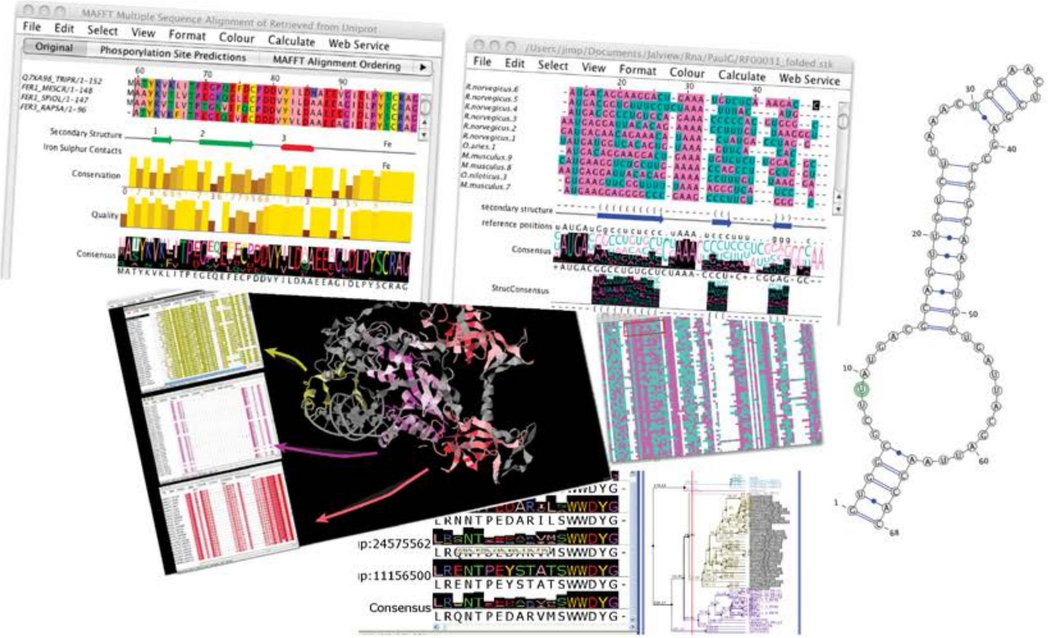
Practical Jalview



Jim Procter
University of Dundee
24th October 2014

supported by
wellcome trust

BBSRC bioscience for the future

Jalview www.jalview.org

THE BARTON GROUP

UNIVERSITY OF DUNDEE

open source initiative

9.00-9.15am. Overview of the day

9.15am - 10.30am. Session 1. Introduction to Jalview

- starting the application, importing alignments, basic editing and creating figures.

10.30-11am. Coffee

11am – 11.30am Geoff Barton: Multiple alignment and Analysis

11.30am - 12.30pm. Session 2: Alignment & alignment analysis

- Creating sequence alignments, importing and calculating trees, tree based alignment analysis

12.30pm to 1.30pm. Lunch

1.30pm – 3.00pm. Session 3: Annotating sequences & alignments

- Creating and viewing sequence annotation
- Protein Secondary structure prediction

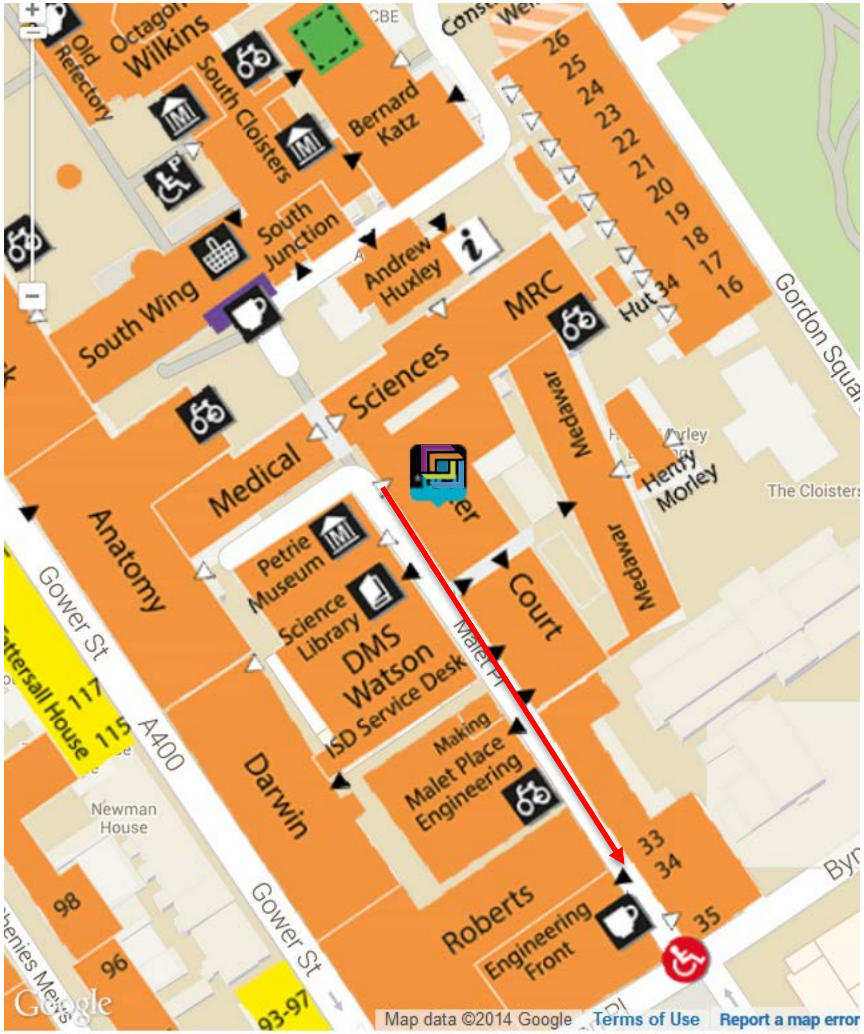
3.00pm – 3.30pm Coffee

3.30pm – 4.30pm. Session 4: Working with molecular structures

- Viewing 3D Structures, superimpositions, mapping disorder and alignment quality
- Viewing RNA Secondary Structure

