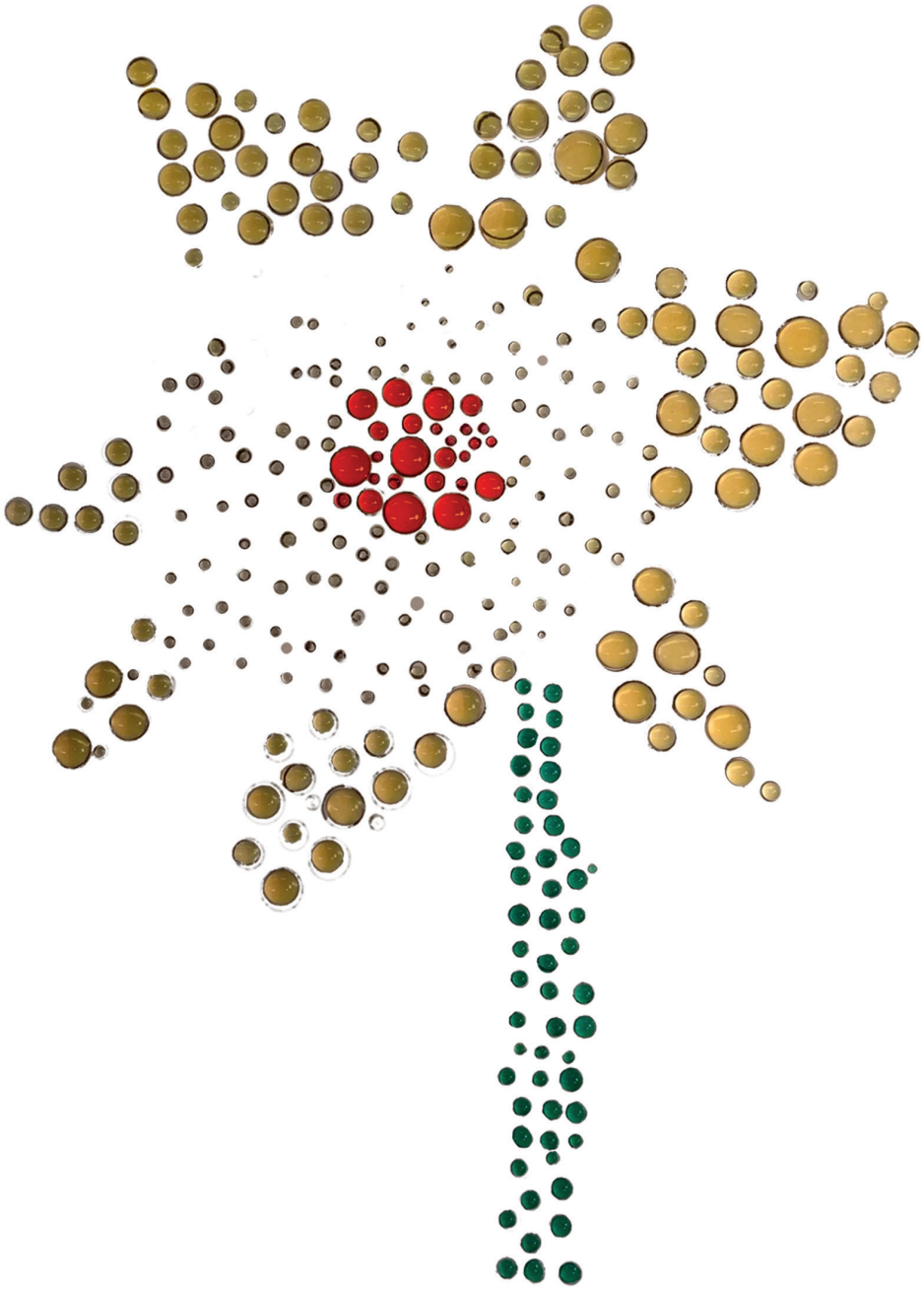
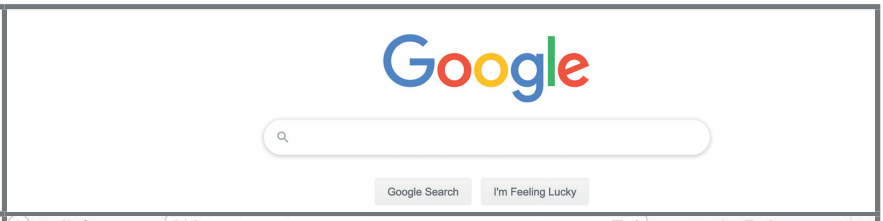


