Using the Artemis Genome Viewer:

Goals:
- To visualize genome records and genome features of interest using the Artemis genome viewer
- To select genes of interest from the genome annotation and create a new file for their independent visualization

General notes on using Artemis
"Artemis is a free genome viewer and annotation tool that allows visualization of sequence features and the results of analyses within the context of the sequence, and its six-frame translation. Artemis is written in Java, and is available for UNIX, Macintosh and Windows systems. It can read EMBL and GENBANK database entries or sequence in FASTA or raw format. Extra sequence features can be in EMBL, GENBANK or GFF format." (from the Sanger site)

Because Artemis requires Java (with more recent versions of Artemis requiring more recent versions of Java), following this tutorial requires access to a computer that either has Java preinstalled or onto which a user can download and install Java.

Once Java is installed, users have two options for visualizing genomes in Artemis:

1. If users have "Java Web Start (included in more recent versions of Java, such as Java SE 6), Artemis can be remotely launched without the need to download the program. Sequences to be viewed can be specified using the Genbank, RefSeq or EMBL accession number or read from local files.

2. If Java Web Start is not available, Artemis can be downloaded and run locally. As in Option #1, sequences to be viewed can be specified using the Genbank, RefSeq or EMBL accession number or read from local files.

The instructions given below are for option number 1. If this option proves unworkable, other approaches will be attempted.

Additional resources can be found at the following sites:

Sanger - Main Artemis page
http://www.sanger.ac.uk/Software/Artemis/

Sanger - Artemis manual
http://www.sanger.ac.uk/Software/Artemis/v10/manual/artemis_manual_complete.html

Sanger - Artemis examples page including a link to an online demo
http://www.sanger.ac.uk/Software/Artemis/Examples/

Pseudomonas-Plant Interaction - Artemis tutorial for PPI users

(representing an expanded version of the tutorial below)
http://www.pseudomonas-syringae.org/artemis_tutorial.htm

1. Starting up with Artemis:

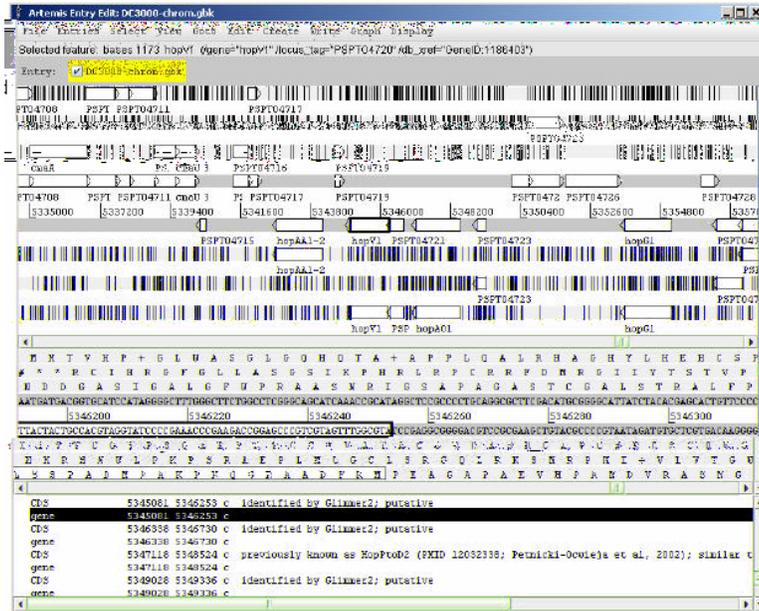Go to: http://www.sanger.ac.uk/Software/Artemis/v10/

Click on the LAUNCH ARTEMIS button. The following window will appear:
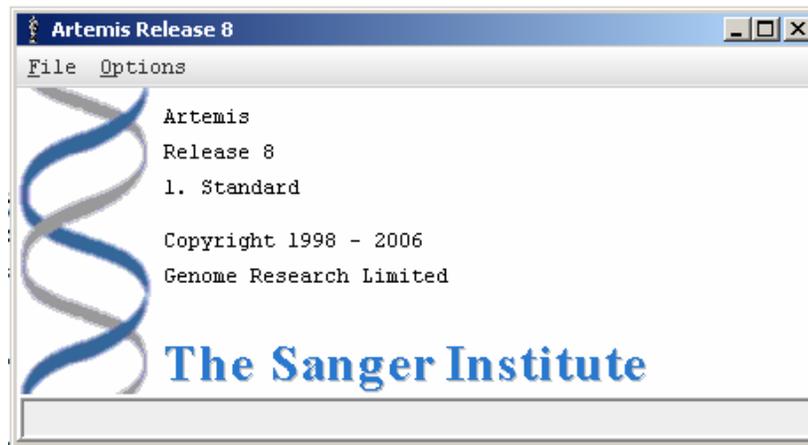
Go to File>Open