4.30pm – 4.45pm. Wrapup – what we didn’t cover today

Tea breaks and Lunch are in Roberts Foyer G02

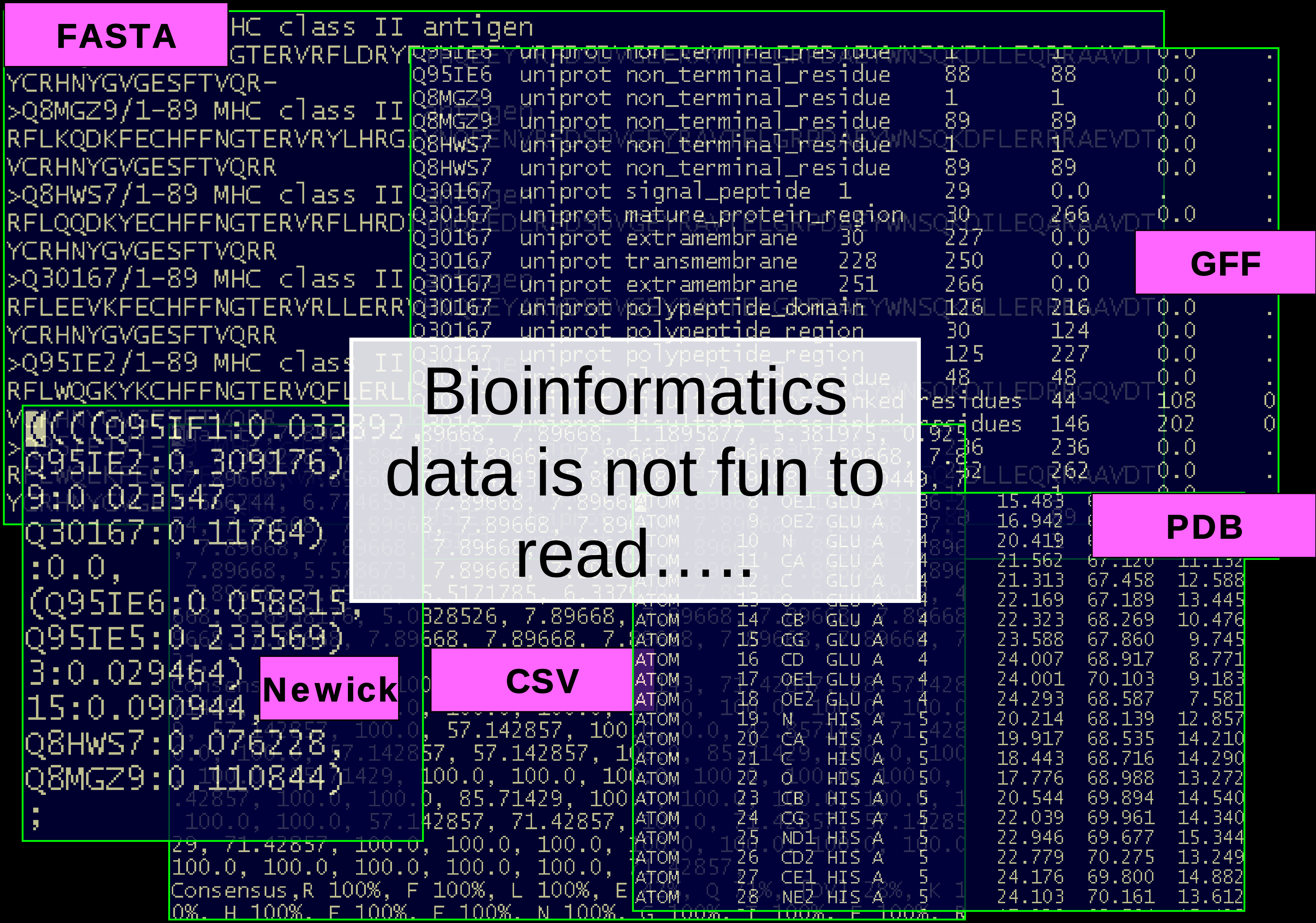


Course materials

Everything is online

http://www.jalview.org/tutorial/training-materials/2014/London/Dec/

- **These slides**
- **Jalview v2.8.2 Manual (v1.5.0)**
  - **Log in and**
  - **Open the manual in your PDF Viewer NOW**



```
HC class II antigen
GTERVRFLDRY
YCRHNYGVGESFTVQR-
>Q8MGZ9/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
VCRHNYGVGESFTVQRR
>Q8HWS7/1-89 MHC class II
RFLQQDKYECHFFNGTERVRFLHRD
YCRHNYGVGESFTVQRR
>Q30167/1-89 MHC class II
RFLEEVKFECHFFNGTERVRLLERR
YCRHNYGVGESFTVQRR
>Q95IE2/1-89 MHC class II
RFLWQGKYKCHFFNGTERVQFLERL
```

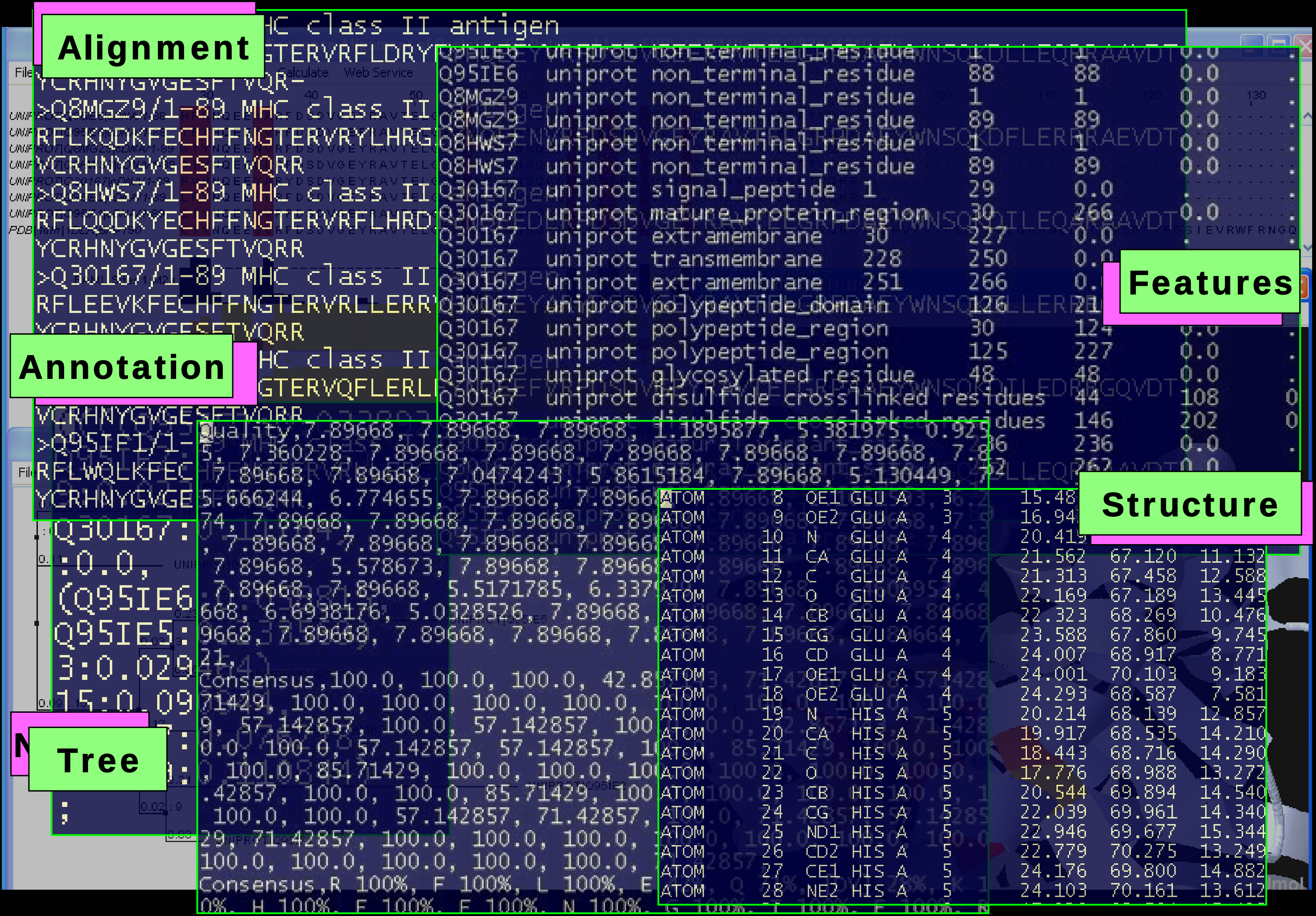
```
Q95IE6  uniprot non_terminal_residue    88      88      0.0     .
Q8MGZ9  uniprot non_terminal_residue    1       1       0.0     .
Q8MGZ9  uniprot non_terminal_residue    89      89      0.0     .
Q8HWS7  uniprot non_terminal_residue    1       1       0.0     .
Q8HWS7  uniprot non_terminal_residue    89      89      0.0     .
Q30167  uniprot signal_peptide  1       29      0.0     .       .
Q30167  uniprot mature_protein_region   30      266     0.0     .
Q30167  uniprot extramembrane   30      227     0.0     .       .
Q30167  uniprot transmembrane   228     250     0.0     .       .
Q30167  uniprot extramembrane   251     266     0.0     .       .
Q30167  uniprot polypeptide_domain      126     216     0.0     .
Q30167  uniprot polypeptide_region      30      124     0.0     .
Q30167  uniprot polypeptide_region      125     227     0.0     .
```

```
((((Q95IF1:0.033892,
Q95IE2:0.309176)
9:0.023547,
Q30167:0.11764)
:0.0,
(Q95IE6:0.058815,
Q95IE5:0.233569)
3:0.029464)
15:0.090944,
Q8HWS7:0.076228,
Q8MGZ9:0.110844)
;
```

```
100.0, 57.142857, 100
57, 57.142857, 1
100.0, 100.0, 10
0, 85.71429, 100
42857, 71.42857,
29, 71.42857, 100.0, 100.0, 100.0,
100.0, 100.0, 100.0, 100.0, 100.0,
Consensus,R 100%, F 100%, L 100%, E
```

```
ATOM     11  CA  GLU A   4      21.562  67.120  11.132
ATOM     12  C   GLU A   4      21.313  67.458  12.588
ATOM     13  O   GLU A   4      22.169  67.189  13.445
ATOM     14  CB  GLU A   4      22.323  68.269  10.476
ATOM     15  CG  GLU A   4      23.588  67.860   9.745
ATOM     16  CD  GLU A   4      24.007  68.917   8.771
ATOM     17  OE1 GLU A   4      24.001  70.103   9.183
ATOM     18  OE2 GLU A   4      24.293  68.587   7.581
ATOM     19  N   HIS A   5      20.214  68.139  12.857
ATOM     20  CA  HIS A   5      19.917  68.535  14.210
ATOM     21  C   HIS A   5      18.443  68.716  14.290
ATOM     22  O   HIS A   5      17.776  68.988  13.272
ATOM     23  CB  HIS A   5      20.544  69.894  14.540
ATOM     24  CG  HIS A   5      22.039  69.961  14.340
ATOM     25  ND1 HIS A   5      22.946  69.677  15.344
ATOM     26  CD2 HIS A   5      22.779  70.275  13.249
ATOM     27  CE1 HIS A   5      24.176  69.800  14.882
ATOM     28  NE2 HIS A   5      24.103  70.161  13.612
```

Bioinformatics data is not fun to read….