Viewing Daffodil Chloroplast DNA Sequences using Jalview

University of Dundee

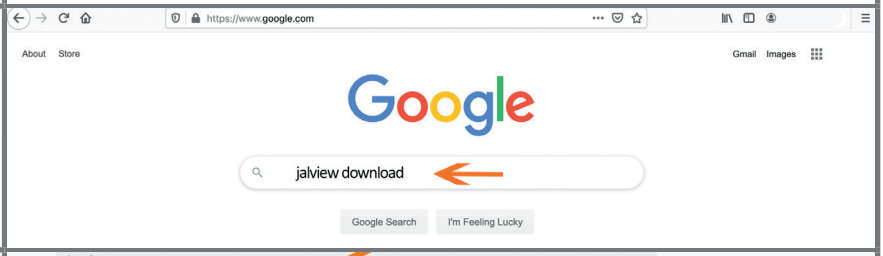
Step 1: Instal Jalview on the Computer

1. **Open a web browser** such as Chrome or Firefox.



2. **Search** using keywords '***Jalview Download***'.

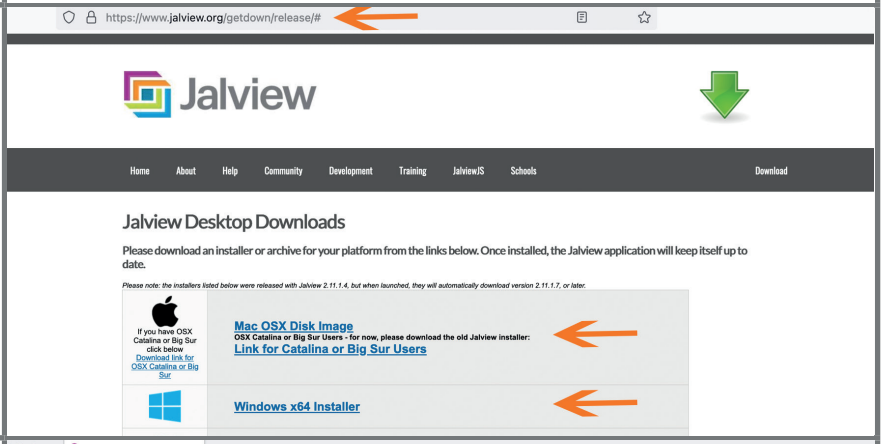
From the list of results, select https://www.jalview.org/getdown/release to open the download webpage.



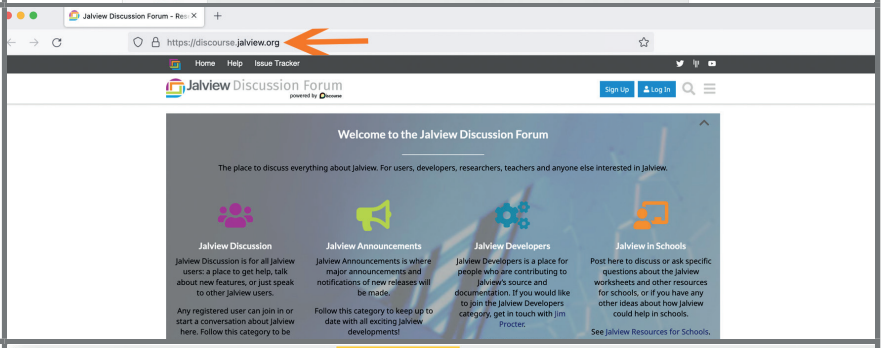
3. To download Jalview, click the link appropriate to your computer's operating system.

This downloads and opens a Jalview webstart file.

Follow the instructions in the dialog boxes that open.

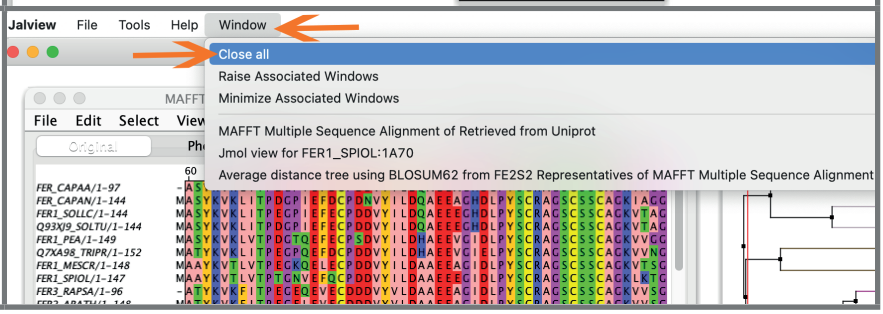


Note: Additional help is available on the Jalview website or at the Jalview Discussion Forum (https://discourse.jalview.org).



