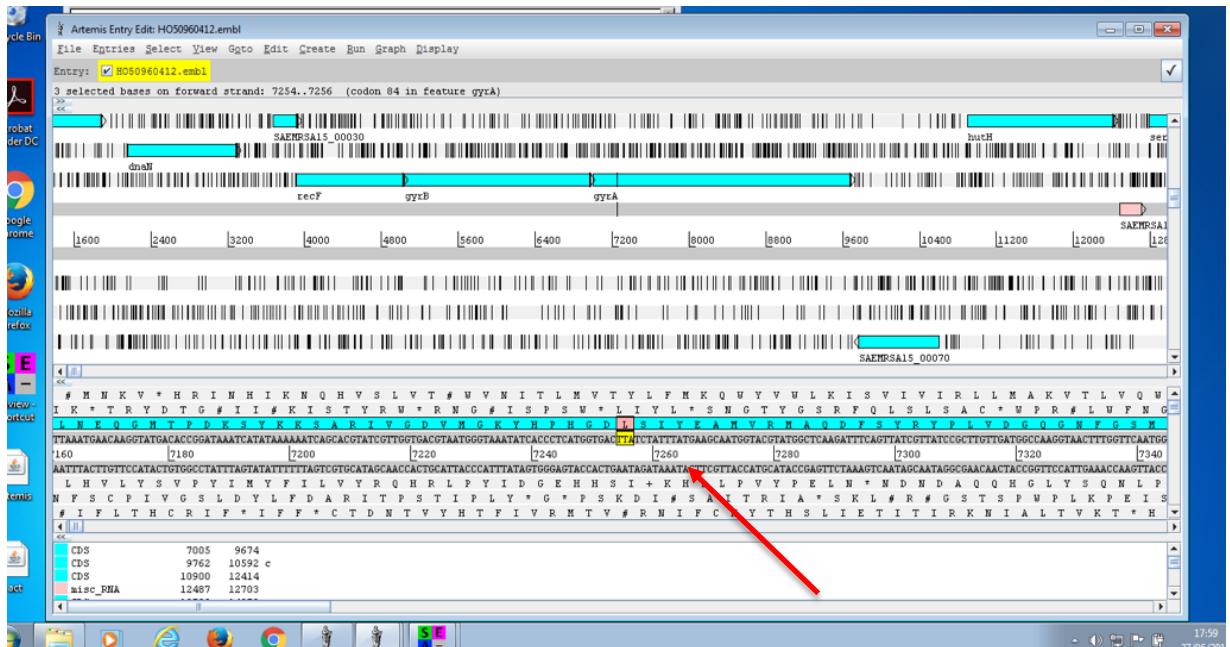
**Step 27:** Click the ‘Goto feature with Gene Name’ and then type ‘gyrA’. If you now click the ‘Goto’ button at the bottom of the box this will select and take you to the *gyrA* gene again.



**Step 28:** If you now go to the ‘Goto’ menu again and select ‘Feature Amino Acid’ from the drop-down menu.



**Step 29:** In the box that appears type '84' – this will take you to amino acid position 84 in the *gyrA* gene.



**Step 30:** You can see here that amino acid 84 is highlighted and you can see that position 84 as we know is L = Leucine.

**Step 31:** Two different amino acid substitutions are known to mediate fluoroquinolone resistance in *S. aureus* ST22 isolates. The second being a Serine (Ser / S) to Phenylalanine (Phe / F) substitution at amino acid position 80 (Ser80Phe) in GrlA. Using what you just learned, can you now go back to the Artemis window and find out if strain HO5906412 also has this substitution?

**Step 32:** Now using what you have learned you can look for some more genes involved in resistance namely the beta-lactamase gene: *blaZ* and the *mecA* the gene that you targeted in your PCRs. Using the ‘Goto’ command search for these two genes. What is strange about the *blaZ* gene? What other genes are close by? When you find the *mecA* gene, what other genes are close by? Does this tell you anything? In the next practical we are going to look more closely at this region.

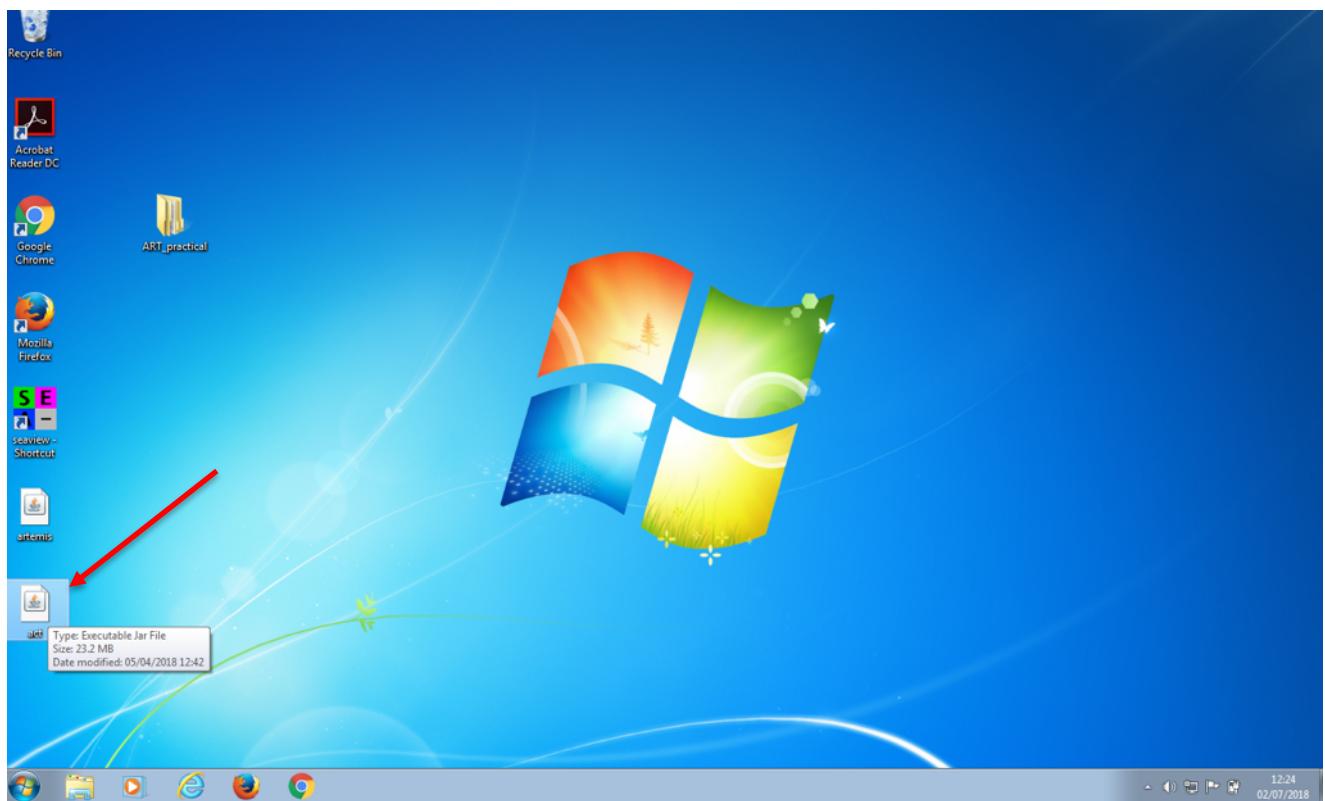
# Computational Practical 14

## Analysis of resistance in genomes

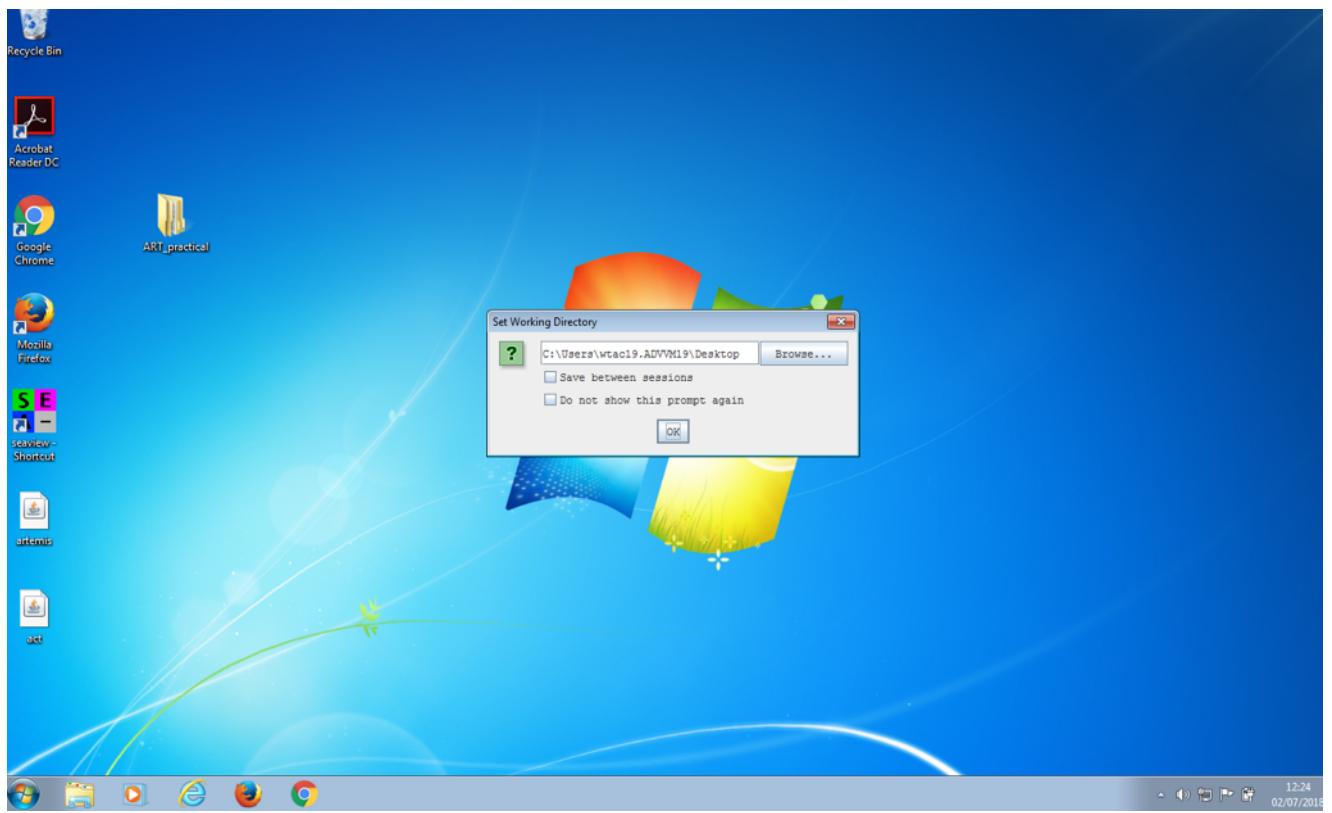
### 14.1 Comparative genomics

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):

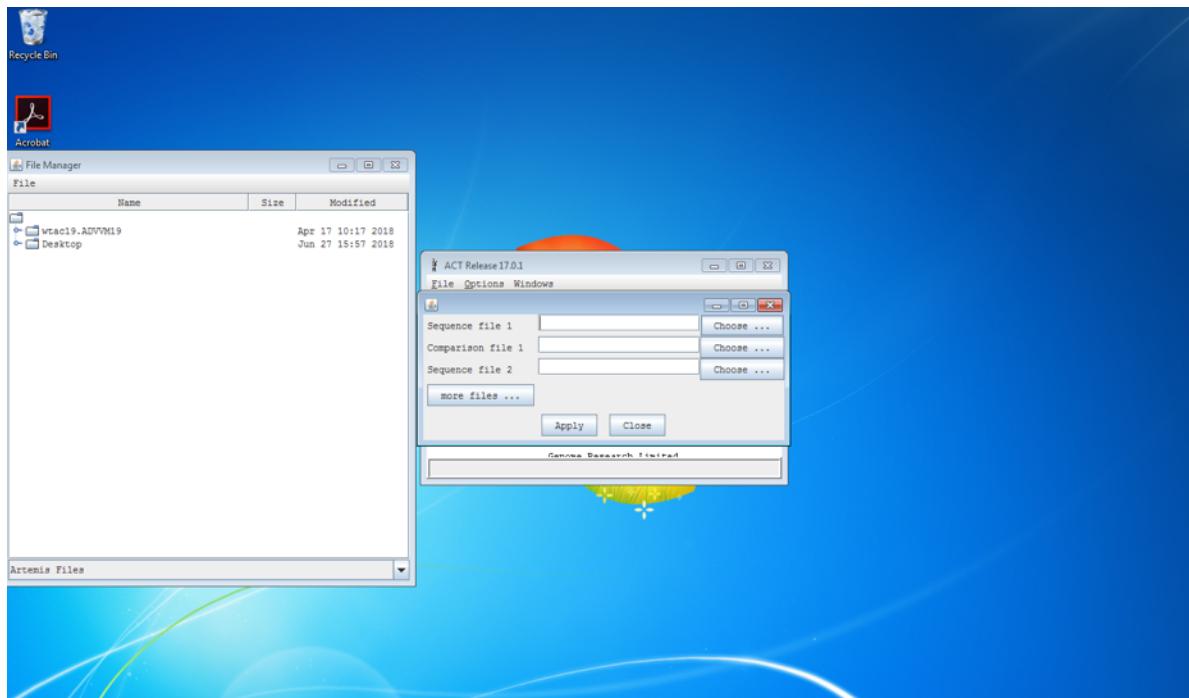
- Artemis Comparison Tool is a genome browser that lets you view comparisons of genomes. It was developed at the Wellcome Sanger Institute (<http://www.sanger.ac.uk/science/tools/artemis>). It is freely available to download for PCs and Mac.
- The practical is designed to give you a basic understanding 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 the 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.