Alignment

```
HC class II antigen
GTERVRFLDRY
YCRHNYGVGESFTVQR-
>Q8MGZ9/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
VCRHNYGVGESFTVQRR
>Q8HWS7/1-89 MHC class II
RFLQQDKYECHFFNGTERVRFLHRD
YCRHNYGVGESFTVQRR
>Q30167/1-89 MHC class II
RFLEEVKFECHFFNGTERVRLLERR
YCRHNYGVGESFTVQRR
HC class II
GTERVQFLERL
VCRHNYGVGESFTVQRR
>Q95IF1/1-
RFLWQLKFEC
YCRHNYGVGE
```

Features

```
Q95IE6  uniprot non_terminal_residue           1    1    0.0 .
Q95IE6  uniprot non_terminal_residue           88   88   0.0 .
Q8MGZ9  uniprot non_terminal_residue           1    1    0.0 .
Q8MGZ9  uniprot non_terminal_residue           89   89   0.0 .
Q8HWS7  uniprot non_terminal_residue           1    1    0.0 .
Q8HWS7  uniprot non_terminal_residue           89   89   0.0 .
Q30167  uniprot signal_peptide  1    29   0.0 .
Q30167  uniprot mature_protein_region  30   266  0.0 .
Q30167  uniprot extramembrane   30   227  0.0 .
Q30167  uniprot transmembrane   228  250  0.0
Q30167  uniprot extramembrane   251  266  0.
Q30167  uniprot polypeptide_domain     126  21
Q30167  uniprot polypeptide_region     30   124
Q30167  uniprot polypeptide_region     125  227  0.0 .
Q30167  uniprot glycosylated residue   48   48   0.0 .
Q30167  uniprot disulfide crosslinked residues  44   108  0
Q30167  uniprot disulfide crosslinked residues  146  202  0
                                                     236  0.0 .
```

Annotation

```
Quality,7.89668, 7.89668, 7.89668, 1.1895877, 5.381975, 0.925
5, 7.360228, 7.89668, 7.89668, 7.89668, 7.89668, 7.89668, 7.8
 7.89668, 7.89668, 7.0474243, 5.8615184, 7.89668, 5.130449, 7
5.666244, 6.774655, 7.89668, 7.8966
74, 7.89668, 7.89668, 7.89668, 7.89
, 7.89668, 7.89668, 7.89668, 7.8966
 7.89668, 5.578673, 7.89668, 7.8966
 7.89668, 7.89668, 5.5171785, 6.337
668, 6.6938176, 5.0328526, 7.89668,
9668, 7.89668, 7.89668, 7.89668, 7.
21,
Consensus,100.0, 100.0, 100.0, 42.8
71429, 100.0, 100.0, 100.0, 100.0,
9, 57.142857, 100.0, 57.142857, 100
0.0, 100.0, 57.142857, 57.142857, 10
, 100.0, 85.71429, 100.0, 100.0, 10
.42857, 100.0, 100.0, 85.71429, 100
 100.0, 100.0, 57.142857, 71.42857,
29, 71.42857, 100.0, 100.0, 100.0,
100.0, 100.0, 100.0, 100.0, 100.0,
Consensus,R 100%, F 100%, L 100%, E
0%, H 100%, F 100%, F 100%, N 100%,
```

Tree

```
Q30167:
:0.0,
(Q95IE6
Q95IE5:
3:0.029
15:0.09
:
:
;
```

Structure

```
ATOM      8  OE1 GLU A   3      15.48
ATOM      9  OE2 GLU A   3      16.94
ATOM     10  N   GLU A   4      20.415
ATOM     11  CA  GLU A   4      21.562  67.120  11.132
ATOM     12  C   GLU A   4      21.313  67.458  12.588
ATOM     13  O   GLU A   4      22.169  67.189  13.445
ATOM     14  CB  GLU A   4      22.323  68.269  10.476
ATOM     15  CG  GLU A   4      23.588  67.860   9.745
ATOM     16  CD  GLU A   4      24.007  68.917   8.771
ATOM     17  OE1 GLU A   4      24.001  70.103   9.183
ATOM     18  OE2 GLU A   4      24.293  68.587   7.581
ATOM     19  N   HIS A   5      20.214  68.139  12.857
ATOM     20  CA  HIS A   5      19.917  68.535  14.210
ATOM     21  C   HIS A   5      18.443  68.716  14.290
ATOM     22  O   HIS A   5      17.776  68.988  13.272
ATOM     23  CB  HIS A   5      20.544  69.894  14.540
ATOM     24  CG  HIS A   5      22.039  69.961  14.340
ATOM     25  ND1 HIS A   5      22.946  69.677  15.344
ATOM     26  CD2 HIS A   5      22.779  70.275  13.249
ATOM     27  CE1 HIS A   5      24.176  69.800  14.882
ATOM     28  NE2 HIS A   5      24.103  70.161  13.612
```