4. When Jalview launches, a demo file opens in the Jalview desktop window.

5. To close the windows, select the **Window menu** in the Jalview desktop window, then select **Close all**.



Step 2: Reading Sequences into Jalview

Background

Jalview can read a range of file formats such as FASTA (.fa, .fasta); Clustal (.aln); Pfam (.pfam); Stockholm (.stk, .sto); GenBank (.gb,.gbk); JSON (.json).

Jalview has a number of different ways to read these files. Three of the most common methods are described below.

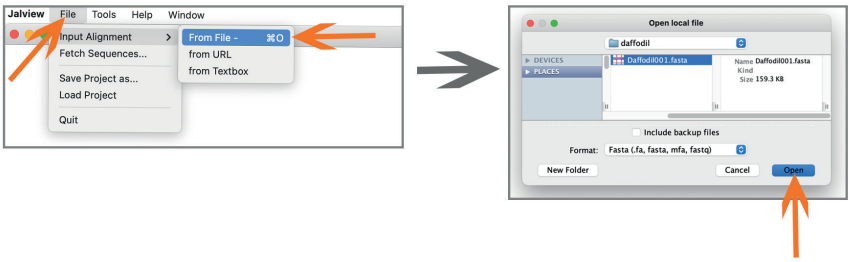
1. **Load a file from a folder**:

Select **File menu** from the Jalview desktop window.

Select **Input Alignment** and **From File -**.

Select the **filename** in the required folder and click **Open**.

Note: If you can't see the file, check the **Format drop-down menu** has the correct format selected.



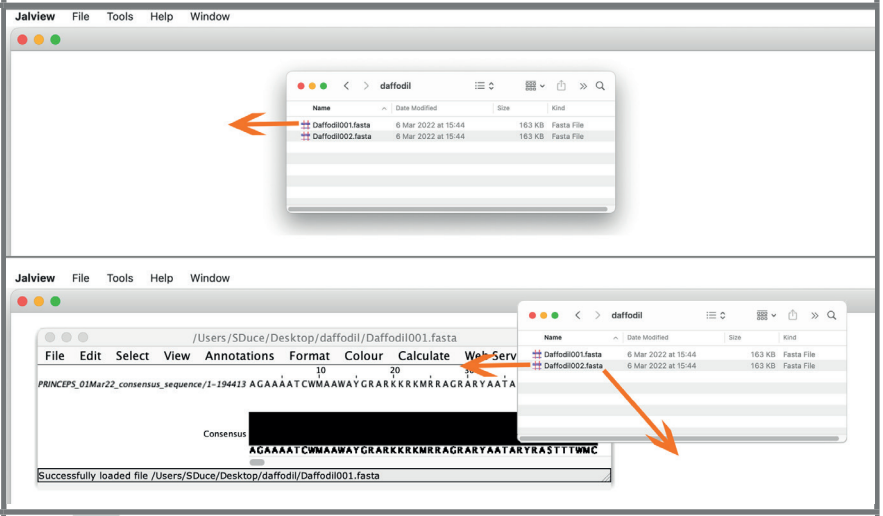
2. **Load by Drag & Drop:**

Drag the FASTA file from its folder and drop it onto the Jalview desktop window.

The alignment will open.

Dragging another file onto the Jalview alignment window will add new sequences to the alignment.

Whilst dragging a file onto the Jalview desktop window will open a new alignment window.



3. **Load from Public Databases:**

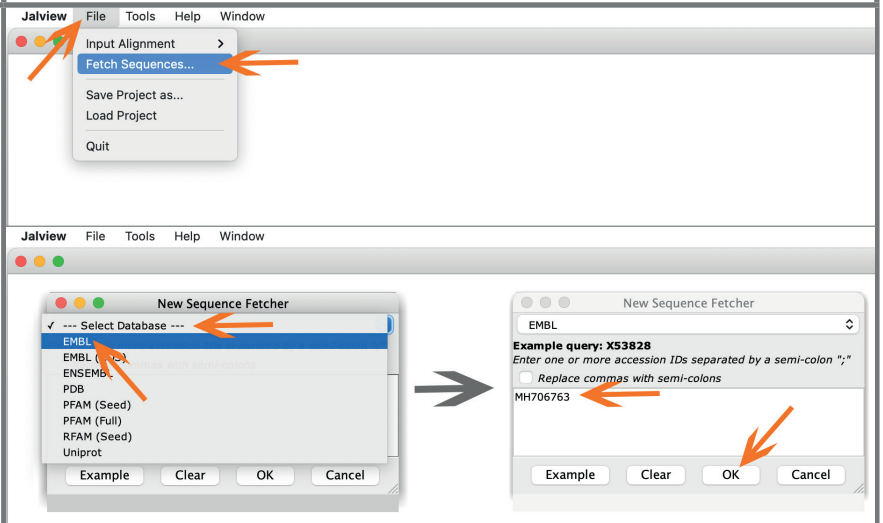
Select **File menu** from the Jalview desktop window.