Navigate to the location of the genome sequence file on your computer

A window similar to the one shown will appear (this may take a minute or two and colors may vary):
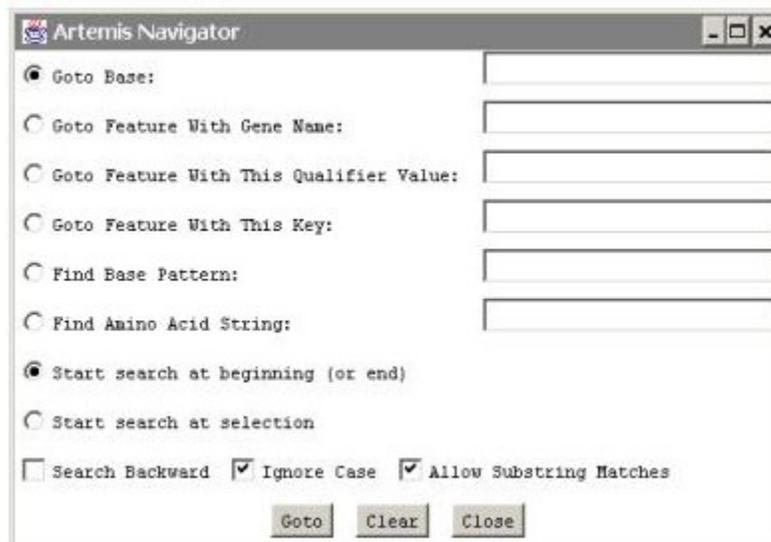
2. A few basic commands for using Artemis

Changing the window size:
Clicking on the scroll bars to the side of the Overview window will expand or reduce the size of the genome region that you are looking at.

To go directly to features of interest:

Bring up the Navigation window using Goto>Navigator or Ctl-G. Enter a search term and click on Goto

<u>Selection and navigation</u>
Double clicking on a feature will select it in all three windows (hold down the shift key to select more than one feature).

Go to the beginning and end of of selected feature(s) using Ctl-left arrow and Ctl-right arrow

Go to the beginning and end of the genome using the Ctl-up arrow and Ctl-down arrow

3. Load files that specify features of special interest

Artemis can read files in different formats including EMBL, Genbank, or GFF.
A list of feature files has been provided for you on your computer

<u>Load feature files into the Artemis display using the following command:</u>
File>Read an entry…

Note if at any time Artemis gives you the following window, click "NO" and continue:
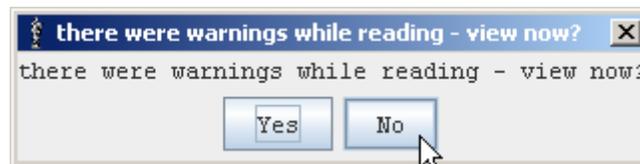
1. The first of these feature files to load is called:

       Pto-virulence.tab
       (Contents of this file and their color coding are shown below)

brown	Type III secretion system and chaperone proteins
red	Type III effectors
orange	siderophores (required for iron acquisition)
blue- light blue	small compounds toxic to the host plant
green	virulence factors associated with the bacterial cell wall including cell wall polysaccharides and secretion pathways

If you want see these more clearly without the complete genome annotation, de-select the first entry listed in the entry line near the top of the Artemis window

Notes on what you can see with this file:
Scan through the genome noting differences in virulence gene distribution and size for different gene classes. Genes encoding secretion pathways such as the Type II and Type III secretion pathways (PSPTO_1379-1403) (PSPTO_3306-3317) are clustered together as are those involved in biosynthesis of the polysaccharide alginate (PSPTO_1232-

1243). Though many Type III effector proteins are isolate from one another, some of these are also clustered (PSPTO_4588-4599).

The average gene size for the pseudomonads is approximately 1 kb. However, you may notice that some of the virulence genes are over 10 kb (PSPTO2135-2150, PSPTO_2601-2602, PSPTO_4686-4687 and PSPTO_4699). These genes encode non-ribosomal peptide synthases and polyketide synthases required for synthesis of siderophores and toxins.