Jalview

Java

Standalone or
web based

alignment
&
annotation,
tree, and
structure

viewer
and
editor

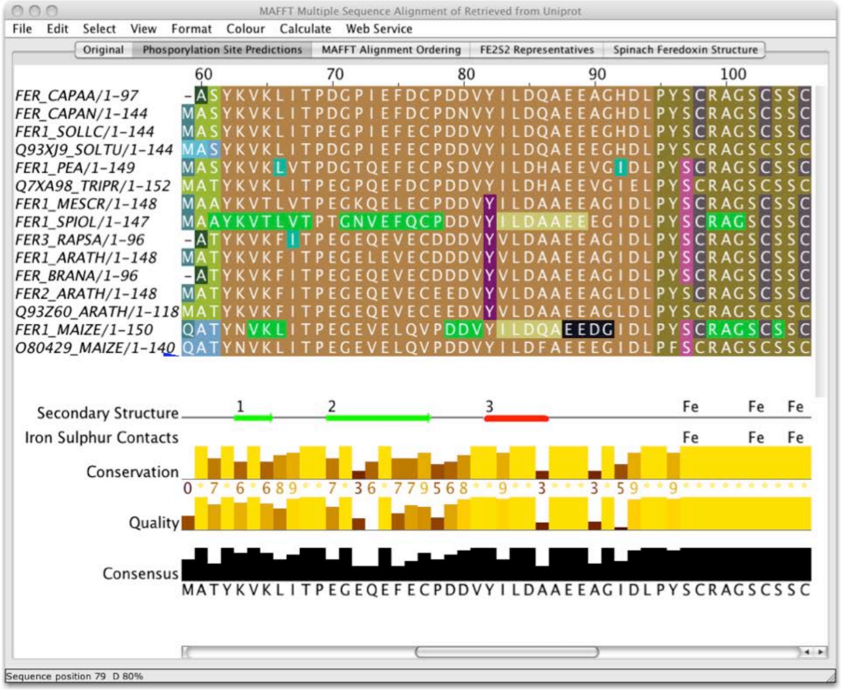
Available at

www.jalview.org

GPLv3 Free Software

THE BARTON GROUP

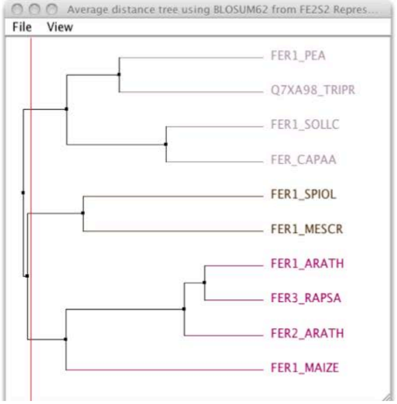
www.compbio.dundee.ac.uk

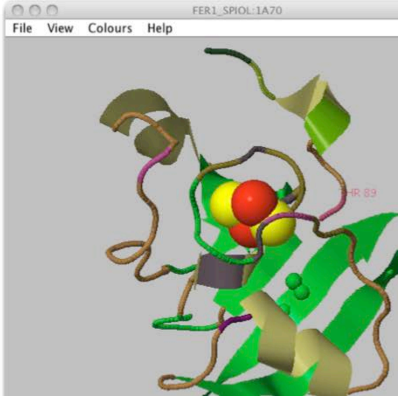


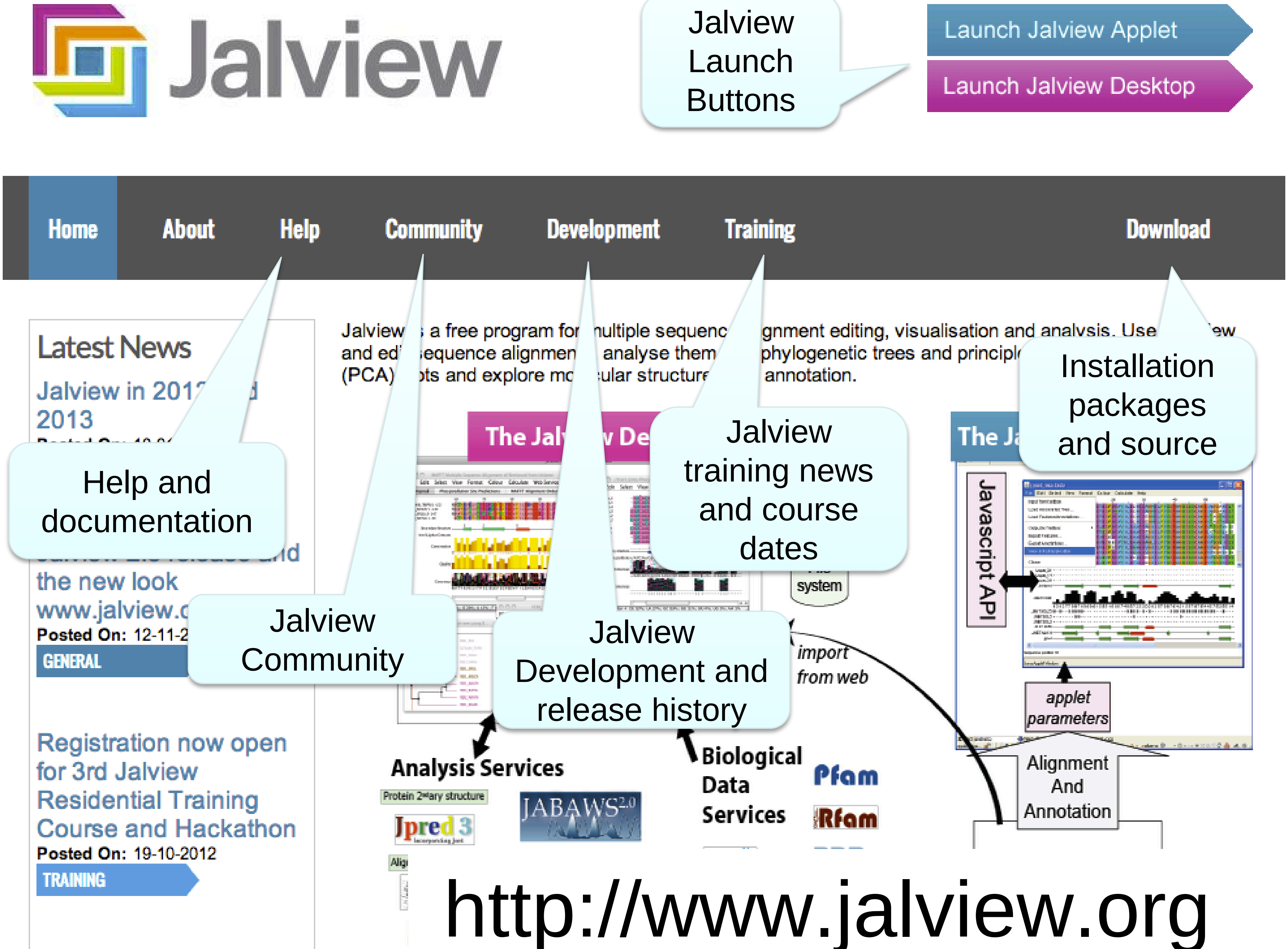
```
FER_CAPAA/1-97      -ASYKVKLITPDGPIEFDCPDDVYILDQAEEAGHDLPYSCRAGSCSSC
FER_CAPAN/1-144     MASYKVKLITPDGPIEFDCPDNVYILDQAEEAGHDLPYSCRAGSCSSC
FER1_SOLLC/1-144    MASYKVKLITPEGPIEFECPDDVYILDQAEEEGHDLPYSCRAGSCSSC
Q93XJ9_SOLTU/1-144  MASYKVKLITPDGPIEFECPDDVYILDQAEEEGHDLPYSCRAGSCSSC
FER1_PEA/1-149      MASYKVKLVTPDGTQEFECPSDVYILDHAEEVGIDLPYSCRAGSCSSC
Q7XA98_TRIPR/1-152  MATYKVKLITPEGPQEFDCPDDVYILDHAEEVGIELPYSCRAGSCSSC
FER1_MESCR/1-148    MAAYKVTLVTPEGKQELECPDDVYILDAAEEAGIDLPYSCRAGSCSSC
FER1_SPIOL/1-147    MAAYKVTLVTPTGNVEFQCPDDVYILDAAEEEGIDLPYSCRAGSCSSC
FER3_RAPSA/1-96     -ATYKVKFITPEGEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER1_ARATH/1-148    MATYKVKFITPEGELEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER_BRANA/1-96      -ATYKVKFITPEGEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER2_ARATH/1-148    MATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSC
Q93Z60_ARATH/1-118  MATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSC
FER1_MAIZE/1-150    QATYNVKLITPEGEVELQVPDDVYILDQAEEDGIDLPYSCRAGSCSSC
O80429_MAIZE/1-140  QATYNVKLITPEGEVELQVPDDVYILDFAEEEGIDLPFSCRAGSCSSC

Consensus           MATYKVKLITPEGEQEFECPDDVYILDAAEEAGIDLPYSCRAGSCSSC
```

Sequence position 79 D 80%







Latest News

Jalview in 201 d 2013

the new look www.jalview.c
Posted On: 12-11-2

Registration now open for 3rd Jalview Residential Training Course and Hackathon
Posted On: 19-10-2012

Jalview ; a free program for nultiple sequenc gnment editing, visualisation and analysis. Use ew and ed sequence alignmen analyse them phylogenetic trees and principl (PCA) ts and explore mo ular structure annotation.