Select **Fetch Sequences...** .

A 'New Sequence Feature' dialog box opens, click on **---Select Database---** .

Select the database required from the drop-down list eg **EMBL**.

Enter the **sequence ID** eg **MH706763** and click **OK**.



Note: Change the appearance of the windows:

(i) **To move a window**, place mouse on the title panel at the top of the window, then click & drag.

(ii) **To enlarge a window**, place mouse on the lower right-hand corner of the window, then click & drag.

(iii) **To close a window,** click the 'X' in the top right-hand corner of the window.

Step 3: Aligning & Colouring Sequences

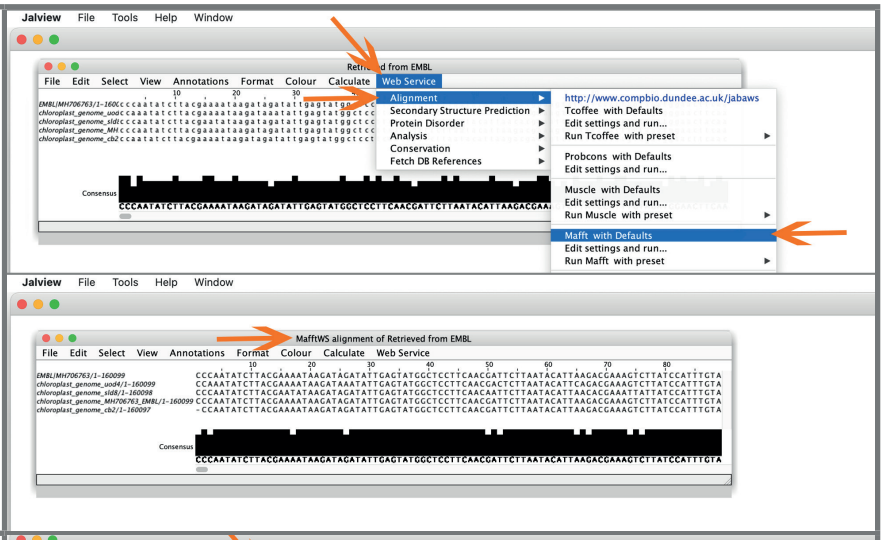
1. Press the **[ESC] key** to deselect any bases or sequences.

Select **Web Service menu** in the alignment window.

Select **Alignment** and **MAFFT with Defaults**.

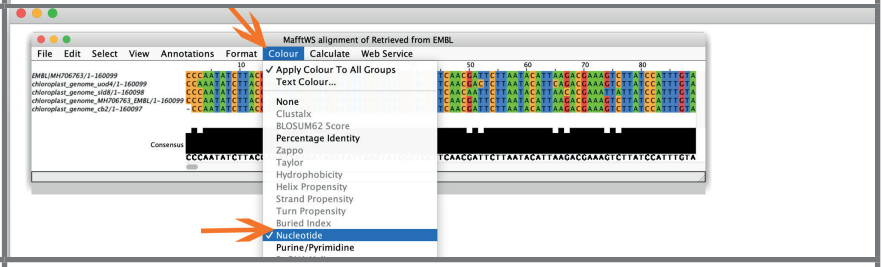
A window will open giving the job status.

Depending on internet speeds and the size of the alignment, this can take some time. Eventually, another alignment window will open containing the aligned sequences.

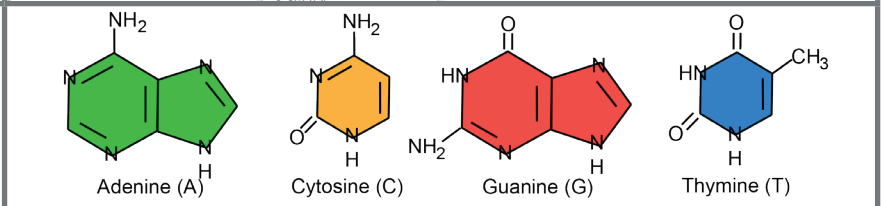


2. To colours the bases, select **Colour menu** in the alignment window.

Select **Nucleotide**.