2. Load the second feature file (specifying binding sites for the HrpL sigma factor in green):

      Pto-Hrp-boxes.tab

Scroll through the genome to see which classes of virulence factors are preferentially associated with hrp boxes

3. Load the third feature file (specifying mobile genetic elements in purple):

      Pto-mobile_DNA.tab

To scroll through selected classes of mobile elements such as the prophages, open the navigation window using Ctl-G and type "prophage" into the "Goto feature with this qualifier" field. You should find 5 annotated prophage features. Note the association of PHAGE05 with Type III effector AvrPto

4. Create a new file with features of interest

One of the most useful aspects of the Artemis genome viewer is that users can create entry files containing genes or other features of individual interest. The following example describes creation of a feature file containing all genes annotated as ABC transporters

About ABC transporters:
ABC transporters are protein complexes in the bacterial envelope that function to export undesirable substances such as drugs and import desired substrates including many critical nutrients. ABC transporters have also been shown to mediate secretion of virulence proteins across the bacterial envelope. The proteins that make up the ABC transporters are well conserved and most are readily identified from sequence data alone. The ABC stands for ATP-Binding Cassette, referring to the ATP binding domain found in one of the components of the machinery. Though not always associated directly with virulence, ABC transporters are critical to bacterial survival in diverse niches.

4a. Creating and naming a novel entry file

From the main menu:

Create>"New entry"
(a new entry entitled "no name" will appear on the entry line near the top of the Artemis display)

Entries>"Set name of entry".
Select the entry "no name" and type in "ABC-transporter.tab"

File>"Save an entry" (selecting "ABC-transporter.tab" for saving)

You now have an empty entry file to which you can save features of your choosing

4b. Select and move features into your new entry

Large numbers of features can be selected using the Feature selector option:

Select coding sequences for ABC transporters:

Select>"Feature selector.."

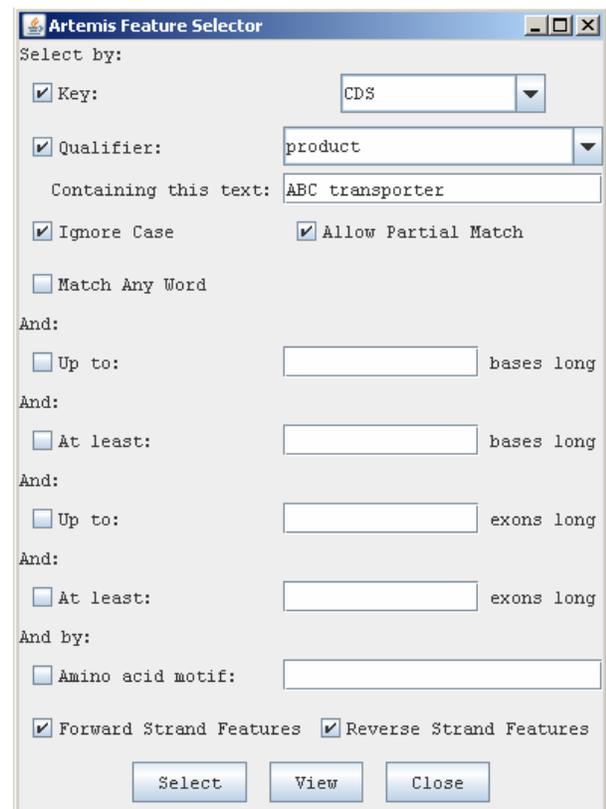In the "Key" field, select CDS from the pulldown menu

For the "Qualifier" field, select "product" from the pulldown menu

In the text box type "ABC transporter"

Edit>"Copy selected features"
Select your "ABC-transporter.tab" file when a specific file is requested

File>"Save an entry" (selecting "ABC-transporter.tab")

4c. Change the color of features in the new entry file:

De-select all active entry files except "ABC-transporter.tab"

Select>All
Edit>Change qualifiers of selected…

A new window will appear

From the "Insert qualifier" pulldown menu select "color"
Click the "insert qualifier" button
Type the following RGB coordinates for the color magenta after "/color="

255 0 255

Click "Add"

File>"Save an entry" (selecting "ABC-transporter.tab")

You will now be able to scroll through the genome and rapidly identify locations of ABC transporter components, of which there are over 300. Select some of these features and use Ctl-V to look at the detailed descriptions of the encoded products. The proteins that make up ABC transporters tend to fall into a small number of discrete components: (i) permeases, (ii) ATP-binding protein, (iii) periplasmic substrate binding protein. Genes clustered together usually encode components of a discrete transporter, though some of the clusters encode more than one of these components.