Jalview Launch Buttons

Help and documentation

Jalview Community

Jalview Development and release history

Jalview training news and course dates

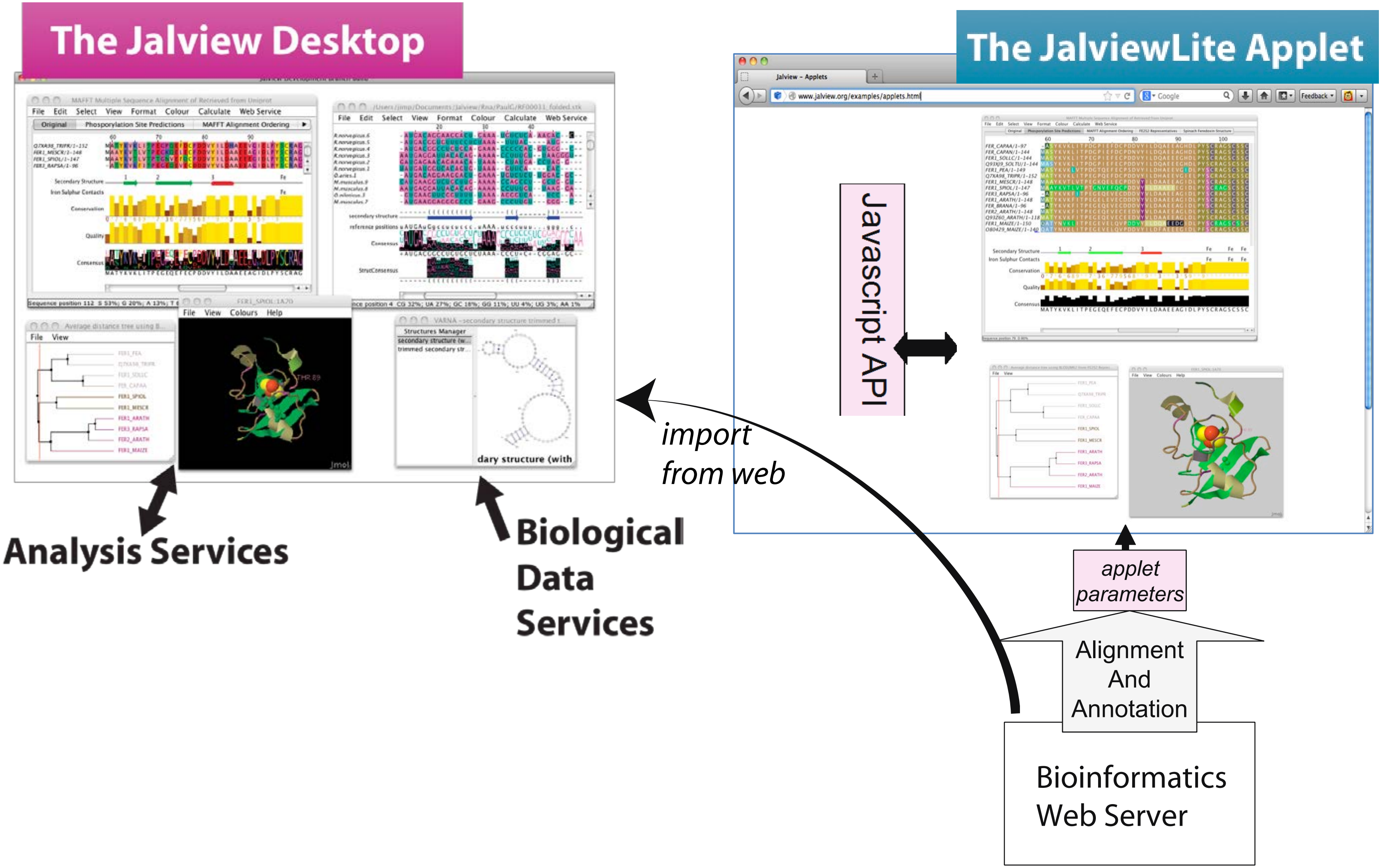
Installation packages and source

http://www.jalview.org

Jalview comes in two flavours

Launch Jalview Applet

Launch Jalview Desktop



The Jalview Desktop

The JalviewLite Applet

Javascript API

import from web

Analysis Services

Biological Data Services

applet parameters

Alignment And Annotation

Bioinformatics Web Server

Ex 1 – starting
The Jalview Dekstop

PAGE 7

Use the ‘Latest Build of Current Release’

