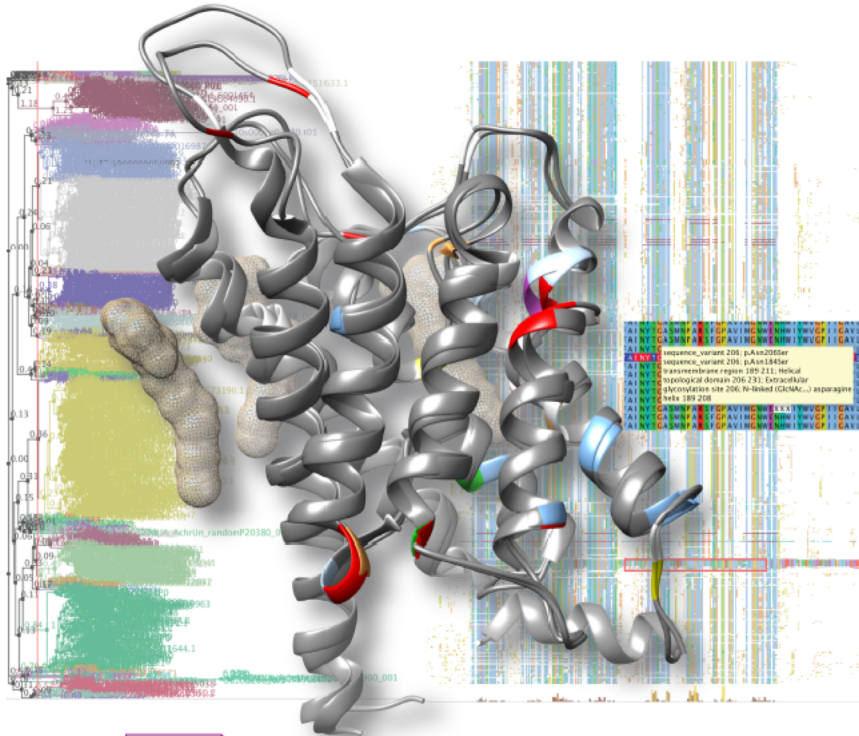


Day 1 – Course Materials

<http://bit.ly/jv2018liv>



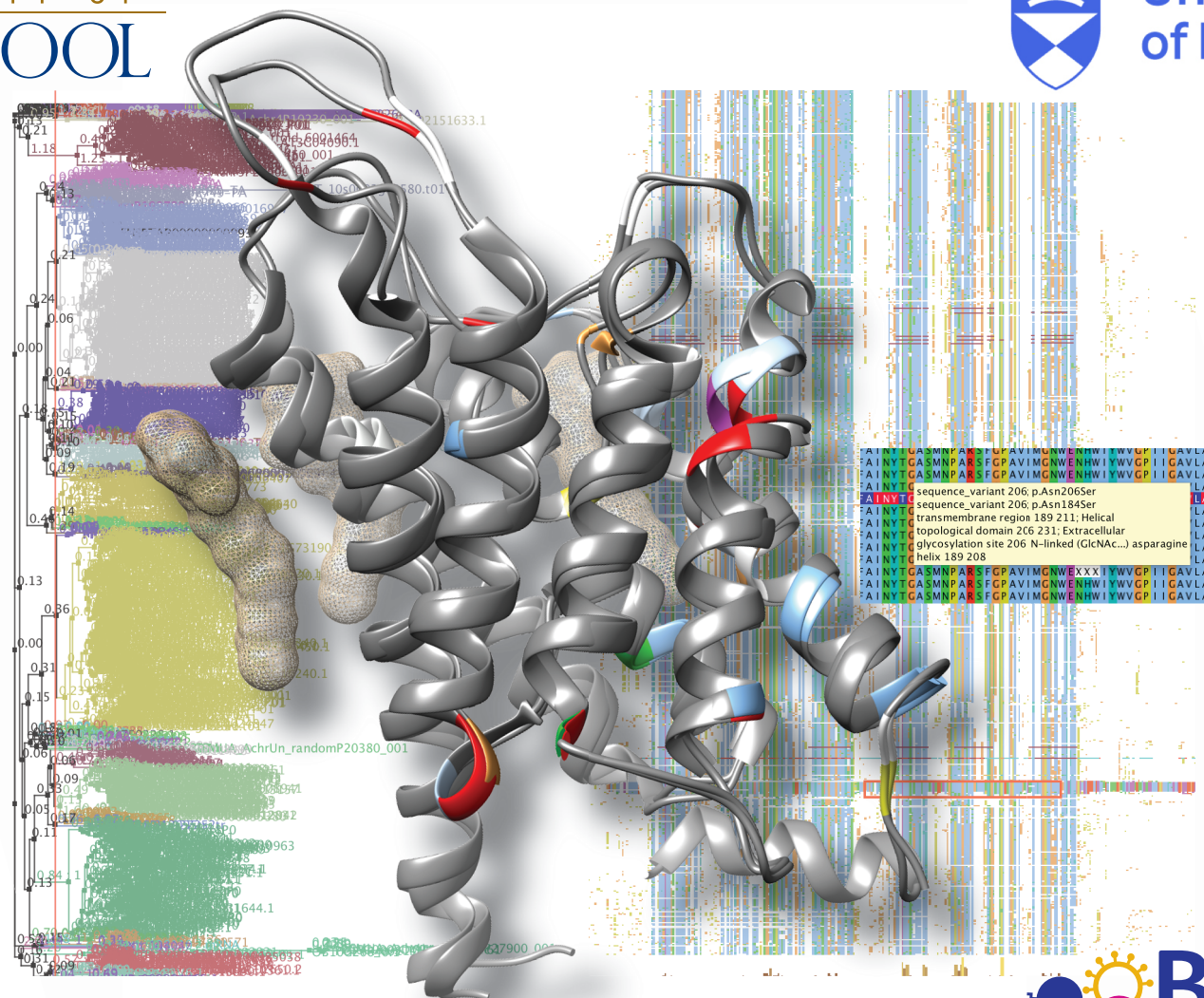
Please

- Login to a computer
- or*
- Connect to WiFi

And open the course materials page



THE BARTON GROUP



elixir
UNITED
KINGDOM



Mon 14th May

9.15am Introduction to Jalview

Session 1

Loading sequences, structures and alignments

10.30 - 11am Coffee

Session 2

Multiple alignment Creation and Alignment Reliability

12.30 - 1.30pm Lunch

Session 3

Trees & subfamily analysis of Proteins & CDS

3 – 3.30pm Coffee

Session 4

Protein 2ndary structure & disorder, and annotation.

Tuesday 15th May

9.15am – Day 2 Start

Session 1: Introduction to Protein 3D Structure

Viewing and Comparing 3D structures with Jalview and UCSF Chimera, Mapping sequence conservation to Biological Assemblies

10.30 - 11am Coffee

Session 2: Functional analysis of Protein Structures

12.30 - 1.30pm Lunch

Session 3: 3D Modelling for Functional Analysis

Homology modelling and model quality evaluation,

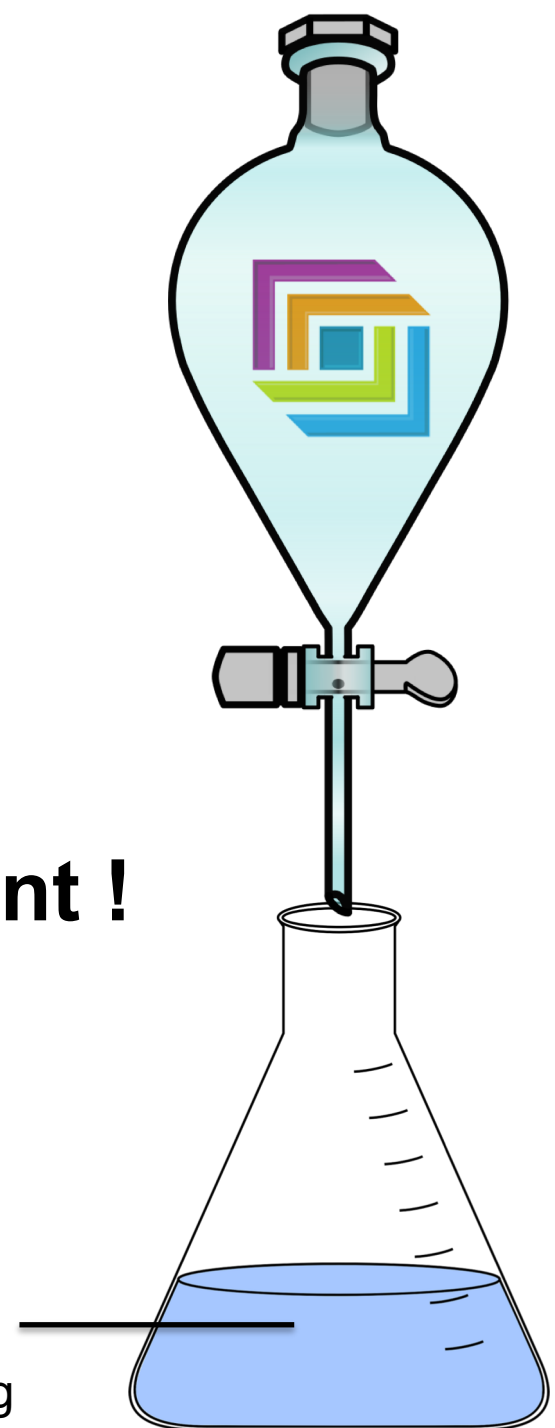
Ab initio models and searching for fold matches

3.30 – 4pm Coffee

Analyse your own protein sequences and structures

This course is an experiment !

Functional Analysis
of 3D Structure
3D ab Initio &
Homology Modelling



Day 1 – Course Materials

<http://www.jalview.org/tutorial/training-materials/2018/Liverpool/May/>

- (some of) These slides
- Jalview v2.10.3 Manual (v1.9.1)
- Additional exercises



Jalview

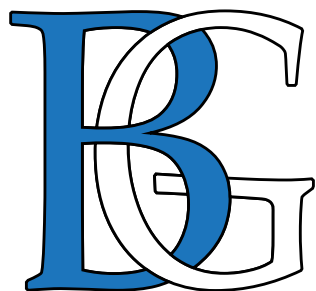
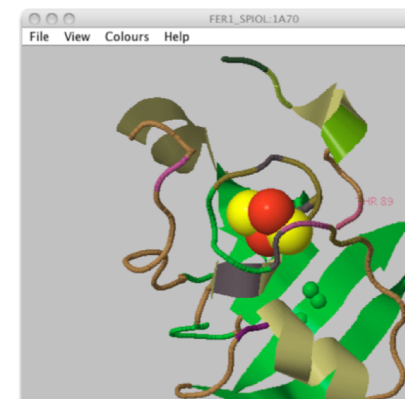
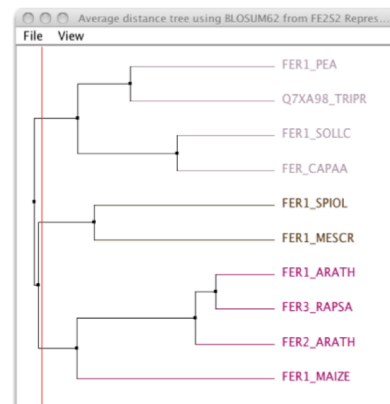
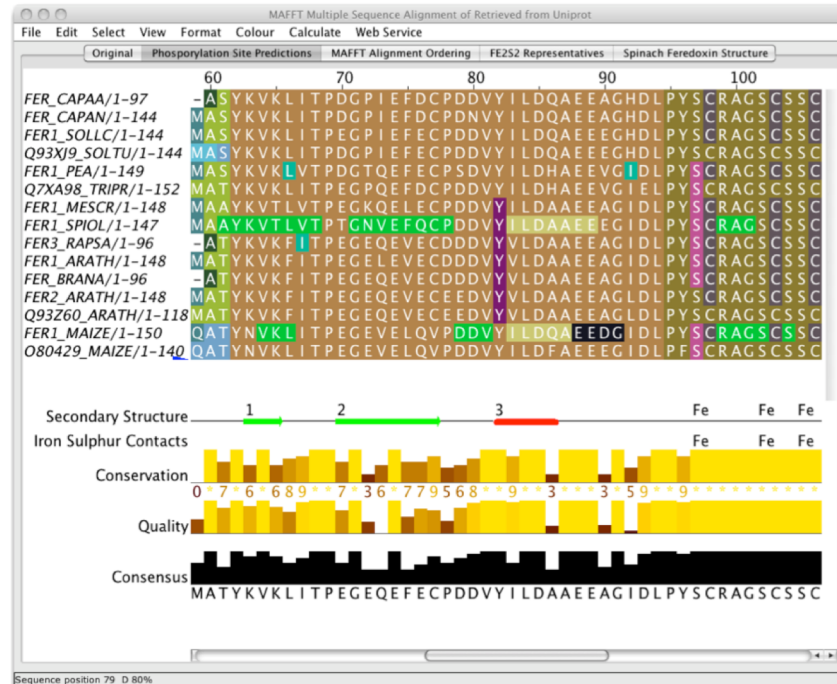
Java

Standalone or
web based

alignment
&
annotation,
tree, and
structure
viewer
and
editor

Available at

www.jalview.org



THE BARTON GROUP

www.compbio.dundee.ac.uk

The Jalview Core Team



Geoff Barton



Tochukwu
(Charles)
Ofoegbu
*Now at
Imperial College*



Kira
Mourao



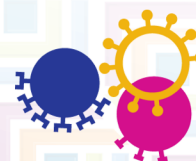
Mungo
Carstairs



Suzanne Duce
Jalview &
Dundee
Protein
Resource

supported by

wellcometrust



BBSRC
bioscience for the future

Bioinformatics
data is not fun to
read....