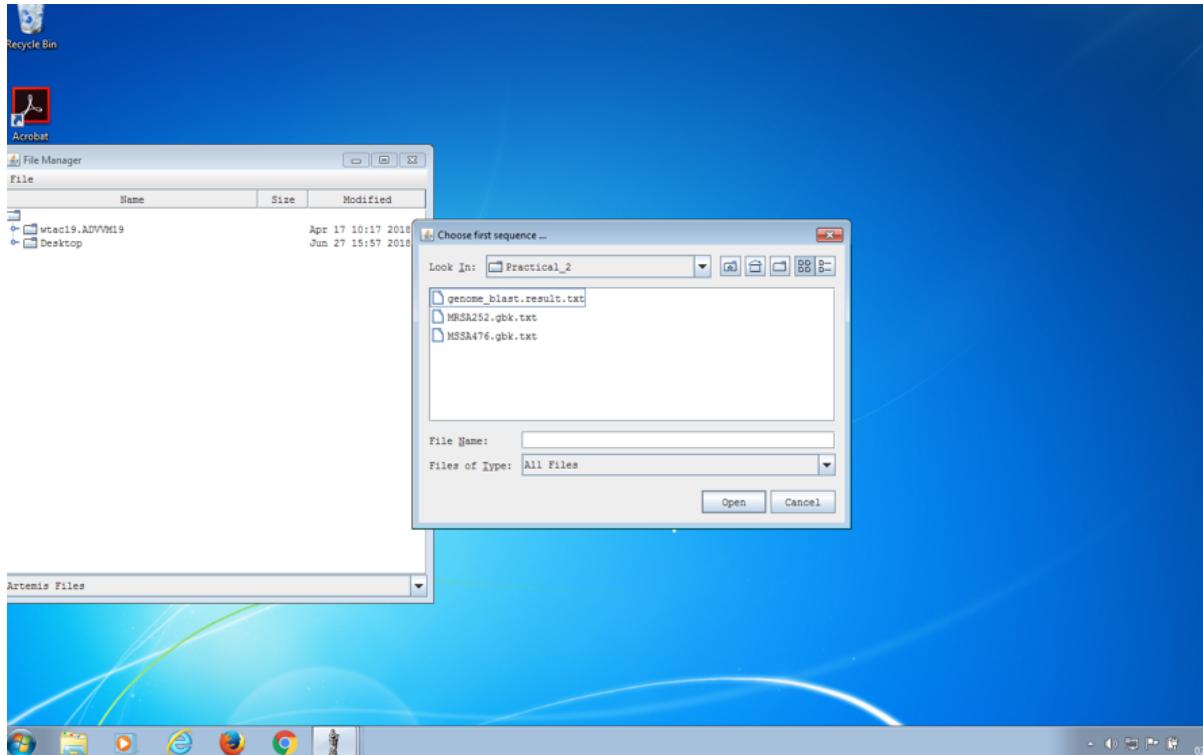


**Step 2:** A dialog box will appear. Click ‘OK’.

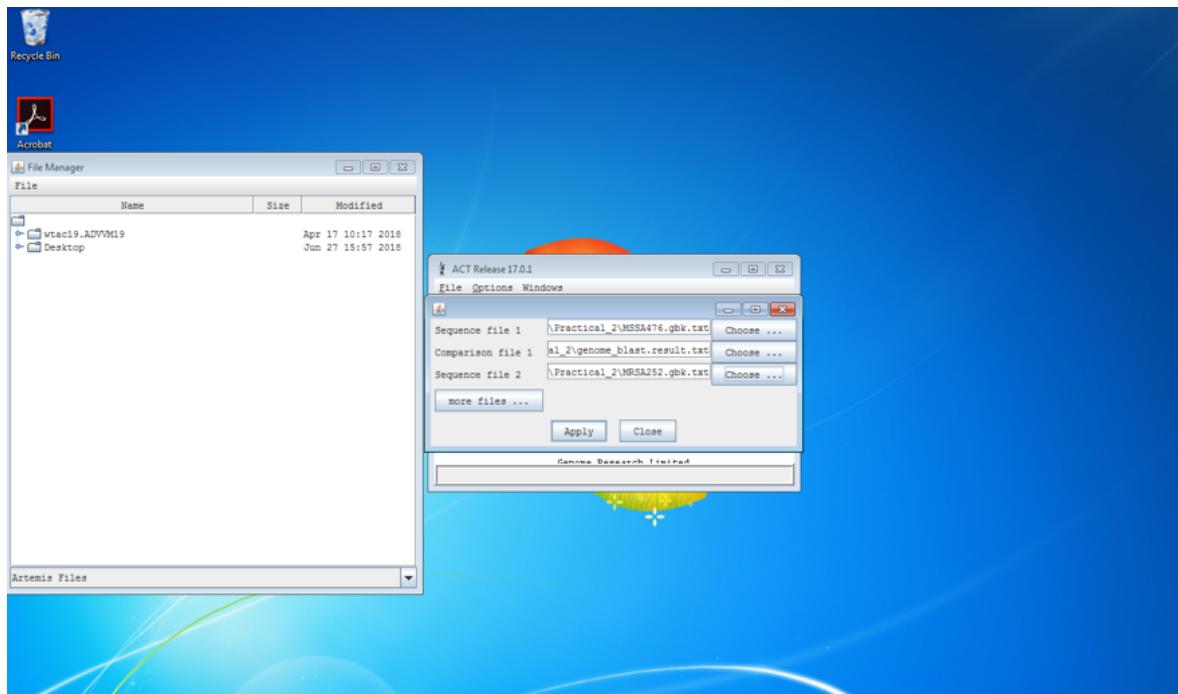


**Step 3:** We are now going to look at two genomes of *Staphylococcus aureus*; MSSA476 - a methicillin-sensitive strain and MRSA252 – methicillin-resistant strain. Click ‘File’ and then select ‘Open ...’. A box will open like the one shown above. For

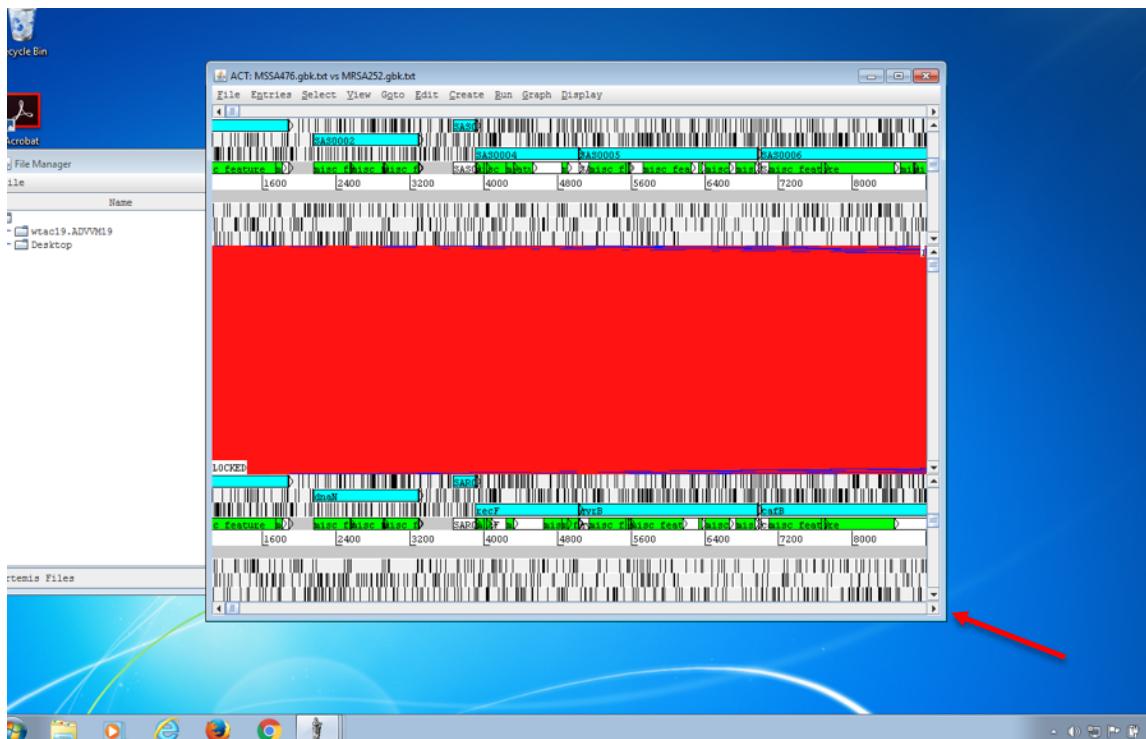
Sequence file 1 – click the ‘Choose ...’ box. Navigate to ~/course/ directory and then click the ‘cp14’ folder.



**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’.



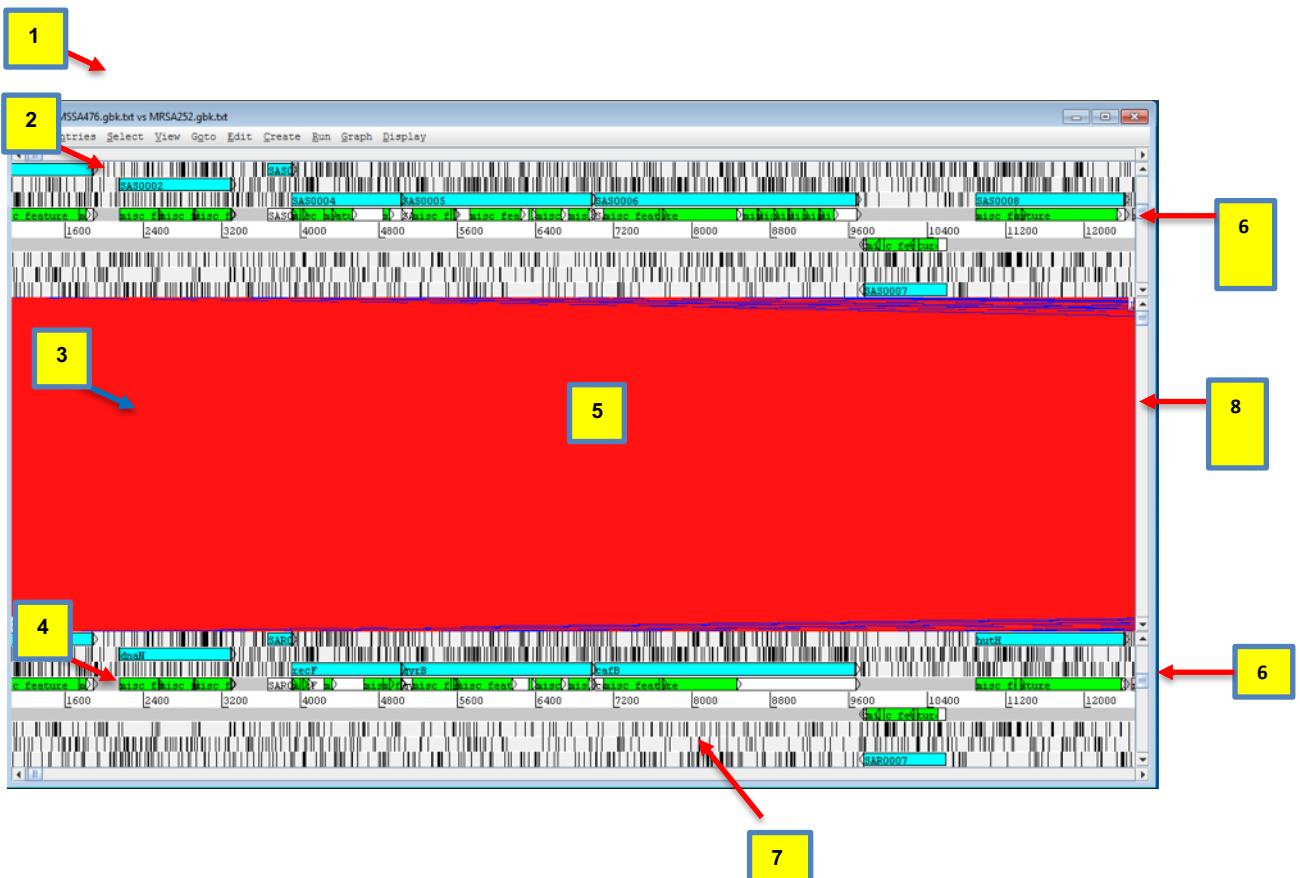
**Step 5:** Now click the ‘Choose ...’ button for ‘Comparison file 1’ and then select MSSA476\_vs\_MRSA252\_blast\_results.txt and click ‘open’. Now do the same for Sequence file 1 selecting the file MRSA252.gbk.txt. Now press the ‘Apply’ button. A message will appear asking if you want view errors. Just click yes.



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



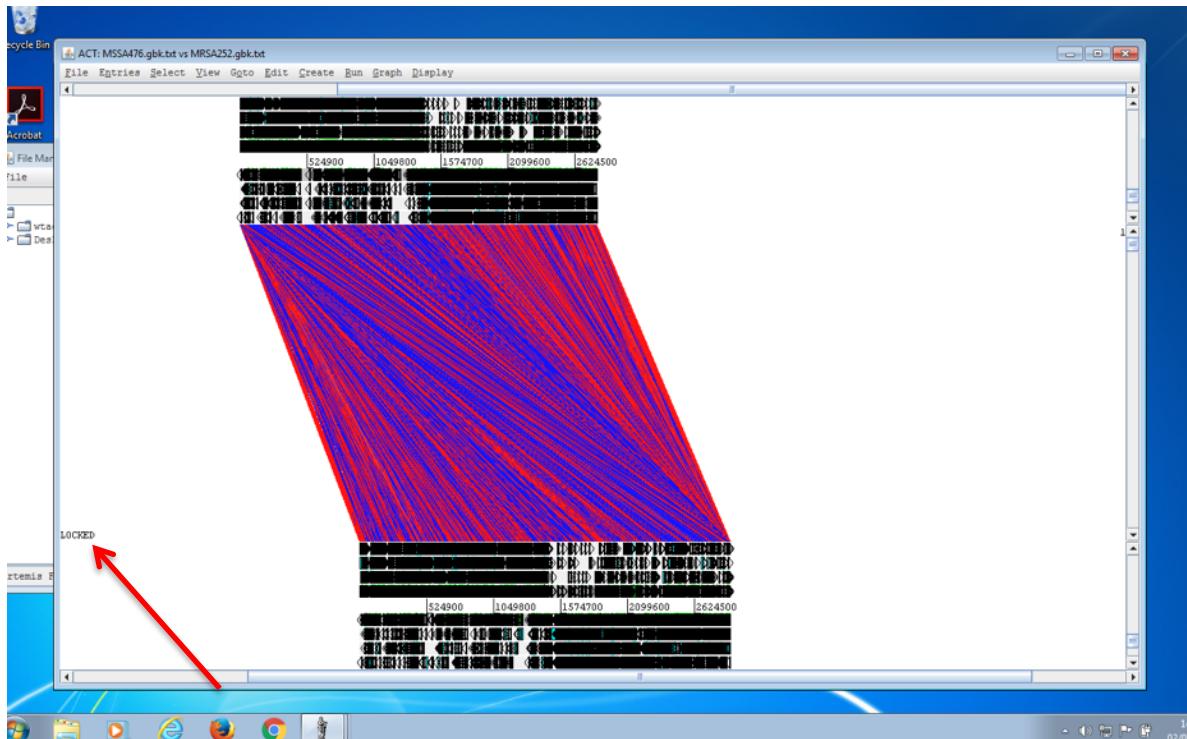
**Step 7:** You should now have a view like the one above. As you can see it is very similar to the view from Artemis. But you have two sequences on top of each other – the top one being MSSA476 and the bottom one MRSA252. The red colour in the middle indicates that these sequences are the same (conserved) 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).



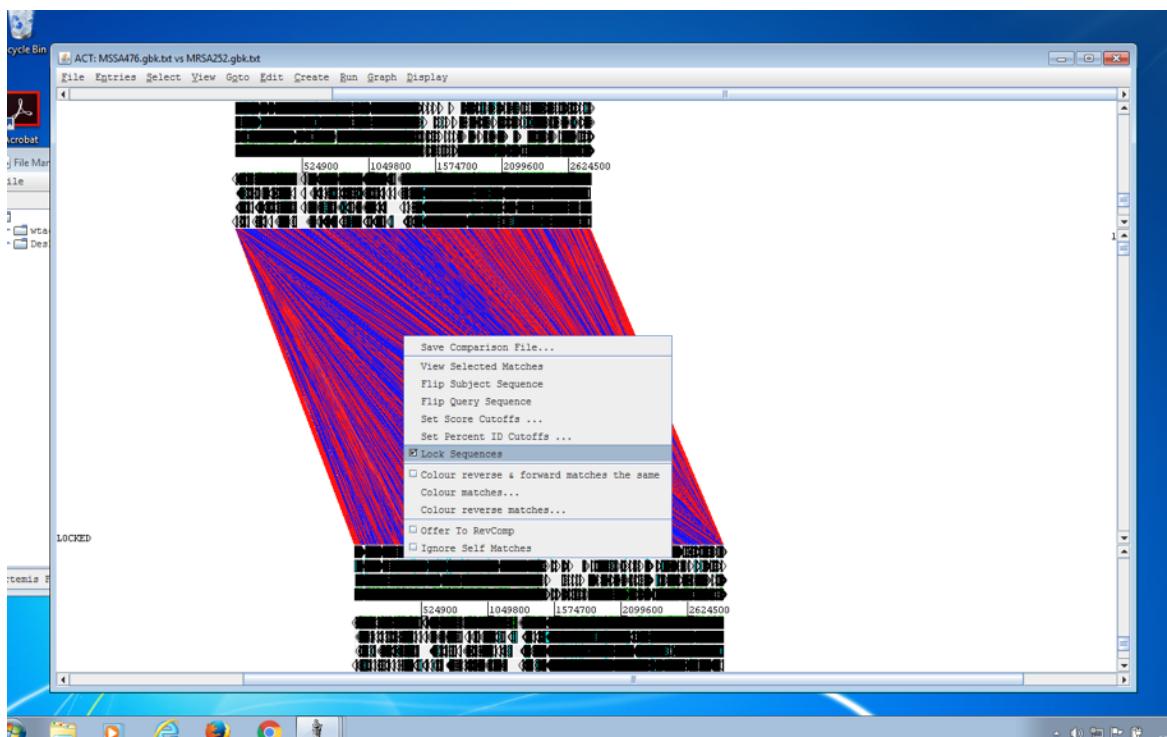
**Step 8: The ACT window is very similar to the Artemis window we used earlier but gives you a comparative view.**

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.