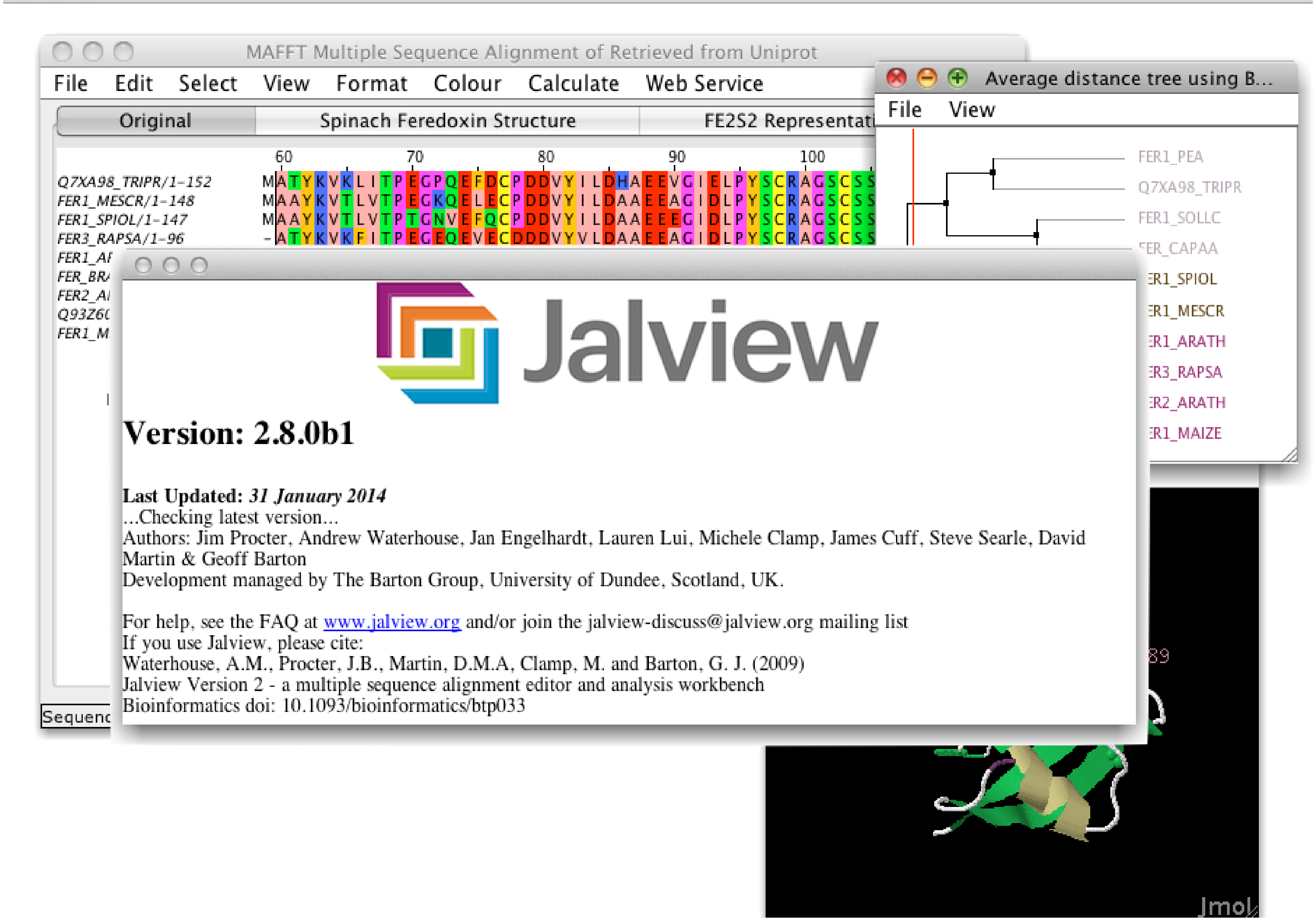
via the development page

http://www.jalview.org/development/development-builds

Webstart launch link is

http://www.jalview.org/builds/release/webstart/jalview_2G.jnlp

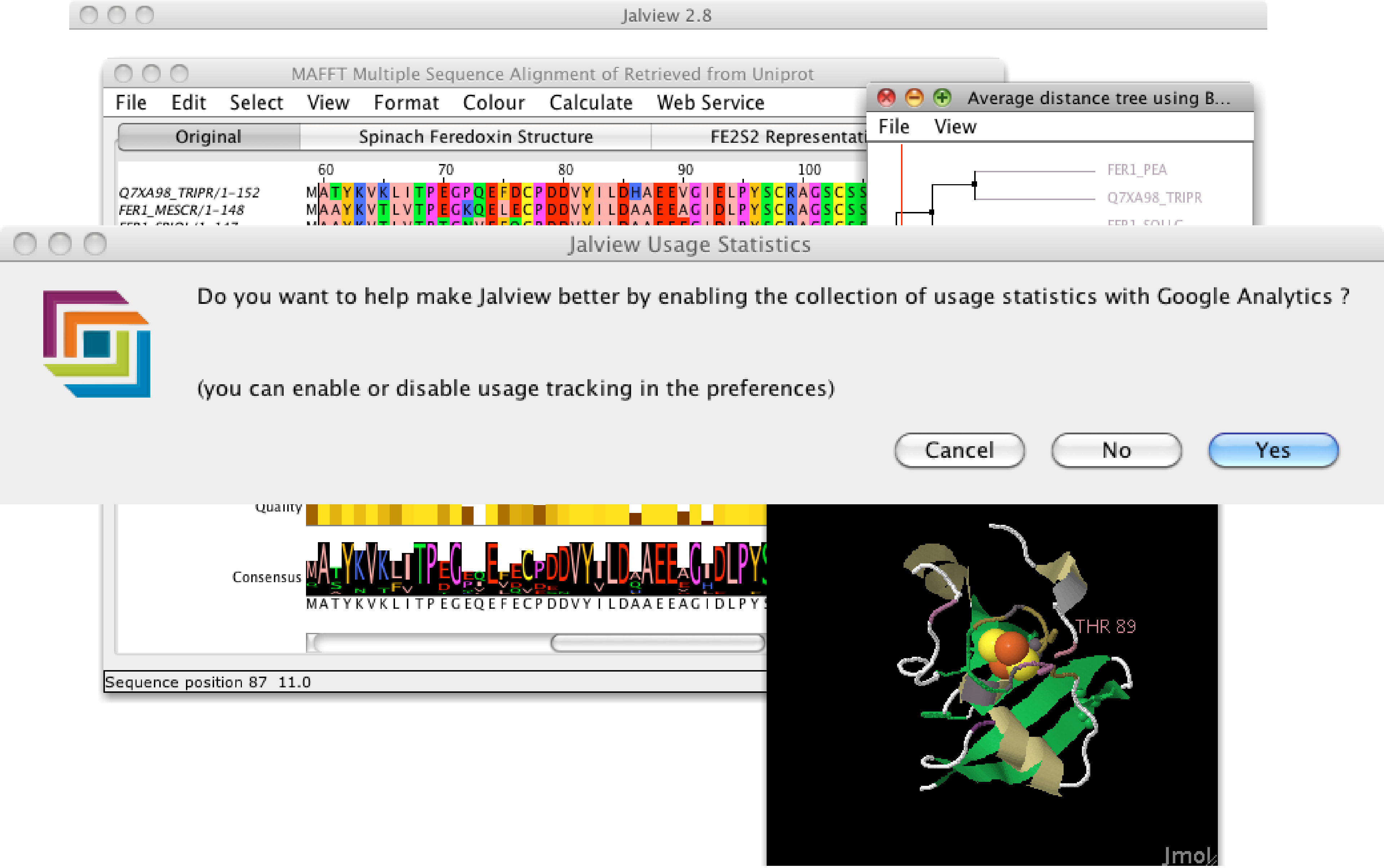
Launching the jalview desktop



Version: 2.8.0b1

Last Updated: ***31 January 2014***
...Checking latest version...
Authors: Jim Procter, Andrew Waterhouse, Jan Engelhardt, Lauren Lui, Michele Clamp, James Cuff, Steve Searle, David Martin & Geoff Barton
Development managed by The Barton Group, University of Dundee, Scotland, UK.

For help, see the FAQ at www.jalview.org and/or join the jalview-discuss@jalview.org mailing list
If you use Jalview, please cite:
Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009)
Jalview Version 2 - a multiple sequence alignment editor and analysis workbench
Bioinformatics doi: 10.1093/bioinformatics/btp033



Do you want to help make Jalview better by enabling the collection of usage statistics with Google Analytics ?

(you can enable or disable usage tracking in the preferences)

Sequence position 87 11.0

Jalview News

brought to you by JSwingReader (jswingreader.sourceforge.net)

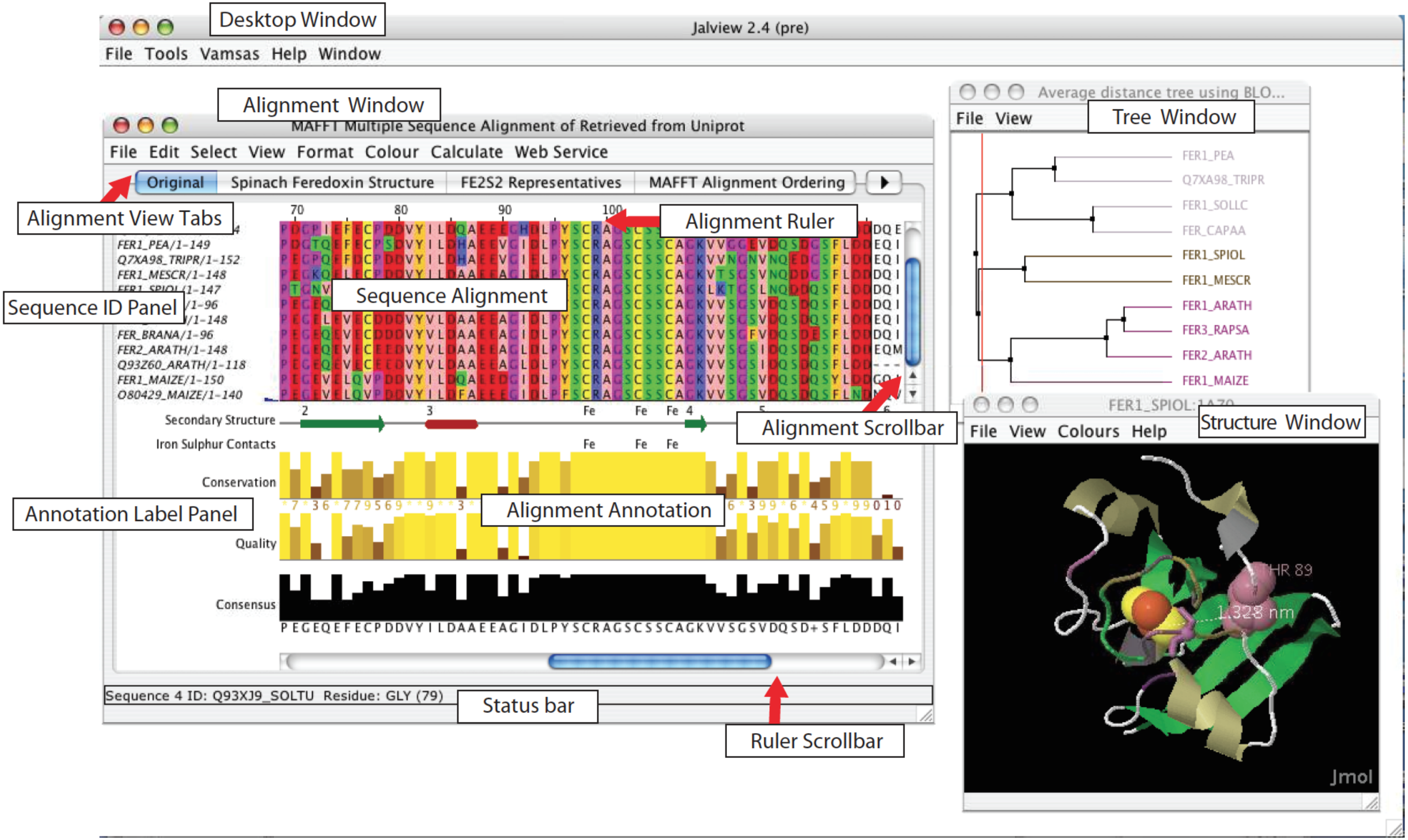
- Jan 4, 2013 Jalview in 2012 and 2013
- **Nov 12, 2012 Jalview 2.8 release and the new look www.jalview.org**
- **Oct 18, 2012 Registration now open for 3rd Jalview Residential Training Course anc**
- **Sep 22, 2011 Welcome to the Jalview Desktop news channel**

Jalview in 2012 and 2013
http://www.jalview.org/General/General-news/Jalview-in-2012-and-2013

2012 was quite a year here at jalview.org. A number of long running projects finally bore fruit with the launch of our new website and logo, and the release of Jalview Version 2.8. The November release of Jalview was the first to support JABAWS 2, which was launched in December 2011, and to include RNA visualization features developed by our 2010 and 2011 Google Summer of Code students: Lauren Lui and Jan Engelhart.

http://www.jalview.org/feeds/desktop/rss

Anatomy of Jalview: Figure 1.7



Sequence 4 ID: Q93XJ9_SOLTU Residue: GLY (79)

Ex 1 – starting Jalview

- Tasks
  - Modify user preferences
  - Test that you can load the example file manually

    http://www.jalview.org/examples/exampleFile_2_7.jar

Ex 1 – starting Jalview

- Tasks
  - Modify user preferences
  - Test that you can load the example file manually
    - http://www.jalview.org/examples/exampleFile_2_7.jar
- Questions
  - Where to find help ?
  - How to report a bug ?

Jalview Community

- Mailing lists
  - Discussion forum and developers forum
  - Links from http://www.jalview.org/community
- Jalview bug database
  - http://issues.jalview.org
    - Also indexed on google
- Jalview development info
  - http://www.jalview.org/development

Ex 2 - Navigation

- Tasks
  - Open the overview window for a view
  - Jump to a specific row and column with keyboard mode

Ex 2 - Navigation

- Tasks
  - Open the overview window for a view
  - Jump to a specific row and column with keyboard mode
- Questions
  - How do you locate a sequence or sequence position if you don't know its row/column ?
  - How do you find a sequence motif ?

Ex 3 Getting data into Jalview

- Tasks
  - Importing an alignment via a url, local file, or cut' n' paste
  - Getting an alignment from Pfam

Ex 3 Getting data into Jalview

- Tasks
  - Importing an alignment via a url, local file, or cut’ n’ paste
  - Getting an alignment from Pfam
- Questions
  - What happens when you drag a file onto an existing alignment ?
  - What is different about the alignment retrieved from Pfam ?
  - What if you want to load a *really* big alignment ?

Ex 4. Saving alignments

- Tasks
  - Save alignments in different formats

Ex 4. Saving alignments

- Tasks
  - Save alignments in different formats
- Questions
  - What’s the biggest difference between a BLC file and a pileup file ?
  - Why are Jalview projects useful ?