Bioinformatics
data is not fun to
read....

FASTA

HC class II antigen

[illegible]

GFF

PDB

Newick

CSV

HC class II antigen

Sequence	Feature	Start	End	Score	Probability
Q95IE6	uniprot non_terminal_residue	1	88	0.0	0.0
Q95IE6	uniprot non_terminal_residue	88	88	0.0	0.0
Q8MGZ9	uniprot non_terminal_residue	1	1	0.0	0.0
Q8MGZ9	uniprot non_terminal_residue	89	89	0.0	0.0
Q8HWS7	uniprot non_terminal_residue	1	1	0.0	0.0
Q8HWS7	uniprot non_terminal_residue	89	89	0.0	0.0
Q30167	uniprot signal_peptide	1	29	0.0	0.0
Q30167	uniprot mature_protein_region	30	266	0.0	0.0
Q30167	uniprot extramembrane	30	227	0.0	0.0
Q30167	uniprot transmembrane	228	250	0.0	0.0
Q30167	uniprot extramembrane	251	266	0.0	0.0
Q30167	uniprot polypeptide_domain	126	211	0.0	0.0
Q30167	uniprot polypeptide_region	30	124	0.0	0.0

HC class II

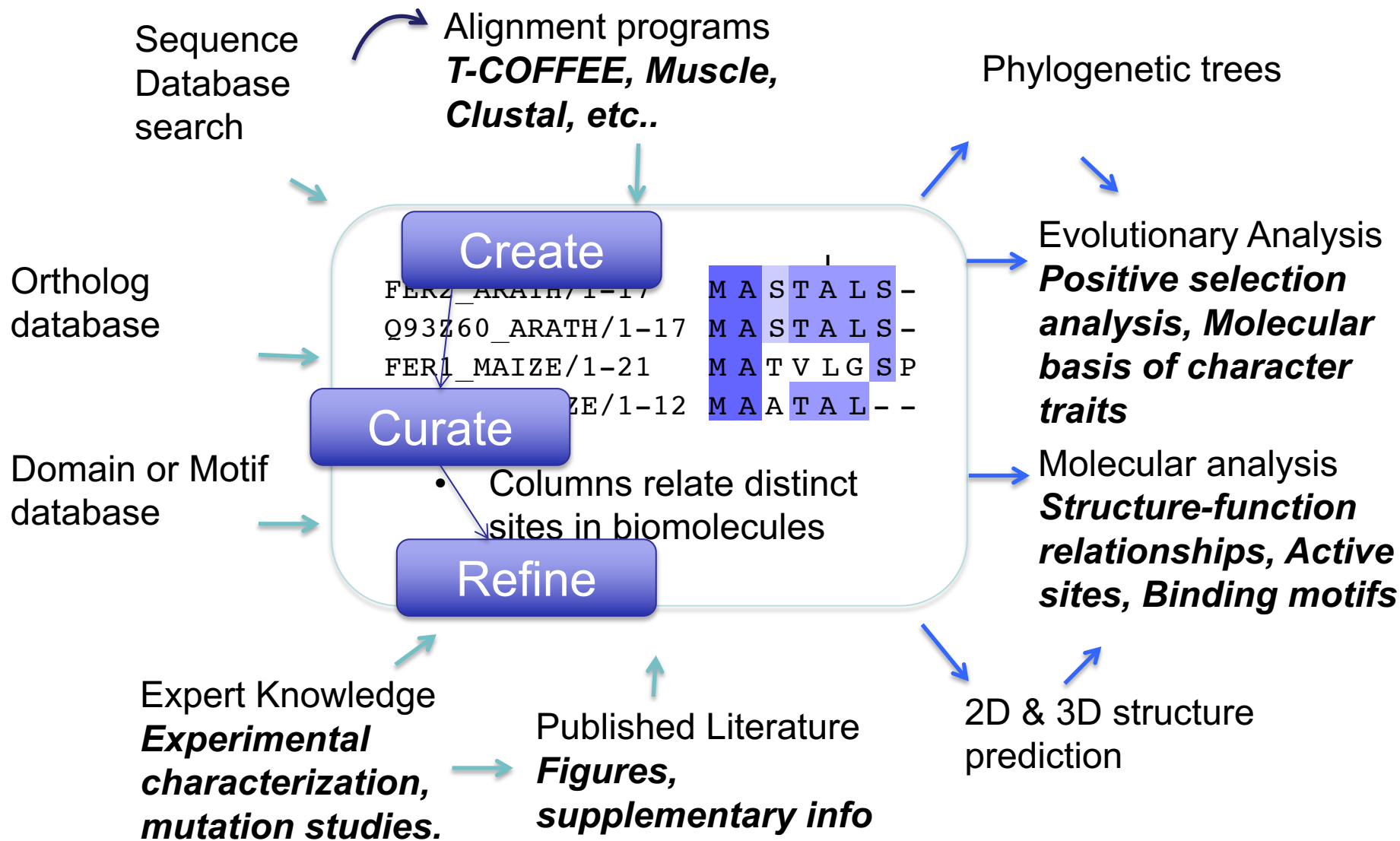
Protein	Uniprot	Modification	Count	Score	Value	Value	Value	Value	Value	Value
Q30167	uniprot	glycosylated residue	48	48	0.0	0.0	0.0	0.0	0.0	0.0
Q30167	uniprot	disulfide crosslinked residues	44	108	0.0	0.0	0.0	0.0	0.0	0.0
Q30167	uniprot	disulfide crosslinked residues	146	202	0.0	0.0	0.0	0.0	0.0	0.0
Q30167	uniprot	disulfide crosslinked residues	236	262	0.0	0.0	0.0	0.0	0.0	0.0

Structure

Structure						
ATOM	9	OE1	GLU A 3, 36.2	15.48		
ATOM	10	OE2	GLU A 3, 37.5	16.94		
ATOM	11	N	GLU A 4, 38.96	20.415		
ATOM	12	CA	GLU A 4, 39.6	21.562	67.120	11.132
ATOM	13	C	GLU A 4, 39.6	21.313	67.458	12.588
ATOM	14	O	GLU A 4, 39.5	22.169	67.189	13.445
ATOM	15	CB	GLU A 4, 39.6	22.323	68.269	10.476
ATOM	16	CG	GLU A 4, 39.6	23.588	67.860	9.745
ATOM	17	CD	GLU A 4, 39.6	24.007	68.917	7.771
ATOM	18	OE1	GLU A 4, 42.8	24.001	70.103	9.183
ATOM	19	OE2	GLU A 4, 42.8	24.293	68.587	7.581
ATOM	20	N	HIS A 5, 42.8	20.214	68.139	12.857
ATOM	21	CA	HIS A 5, 42.8	19.917	68.535	14.210
ATOM	22	C	HIS A 5, 42.8	18.443	68.716	14.290
ATOM	23	O	HIS A 5, 42.8	17.776	68.988	13.272
ATOM	24	CB	HIS A 5, 42.8	20.544	69.894	14.540
ATOM	25	CG	HIS A 5, 42.8	22.039	69.961	14.340
ATOM	26	ND1	HIS A 5, 42.8	22.946	69.677	15.344
ATOM	27	CD2	HIS A 5, 42.8	22.779	70.275	13.249
ATOM	28	CE1	HIS A 5, 42.8	24.176	69.800	14.882
ATOM	29	NE2	HIS A 5, 42.8	24.103	70.161	13.612

7.1.

[illegible]

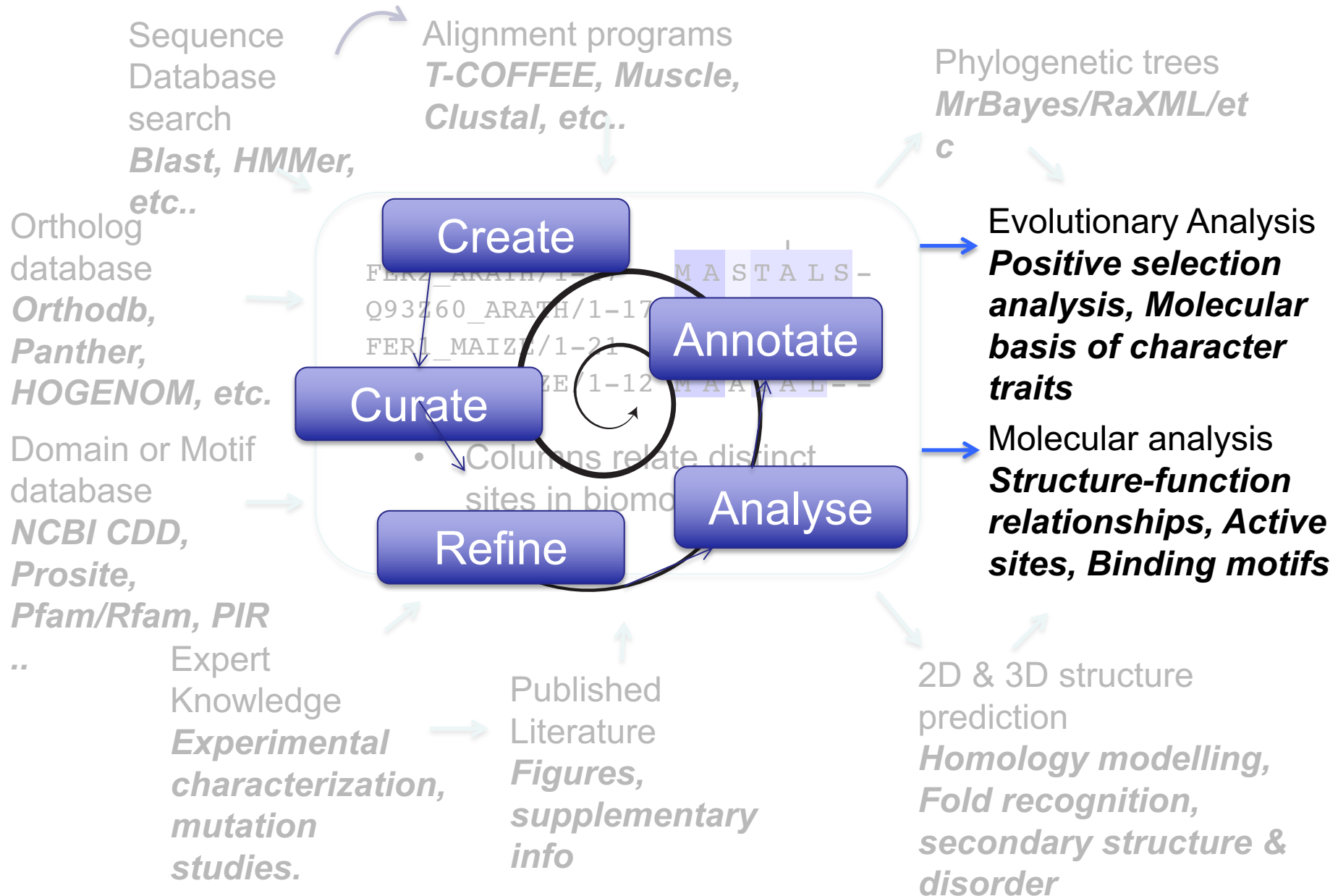


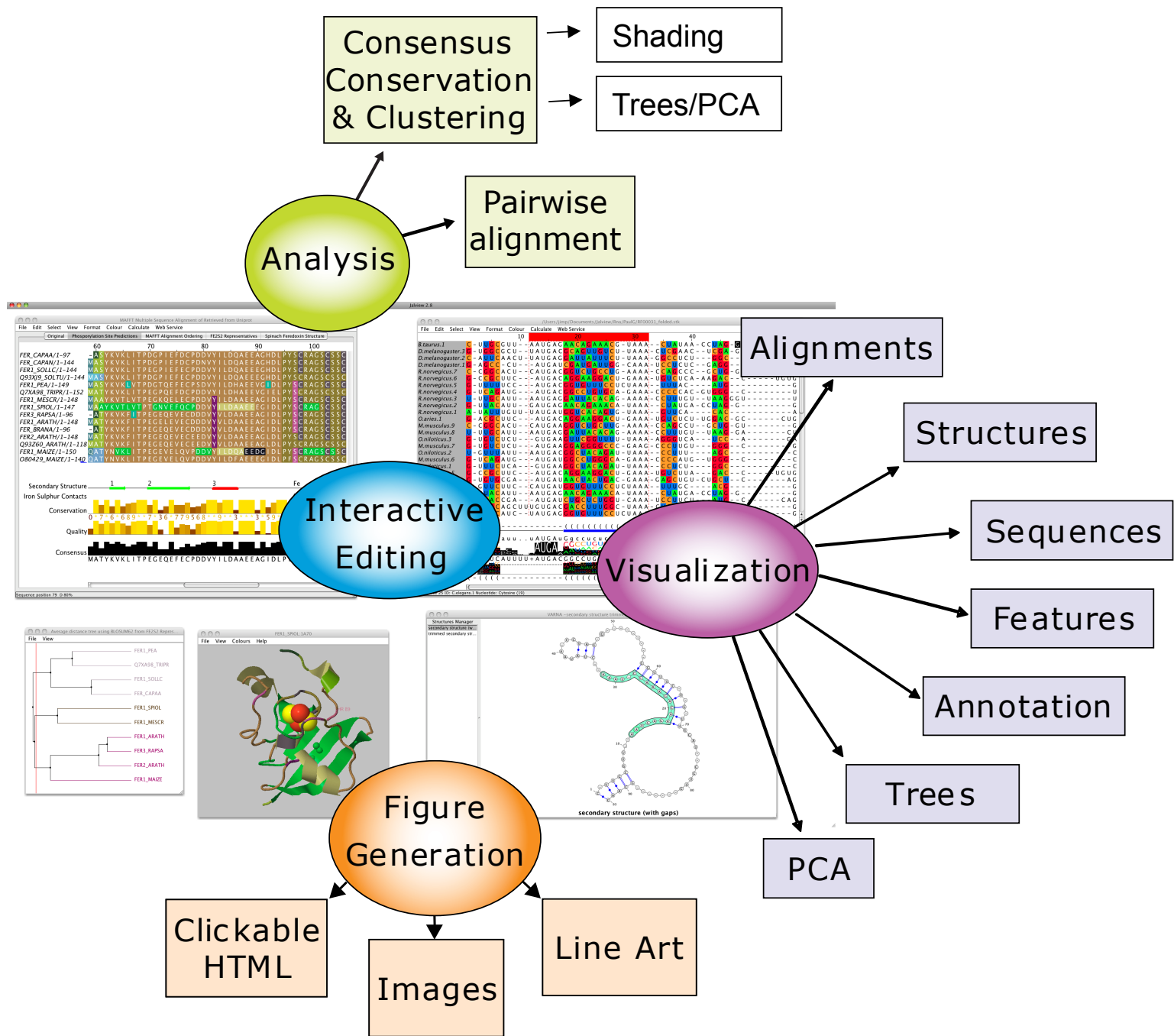
Curating Alignments

- Alignments sometimes need manual curation
- Correct alignments conserve common properties
- Shading can highlight differences