The Jalview nucleotide colour scheme: adenine bases are green, cytosine bases are yellow, guanine bases are red and thymine bases are blue.



3. There are several ways to modify the appearance of the alignment.

(a) The font style and size can be changed.

(b) The alignment can be wrapped.

(c) The annotation rows can be toggled on and off.

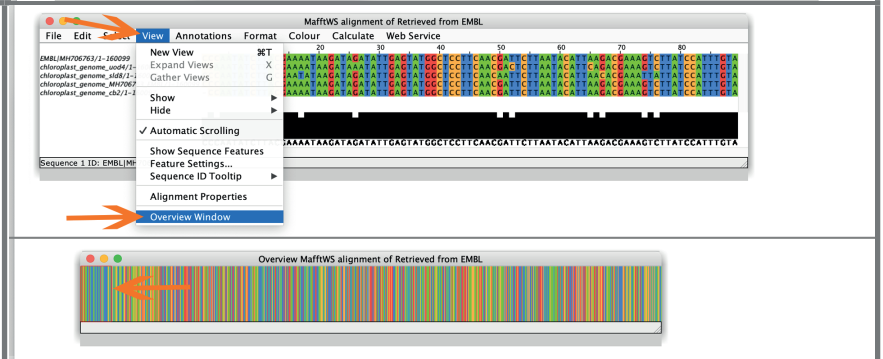
Re-scale the window by placing the cursor on right-hand botttom corner then click & drag.

Move the window by placing the cursor on the title panel at the top of the window, then click & drag.

4. To open the overview window, select the **View menu** in the alignment window.

Select **Overview Window**.

Note: The window contains a red box that reflects the bases displayed in the alignment window. Click & drag the red box to scroll along the sequences in the alignment window.



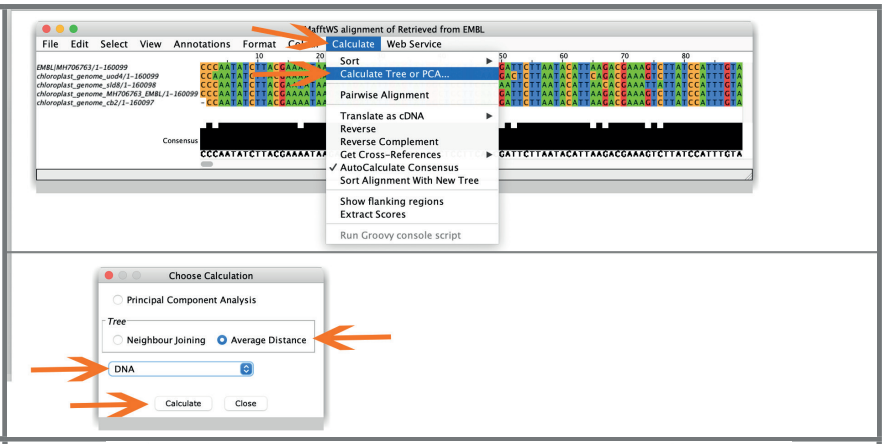
Step 4: Creating a Tree & Reorder Sequences

1. Go to the **Calculate menu** in the alignment window.

Select **Calculate Tree or PCA...** .

A 'Choose Calculation' dialog box opens, select **Average Distance** and **DNA** options.

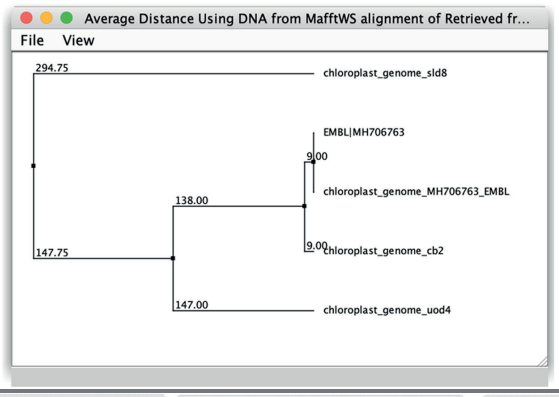
Click the **Calculate** button.



2. A tree window opens, the tree reflects the similarity between the sequences in the alignment window.

The numbers on the trees are similarity scores. They can be added together as a measure of sequence similarity.