Ex **5**
selecting regions

- Task
  - Get used to marking parts of the alignment with the mouse and keyboard
  - Learn how to create and remove groups

Ex **5**
selecting regions

- Task
  - Get used to marking parts of the alignment with the mouse and keyboard
  - Learn how to create and remove groups

Question:

- What happens if you press the up/down arrow keys ?

Ex **7**
hiding & showing

Task

- Learn about hiding columns and rows in the alignment

Ex **7**
hiding & showing

Task

- Learn about hiding columns and rows in the alignment

Question:

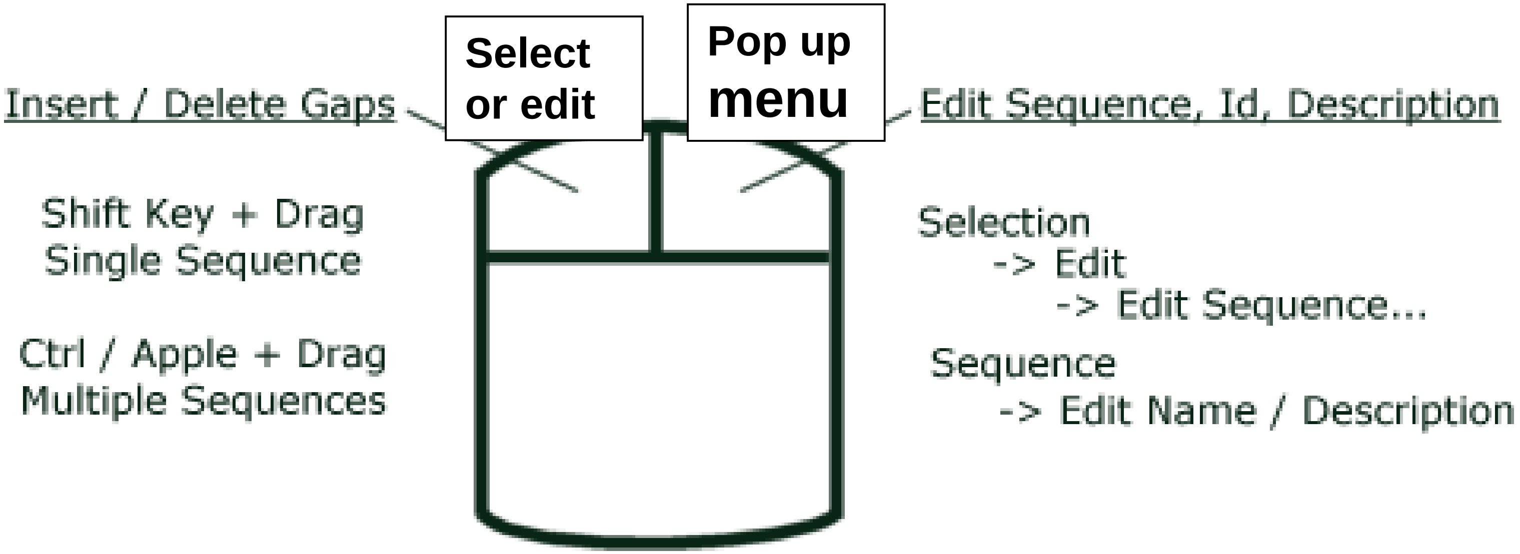
- Why would you create representative sequences ?

Ex 8 and 9
editing

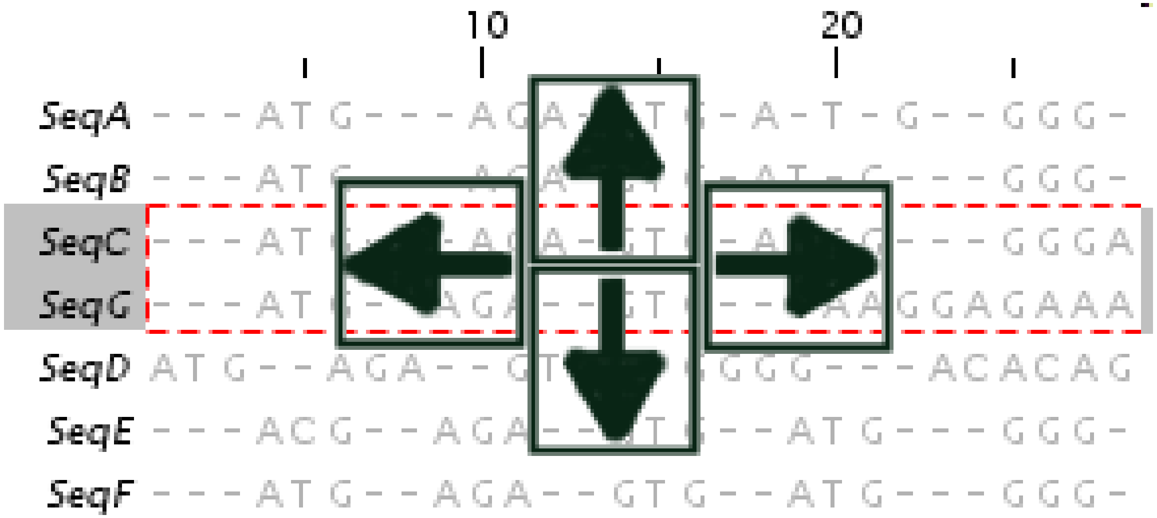
- Tasks
  - Get used to the mouse and keyboard based selection and alignment editing controls
  - Learn how to work on specific parts of an alignment

- This is homework !

How do I edit sequences in Jalview
http://www.jalview.org/examples/editing.html

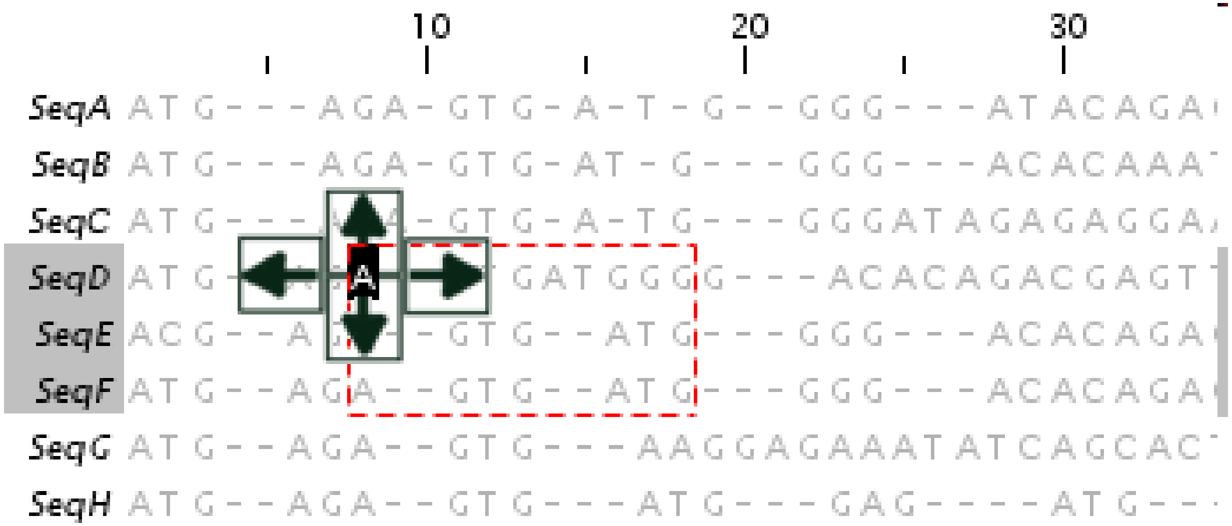


Selected sequences can be moved up and down or slid from left to right



http://www.jalview.org/examples/editing.html

F2 enables/disables keyboard mode



Cursor Keys - Move Cursor
Alt + Cursor Keys - Move Sequence
[X] Space - Insert [X] gap(s)
[X] Delete / Backspace - Delete [X] gap(s)
8 C - Move to Column 8
4 S - Move to Sequence 4
8,4<return> - Move to column 8, sequence 4
6 P - Move to Position 6
Q - Define the top left corner of selection area
M - Define the bottom right corner of selection area

__Windows 7 Users:__

Group Insert - use

Shift+Space

(not CTRL+Space)

Ex 10: Colouring

- Tasks
  - Learn how to colour all, or part of the alignment by
    - Amino acid property
    - Alignment conservation

Ex 10 : Colouring

- Tasks
  - Learn how to colour all, or part of the alignment by
    - Amino acid property
    - Alignment conservation
- Questions
  - Why is colouring the alignment useful ?
  - How would you highlight acidic residues in your alignment ?