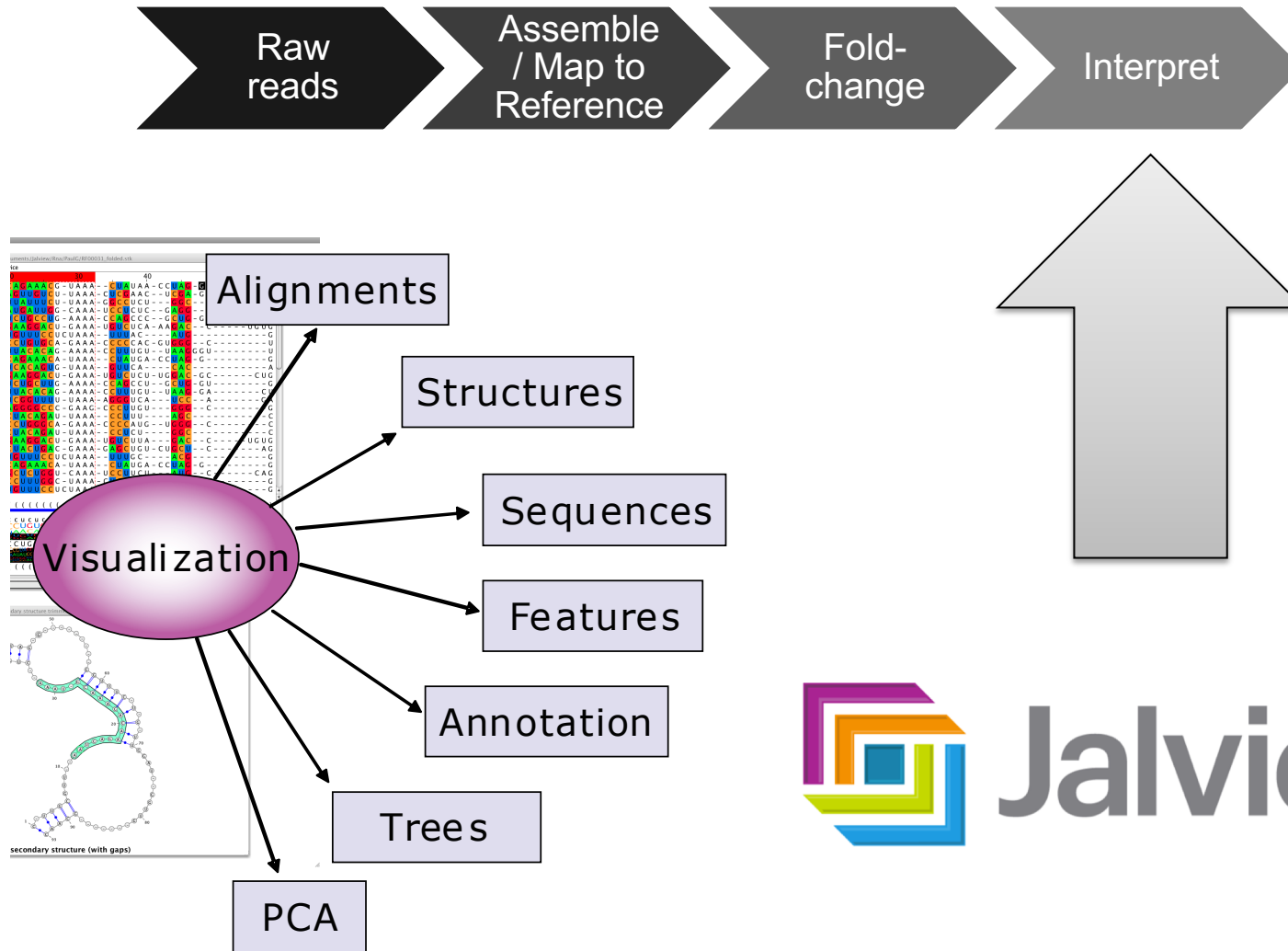
FER_CAPAA/28-82	AEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDQLEEGWVLTCAVAYPQ
FER_CAPAN/75-129	AEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDQLEEGWVLTCAVAYPQ
FER1_LYCES/75-129	AEEEGHDLPYSCRAGSCSSCAGKVTAGSVDQSDGNFLDEDQEAAGFVLTCVAYPK
Q93XJ9_SOLTU/75-129	AEEEGHDLPYSCRAGSCSSCAGKVTAGTVDQSDGKFLDDQEAAGFVLTCVAYPK
FER1_PEA/80-134	AEEVGIDLPYSCRAGSCSSCAGKVVGGEVDQSDGSFLDDEQIEAGFVLTCVAYPT
Q7XA98_TRIPR/83-137	AEEVGIELPYSCRAGSCSSCAGKVVGNGVNGEDGSFLDDEQIEGGWVLTCAVAFPT
FER1_MESCR/79-133	AEEAGIDLPYSCRAGSCSSCAGKVTSGSVNQDDGSFLDDQIEKGWVLTCAVAYPT
FER1_SPIOL/78-132	AEEEGIDLPYSCRAGSCSSCAGKLTGSLNQDDQSFLLDDQIDEGWVLTCAAYPV
FER3_RAPSA/28-82	AEEAGIDLPYSCRAGSCSSCAGKVVS GSVDQSDQSFLDDQIAEGFVLTCAYPT
FER_ARATH/80-134	AEEAGIDLPYSCRAGSCSSCAGKVVS GSVDQSDQSFLDDEQIEGEGFVLTCAYPT
FER_BRANA/28-82	AEEAGIDLPYSCRAGSCSSCAGKVVS GFVDQSDQSFLDDQIAEGFVLTCAYPT
FER2_ARATH/80-134	AEEAGLDLPYSCRAGSCSSCAGKVVS GSIDQSDQSFLDDEQMSEGYVLTCVAYPT
FER1_MAIZE/80-134	AEEDGIDLPYSCRAGSCSSCAGKVVS GSVDQSDQSYLDDGQIADGWVLTCHAYPT
O80429_MAIZE/72-126	AEEEGIDLPFSCRAGSCSSCAGKVVS GSVDQSDQSFLNDNQVADGWVLTCAAYPT

Alignments are central ...





Jalview is designed for gene-centric analysis



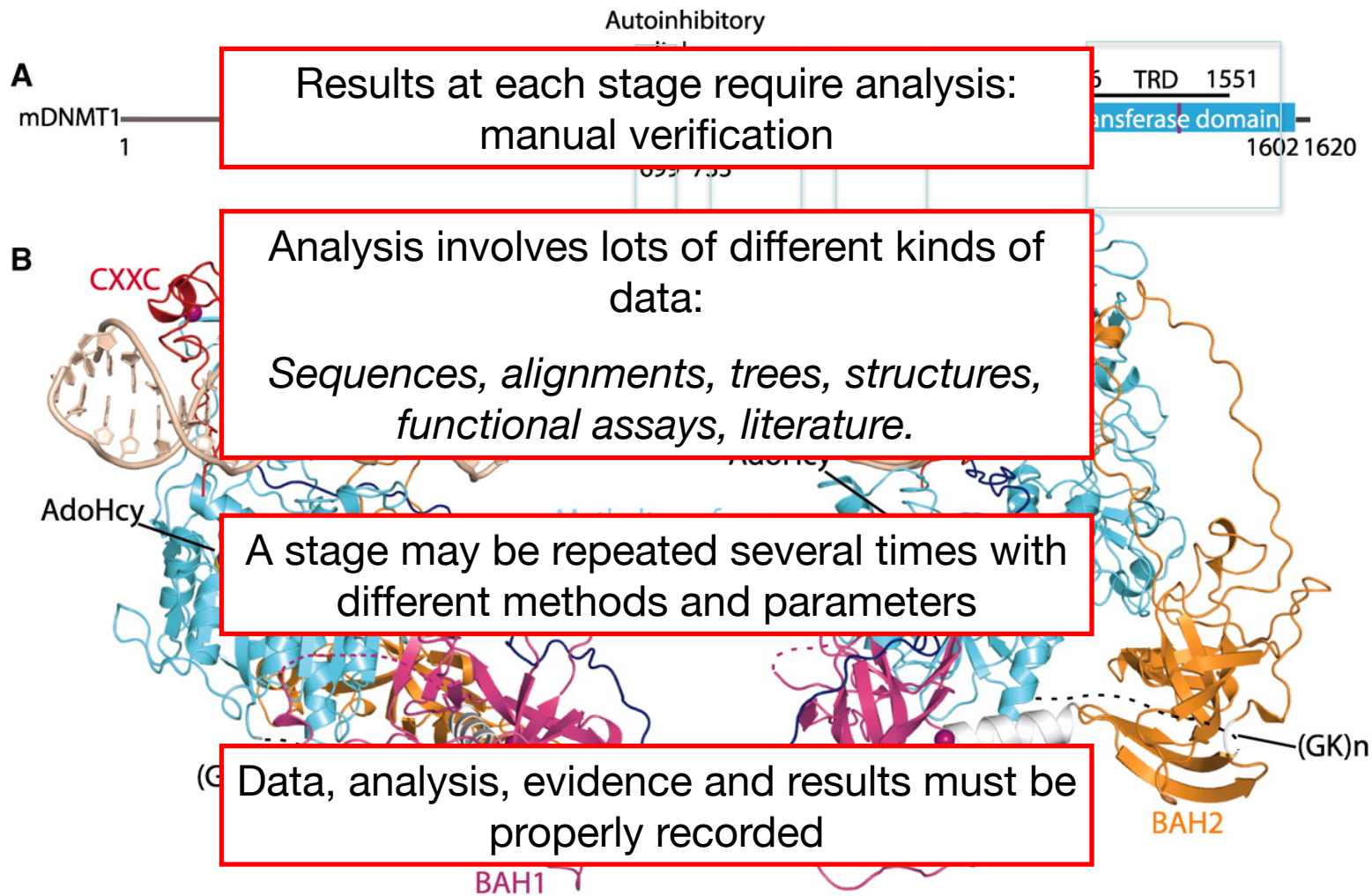


Fig. 1 Structural overview of mDNMT1(650–1602)–DNA 19-nucleotide oligomer complex with bound AdoHcy.