Using the Artemis Comparison Tool (ACT):

Goals:
- To visualize comparisons of related genomes using pre-generated and de novo comparison at WebACT
- To visualize feature files in the ACT context and practice navigating among specified features

General notes on using ACT

"ACT (Artemis Comparison Tool) is a DNA sequence comparison viewer based on Artemis. In common with Artemis, ACT is written in Java and runs on UNIX, GNU/Linux, Macintosh and MS Windows systems. It can read complete EMBL and GENBANK entries or sequence in FASTA or raw format. Extra sequence features can be in EMBL, GENBANK or GFF format. The sequence comparison displayed by ACT is usually the result of running a blastn or tblastx search." (from the Sanger site)

As with Artemis, ACT requires that the computer to be used either has Java pre-installed or that the user is able to download and install Java.

Using ACT requires that a comparison file be generated. This can be accomplished using the perl scripts provided by Sanger. But fortunately, for researchers without programming expertise, comparisons can be automatically generated and viewed using WebACT (http://www.webact.org/WebACT/home).

The instructions given below are for viewing pre-computed or user-generated comparisons obtained through WebACT (http://www.webact.org/WebACT/home).

Additional resources can be found at the following sites:

Sanger - Main ACT page
http://www.sanger.ac.uk/Software/ACT/

Sanger - Artemis manual
 http://www.sanger.ac.uk/Software/ACT/v7/manual/

Sanger - ACT examples page including a link to an online demo
http://www.sanger.ac.uk/Software/ACT/Examples/

Pseudomonas-Plant Interaction - ACT tutorial for PPI users
http://www.pseudomonas-syringae.org/pst_gen_analy.htm#ACT (end of page)

1. Starting up with ACT

Use the precomputed comparisons at WebACT

a. Go to WebACT Pre-Computed (http://www.webact.org/WebACT/prebuilt)

b. Select the genomes you wish to align (ACT works well with up to 3 genomes, but for this tutorial, aligning only 2 is preferred). For this tutorial, select the genomes for *P. syringae* pv. *tomato* DC3000 and *P. syringae* pv. *phaseolicola* 1448A.

c. Follow the prompts on the site until you get to **Start ACT** which allows the user to web launch ACT and view comparisons. If you do not have Java Web Start (included in more recent versions of Java), you can download the comparison file from WebACT, and read it using ACT downloaded from Sanger and run locally.

**Generate de novo comparisons using WebACT**

WebACT can also generate blastn or tblastx comparison files from either sequence/annotation files uploaded from your hard drive or specified with accession numbers. Generating de novo comparisons is sometimes preferable to using pre-computed comparisons, particularly if the pre-computed comparisons are out of date or if WebACT has not yet generated comparisons for more recently deposited sequences. If you choose to use this approach, follow the steps below.

a. Go to the WebACT Generate site (http://www.webact.org/WebACT/generate). Upload sequence/annotation files from your hard drive or paste in the RefSeq accession number (shown in the table above). Generate the comparison file as instructed.
b. Upon completion of the comparison, WebACT gives the option to download files or start ACT. Launching ACT from the WebACT site is the fastest means of viewing comparison results, but if you wish to save the comparison files beyond the 7 days for which they are accessible on the WebACT server, they can be downloaded and subsequently viewed by launching ACT from the Sanger ACT site.

The display below shows a comparison of a part of the *P. syringae tomato* DC3000 genome (top) with that of P. syringae phaseolicola 1448A.

Genome window

Comparison window

Genome window

The display below shows a comparison of all three *P. syringae* genomes (note that when 3 or more genomes are compared, the translation frames are omitted to save space)