Ex 12,13 – alignment layout and export

- Tasks
    - Adjust the alignment formatting options
        - Wrap
        - Sequence id margin
    - Export the alignment as a figure
        - HTML, EPS, SVG and PNG

Ex 12,13 – alignment layout and export

- Tasks
  - Adjust the alignment formatting options
    - Wrap
    - Sequence id margin
  - Export the alignment as a figure
    - HTML, EPS, SVG and PNG
- Questions
  - How do you control the number of columns shown in wrapped mode ?
  - How can you easily experiment with different alignment figure layouts ?
  - What programs can edit EPS files ?

- End of Session 1
  - Loading/saving
  - Navigation/Selecting
  - Colouring & Figures

- Session 2
  - Alignment & analysis
    - Trees
    - PCA
    - Subfamily analysis

Jalview alignment
Section 2.4 Exercise 25

- Tasks
  - Align sequences using different methods
    - Use the Webservices' 'alignment' submenu
  - Explore how hidden regions affect alignment jobs.

Jalview alignment
Section 2.4 Exercise 25

- Tasks
  - Align sequences using different methods
    - Use the Webservices' ‘alignment’ submenu
  - Explore how hidden regions affect alignment jobs.
- Questions
  - Why does jalview run several jobs if the input includes hidden regions ?
  - What does ‘re-alignment’ mean ?

Alignment Job Parameter Settings

Browse or edit to change name of set

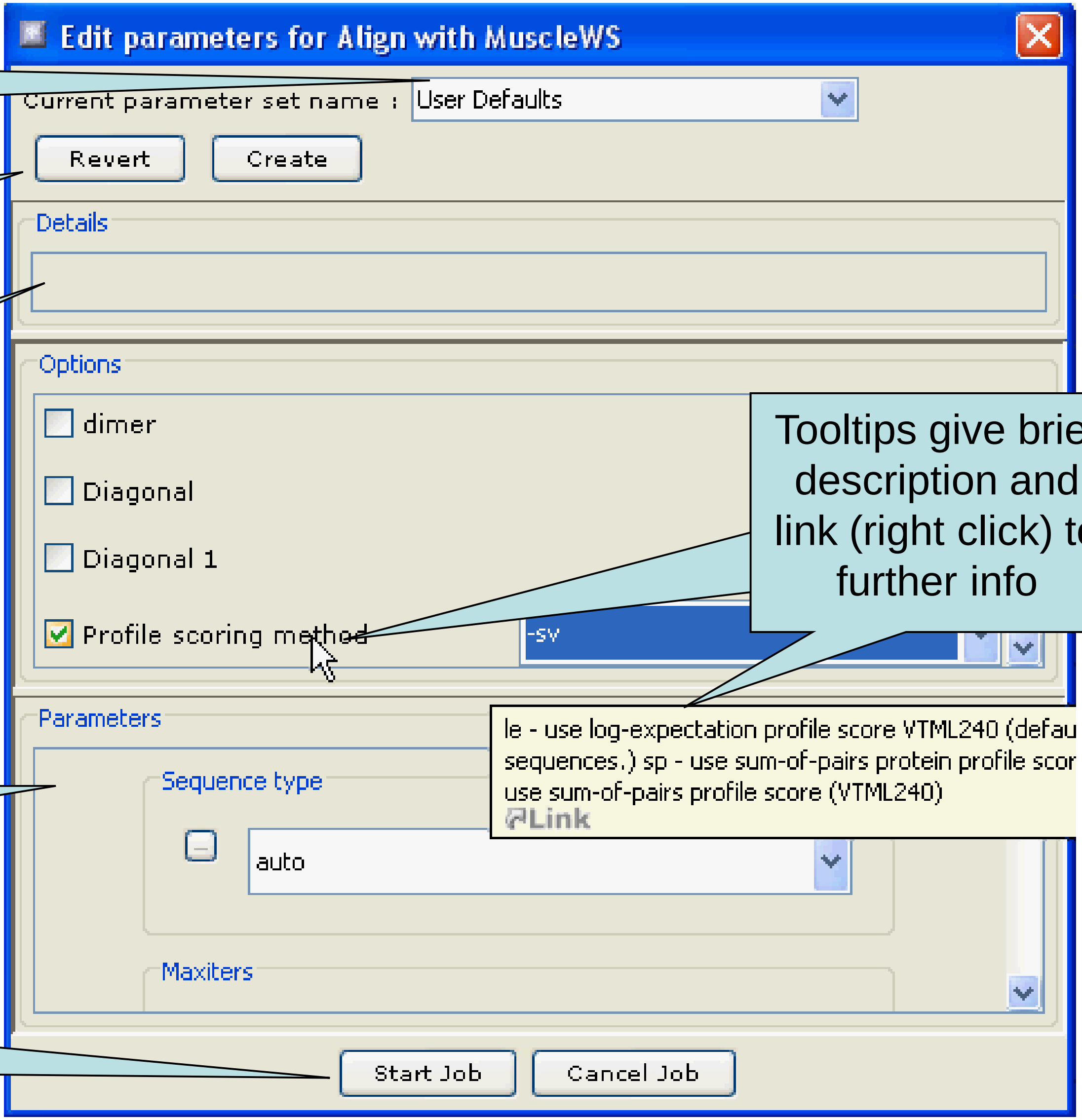
Buttons appear to create, update, rename or delete user settings.

text box to add notes for the parameter set

Parameters contains more complex settings

Start job with current settings or cancel.

Tooltips give brief description and link (right click) to further info



Edit parameters for Align with MuscleWS

Current parameter set name : User Defaults

☐ dimer

☐ Diagonal

☐ Diagonal 1

☑ Profile scoring method -sv

auto

le - use log-expectation profile score VTML240 (defau
sequences.) sp - use sum-of-pairs protein profile scor
use sum-of-pairs profile score (VTML240)
Link

Why change alignment parameters ?

Jaba Alignment Exercise

- Task
  - Run the alignment from step ***b*** of ex. 25 using the JABA clustalW service
    1. Run with default settings
    2. Use the 'Edit parameters' dialog to run an alignment with the following:
       - Gap opening (internal and end gaps) = 3
       - Gap Extension = 0.05
  - Compare the two alignments. You may want to save them for later, too.

- Questions
  - What effect has modifying the gap penalties had on the feredoxin alignment ?

- Session 2
    - Alignment & analysis
        - Trees
        - PCA
        - Subfamily analysis

Alignment analysis – Section 2.2

- Principal component analysis
- Phylogenetic trees
- Redundancy removal
- Tree based conservation analysis
- Subdividing alignment by mutation

Phylogenetic analysis and Jalview

- Built in tree methods
  - UPGMA
    - Fast, simple, but not reliable for phylogenetic inferrence
  - Neighbour joining
    - Slower than UPGMA
    - Useful for a first approximation
      - NJ does not work well for very divergent sequence sets
        - Need to add in close relatives to get an idea of topology
- Import trees from elsewhere
  - Load a Newick format tree file onto an alignment from another program

PCA and Phylogeny Section 2.2 Exercise 17 and 18

- Tasks
  - Calculate Principal component analyses (PCAs) and trees on the feredoxin alignment
  - Explore the use of the interactive tree viewer
    - Use it to select subgroups on the alignment.

PCA and Phylogeny Section 2.2 Exercise 17 and 18

- Tasks
  - Calculate Principal component analyses (PCAs) and trees on the feredoxin alignment
  - Explore the use of the interactive tree viewer
    - Use it to select subgroups on the alignment.
- Questions
  - What is the role of BLOSUM62, PAM250 and ‘Percentage Identity’ in the tree building process ?

Tree based conservation analysis
Sect. 2.2.3 Exercise 19

- “Poor man’s” character inference analysis
  - Compare conservation patterns within and between branches of a tree
- Task
    - Use interactive tree viewer to subdivide alignment and identify difference in conservation pattern

Tree based conservation analysis
Sect. 2.2.3 Exercise 19

- “Poor man’s” character inference analysis
  - Compare conservation patterns within and between branches of a tree
- Task
    - Use interactive tree viewer to subdivide alignment and identify difference in conservation pattern
- Question
  - How can you tell which differences are important ?

Sub-groups and Sub-group Annotation Exercise 21

- Task
  - Use the group consensus sequence logos to more easily compare tree subgroups
  - Use ‘Make groups for selection’ to subdivide groups by specific mutation

Sub-groups and Sub-group Annotation
Exercise 21

- Task
  - Use the group consensus sequence logos to more easily compare tree subgroups
  - Use ‘Make groups for selection’ to subdivide groups by specific mutation
- Question
  - How can you navigate the sub-groups of a large alignment ?