J Song et al. Science 2011;331:1036-1040

• Key qualities for a workbench

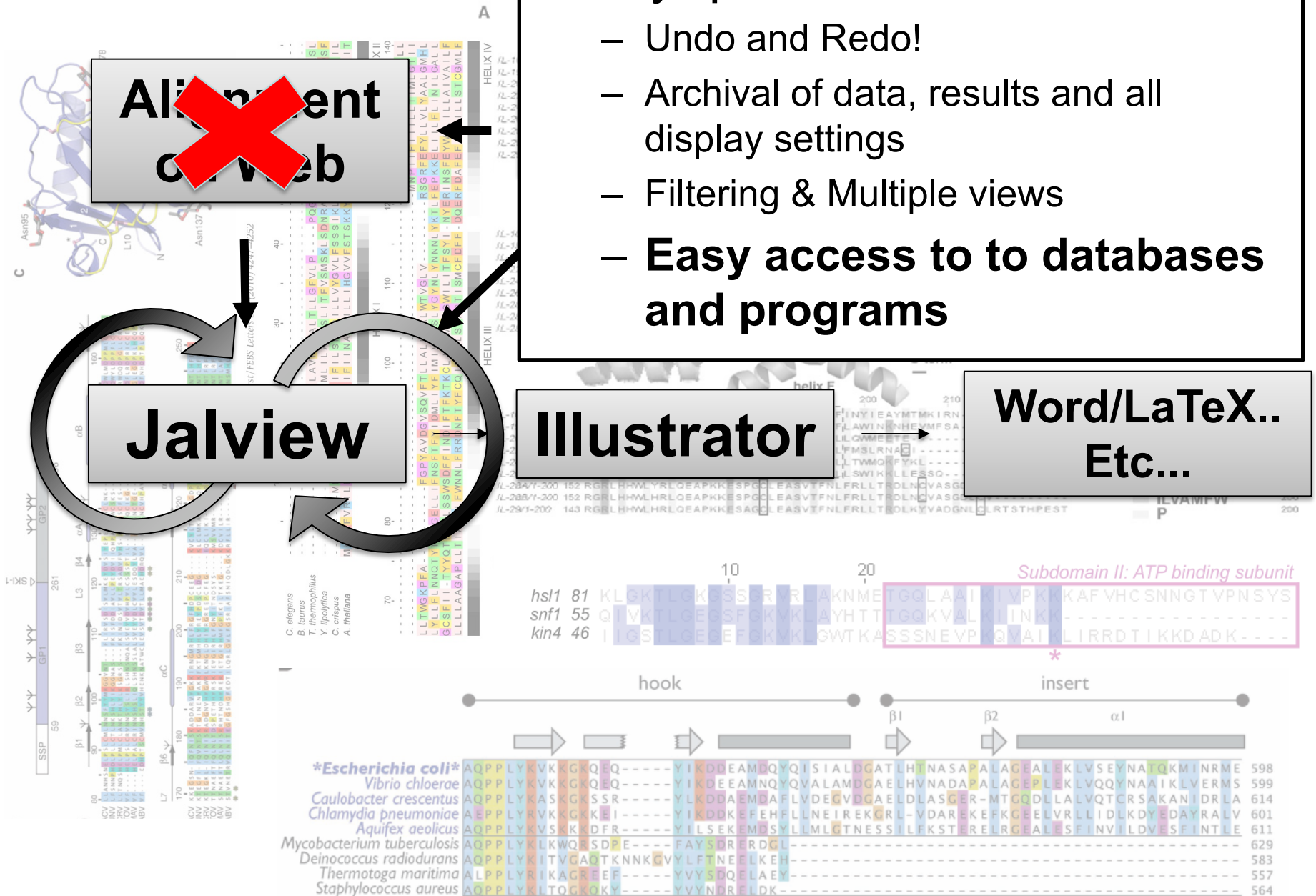
- Undo and Redo!
- Archival of data, results and all display settings
- Filtering & Multiple views
- **Easy access to to databases and programs**

Alignment
on web

Jalview

Illustrator

Word/LaTeX..
Etc...



Jalview
Launch
Buttons

Launch Jalview Applet

Launch Jalview Desktop

Home

About

Help

Community

Development

Training

Download

Latest News

Jalview in 2017
2013

Posted On: 10-07-2017

Help and
documentation

the new look

www.jalview.org

Posted On: 12-11-2012

GENERAL

Registration now open
for 3rd Jalview
Residential Training
Course and Hackathon

Posted On: 19-10-2012

TRAINING

Jalview is a free program for multiple sequence alignment editing, visualisation and analysis. Use Jalview to view sequence alignment, analyse them using phylogenetic trees and principal component analysis (PCA) plots and explore molecular structure and annotation.

Installation
packages
and source

Jalview
training news
and course
dates

Jalview
Community

Jalview
Development and
release history

Analysis Services

Protein 2ndary structure
Jpred3

JABAWS^{2.0}

Biological
Data
Services

Pfam

Rfam

The Jalview
Javascript API

applet
parameters

Alignment
And
Annotation

<http://www.jalview.org>



Jalview

Online Training



Jalview Online Training

Subscribe 112

Home

Videos

Playlists

Channels

Discussion

About



Already subscribed



The Jalview Team based at School of Life Scien...



Jalview

The Jalview Team based at School of Life Sciences, University of Dundee, UK

487 views 1 year ago

Jalview Online Training YouTube Channel is a free bioinformatics educational course suitable for Undergraduate and PhD students. Playlists include (a) Getting Started; (b) Selecting and Editing Sequences; (c) Colour, Appearance and Figure Generation; (d) Jalview Dundee.

This video features Dr Suzanne Duce, Dr Jim Procter, Mr Mungo Carstairs, Mr Charles Tochukwu Ofoegbu, Dr Alexey Drozdetskiy, Mr Dmitry Finkelbergs and Pr...



0:04 / 2:38



Getting Started

▶ Play all



What is Jalview?

2:03

What is Jalview multiple sequence alignment software?

Jalview Online Training

3,415 views • 2 years ago

Navigating the Jalview Interface

WATCHED



5:28

Navigating the Jalview desktop with interactive alignment, tree,...

Jalview Online Training

2,977 views • 2 years ago



Launching Jalview



3:52

How to launch the multiple sequence alignment viewer...

Jalview Online Training

1,914 views • 2 years ago



5:08

Loading protein sequences and alignments into Jalview

Jalview Online Training

2,780 views • 2 years ago

96

Selecting & Editing

Selecting residues, sequences and columns of proteins, RNA...

Jalview Online Training

783 views • 1 year ago

Selecting and exporting groups of protein, RNA or DNA residue...

Jalview Online Training

411 views • 1 year ago

Reordering and hiding residues in sequence alignment

Jalview Online Training

387 views • 1 year ago

Editing groups in proteins, RNA and DNA alignments using Jalview

Jalview Online Training

320 views • 1 year ago

Colour, Appearance and Figure Generation

Colouring proteins, RNA and DNA residues and sequences in an...

Jalview Online Training

576 views • 8 months ago

Customising the appearance of the Jalview alignment window f...

Jalview Online Training

367 views • 8 months ago

Exporting data and saving images in Jalview

Jalview Online Training

299 views • 8 months ago

Exporting data and Saving image

3:21

minutes

Analysis and annotation

Multiple Sequence Alignment and Analysis with Jalview

Jalview Online Training

184 views • 2 months ago

Principal components analysis of protein and nucleotide...

Jalview Online Training

54 views • 1 month ago

Calculating and Displaying Trees in Jalview

Jalview Online Training

65 views • 4 weeks ago

Calculating Trees in Jalview

6:44

Accessing Jmol and UCSF Chimera 3D Molecular Structure Viewers in Jalview

Accessing Jmol and UCSF Chimera 3D molecular structur...

Jalview Online Training

48 views • 4 weeks ago

4:45

Printed Manual & Online Videos

- Basics
- Selecting and Editing Sequences
- Colouring Sequences and Figure Generation
- Annotation and Features
- Multiple Sequence Alignment
- Analysis of Alignments
- Working with 3D structures
- Protein sequence analysis and structure prediction
- DNA and RNA Sequences
- Webservices

Getting started

Selecting & Editing

Colour, Appearance and Figure Generation

Analysis and annotation

Not covered today – some exercises out of date

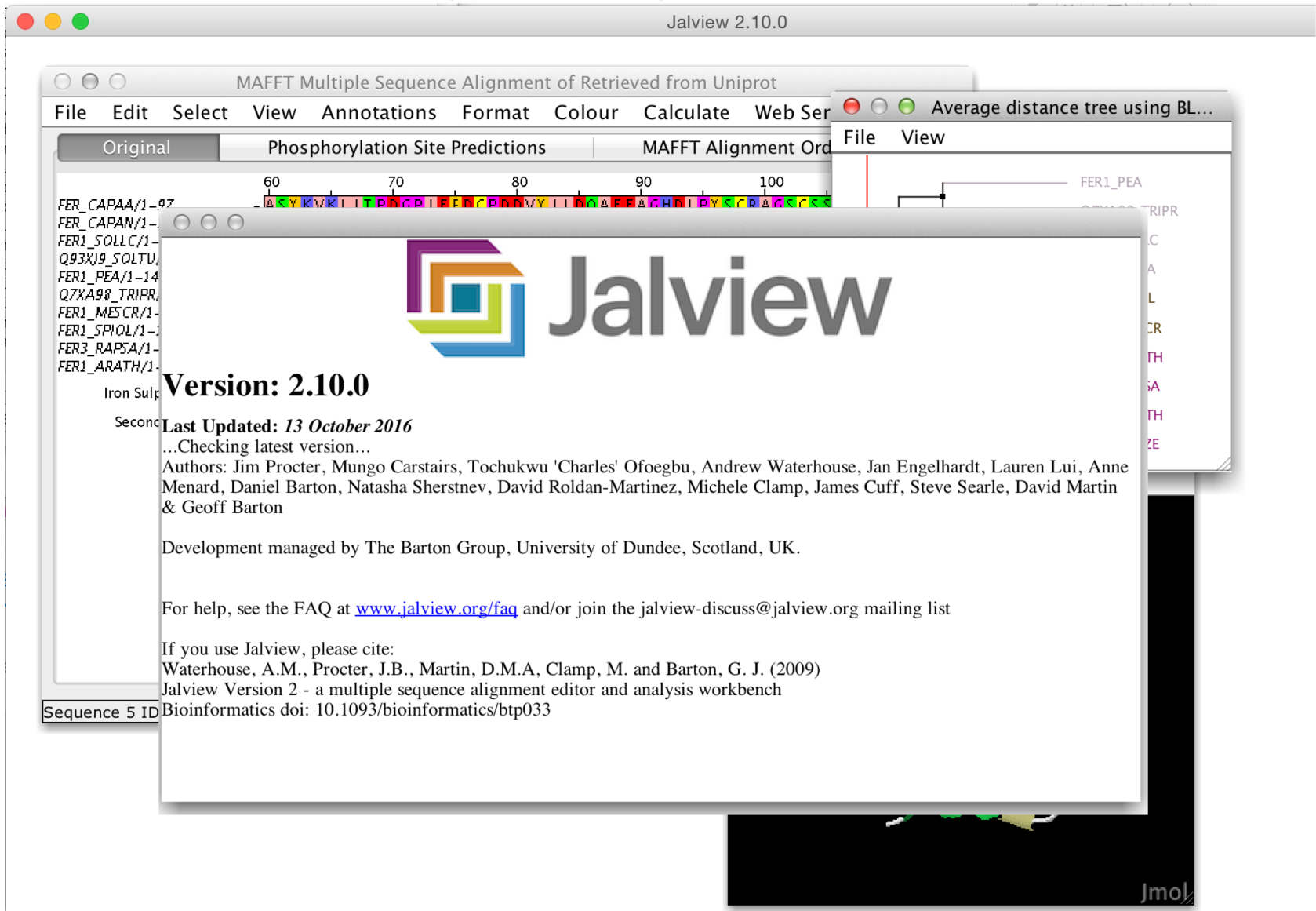
Ex 1 – starting The Jalview Desktop

PAGE 7

Pick the webstart link on the tutorial page
that works best for your computer !

~ 1G less than your physical memory
Or ~45% on older O.Ss (Java 1.7)

Launching the jalview desktop



Jalview 2.10.0

MAFFT Multiple Sequence Alignment of Retrieved from Uniprot

File Edit Select View Annotations Format Colour Calculate Web Services

Original Phosphorylation Site Predictions MAFFT Alignment Order

60 70 80 90 100

FER_CAPAA/1-97
FER_CAPAN/1-
FER1_SOLLIC/1-
Q93XJ9_SOLTU/
FER1_PEA/1-14
Q7XA98_TRIPR/
FER1_MESCR/1-
FER1_SPIOL/1-
FER3_RAPSA/1-
FER1_ARATH/1-

Iron Sulphur
Second

Version: 2.10.0
Last Updated: 13 October 2016
...Checking latest version...
Authors: Jim Procter, Mungo Carstairs, Tochukwu 'Charles' Ofoegbu, Andrew Waterhouse, Jan Engelhardt, Lauren Lui, Anne Menard, Daniel Barton, Natasha Sherstnev, David Roldan-Martinez, 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/faq 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

Sequence 5 ID

FER1_PEA
TRIPR
C
A
L
CR
TH
SA
TH
ZE

Jmol

Jalview 2.10.0

MAFFT Multiple Sequence Alignment of Retrieved from Uniprot

File Edit Select View Annotations Format Colour Calculate Web Ser

Original

Phosphorylation Site Predictions

MAFFT Alignment Ord

FER_CAPAA/1-97
FER_CAPAN/1-144

60 70 80 90 100
-ASYKVKLITPDGPIEFDCPDVYILDAEEAGHDLPHYSCRAGSCSS
MASYKVKLITPDGPIEFDCPDVYILDAEEAGHDLPHYSCRAGSCSS

Average distance tree using BL...