The smaller the number, the greater the sequence similarity.



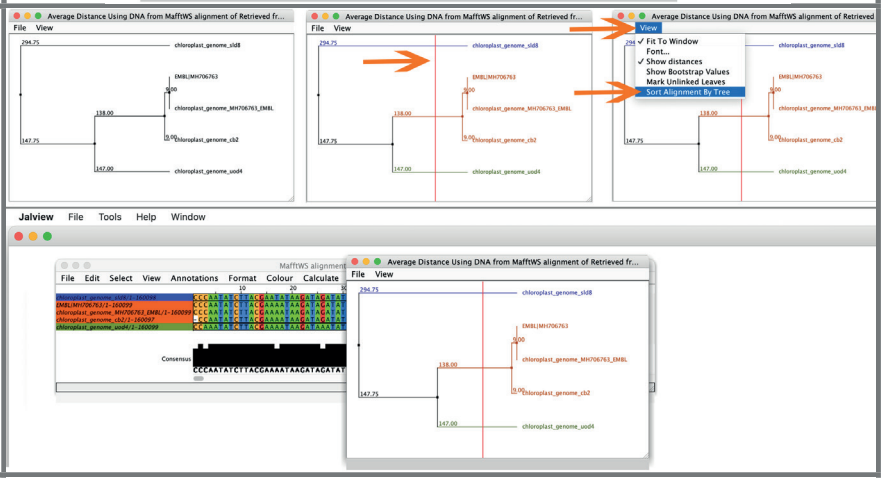
3. **Click the mouse on the tree**, and a red vertical line appears.

The **red line** groups sequences within the tree. Each group has its own randomly generated colour.

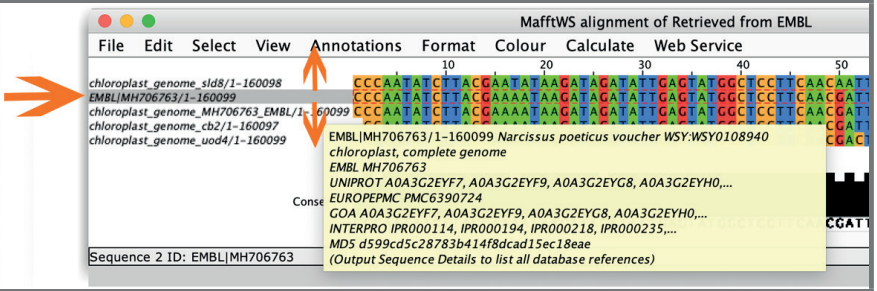
Go to the **View menu** in the tree window.

Select **Sort Alignment By Tree**.

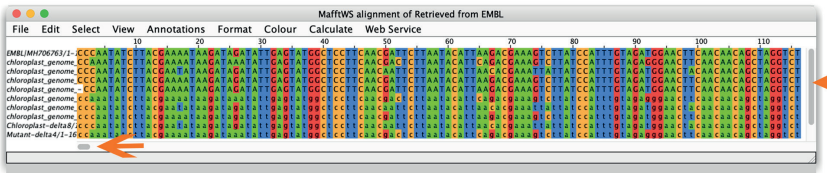
The sequences in the alignment window are reordered to reflect the sequence order in the tree.



4. The order of sequences in the alignment can be changed manually by clicking on the **sequence ID name** and the use then **up and down arrows** on the keypad.

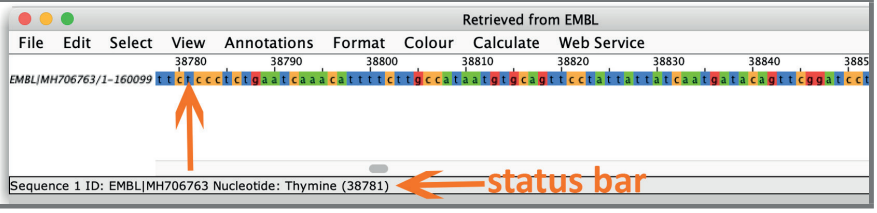


Note: The **vertical and horizontal scroll bars** can help navigate around the sequences.



5. Place the **mouse cursor on a base** and view the **status bar** in the lower left-hand-corner.

This provides additional information about the base.