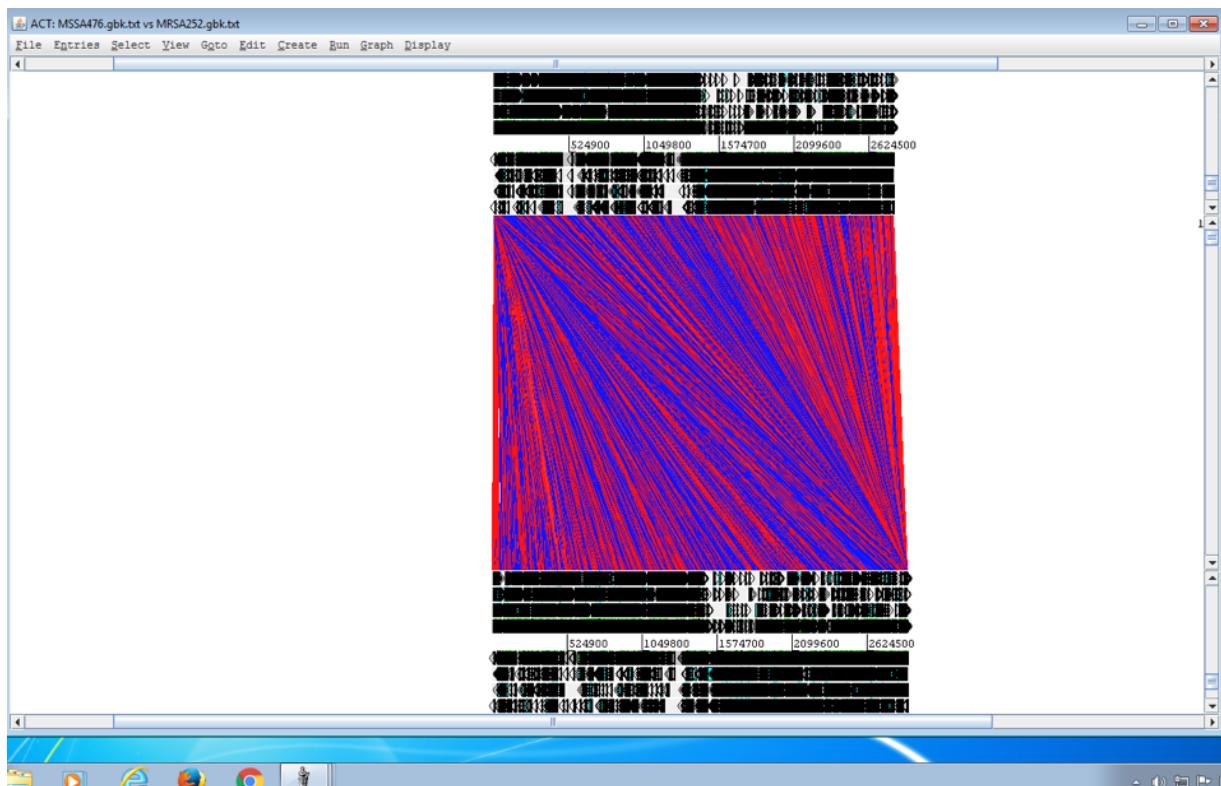
**Step 9:** Use one of the zooming sliders (marked 6) and zoom the view all the way out to the maximum.



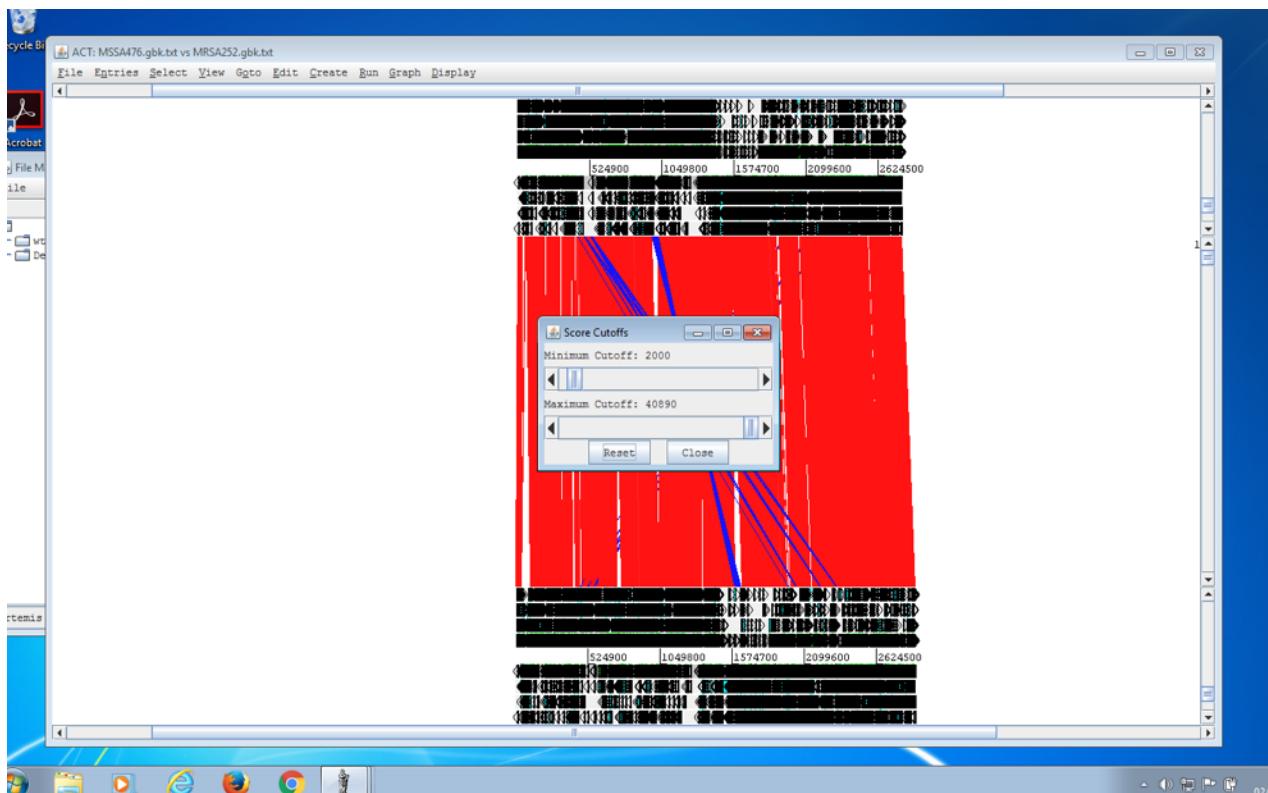
**Step 10:** You will notice here that it says 'LOCKED'. If you use one of the sliders to move down the genome – both genomes will move together. We are going to turn this off so the genomes move independently.



**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'.

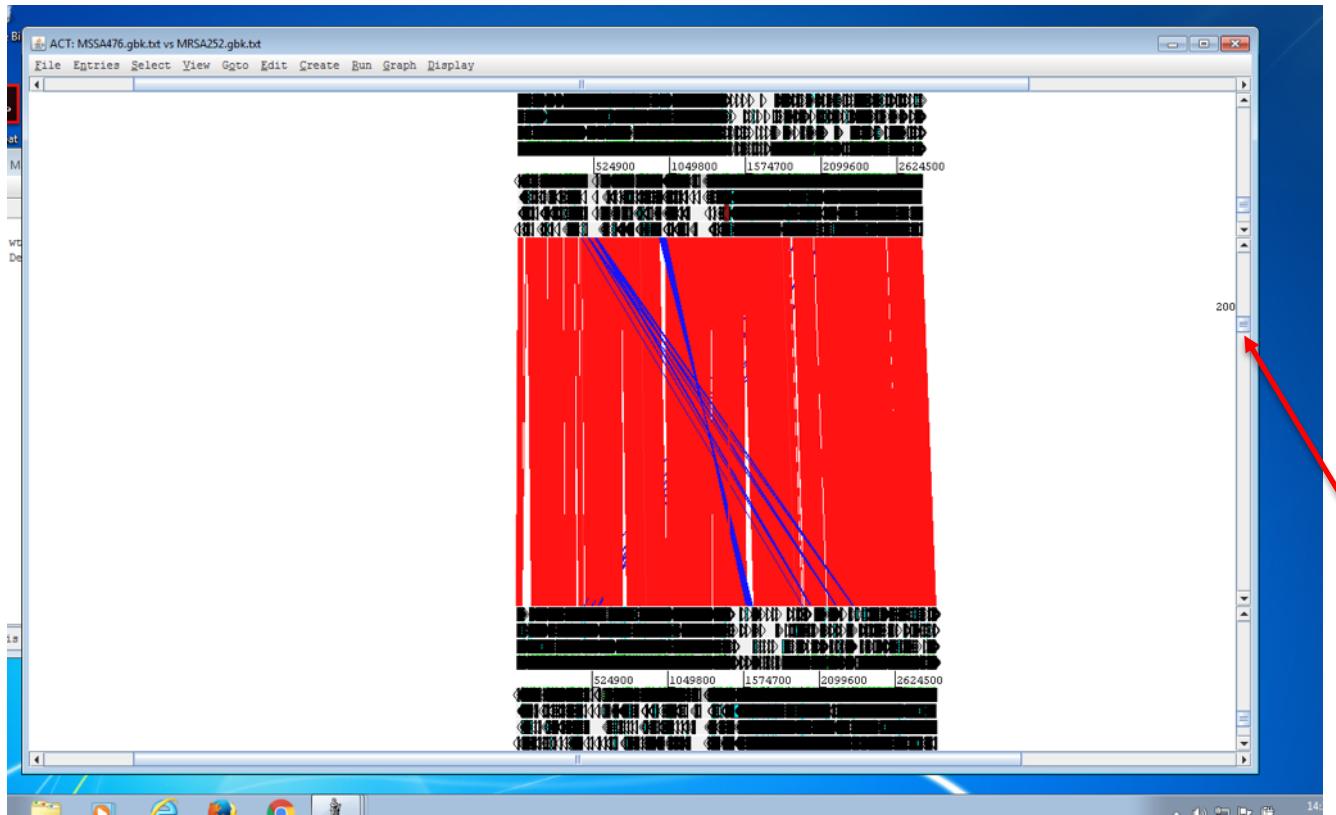


**Step 12:** The sequences will now move independently. Try moving them around yourself. Once you have tried, line them up again so it looks like the picture above.

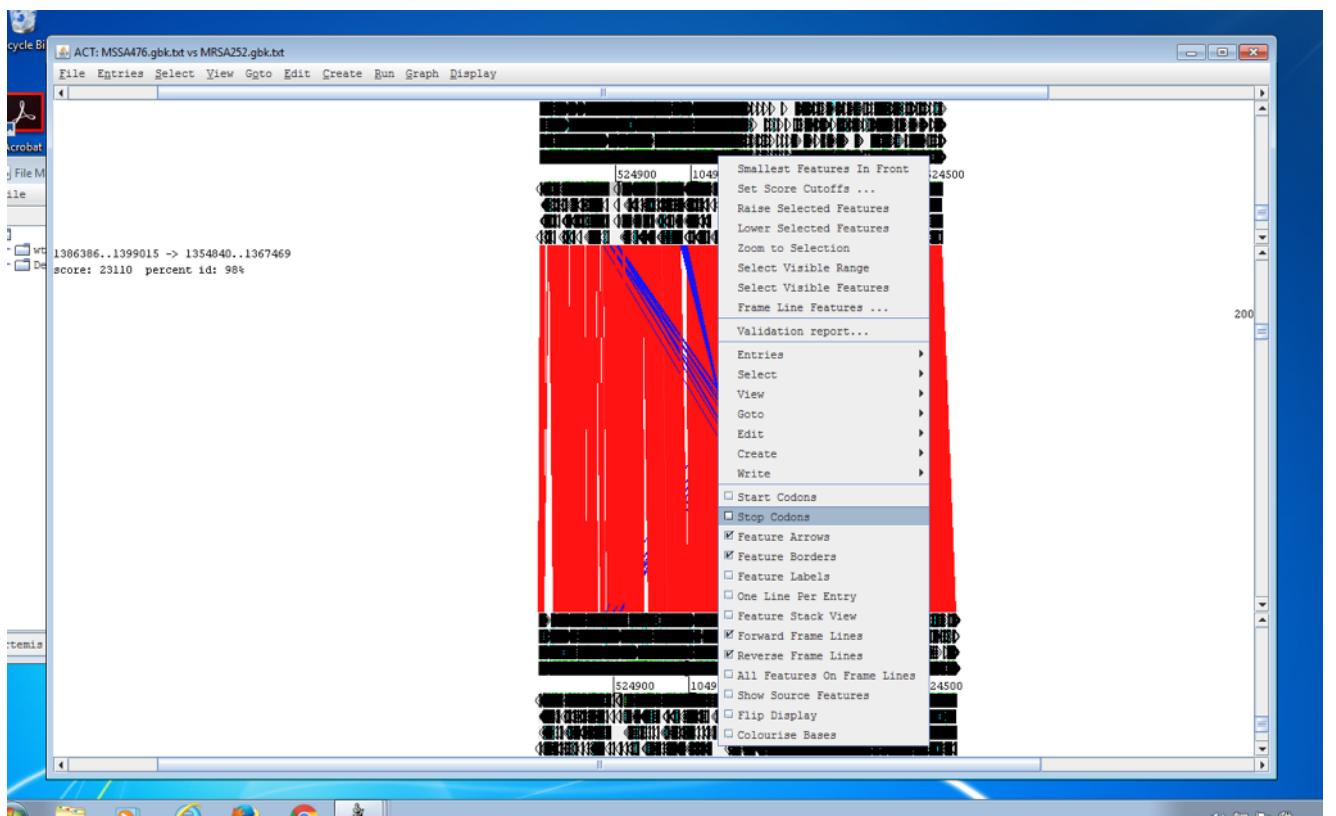


**Step 13:** Now, right click the mouse anywhere in the comparison area (the red / blue). Select the 'Set score cutoff'. Move the upper slider in the Score Cutoff of menu to 2000. This will remove any BLAST hits from the view with a score of less than 2000. Try moving the cutoff up and down to see what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

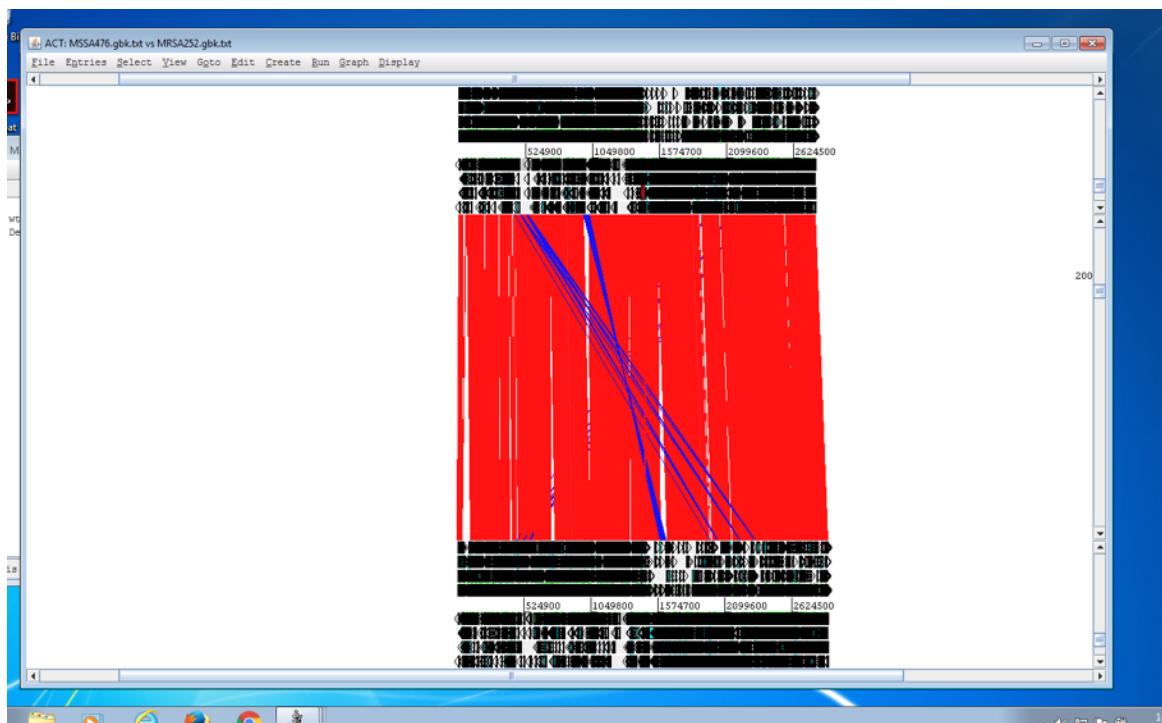
and click anywhere outside in the white to hide the menu or just move it out of the main window.



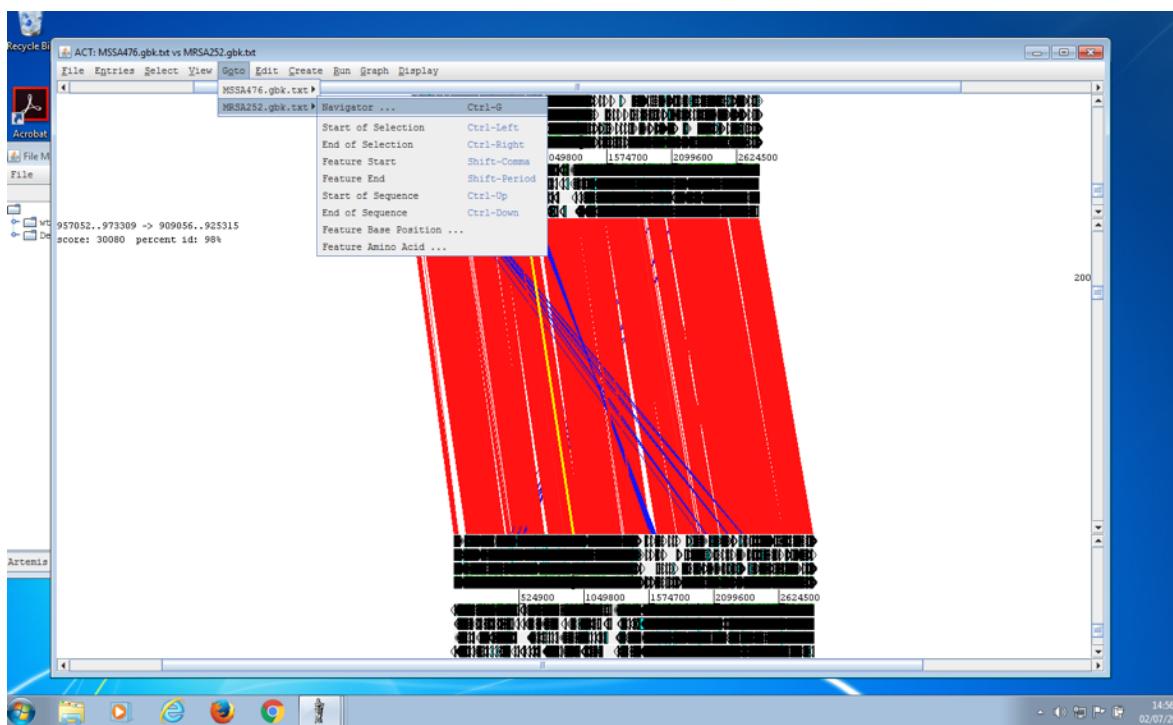
**Step 14:** Now slide the length of the sequence slider to 200. The window will now only show BLASTN hits longer than 200bp.