File View



Jalview Usage Statistics



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)

Cancel

No

Yes

Quality

Consensus

Quality: [Bar chart showing quality scores for each residue]
Consensus: MATYKVKLITPEGEQEFECPPDDVYILDAEEAGIDLPHY

Sequence 5 ID: FER1_PEA Residue: ALA (80)

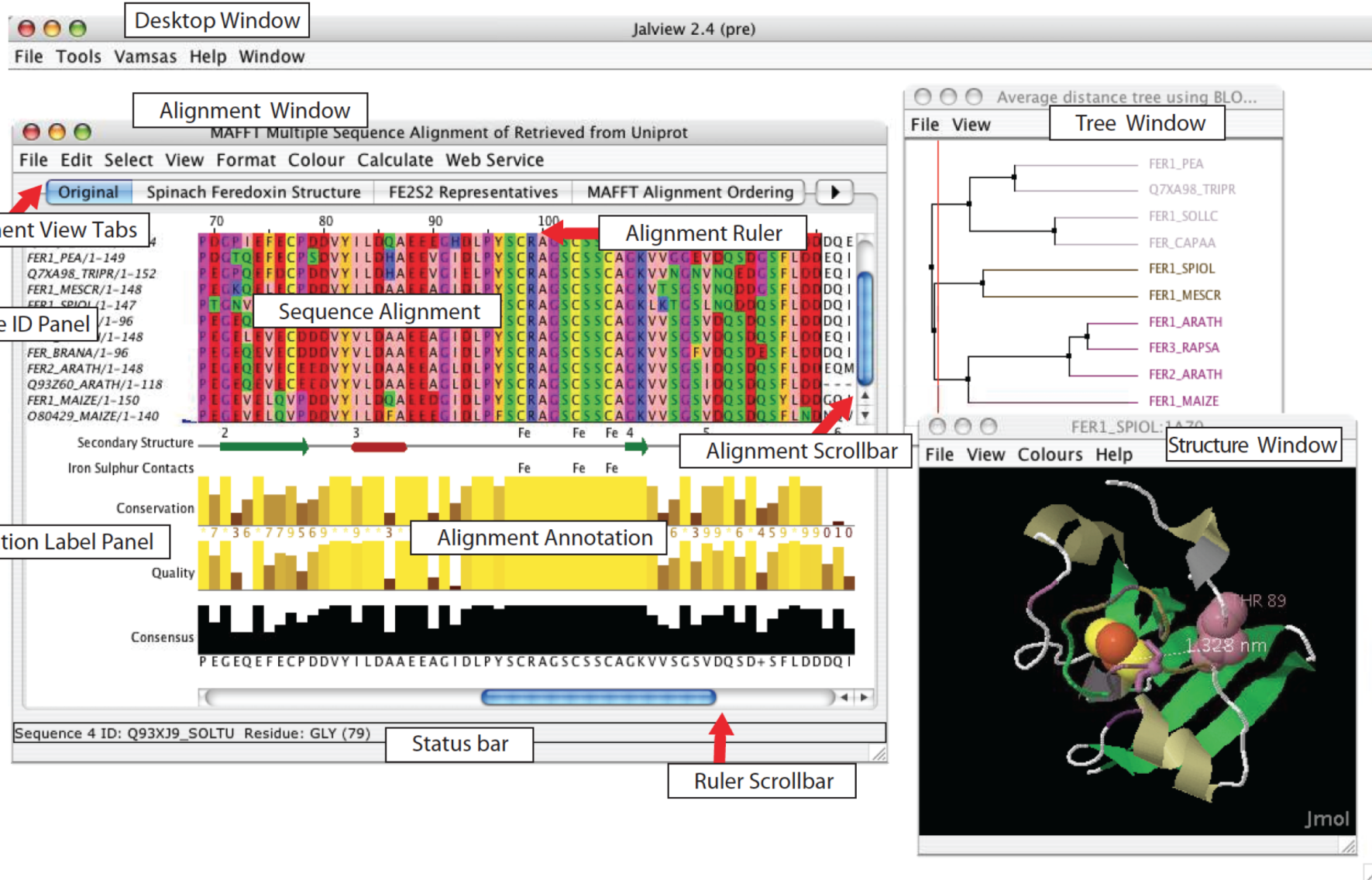


Jmol

Jalview News



Anatomy of Jalview: Figure 1.7



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 ?

Where to Find Help?

Help

Getting Started

FAQ

Documentation

Discussion Mailing List

Other Resources

Community

News Mailing List

Discussion Mailing List

Links

Development

Jalview Bug Tracker

Development Version

Jalview Git Web

Version Archive

Release History

Training

Training Courses

Training Videos

Training News

Jalview & JABA Slides

The screenshot shows the Jalview website interface. A green arrow points from the 'Development' section to the 'Launch Jalview Desktop' button. A blue arrow points from the 'Training' section to the 'The JalviewLite Applet' section. An orange arrow points from the 'Community' section to the 'Latest Training Courses' section. The website content includes the Jalview logo, navigation menu (Home, About, Help, Community, Development, Training, Download), and sections for 'Latest Training Courses', 'Poll', 'What is your favourite Jalview analysis tool?', 'Latest News', and 'Analysis Services'. The 'Analysis Services' section lists various biological data services like Pfam, Rfam, PDB, and Jpred3.

Ex 2 – Navigation (Page 11)

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

Alignment editing modes

[F2] (PC)

toggle between **Normal**
and **Cursor** Mode

[F2] & Fn (Mac)

toggle between **Normal**
and **Cursor** 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 ?

'Split Frame' cDNA & Protein MSAs

Opened with any linked cDNA/Protein data

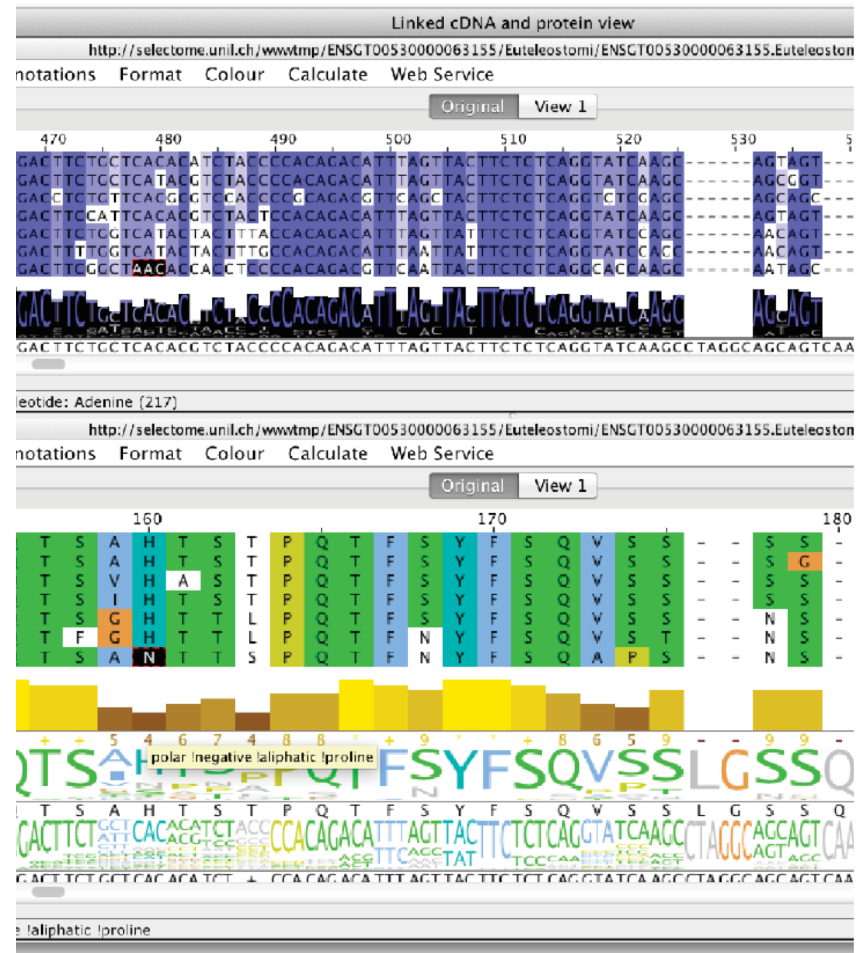
e.g.

- ENA records
- Flat file
- cDNA translation

cDNA/Protein Alignment Reconstruction:

- Gaps inserted according to 'reference alignment'

Layout and display are configurable.



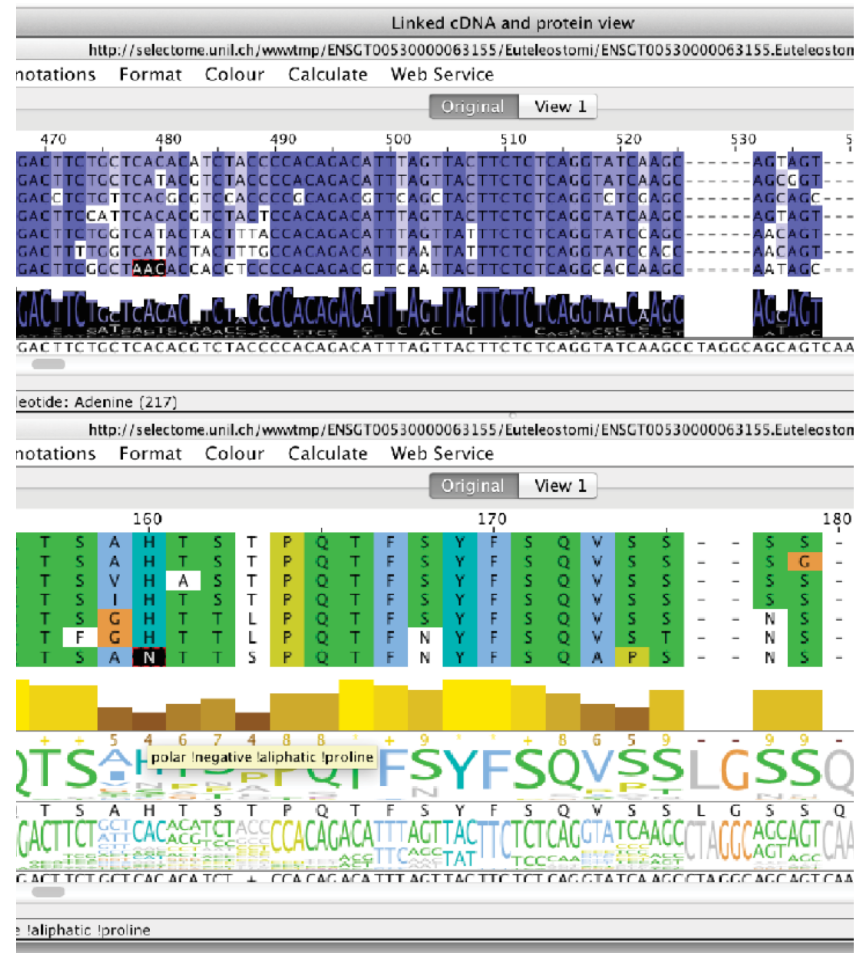
'Split Frame' cDNA & Protein MSAs

Linked alignments share:

- Selections
- Gap editing
- Multiple Views
- Tree sub-groups

Additional Annotation

- Codon conservation



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 ?

**Annotated Contigs,
cDNA and products**

**Genes, transcripts,
Products, and Variants**

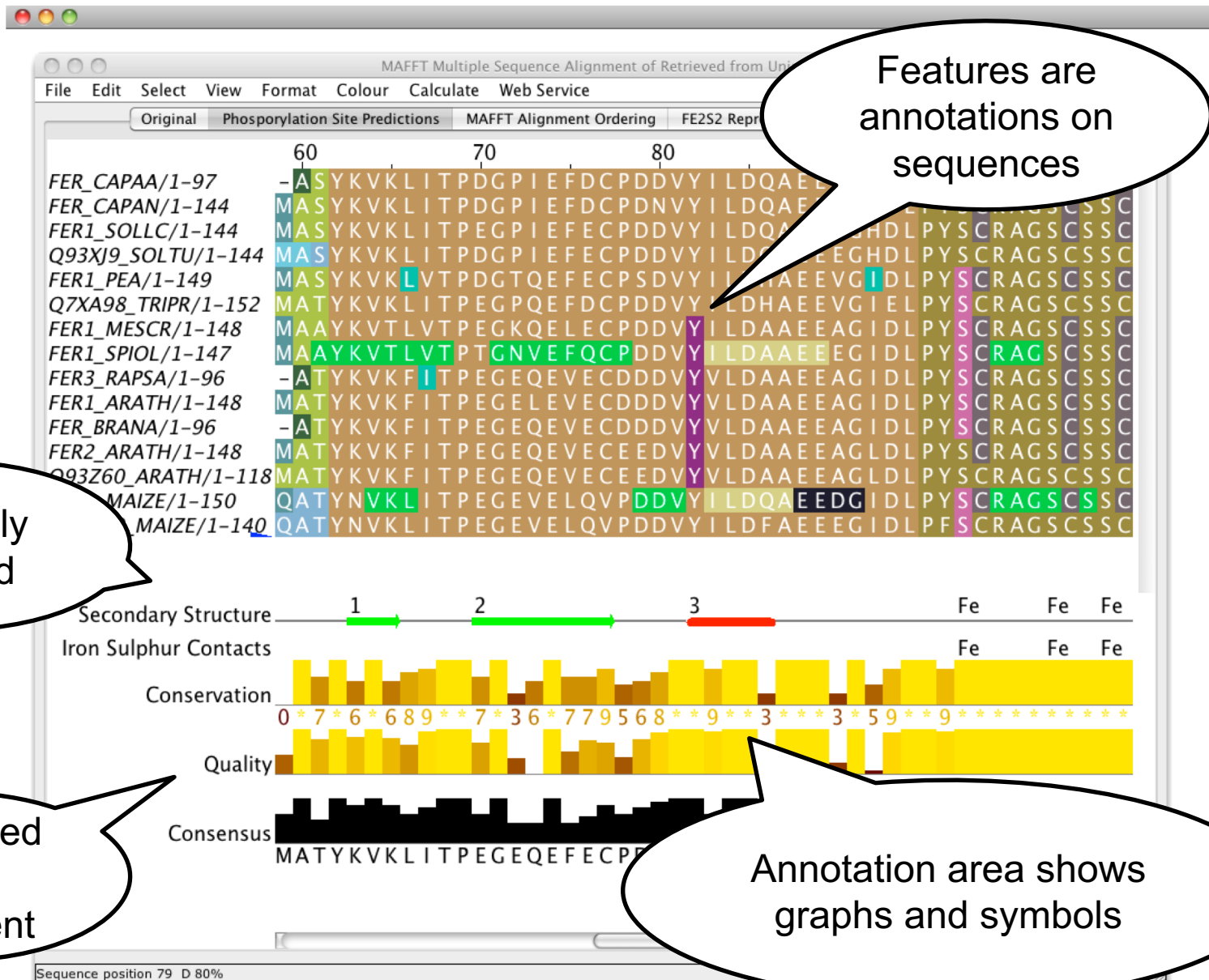
**3D Structures
Features & Annotation**

**Annotated Protein
Domain Alignments**



**Annotated Protein
Sequences**

**Annotated RNA
Domain Alignments**



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 (or MSF) file ?
 - Why are Jalview projects useful ?