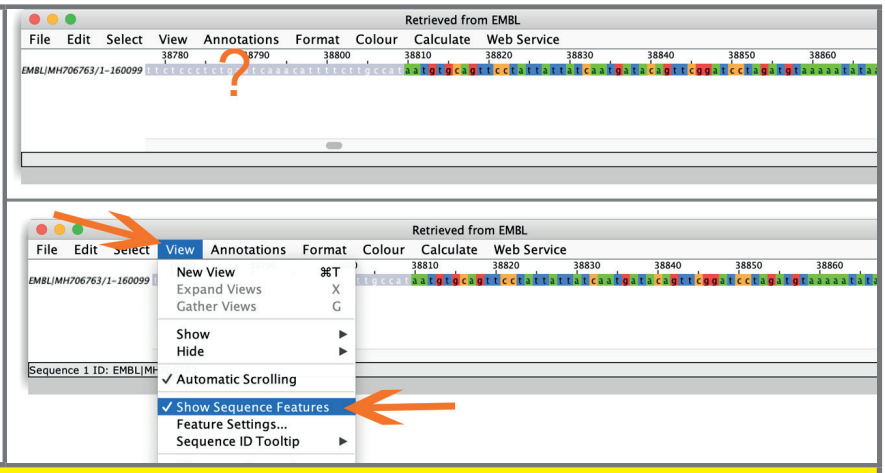


A Common Colouring Base Issue

Reading sequences from public databases may load additional sequence features.

These features can mask the sequence colouring, as shown in the figure.

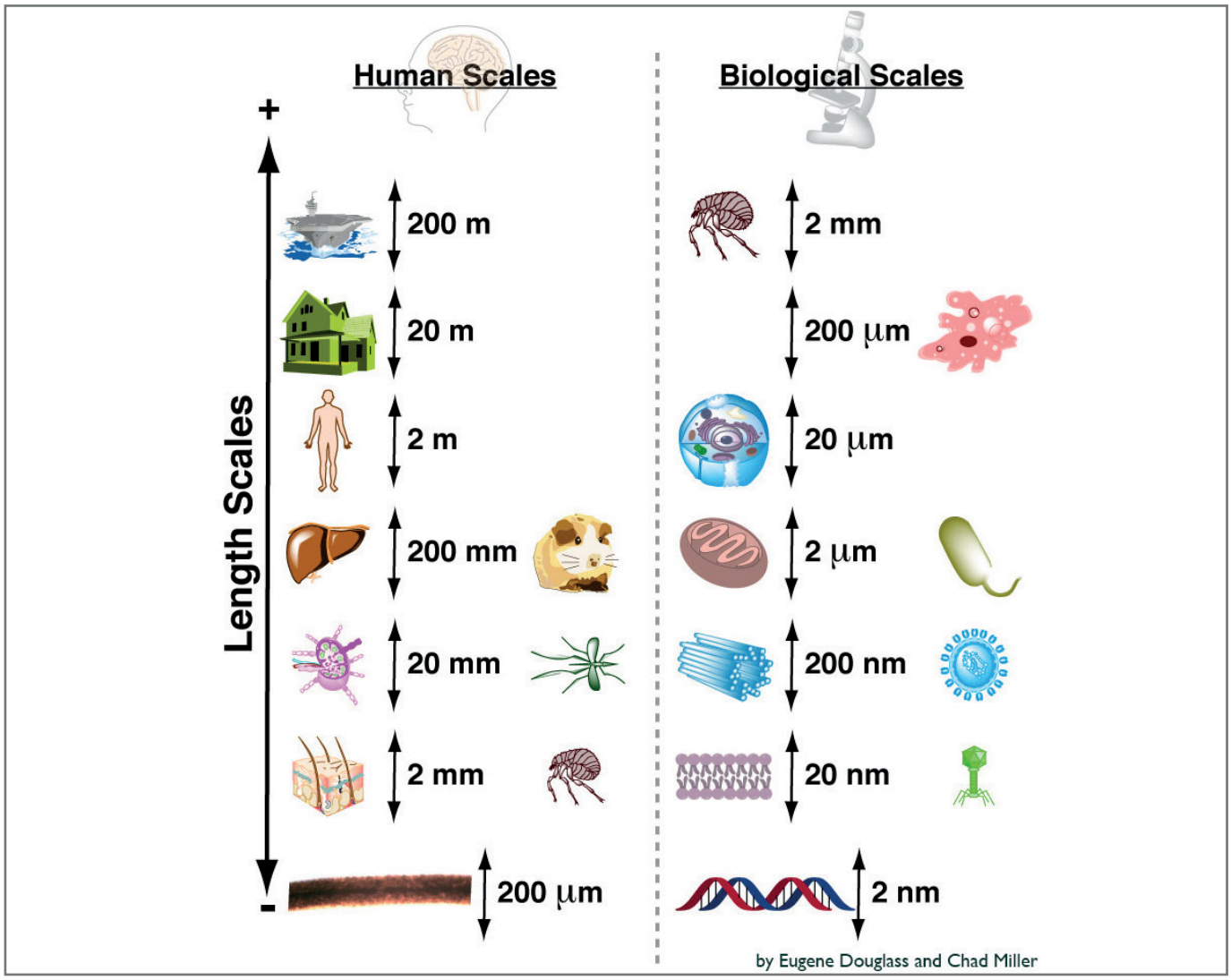
The sequence features can be toggled off by clicking the **View menu** in the alignment window. Then untick **Show Sequence Features**.