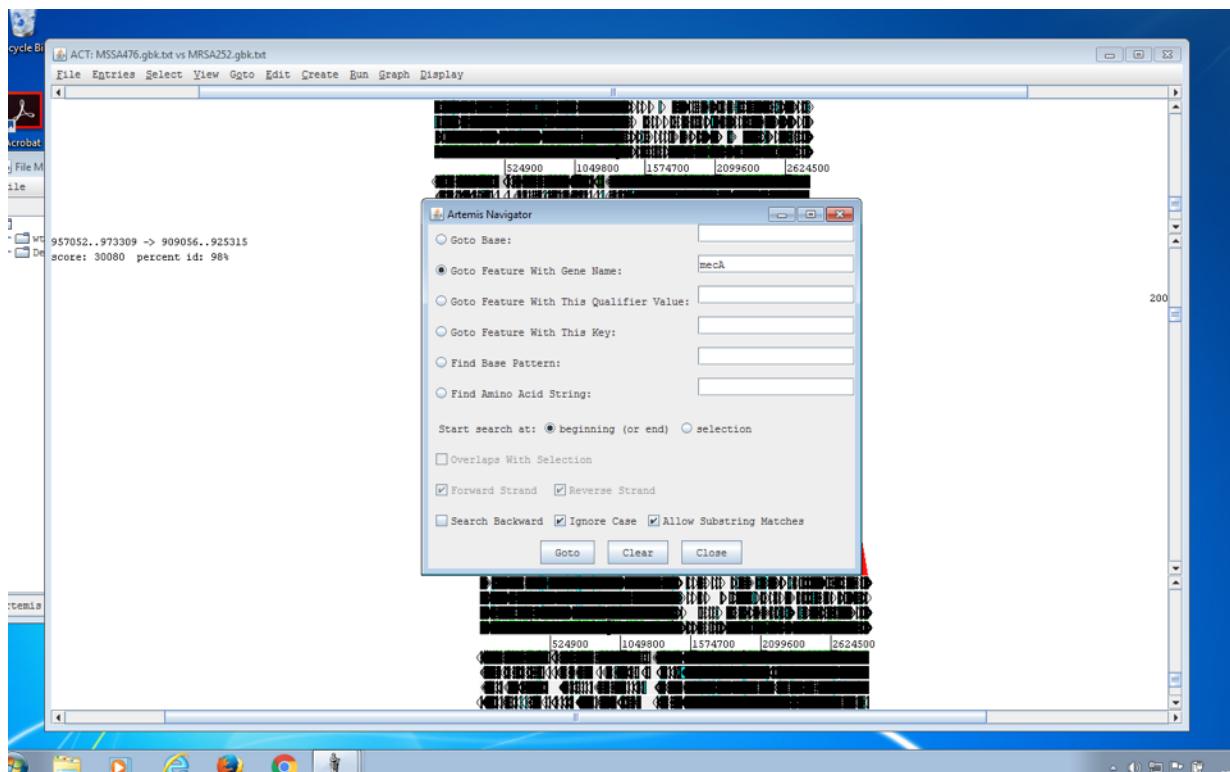
**Step 15:** Next, right click on the sequence viewer for the top sequence (MSSA476) and un-tick the ‘Stop codons’ box. Now do the same for the bottom sequence (MRSA252). We are now ready to explore the sequences.



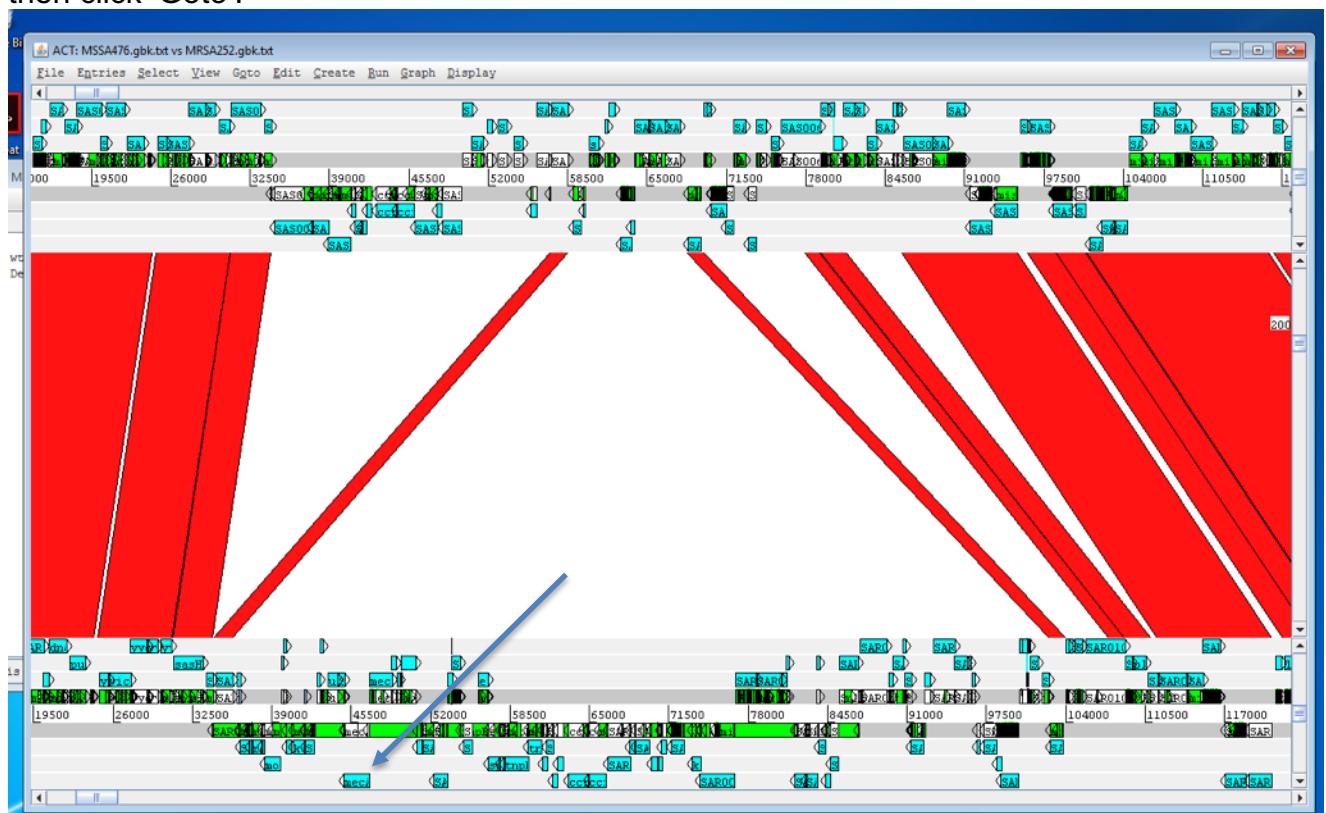
**Step 16:** We can see that the two genomes share a great deal of conservation and synteny (genes in the same order). You can also see a number of regions of difference between the two areas that are white in one genome but present in the other. If you use the slide bar we can zoom in to take a closer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



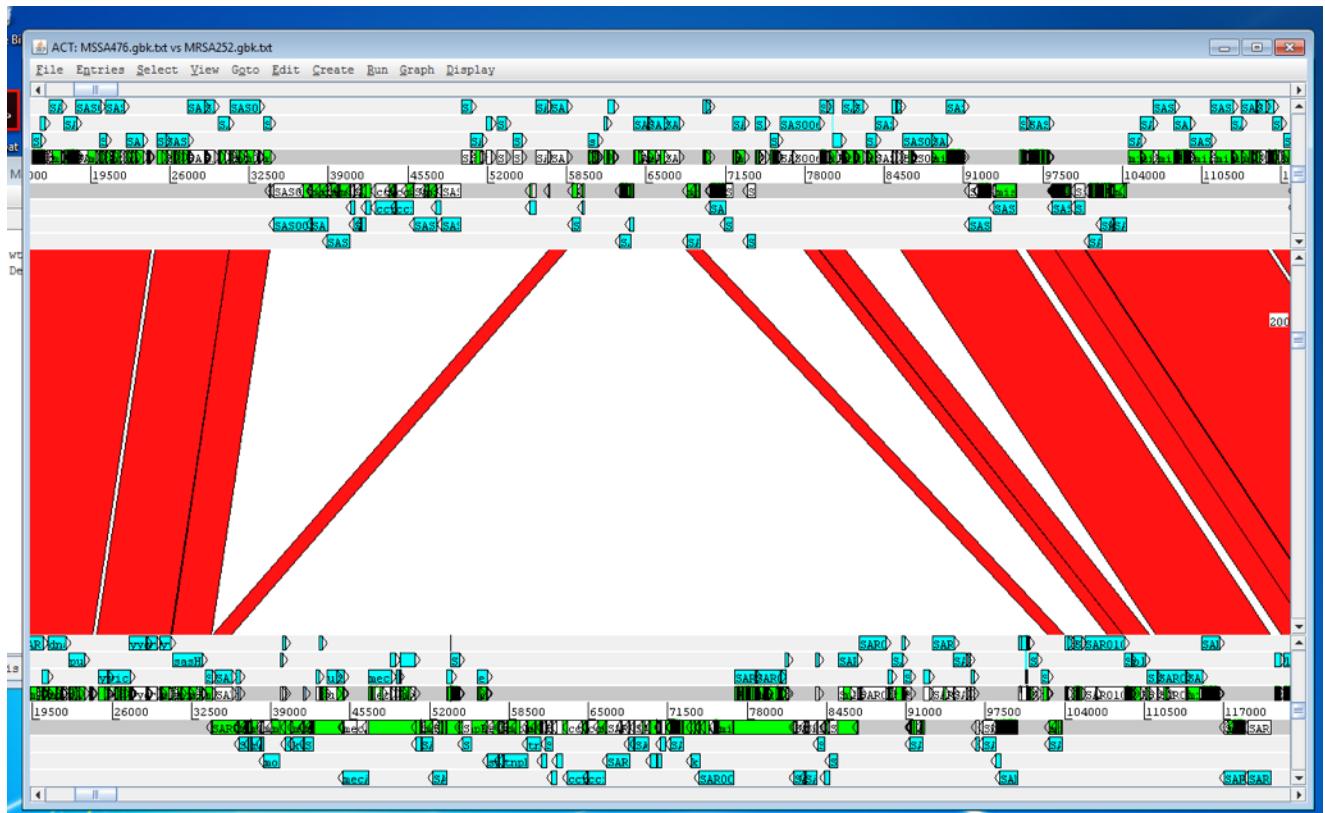
**Step 17:** We are now going to take a look at some of these differences that contain antibiotic resistance genes. Click on the ‘Goto’ menu at the top and select the lower sequence ‘MRSA252.gbk.txt’ and then click ‘Navigator’.



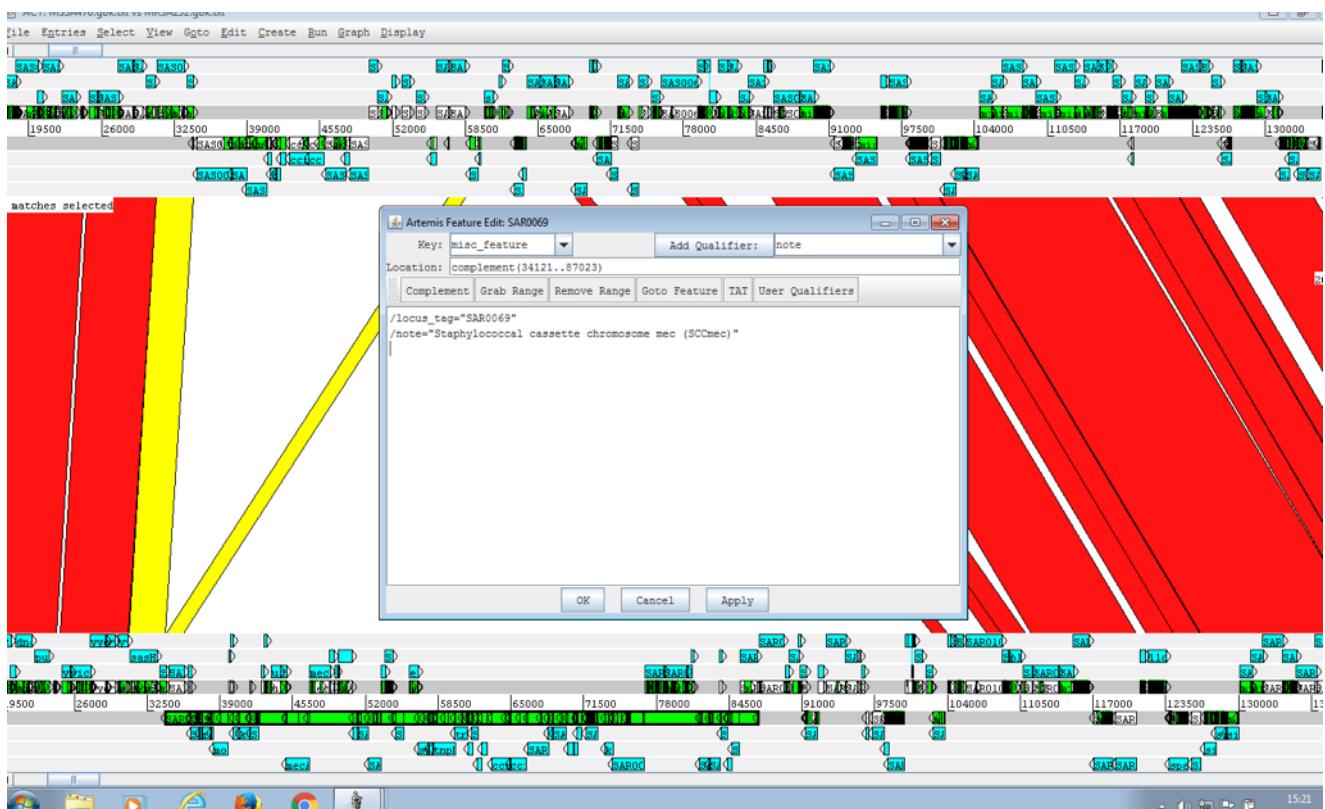
**Step 18:** Click the ‘Goto Feature With Gene Name’ and type ‘*mecA*’ in the box and then click ‘Goto’.



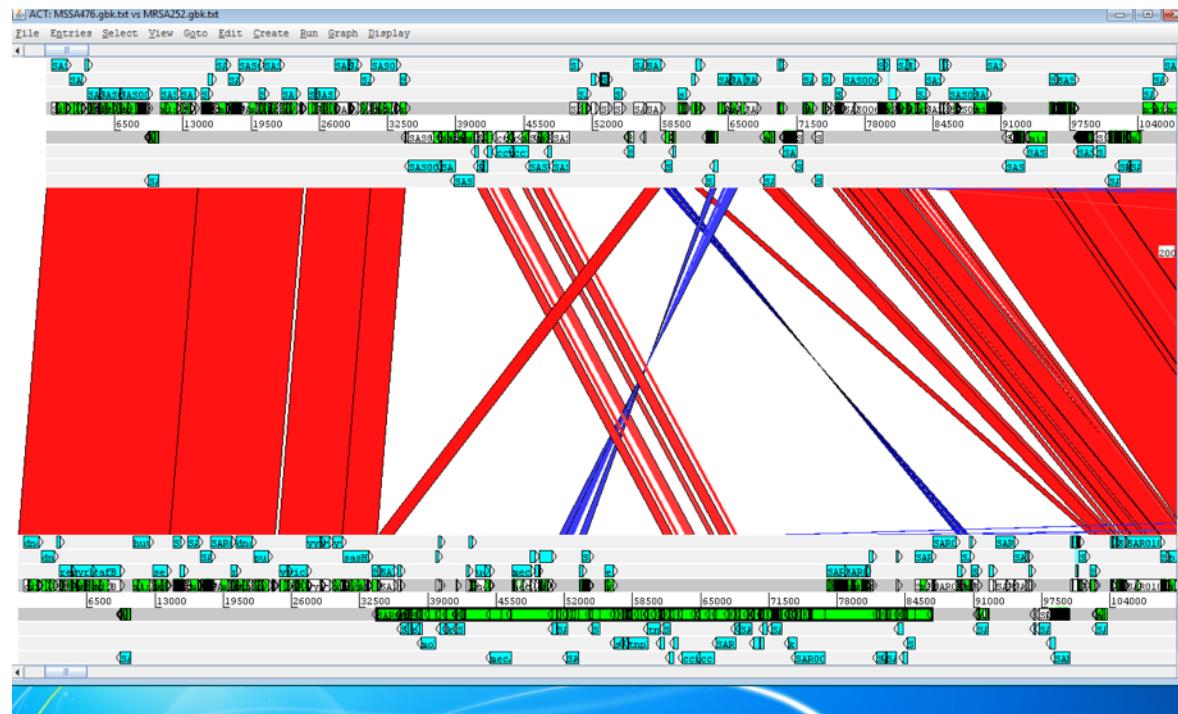
**Step 19:** Using the scrolling bars and zoom move the viewer until the window looks something like this. *mecA* should be highlighted (shown here with the arrows).