END of Session 1: Getting Started

- Launching the Jalview Desktop
- Navigation
 - Normal and Cursor Mode
- Getting data into Jalview
- CDS/Protein view
- Saving alignments and projects

Session 2:

Multiple Sequence Alignment

- Selecting parts of an alignment
- Jalview's web services
- Performing alignments
- Creating hidden regions
- Understanding how hidden regions affect alignment
- Changing alignment parameter settings

Ex 5

selecting regions & making groups

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

An online video about selection and creating groups is available at

<http://www.jalview.org/training/Training-Videos>

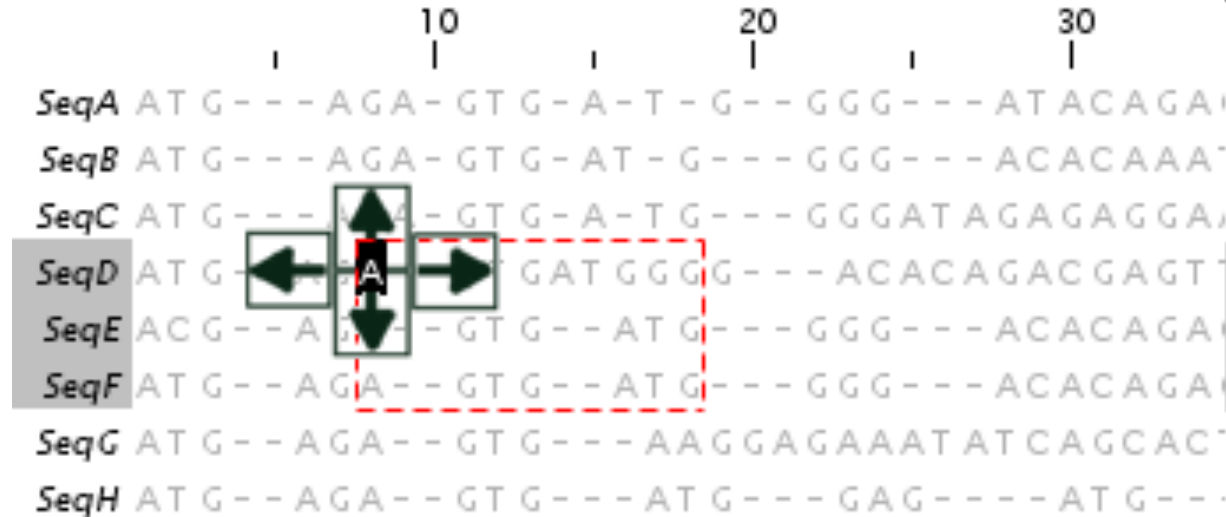
Normal Mode

Original Phos	
/1-148	MAST - - - - A
/1-150	MATV LGSP R
/1-148	MAAT - - TAA
-149	MATT - - - P A
/1-144	MA - - - - - S

Cursor Mode

Original Phos	
/1-148	MAST - - - - A
/1-150	MATV LGSP I
/1-148	MAAT - - TAA
149	MATT - - - P A
1-144	MA - - - - - S

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:

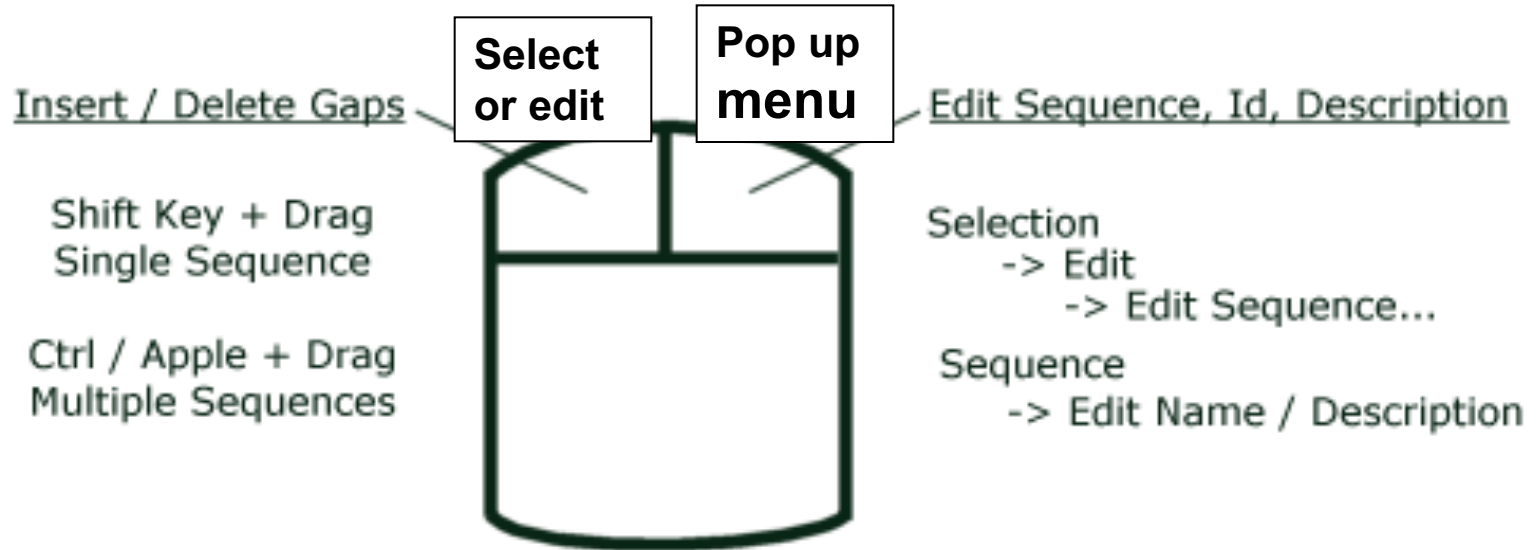
Group Insert - use

Shift+Space

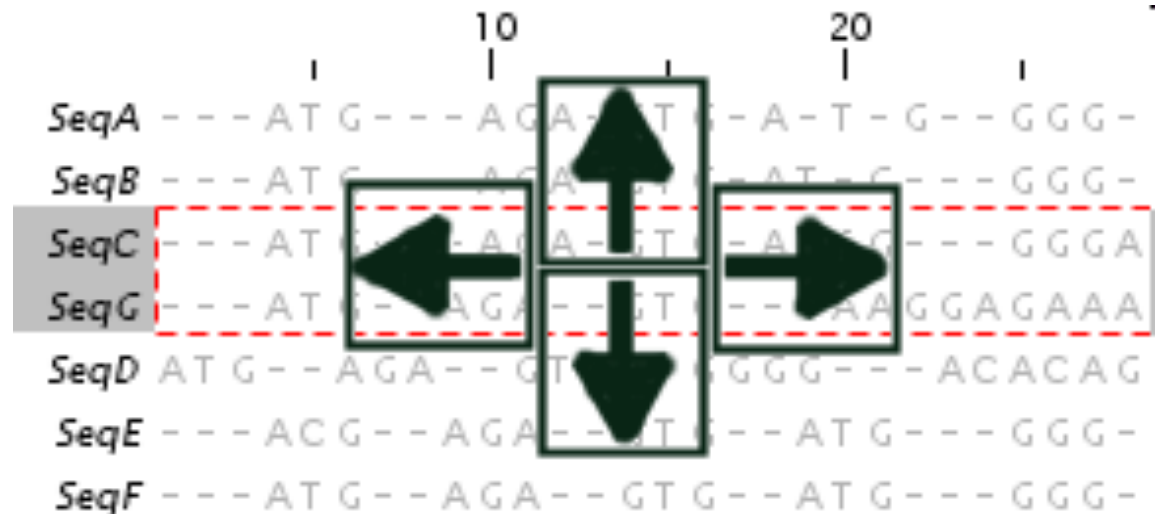
(not CTRL+Space)

How do I edit sequences in Jalview

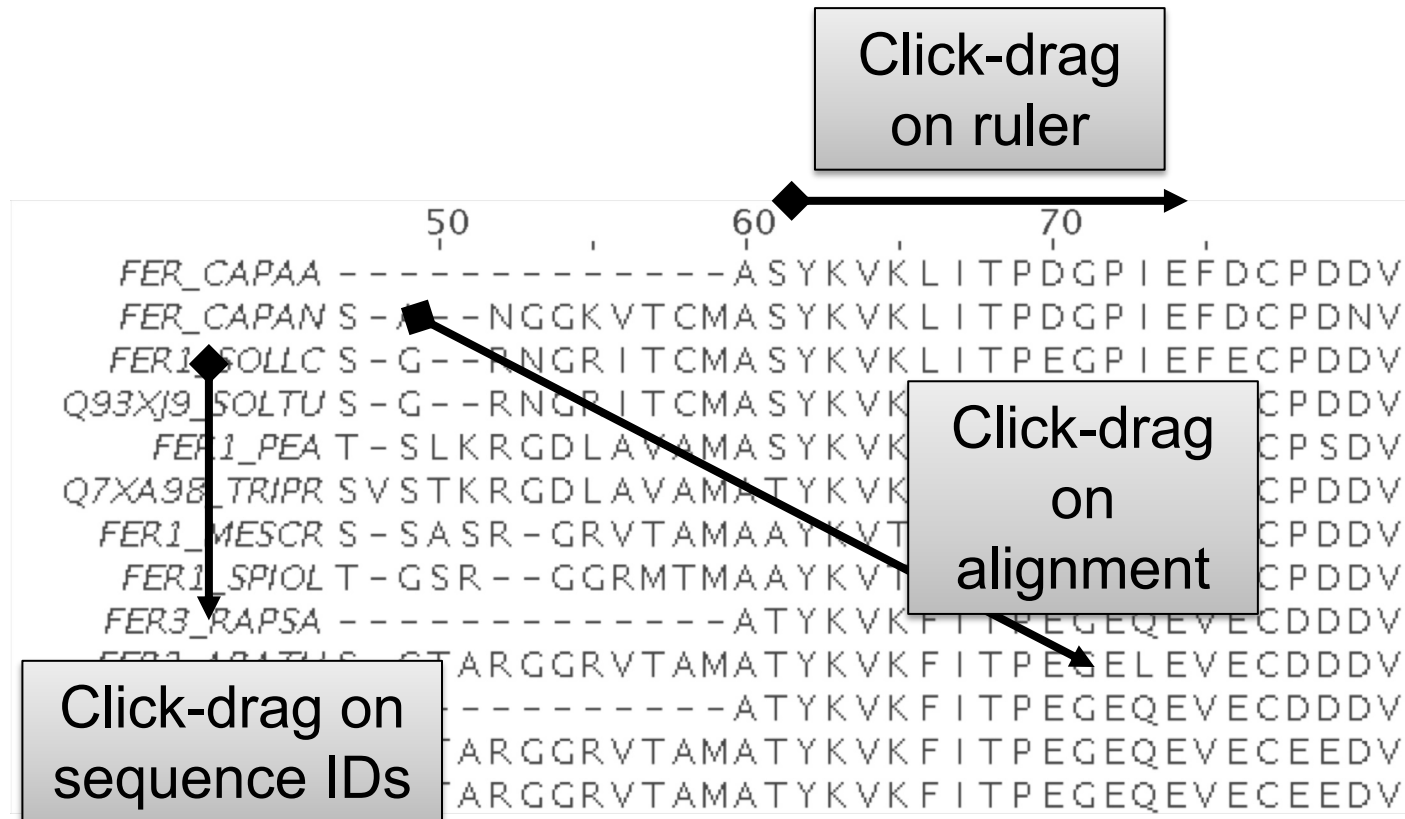
<http://www.jalview.org/examples/editing.html>