If you have a question or comment, please drop the Jalview team a line at schools@jalview.org. Also check the comments on the Jalview Discussion Forum. Feel free to contributes, there is a schools section https://discourse.jalview.org.

Guide to Biological and Human Scales

Molecular and cellular biology scales compared to human scales (approx 5 orders of magnitude per column)



Human Scale Key: Aircraft Carrier (200m), 3-story House (20m), Human (2m), Liver/Gerbil (200mm), Lymph Node/Mosquito (20mm), Skin/Flea (2mm), Hair thickness (200um)

Biological Scale Key: Flea (2mm), Amoeba (200um), Eukaryotic Cell (20um), Mitochondria/Bacteria (2um), Centriol/Large Virus (200nm), Cell Membrane/Small Virus (20nm), DNA thickness (2nm)

https://www.practicallyscience.com/intuiting-biological-scales-using-human-scales/#more-53

Glossary

Amino acid:- molecular sub-units of peptides and proteins.

Bioinformatics:- the application of computer and statistical techniques to the management of biological data.

cDNA (complementary DNA):- cDNA sequence is synthesized from an RNA template by reverse transcription. It contains 5' and 3' untranslated regions (UTRs) as well as coding regions.

CDS (protein-coding sequence):- the portion of the mRNA transcript that is translated by ribosomes into proteins.

Chromosome:- located in the cell nucleus, it contains the cellular DNA along with a number of proteins (eg histones) that compact and package the DNA.

Codon:- a set of three adjacent nucleotides (triplet) that code for a specific amino acid residue during protein synthesis.

DNA (deoxyribonucleic acid):- the molecule that encodes genetic information. It carries the instructions for all aspects of an organism's functions such as growth, metabolism and reproduction. These chains can be over 100,000,000 molecules in length.

Exon:- the sections of a gene that are translated into proteins, they remain in the transcript (mRNA) after introns have been spliced out of the genomic sequence.

Gene:- a region of DNA that encodes a specific protein or protein subunit.

Genetic code:- sets of triplet nucleotides that encode specific amino acids.

Genome:- all the genetic material in the chromosomes of a particular organism.

Genomic DNA (gDNA):- all the DNA residing in the chromosomes.

Genotype:- all the genes in a particular individual.

Intron:- the noncoding part of the genome that is transcribed then spliced out of the RNA.

Phenotype:- the observable characteristics or features of a living organism.

Phylogenetic tree:- an evolutionary tree for organismal species or cellular macromolecules that is built using inheritance or molecular sequence information.

Protein:- a biological macro-molecule composed of a string of amino acids joined together by peptide bonds.

Protein sequence:- the sequence of amino acids in a protein.

Nucleoside:- nucleotides without a phosphate group.

Nucleotide:- building blocks of RNA and DNA made up of a nitrogenous base, a molecule of sugar and phosphoric acid.

Multiple sequence alignment:- an alignment of three or more sequences with gaps inserted in the sequences such that residues with common structural positions and/or ancestral residues are aligned in the same column.

RNA (ribonucleic acid):- RNA are similar to DNA but contains the ribose sugar rather than deoxyribose sugar and the base uracil (U) rather than thymine (T). Typically they are single-stranded.

Replication:- process by which DNA makes a copy of itself during cell division.

Sequence alignment:- arranging the sequences of protein, RNA or DNA to identify regions of similarity. The similarity could be a consequence of functional, structural, or evolutionary relationships.

Translation:-process where mRNA is decoded by ribosomes to produce specific amino acids and polypeptides.

Transcription:- process where a segment of DNA is copied into RNA by the enzyme RNA polymerase.

For more information, please visit the Jalview website www.jalview.org

Dr Suzanne Duce
Division of Computational Biology, School of Life Sciences,
University of Dundee, Dundee DD1 5EH, UK