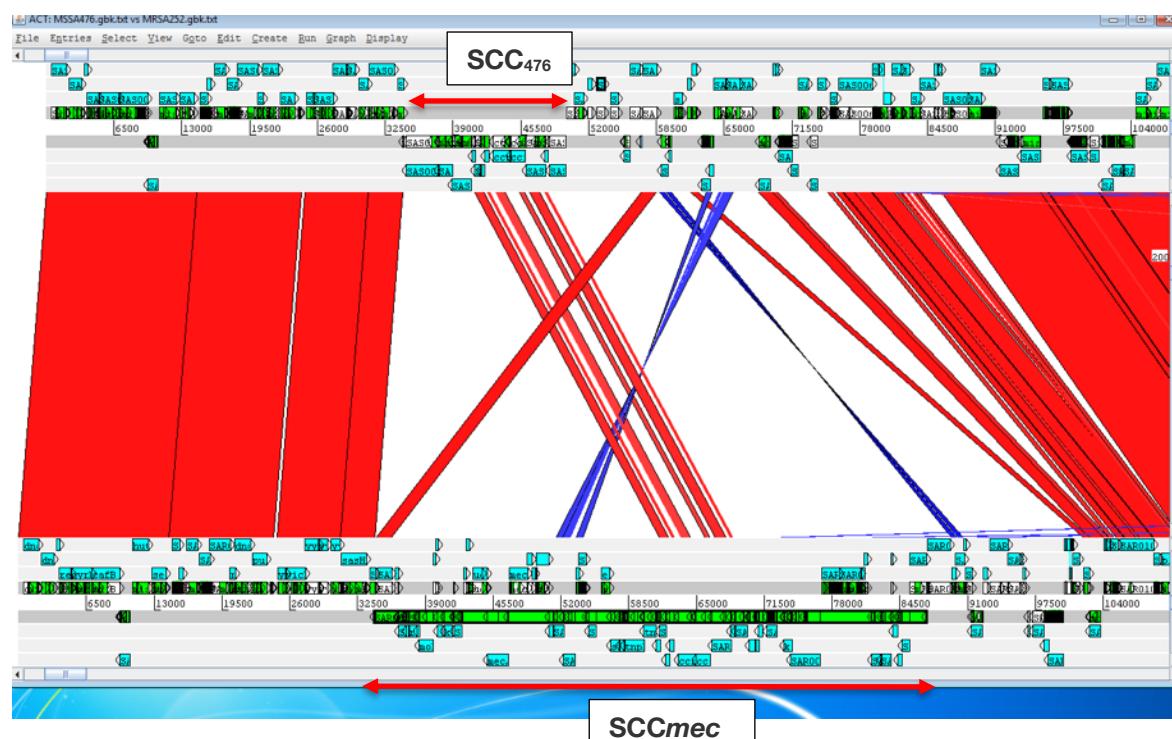
**Step 20:** If you click on the green region in the bottom genome, and then right click and select 'View' and 'Selected feature'.



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

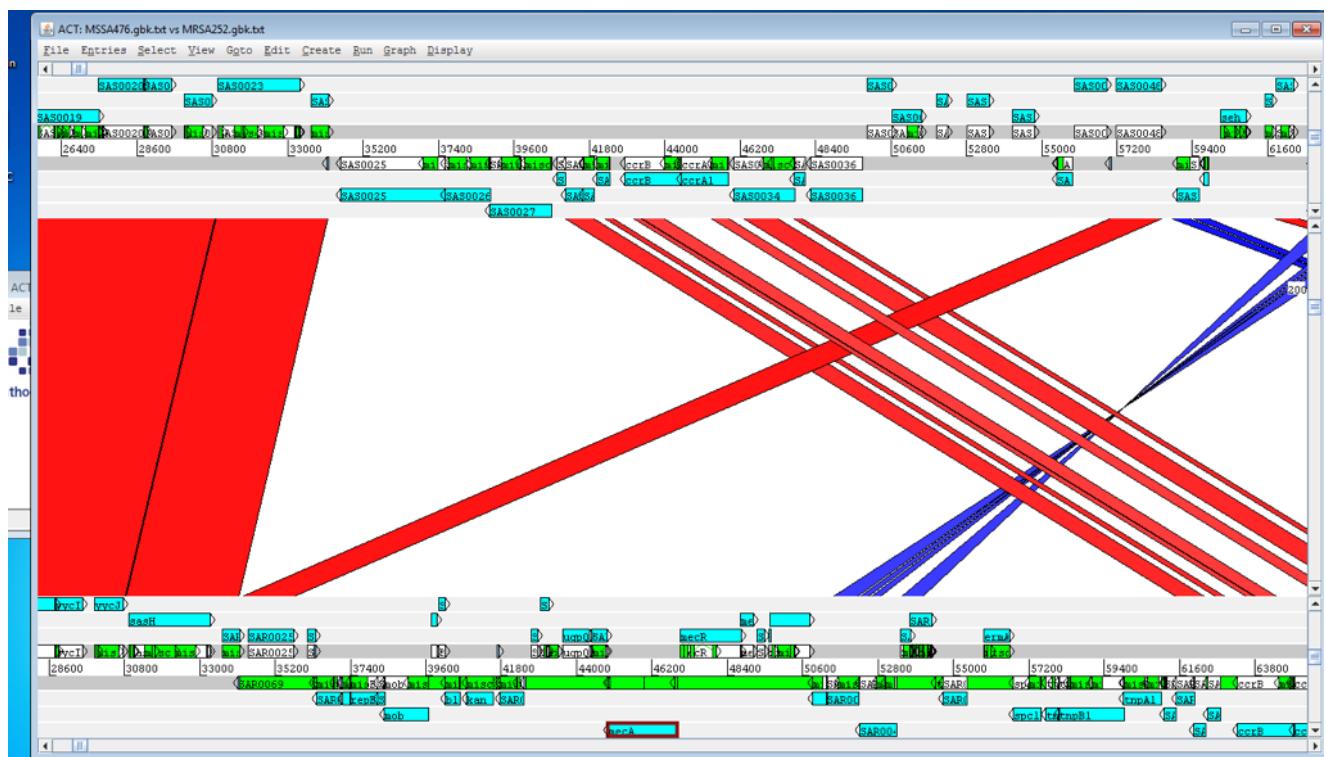


**Step 22:** If right click the mouse anywhere in the comparison area (the red / blue). Select the ‘Set score cutoff’. Move the slider down in the Score Cutoff of menu to back to 0. You should then see the above view. You can now see that in both genomes something has inserted in the chromosome and is 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.

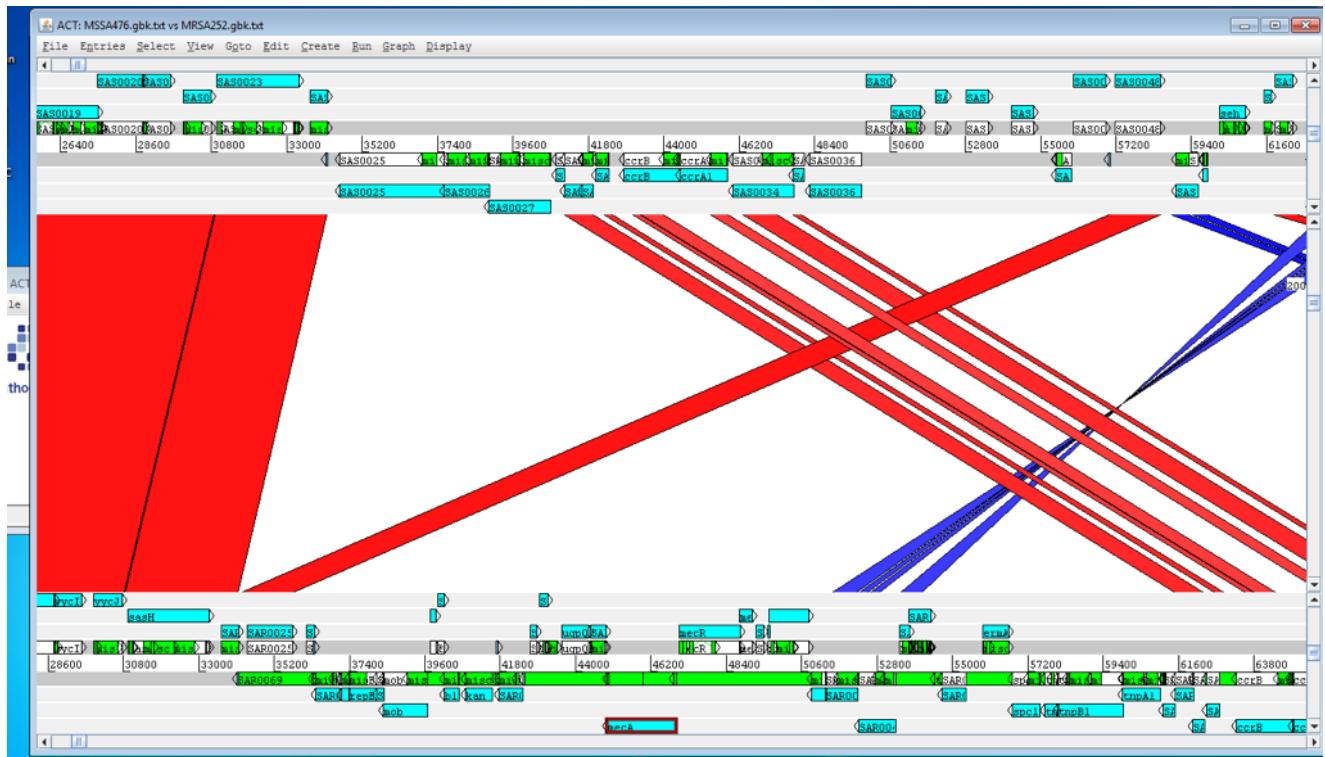


**Step 23:** You can see that there is some degree of similarity between the SCCmec element in MRSA252 (bottom) genome and a region in the MSSA476 (top) genome. This is because MSSA476 has another mobile element of the same family called SCC<sub>476</sub> (see highlighted region above). Click some of 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.

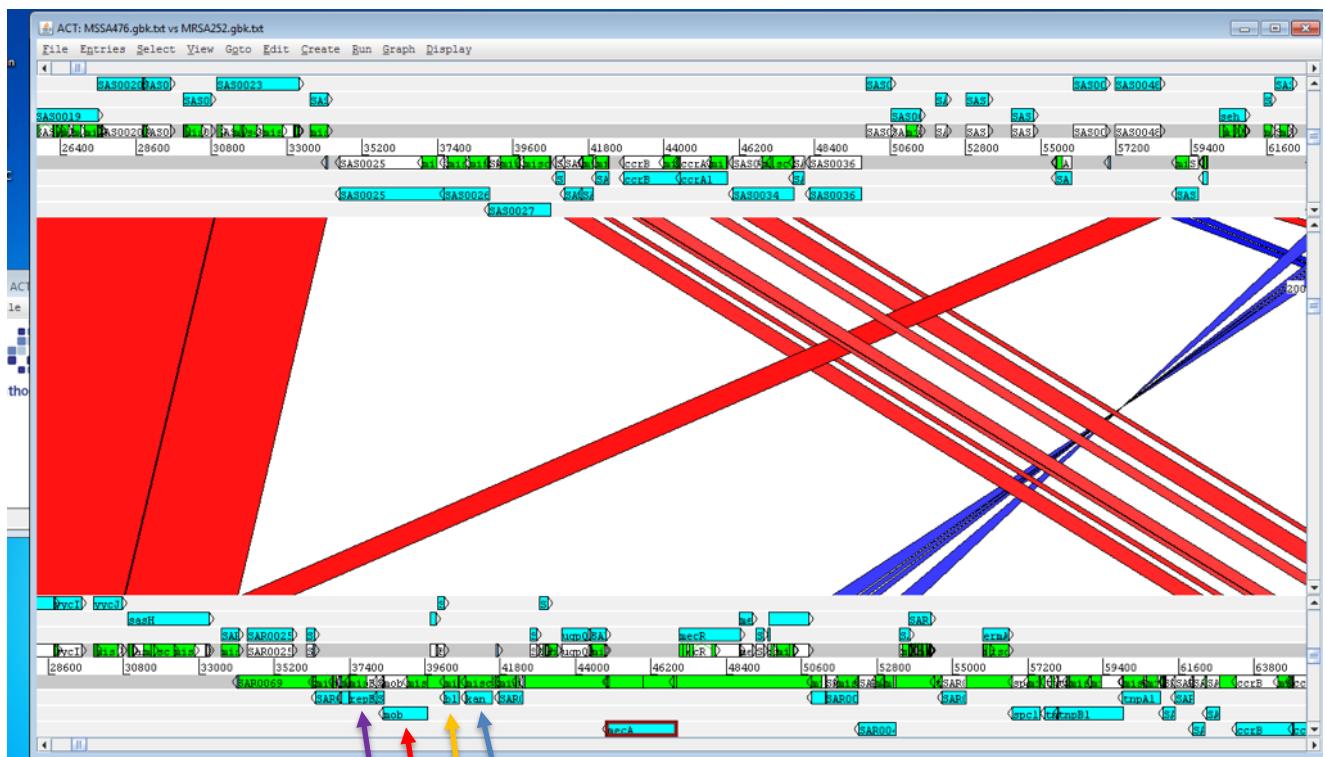
**Step 24:** You should have found that in both genomes – there are some genes annotated as *ccrA* and *ccrB*. These are the site-specific recombinase genes – that mediated the site-specific integration and excision from the genome, enabling horizontal transfer. Finding these genes alongside resistance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



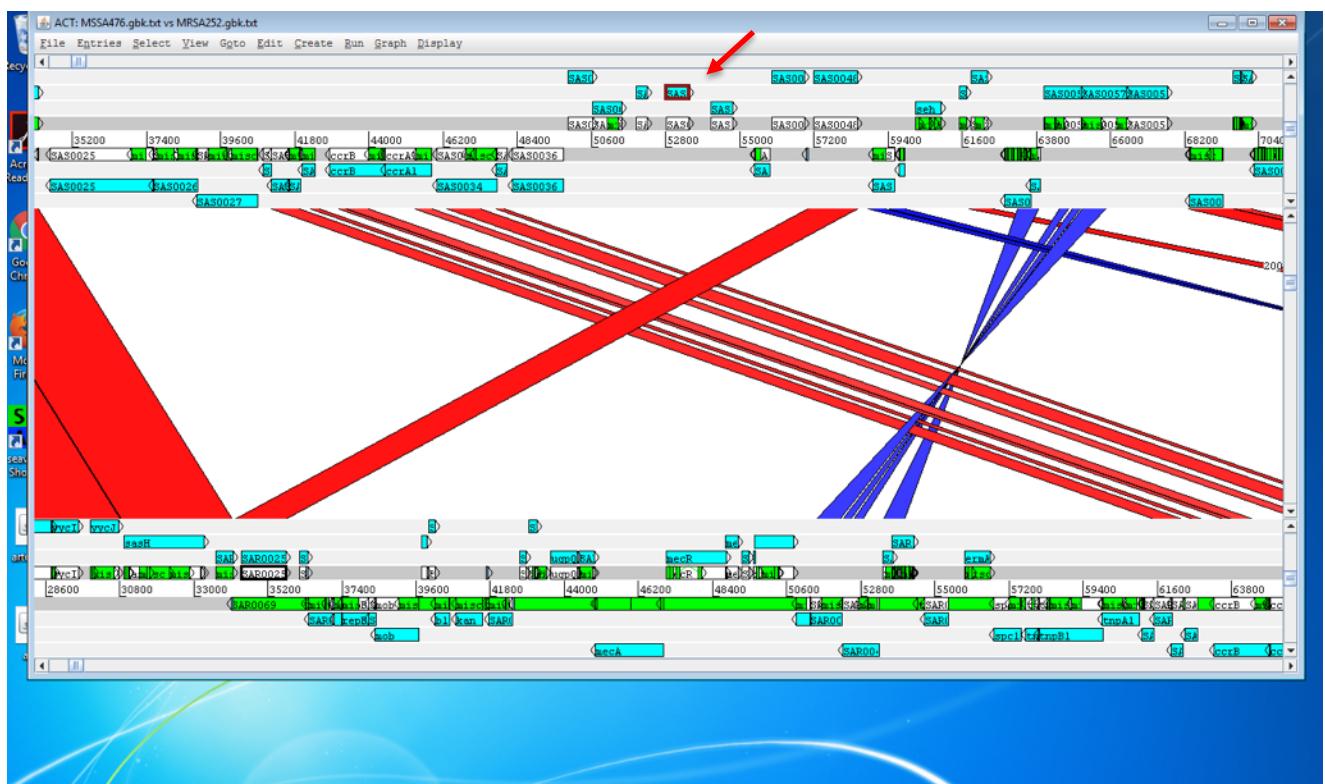
**Step 26:** Next zoom into the region around the *mecA* gene a bit closer. (If you need to find *mecA* again - Click on the 'Goto' menu at the top and select the lower sequence 'MRSA252.gbk.txt' and then click 'Navigator'. Click the 'Goto Feature With Gene Name' and type '*mecA*' in the box and then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