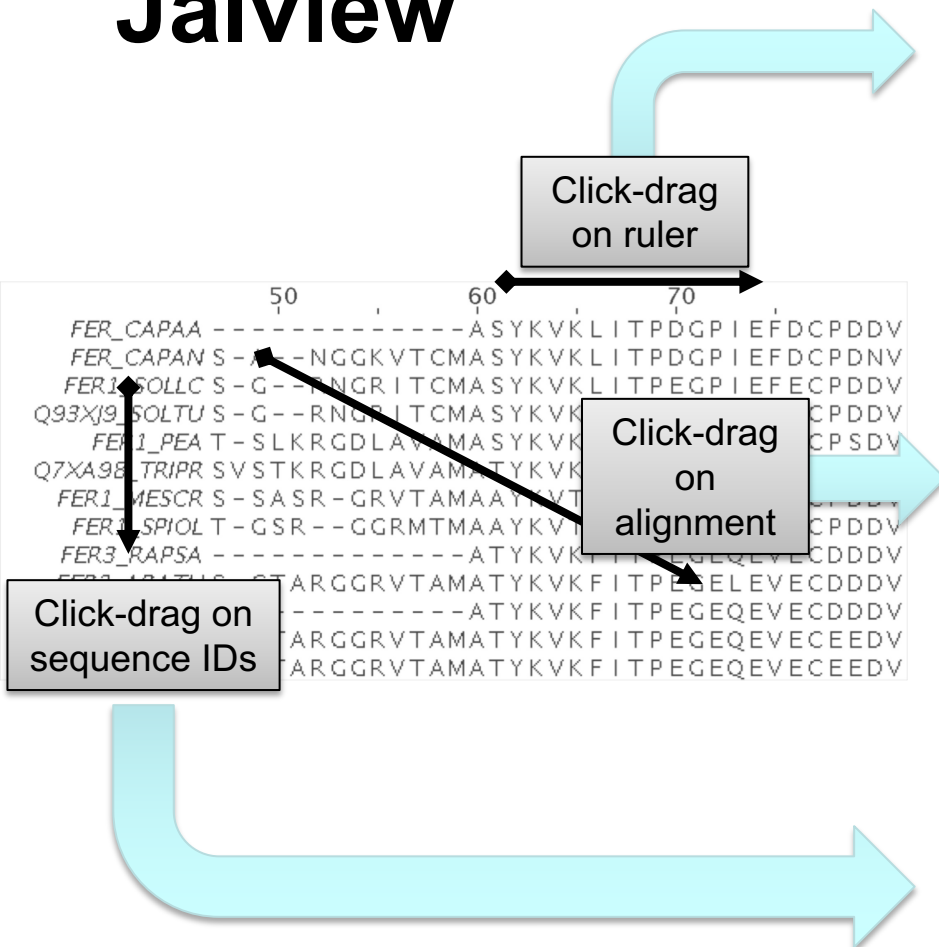
Selected sequences can be moved up and down or slid from left to right



Creating selections with the mouse



Types of selection in Jalview



50 60 70

Selected columns

```
FER_CAPAA -----ASYKVKLITPDGPIEFDCPDDV
FER_CAPAN S-A--NGGKVTCMASYKVKLITPDGPIEFDCPDNV
FER1_SOLLC S-G--RNGRITCMASYKVKLITPDGPIEFDCPDDV
Q93XJ9_SOLTU S-G--RNGRITCMASYKVKLITPDGPIEFDCPDDV
FER1_PEA T-SLKRGLAVAMATYKVKLITPDGPIEFDCPDDV
Q7XA98_TRIPR SVSTKRGLAVAMATYKVKLITPDGPIEFDCPDDV
FER1_MESCR S-SASR-GRVTAMATYKVKLITPDGPIEFDCPDDV
FER1_SPIOL T-GSR--GGRMTMAAYKVKLITPDGPIEFDCPDDV
FER3_RAPSA -----ATYKVKFITPEGEQVEECDDDV
FER2_ARATH S-GTARGGRVTAMATYKVKFITPEGEQVEECDDDV
FER_BRANA -----ATYKVKFITPEGEQVEECDDDV
FER1_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECDDDV
Q93Z60_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECDDDV
```

50 60 70

Selected region

```
FER_CAPAA -----ASYKVKLITPDGPIEFDCPDDV
FER_CAPAN S-A--NGGKVTCMASYKVKLITPDGPIEFDCPDNV
FER1_SOLLC S-G--RNGRITCMASYKVKLITPDGPIEFDCPDDV
Q93XJ9_SOLTU S-G--RNGRITCMASYKVKLITPDGPIEFDCPDDV
FER1_PEA T-SLKRGLAVAMATYKVKLITPDGPIEFDCPDDV
Q7XA98_TRIPR SVSTKRGLAVAMATYKVKLITPDGPIEFDCPDDV
FER1_MESCR S-SASR-GRVTAMATYKVKLITPDGPIEFDCPDDV
FER1_SPIOL T-GSR--GGRMTMAAYKVKLITPDGPIEFDCPDDV
FER3_RAPSA -----ATYKVKFITPEGEQVEECDDDV
FER2_ARATH S-GTARGGRVTAMATYKVKFITPEGEQVEECDDDV
FER_BRANA -----ATYKVKFITPEGEQVEECDDDV
FER1_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECDDDV
Q93Z60_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECDDDV
```

50 60 70

Selected sequences

```
FER_CAPAA -----ASYKVKLITPDGPIEFDCPDDV
FER_CAPAN S-A--NGGKVTCMASYKVKLITPDGPIEFDCPDNV
FER1_SOLLC S-G--RNGRITCMASYKVKLITPDGPIEFDCPDDV
Q93XJ9_SOLTU S-G--RNGRITCMASYKVKLITPDGPIEFDCPDDV
FER1_PEA T-SLKRGLAVAMATYKVKLITPDGPIEFDCPDDV
Q7XA98_TRIPR SVSTKRGLAVAMATYKVKLITPDGPIEFDCPDDV
FER1_MESCR S-SASR-GRVTAMATYKVKLITPDGPIEFDCPDDV
FER1_SPIOL T-GSR--GGRMTMAAYKVKLITPDGPIEFDCPDDV
FER3_RAPSA -----ATYKVKFITPEGEQVEECDDDV
FER2_ARATH S-GTARGGRVTAMATYKVKFITPEGEQVEECDDDV
FER_BRANA -----ATYKVKFITPEGEQVEECDDDV
FER1_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECDDDV
Q93Z60_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECDDDV
```

Working with selections

Active selection can be:

- Hidden
- Copy & Pasted
- Grouped & Annotated
- Edited
- Aligned in a new window
- Used to compute a Tree or PCA plot

50 60 70

```
FER_CAPAA -----ASYKVKLITPDGPIEFDCPDDV
FER_CAPAN S-A--NNGKVTCMASYKVKLITPDGPIEFDCPDNV
FER1_SOLLC S-G--RNGRITCMASYKVKLITPEGPIEFECPPDDV
Q93XJ9_SOLTU S-G--RNGRITCMASYKVKLITPDGPIEFECPPDDV
FER1_PFA T-SLKRGDIAYAMASYKVKLVTPDGTQEFECPSDV
Q7XA98_TRI -----FDCPDDV
FER1_MES -----LECPDDV
FER1_SPI -----FQCPDDV
FER3_RAPSA -----ATYKVKFITPEGEQVEECDDDDV
FER2_ARATH S-GTARGGRVTAMATYKVKFITPEGELEVEECDDDDV
FER_BRANA -----ATYKVKFITPEGEQVEECDDDDV
FER1_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECEEDV
Q93Z60_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECEEDV
```

Selected Columns

50 60 70

```
FER_CAPAA -----ASYKVKLITPDGPIEFDCPDDV
FER_CAPAN S-A--NNGKVTCMASYKVKLITPDGPIEFDCPDNV
FER1_SOLLC S-G--RNGRITCMASYKVKLITPEGPIEFECPPDDV
Q93XJ9_SOLTU S-G--RNGRITCMASYKVKLITPDGPIEFECPSDV
FER1_PEA -----EFDCPDDV
Q7XA98_TRIPR -----EFDCPDDV
FER1_MESCR -----LECPDDV
FER1_SPIOL -----FQCPDDV
FER3_RAPSA -----ATYKVKFITPEGEQVEECDDDDV
FER2_ARATH S-GTARGGRVTAMATYKVKFITPEGELEVEECDDDDV
FER_BRANA -----ATYKVKFITPEGEQVEECDDDDV
FER1_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECEEDV
Q93Z60_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECEEDV
```

Selected Region

50 60 70

```
FER_CAPAA -----ASYKVKLITPDGPIEFDCPDDV
FER_CAPAN S-A--NNGKVTCMASYKVKLITPDGPIEFDCPDNV
FER1_SOLLC S-G--RNGRITCMASYKVKLITPEGPIEFECPPDDV
Q93XJ9_S -----ECPDDV
FER1 -----ECPDDV
Q7XA98_ -----ECPDDV
FER1_M -----ECPDDV
FER1_SPIOL T-GSR--GGRMTMAAYKVTLVPTGNVEFQCPDDV
FER3_RAPSA -----ATYKVKFITPEGEQVEECDDDDV
FER2_ARATH S-GTARGGRVTAMATYKVKFITPEGELEVEECDDDDV
FER_BRANA -----ATYKVKFITPEGEQVEECDDDDV
FER1_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECEEDV
Q93Z60_ARATH S-STARGGRVTAMATYKVKFITPEGEQVEECEEDV
```

Selected sequences

Ex 5

selecting regions & making groups

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

An online video about selection and creating groups is available at

<http://www.jalview.org/training/Training-Videos>

Ex 5

selecting regions & making groups

- 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 6

Reordering the alignment

- Task
 - Get used to selecting and reordering sequences

Question:

- What's the quickest way to move the bottom-most sequence to the top of the alignment ?
- What other ways can you reorder sequences ?

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 ?

Practice at home

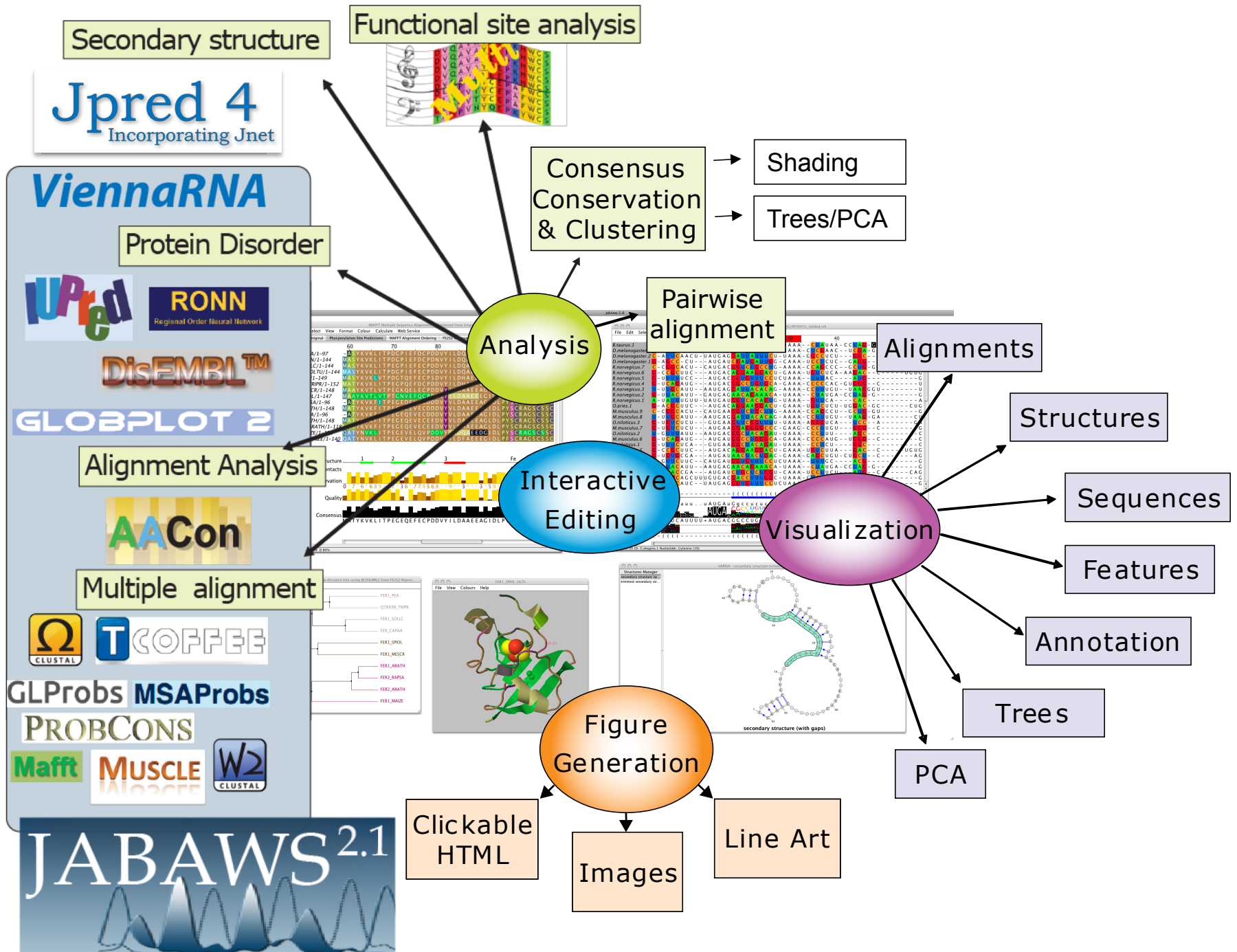
- Exercises 8 and 9
 - Selecting
 - Creating groups
 - Hiding and revealing hidden regions
 - Editing with the keyboard and mouse

Lightning introductions @Coffee

- Who are you ?
- Where are you working ?
- What are you working on ?

LUNCH BREAK

Jalview's web services



www.compbio.dundee.ac.uk/jabaws

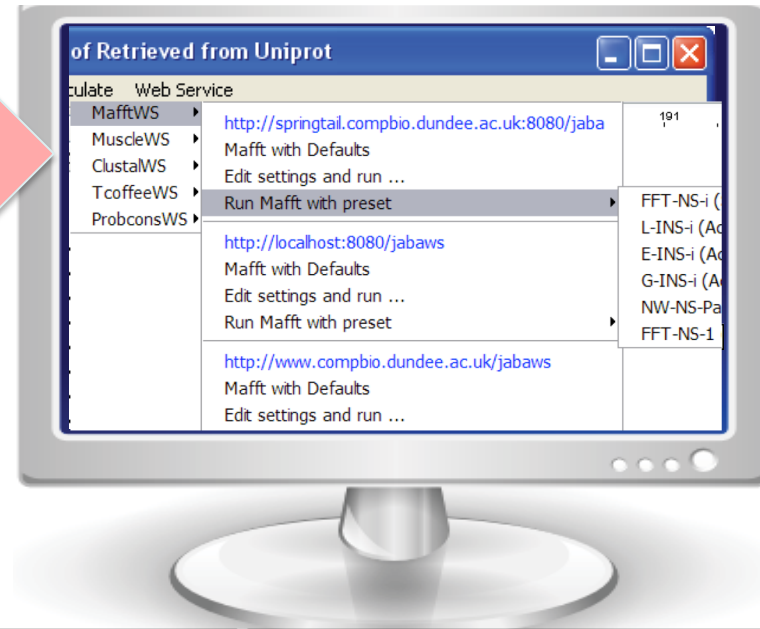
JABAWS:MSA Troshin et al. 2011,
Bioinformatics.