Psp 1448A chrom

1448A-DC3000 comparison

Pst DC3000 chrom

DC3000-B728a comparison

Psy B728a chrom
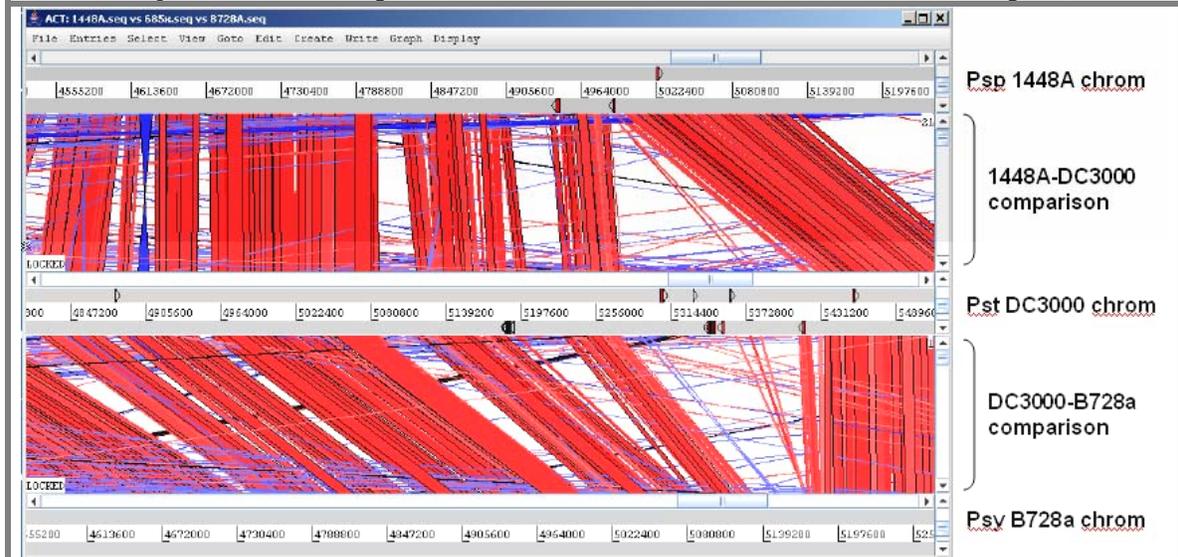
2. A few basic commands for using ACT

The red and blue bars indicate regions of similarity with red bars indicating corresponding regions that are oriented similarly and blue bars indicating regions oriented in opposite directions

To vertically align similar regions, double click on the red or blue connecting them

To ease viewing of inverted regions, right click on one of the genome windows and from the menu that appears check the box beside "Flip display"

Some genome alignments are cluttered with a "spider web" of lines resulting from very short or artifactual BLAST hits. To eliminate these from the display, right click on the comparison window and select "Set score cutoffs…" from the resulting menu. The window below will appear:
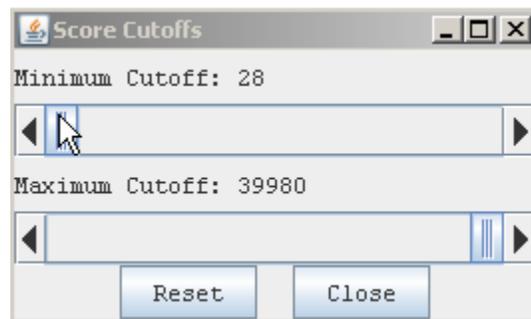
Slide the minimum cutoff bar up to improve the appearance of your display. (I often use a setting around 700 to balance selectivity and sensitivity of the output, but the appropriate setting vries with your needs and with the overall similarity of the genomes being compared)

Feature files can be loaded into specified genomes using the pull-down menu at the top of the main Artemis window.
File>(select genome)>Read an entry

Entries can be de-selected by right clicking on a genome, selecting "Entries" in the window that appears, and clicking on the box next to the entry to be de-selected

3. Viewing selected differences between two genomes

To view feature files in ACT:

File>(select genome)>Read an entry:

      Navigate to the feature file you wish to load

Right click on the Pto DC3000 genome.
In the resulting window select Goto>Navigator… to get the navigator window for that genome.

Type "hopR1" in the "Goto Feature with this qualifier value" line and click on Goto.

Double click on the highlighted bar below hopR1 to align with its orthologs in Pph 1448A. Note how only hopR1 and the upstream region containing the hrpL promoter are conserved between the two regions. The lack of further conservation upstream or downstream suggests independent acquisition of hopR1 by the two genomes.

Now use the navigator window to search for "hopY1"
Double click on the red bars below to bring the corresponding regions of the two genomes into alignment. Note that Genes upstream and down are conserved between the two, but that HopY1 is absent from Pph 1448A. This could indicate either acquisition of hopY1 by Pto DC3000 or loss of hopY1 by ph 1448A. Sequencing of additional *P. syringae* strains is expected to help distinguish between these two possibilities.