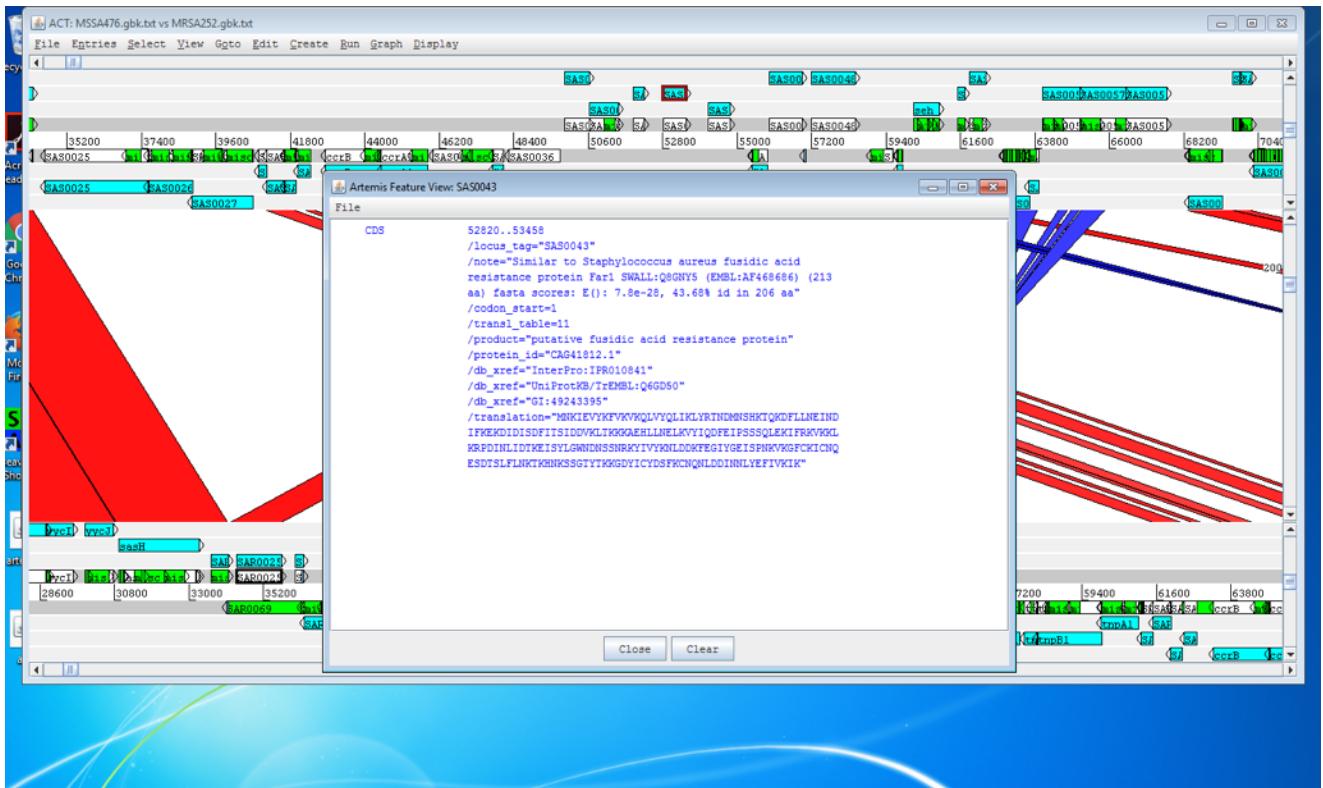
**Step 27:** Try clicking and viewing a few of the genes (use **ctrl-v** or right click – 'View' – 'View selected features') downstream (indicated by arrows – above) of the *mecA* gene. What are they annotated as?

**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.

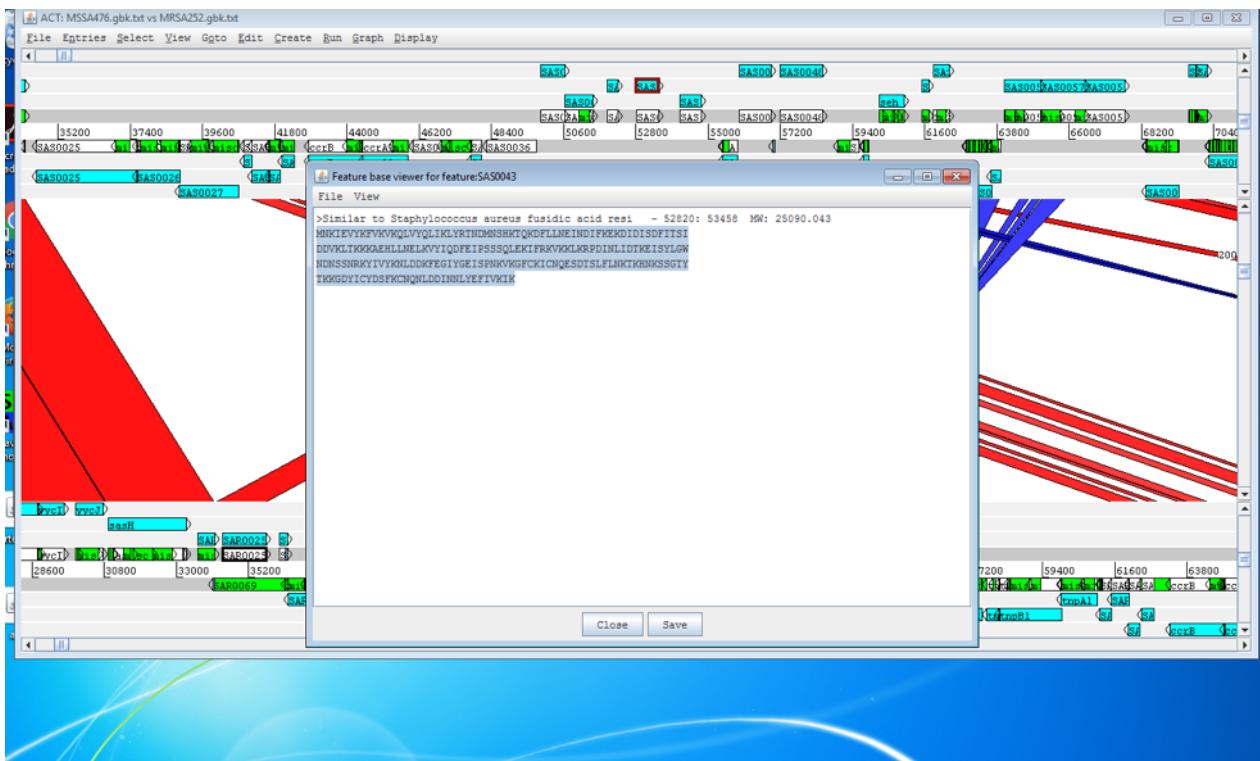
**Step 29:** The discrepancy between what a gene is annotated as and its actual function, is a problem that can occur for various reasons including genomes that were annotated a long time ago, or if a new sequence was annotated by an automated method. In this latter case, we would want to check if the 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.

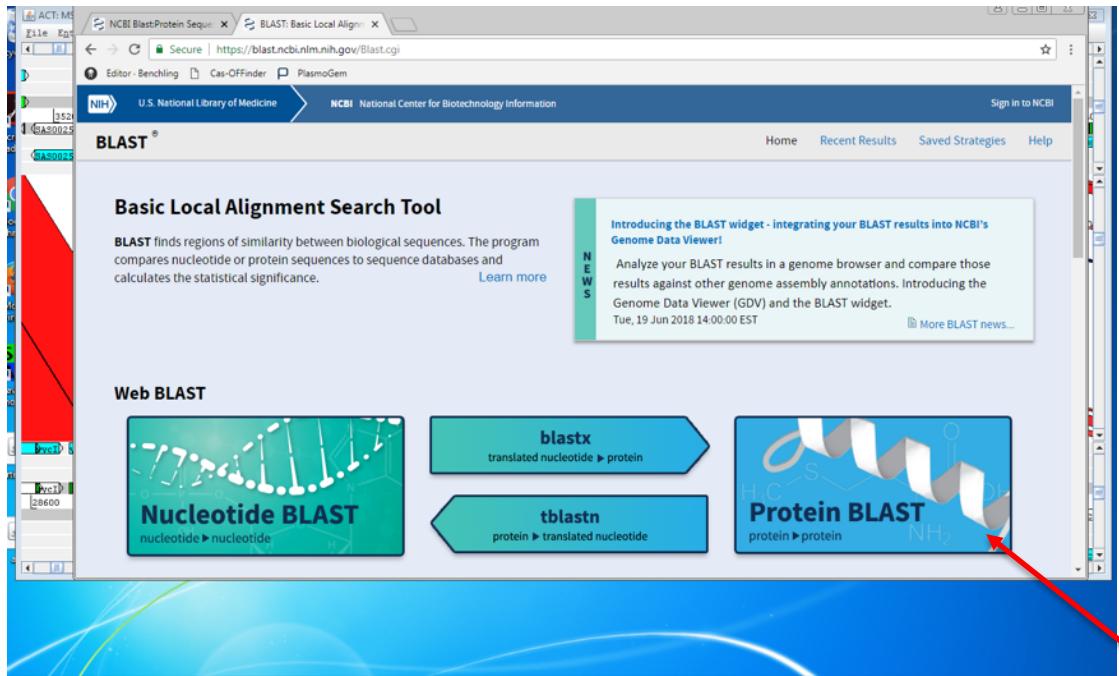


**Step 31:** Use ctrl-v or right click – ‘View’ – Select the upper sequence ‘MSSA476.gbk.txt’ then ‘View selected features’ – As you can see the gene is annotated as ‘Similar to *Staphylococcus aureus* fusidic acid resistance protein’. We will now check this annotation.

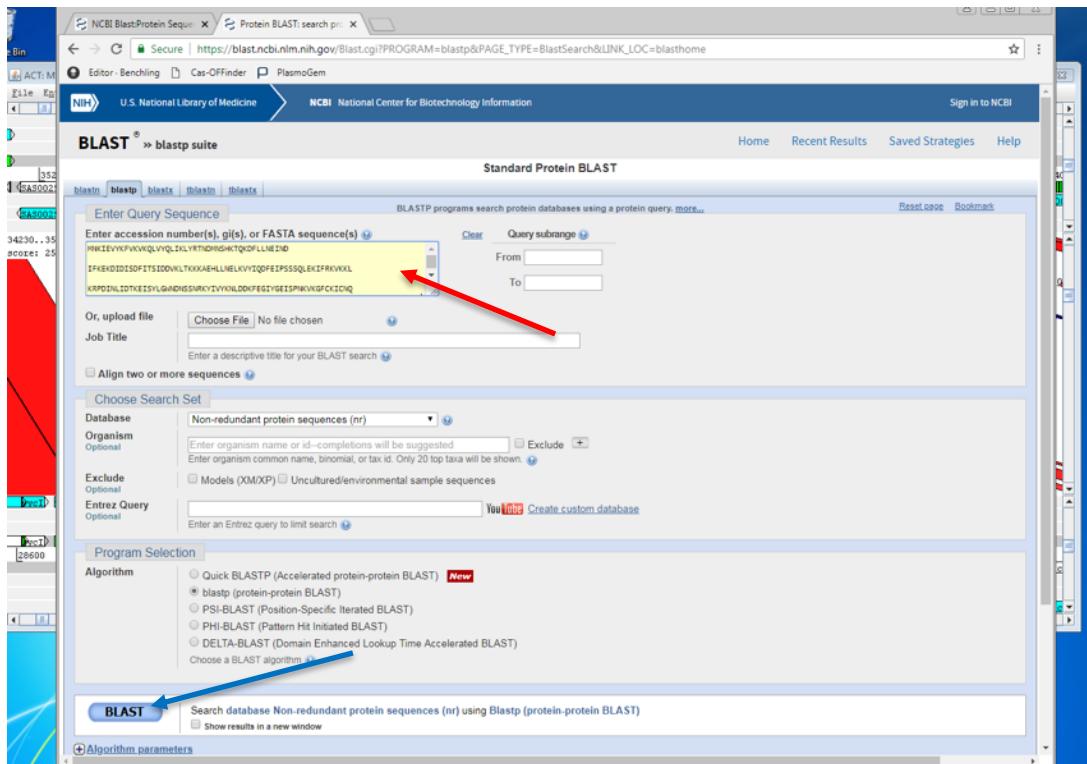


**Step 32:** Now Click on the ‘View’ window and select upper sequence ‘MSSA476.gbk.txt’ ‘Amino Acids’ and then select ‘Amino Acids of Selection As Fasta’

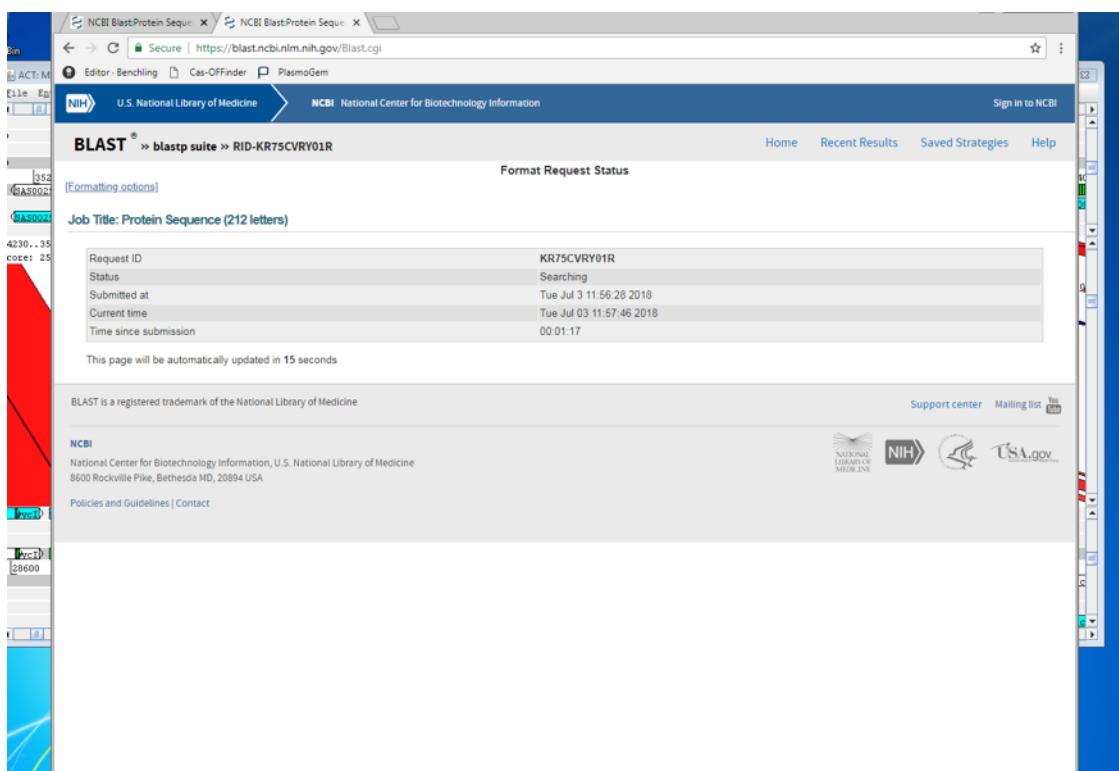
and then highlight and copy the sequence (like shown above) and press **ctrl + C** to copy the amino acid sequence.



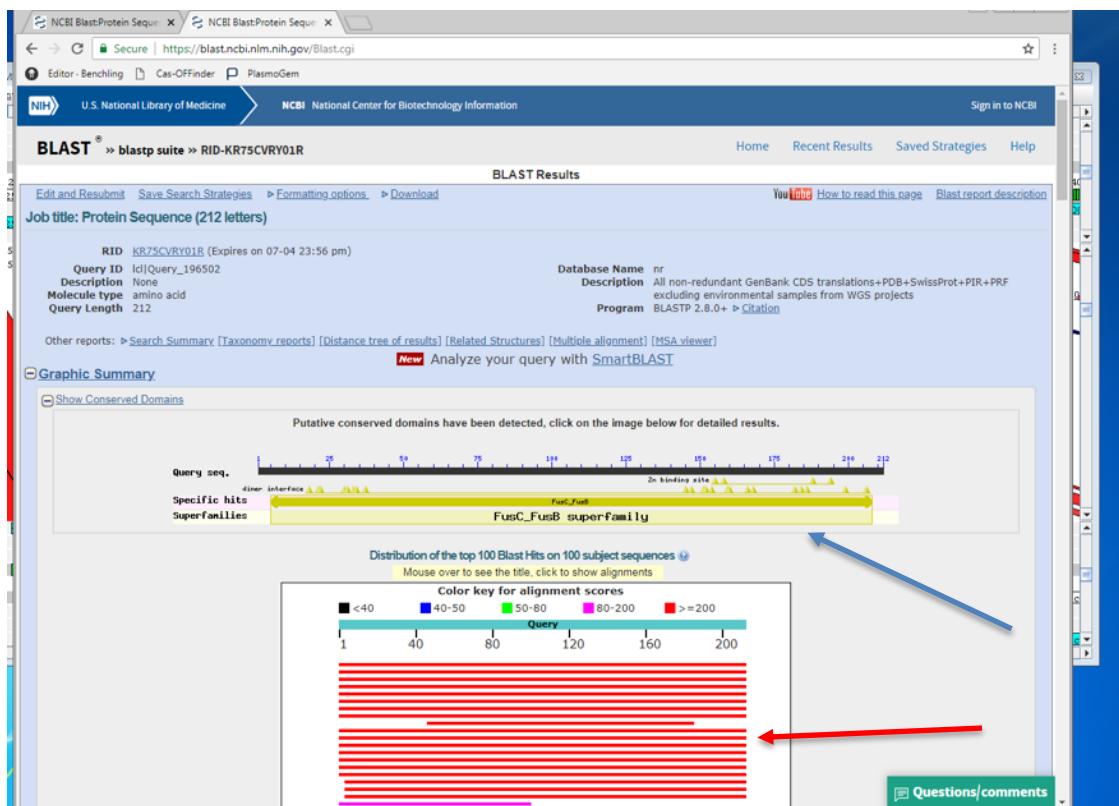
**Step 33:** Now minimise the ACT window and click on the ‘Google chrome’ web browser icon on the desktop. And then type: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> in the web address bar. Then click on ‘Protein blast’ (see red arrow).



**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 '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 by a blue arrow, corresponds to the entry '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%	<a href="#">ADC40001.1</a>
MULTISPECIES: fusidic acid resistance EF-G-binding protein FusC [Staphylococcus]	419	419	100%	5e-148	100%	<a href="#">WP_001033157.1</a>
FusC family: fusidic acid resistance EF-G-binding protein [Staphylococcus cohnii]	419	419	100%	5e-148	100%	<a href="#">ZYB5_A</a>
FusC family: fusidic acid resistance EF-G-binding protein [Staphylococcus haemolyticus]	417	417	100%	3e-147	99%	<a href="#">WP_107386114.1</a>
unnamed protein product [Staphylococcus aureus subsp. aureus]	416	416	100%	7e-147	99%	<a href="#">WP_059747714.1</a>
FusC family: fusidic acid resistance EF-G-binding protein [Staphylococcus aureus]	412	412	100%	3e-145	98%	<a href="#">CCP89052.1</a>
MULTISPECIES: FusC family fusidic acid resistance EF-G-binding protein [Macrococcus]	410	410	100%	8e-145	98%	<a href="#">WP_025176262.1</a>
fusidic acid resistance protein [Staphylococcus haemolyticus]	404	404	100%	2e-142	97%	<a href="#">WP_096077718.1</a>
hypothetical protein [Staphylococcus agnetis]	277	277	65%	6e-93	100%	<a href="#">AY30227.1</a>
hypothetical protein [Staphylococcus agnetis]	265	265	100%	5e-87	60%	<a href="#">WP_095622151.1</a>
elongation factor G-binding protein [Staphylococcus agnetis]	263	263	100%	2e-86	60%	<a href="#">WP_060552383.1</a>
elongation factor G-binding protein [Staphylococcus agnetis]	263	263	100%	2e-86	60%	<a href="#">WP_037566393.1</a>
hypothetical protein [Staphylococcus agnetis]	262	262	100%	4e-86	60%	<a href="#">WP_103346372.1</a>
hypothetical protein [Staphylococcus agnetis]	259	259	100%	6e-85	59%	<a href="#">WP_107391064.1</a>
hypothetical protein [Staphylococcus agnetis]	258	258	100%	2e-84	59%	<a href="#">WP_105994924.1</a>
hypothetical protein [Leptospirillum halophilum]	250	250	100%	2e-81	55%	<a href="#">WP_092594841.1</a>
hypothetical protein [Macrococcus sp. IME1552]	238	238	98%	2e-76	55%	<a href="#">WP_096076399.1</a>
hypothetical protein [Macrococcus goetzii]	225	225	98%	2e-71	57%	<a href="#">WP_099578357.1</a>
hypothetical protein [Macrococcus caseolyticus]	220	220	98%	2e-69	56%	<a href="#">WP_101035271.1</a>
fibronectin-binding domain protein [Staphylococcus aureus subsp. aureus 21304]	194	194	47%	1e-60	100%	<a href="#">EZ190619.1</a>
hypothetical protein [Streptococcus carabensis]	189	189	91%	2e-57	51%	<a href="#">WP_103346991.1</a>
Fibronectin-binding protein (FBP) [Streptococcus carabensis]	189	189	91%	2e-57	51%	<a href="#">PNY119232.1</a>
elongation factor G-binding protein [Enterococcus faecium]	183	183	98%	7e-55	46%	<a href="#">WP_002343616.1</a>
elongation factor G-binding protein [Enterococcus faecium]	183	183	98%	9e-55	46%	<a href="#">WP_002316946.1</a>
elongation factor G-binding protein [Enterococcus faecium]	183	183	98%	9e-55	46%	<a href="#">WP_053542686.1</a>
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: ▾

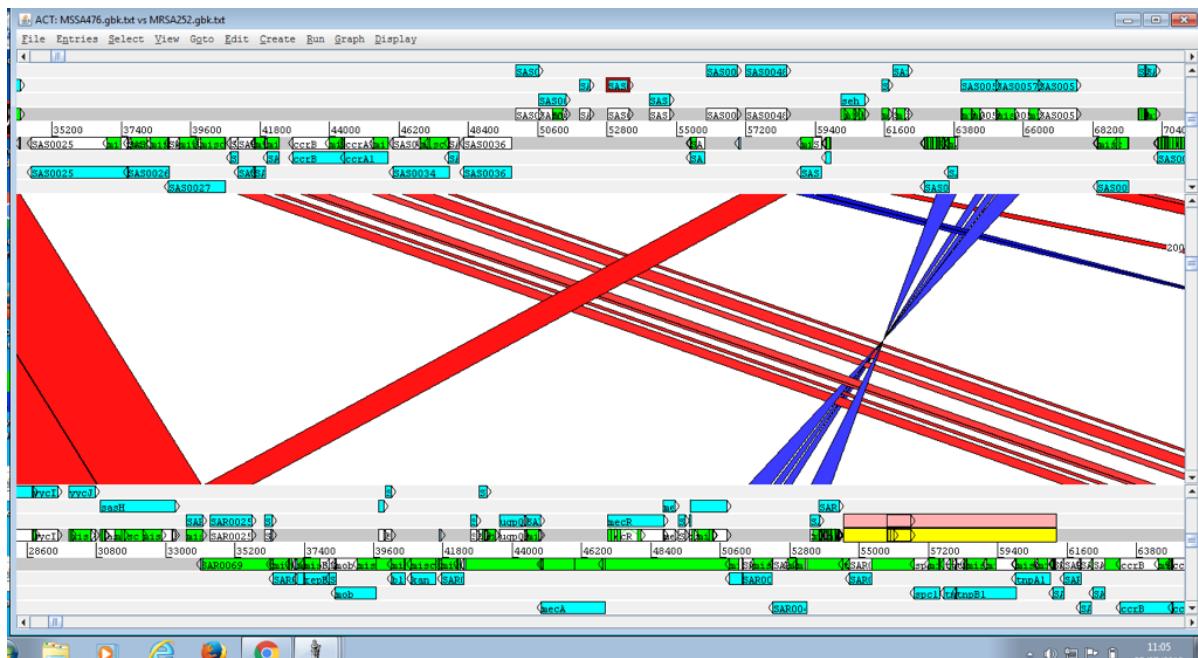
Locus: 2YB5\_A 215 aa linear BCT 23-SEP-2015  
Definition: Chain A, Putative Fusidic Acid Resistance Protein.  
Accession: 2YB5\_A  
Version: 2YB5\_A  
DBSOURCE: pdb: molecule 2YB5, chain 65, release Sep 23, 2015; deposition: Mar 1, 2011; class: Translation; source: Mmdb\_id: 96621, Pdb\_id 1: 2YB5; 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 III: Investigating multidrug resistance

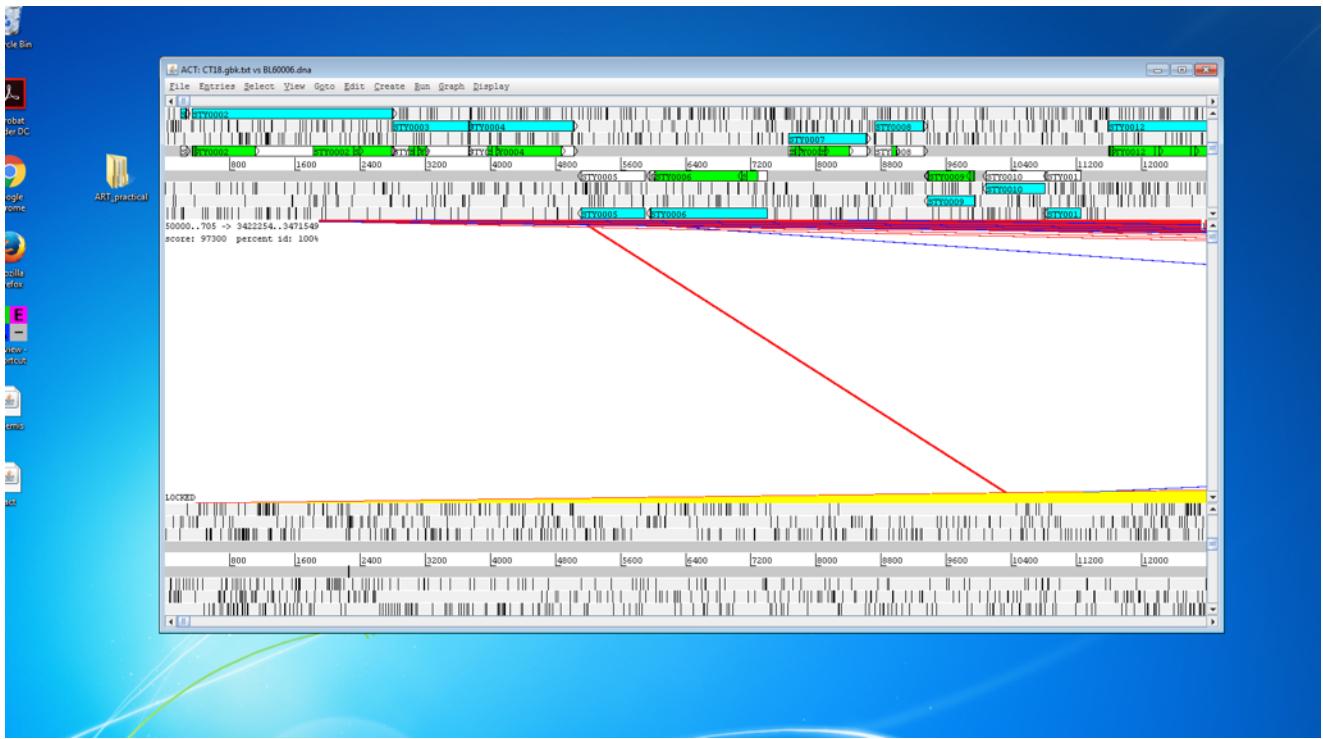
In the previous two parts of this practical you have been looking at completed genomes – that is genomes that have been completely sequenced, aligned into a single chromosome and manually annotated (that is each gene annotation checked manually). But if you generate your own data you will probably end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

In this practical you are going to analyse one the genomes of from the pretend Typhoid outbreak you have been investigating. The isolate you are looking at is a *Salmonella typhi* called BL0006 that comes from a recent outbreak of multidrug resistant typhoid in Pakistan. You are going to use what you have learned to find the different mechanisms for resistance that you identified as been present in Computational Practical 2.

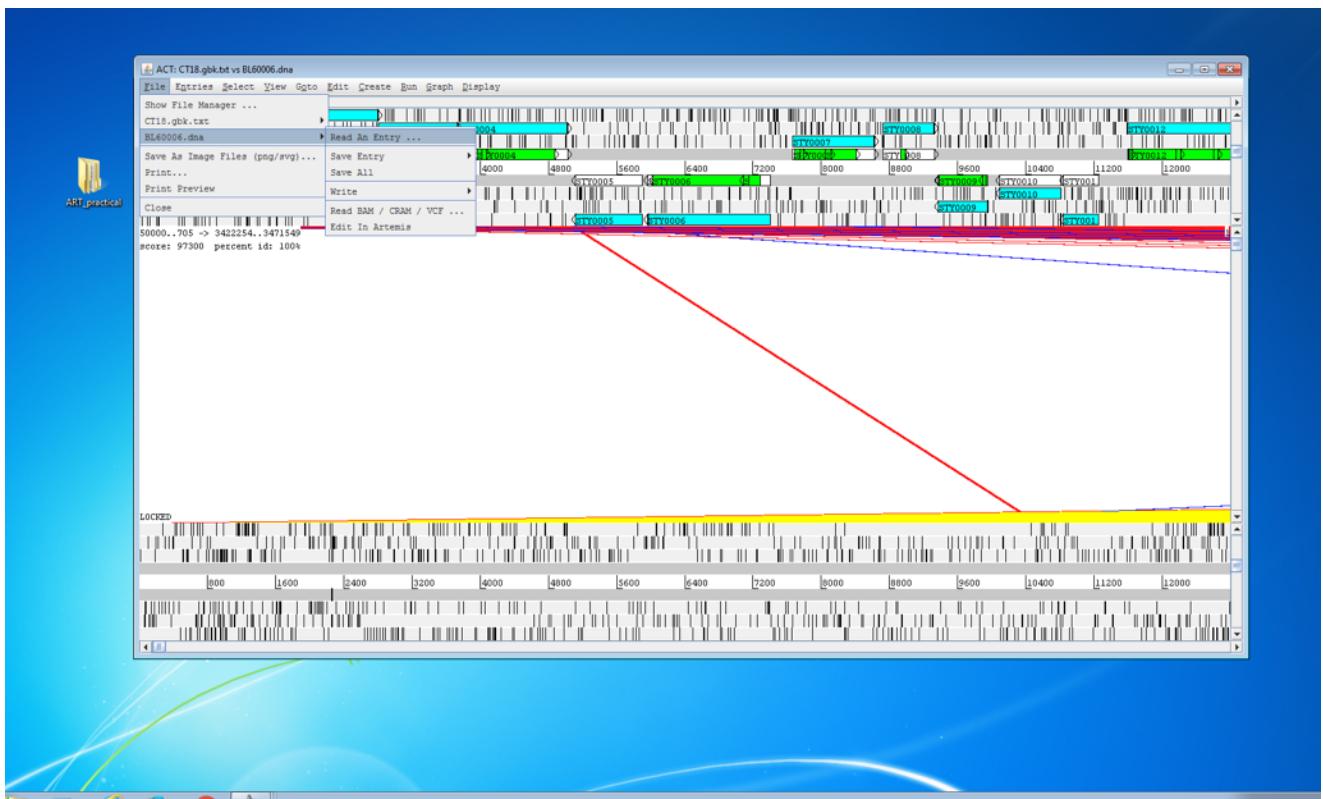
**Step 1:** We are now going to look at BL0006 in comparison to another *Salmonella typhi* reference genome called CT18. Open ACT as before and click ‘File’ and then select ‘Open’. A box will open like the one shown above. For Sequence file 1 – click the ‘Choose ...’ box. Navigate to ~/course/ directory and then select ‘cp14’ folder.

**Step 2:** The folder contains four files, a folder called Annotation (don’t worry about this now we will use this later), the two genome sequences (BL6006.dna, CT18.gbk.txt ) and a blast results file of the two genomes (CT18\_vs\_BL6006\_blast\_results.txt). Click CT18.gbk.txt and click ‘open’.

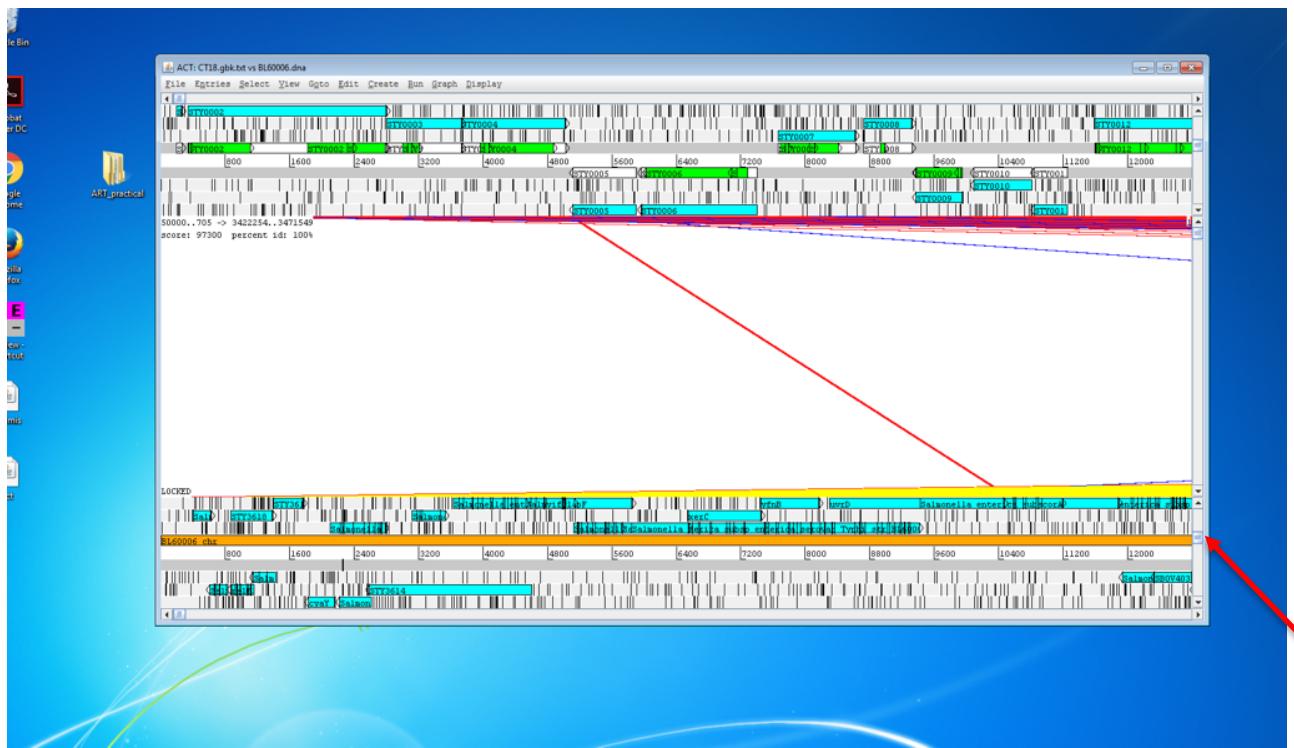
**Step 3:** Now click the ‘Choose ...’ button for ‘Comparison file 1’ and then select CT18\_vs\_BL6006\_blast\_results.txt and click ‘open’. Now do the same for Sequence file 2 selecting the file: BL6006.dna. Now press the ‘Apply’ button. A series of messages will appear asking if you want view errors. Just click yes and okay.



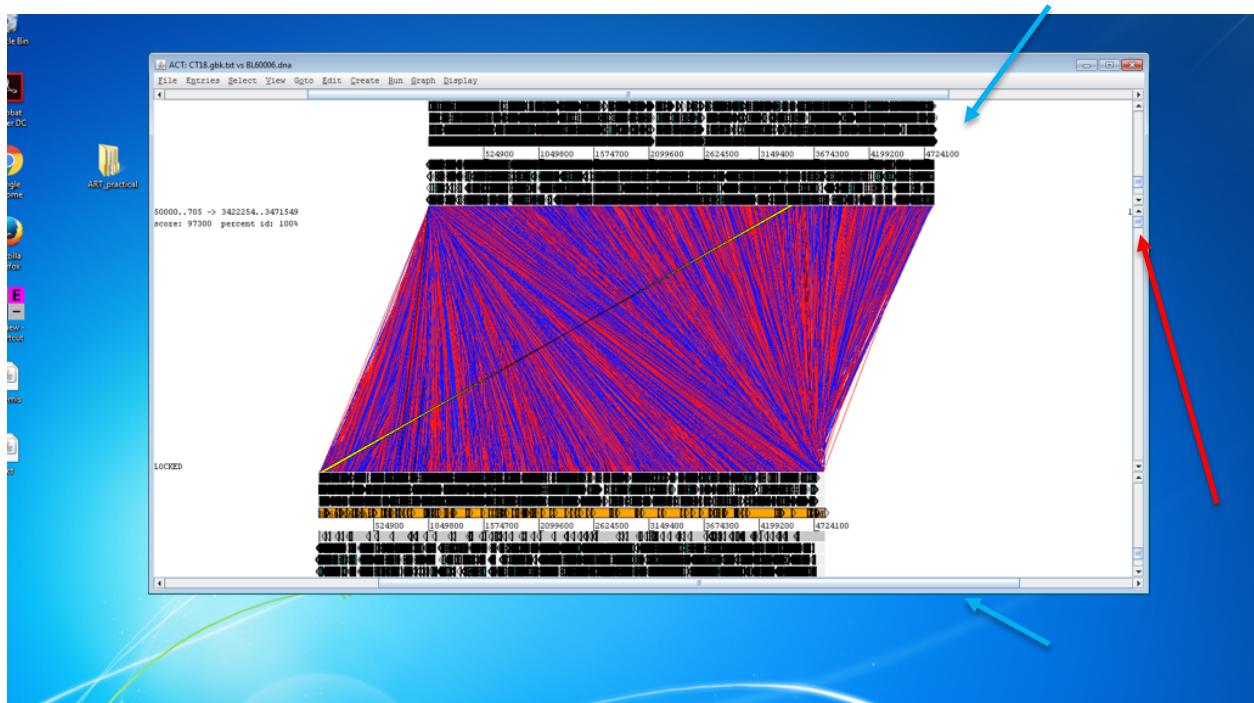
**Step 4:** You should now have a view like the one above. As you can see the bottom sequence contains no annotation information – unlike for the reference genomes like CT18 we now need to add an extra file to view this.



**Step 5:** Now go to the ‘File’ menu and select ‘BL60006.dna’ and then select ‘Read an Entry’. The ‘cp14’ folder should now open and you should click the ‘Annotation’ folder. Inside the folder you should find a file called: ‘Salmonella\_enterica\_subsp\_enterica\_serovar\_Typhi\_str\_BL60006\_v1.1.gff’ click on this and then click ‘open’.

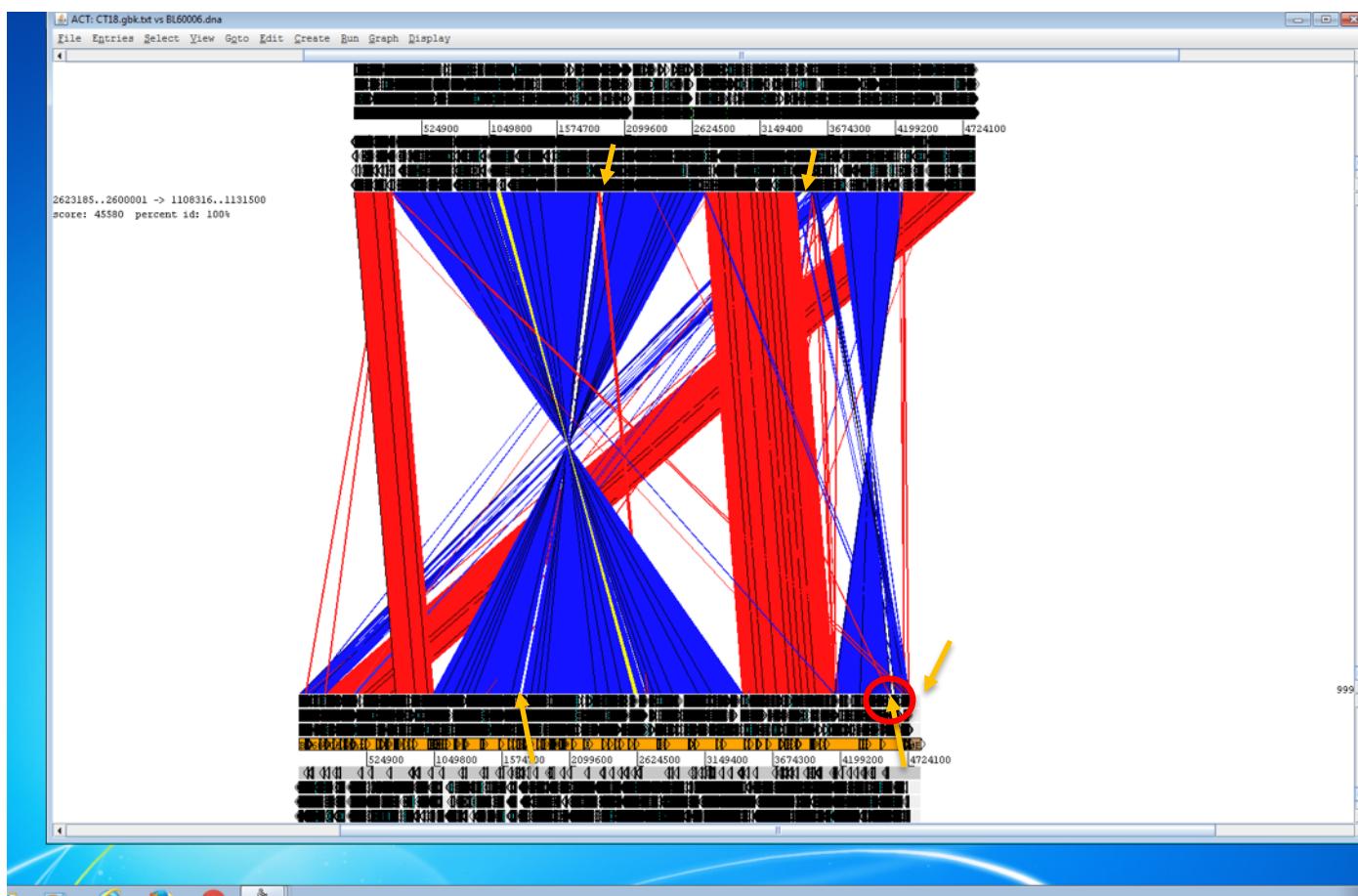


**Step 6:** You should now have a view like the one above with the annotation present for the BL6006 genome (the lower genome). Next, right click the mouse anywhere in the comparison area (the part where the red / blue lines appear). A menu should appear. On the menu un-tick the box that says 'Lock sequences'. Now zoom out using one of the sliders indicated by the arrow above.

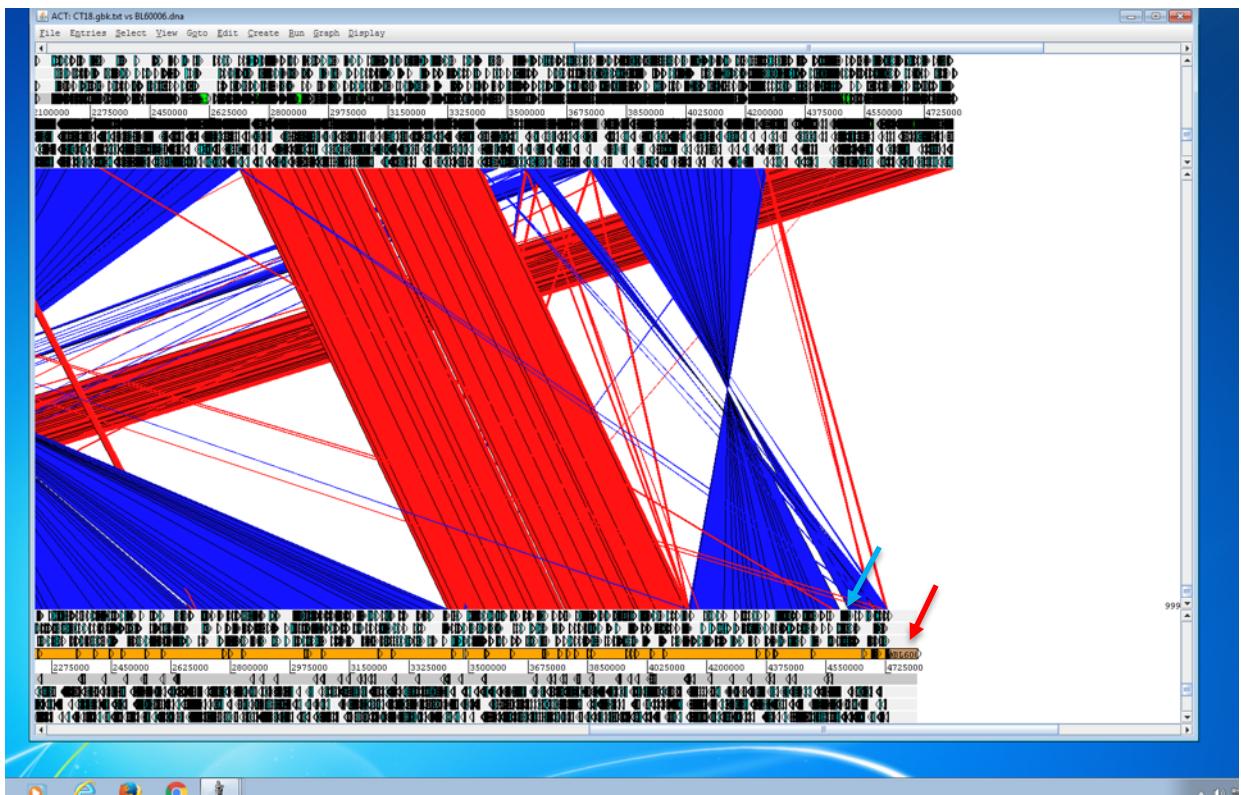


**Step 7:** Adjust the view so using the sliders (blue arrow indicates them above) so you should now have a view like shown above. As you can see there is a lot of red and blue lines indicating BLAST matches. For what we want to do this is too much

detail. Now slide the length of the sequence slider to 999 (indicated with the red arrow). The window will now only show BLASTN hits longer than 999bp.



**Step 8:** You should now have a view something like the one above. If you look across the genome you can see most parts are conserved – the blue sections indicate that these parts are inverted. This is quite common in bacterial genomes. You can see that there are only a few locations where there are gaps (see - orange arrows). If you remember these indicate that this DNA is unique in that genome (e.g. the other genome doesn't have it). As lots of resistance genes are horizontally transferred, we are going to focus our search into these regions. You might have noticed that a very big block of DNA is unique to BL60006 at the right hand (3')- end of the genome, this is close by another block of unique DNA (in the red circle). We are going to focus our attention on this block of unique DNA. Zoom into the region indicated by the red circle.



**Step 9:** You should now have a view like the one above. Now click on the brown arrow right at the end of the BL60006 sequence (indicated with the red arrow) and using the 'View' menu select 'View selection'. This will tell you that this part of the sequence is called pBL60006. This section of the file is not actually chromosomal DNA - it is separate plasmid called pBL60006 that is included in the file so you can view it easily, we will come back to the plasmid later. We are now going to look at the block of unique DNA that is just upstream from this (indicated by blue arrow). Zoom into this region and take a closer look at the genes that are present. (The genome location is: 4591121..4557538 – so if you have trouble navigating there use: 'Goto -> BL6006.dna -> 'Navigator' -> 'Goto Base:' and type 4591121.



**Step 10:** You should now have a view like this. Using what was taught to you in the previous practical - can you find any resistance genes? Based on your analysis which antibiotics should this strain be resistant to?

**Step 11:** Using what you learned earlier can you find out if the strain is resistant to ciprofloxacin? (note: the resistance mutation is S83F in GyrA) – also is strain CT18 resistant?

**Step 12:** Again, using what we learned earlier –open a new Artemis (not ACT) window and navigate to the ~/course/cp14/Annotation folder and open the file: ‘*Salmonella\_enterica\_subsp\_enterica\_serovar\_Typhi*\_str\_BL60006\_v1.1.gff’. Now navigate to the brown plasmid sequence (at the far-right end of the genome – we looked at this above). Zoom in to the region around genome coordinates 4792793..4806592 (you can use ‘Select’ -> ‘Base Range’ and then input the coordinates to highlight the region of interest).

Can you find any resistance genes here? What strikes you about this region in comparison to the chromosomal region you were just looking in Step 8? How do you think this block of DNA got into the plasmid (e.g. can you find any genes associated with mobile genetic elements?)