JABAWS 2 In Preparation.



Jalview Web
Service GUI

JABAWS
Java Client



JABAWS Virtual
Appliance for
your private use.



Native JABAWS installs
on a range of platforms



Apache
Tomcat

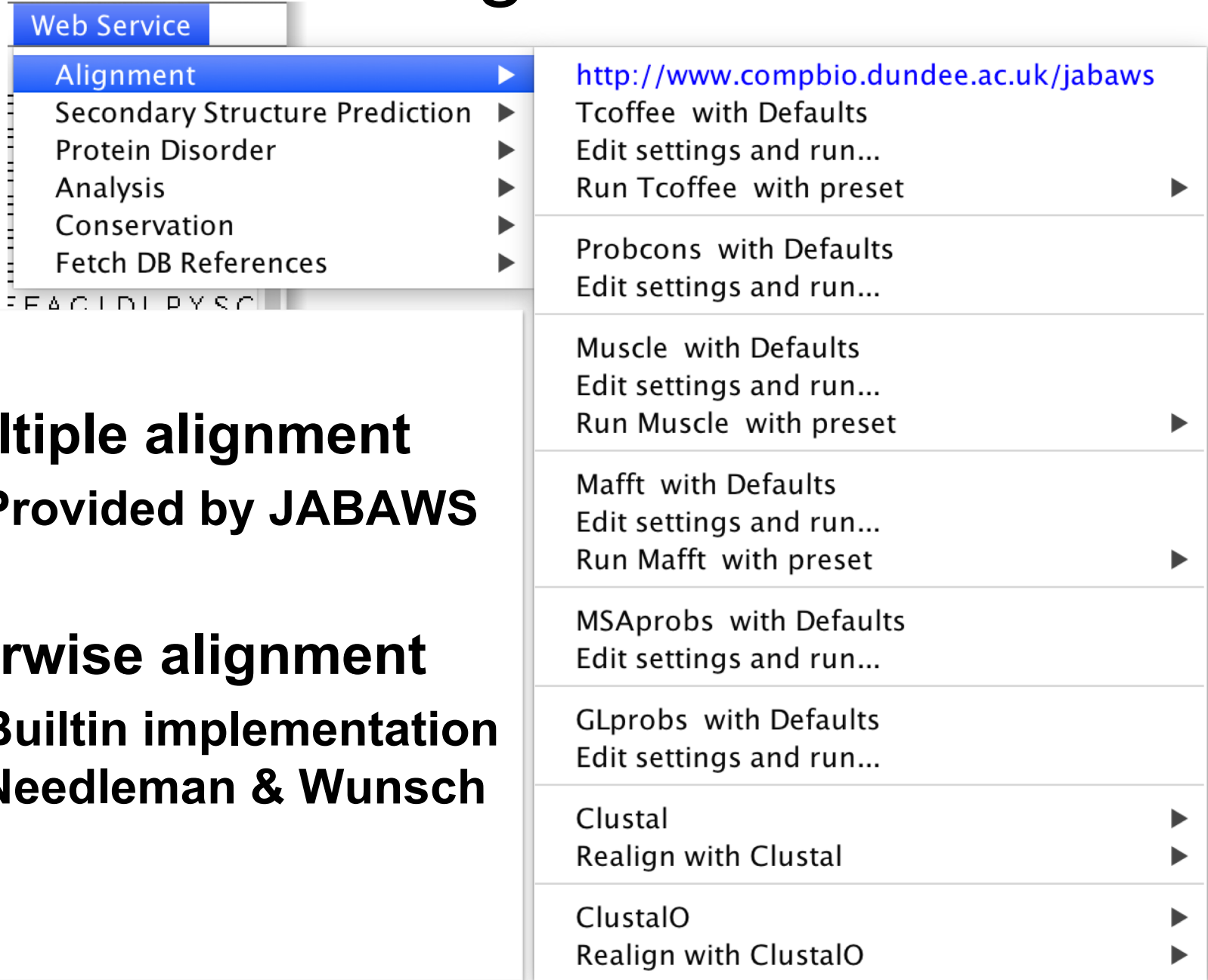


JABAWS Amazon
Machine Image
on EC2



the
cloud
market

Jalview's Alignment Methods



The screenshot shows the 'Web Service' menu in Jalview. The 'Alignment' option is selected, opening a submenu. The submenu lists several alignment methods, each with options to 'Edit settings and run...' or 'Run' with a preset. The methods listed are Tcoffee, Probcons, Muscle, Mafft, MSAProbs, GLprobs, Clustal, and ClustalO.

Method	Options
Alignment	Alignment, Secondary Structure Prediction, Protein Disorder, Analysis, Conservation, Fetch DB References
Tcoffee	with Defaults, Edit settings and run..., Run Tcoffee with preset
Probcons	with Defaults, Edit settings and run...
Muscle	with Defaults, Edit settings and run..., Run Muscle with preset
Mafft	with Defaults, Edit settings and run..., Run Mafft with preset
MSAProbs	with Defaults, Edit settings and run...
GLprobs	with Defaults, Edit settings and run...
Clustal	Realign with Clustal
ClustalO	Realign with ClustalO

- **Multiple alignment**
 - **Provided by JABAWS**
- **Pairwise alignment**
 - **Builtin implementation Needleman & Wunsch**

Why do we provide so many ?

- Different purposes
 - Large alignments
 - Difficult sequence families
 - Special score models

Common types of alignment algorithm

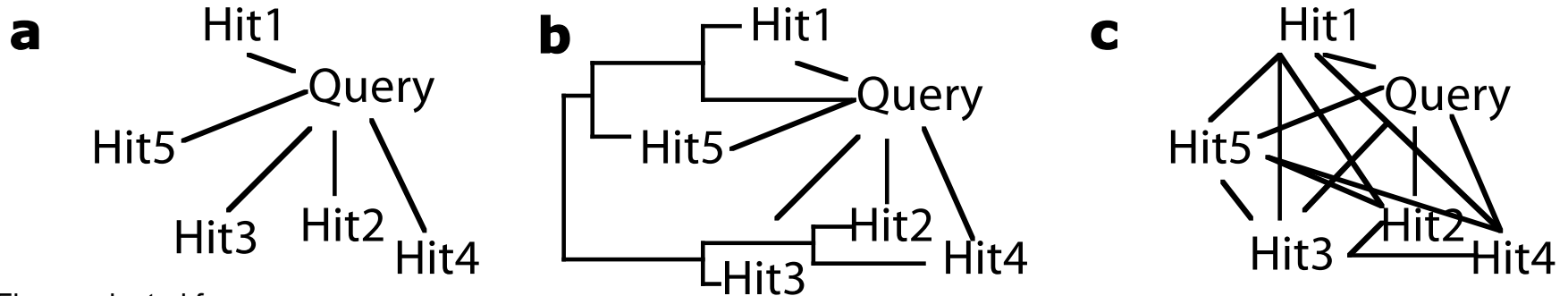


Figure adapted from
Procter et al. (2010) *Nature Methods* 7 S16 - S25

- a. Sequence database searches – optimal alignment between query and hit
e.g. **Blast** (single sequence), **PSI-Blast** and **HMMER**
- b. Progressive – optimise alignment between branches on guide tree
e.g. **ClustalW**
- c. Transitive – optimise MSA to maximise consistency between pairs
e.g. **T-COFFEE**, **ProbCons**

Profile methods – e.g. Muscle and MAFFT are hybrid of **b** and **c**.

Latest methods, e.g. **ClustalO**, also employ sampling strategies to speed up tree building & refinement.

Why do we provide so many ?

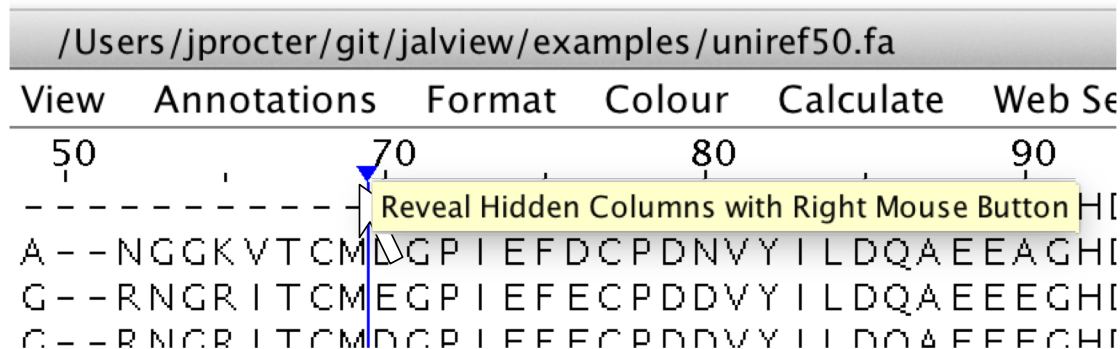
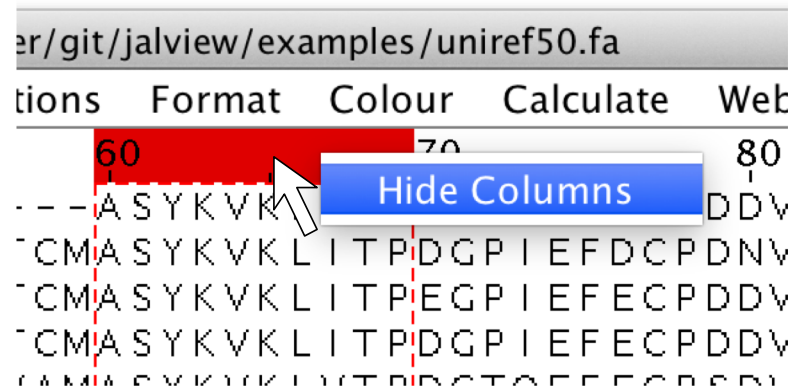
- Different purposes
 - Large alignments
 - Difficult sequence families
 - Special score models
- Reproducibility
 - Older methods still widely used
- Special features

Multiple Sequence Alignment exercise 16 (Section 5.2)

- Tasks
 - Align sequences using different methods
 - Explore how hidden regions affect alignment jobs.

Essential skill for next exercise:

Hiding columns



Multiple Sequence Alignment exercise 16 (Section 5.2)

- Tasks
 - Align sequences using different methods
 - Explore how hidden regions affect alignment jobs.

Multiple Sequence Alignment

exercise 16 (Section 5.2)

- Tasks
 - Align sequences using different methods
 - 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 ?

Realignment with ClustalW & ClustalO

1. Import a good alignment

- E.g. a PFAM seed

Clustal

Realign with Clustal



2. Add more sequences

ClustalO

Realign with ClustalO



- ClustalW

- Preserves original alignment
- Align new sequences to profile

- ClustalO

- Builds profile
- Aligns **ALL** sequences to profile

Custom Analysis Parameters

The screenshot shows a software window titled "Edit parameters for Align with MuscleWS". The window has a blue title bar with a close button (X) in the top right corner. Below the title bar, there is a dropdown menu for "Current parameter set name" set to "User Defaults". Below this are "Revert" and "Create" buttons. The main area is divided into three sections: "Details", "Options", and "Parameters". The "Details" section has a large empty text box. The "Options" section contains four checkboxes: "dimer", "Diagonal", "Diagonal 1", and "Profile scoring method" (which is checked). To the right of these checkboxes is a dropdown menu showing "-SV". The "Parameters" section contains a "Sequence type" dropdown menu set to "auto". Below this is a "Maxiters" section with a dropdown menu. At the bottom of the window are "Start Job" and "Cancel Job" buttons. Several callout boxes with light blue borders and pointers provide additional information: one points to the "Current parameter set name" dropdown, another to the "Revert" and "Create" buttons, a third to the "Details" text box, a fourth to the "Profile scoring method" checkbox, a fifth to the "Sequence type" dropdown, and a sixth to the "Start Job" and "Cancel Job" buttons. A tooltip box on the right side of the window provides a brief description and a link for the "-SV" option.

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

le - use log-expectation profile score VTML240 (default sequences.) sp - use sum-of-pairs protein profile score use sum-of-pairs profile score (VTML240)
[Link](#)

Modifying alignment parameters

?

End of Session 2:

Multiple Sequence Alignment

- Performing alignments with Jalview
- Understanding how hidden regions affect alignment
- Changing alignment parameter settings

Alignment Colouring and Analysis

Using jalview to analyse and
visualise the relationships
between aligned sequences.

Session 3

Colouring, Trees and PCA

- Colouring sequences and alignments
- Calculating PCA and Trees
- Colouring subgroups to highlight patterns

Ex 10: Colouring (Section 3.2)

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

Ex 10: Colouring (Section 3.2)

- 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 ?

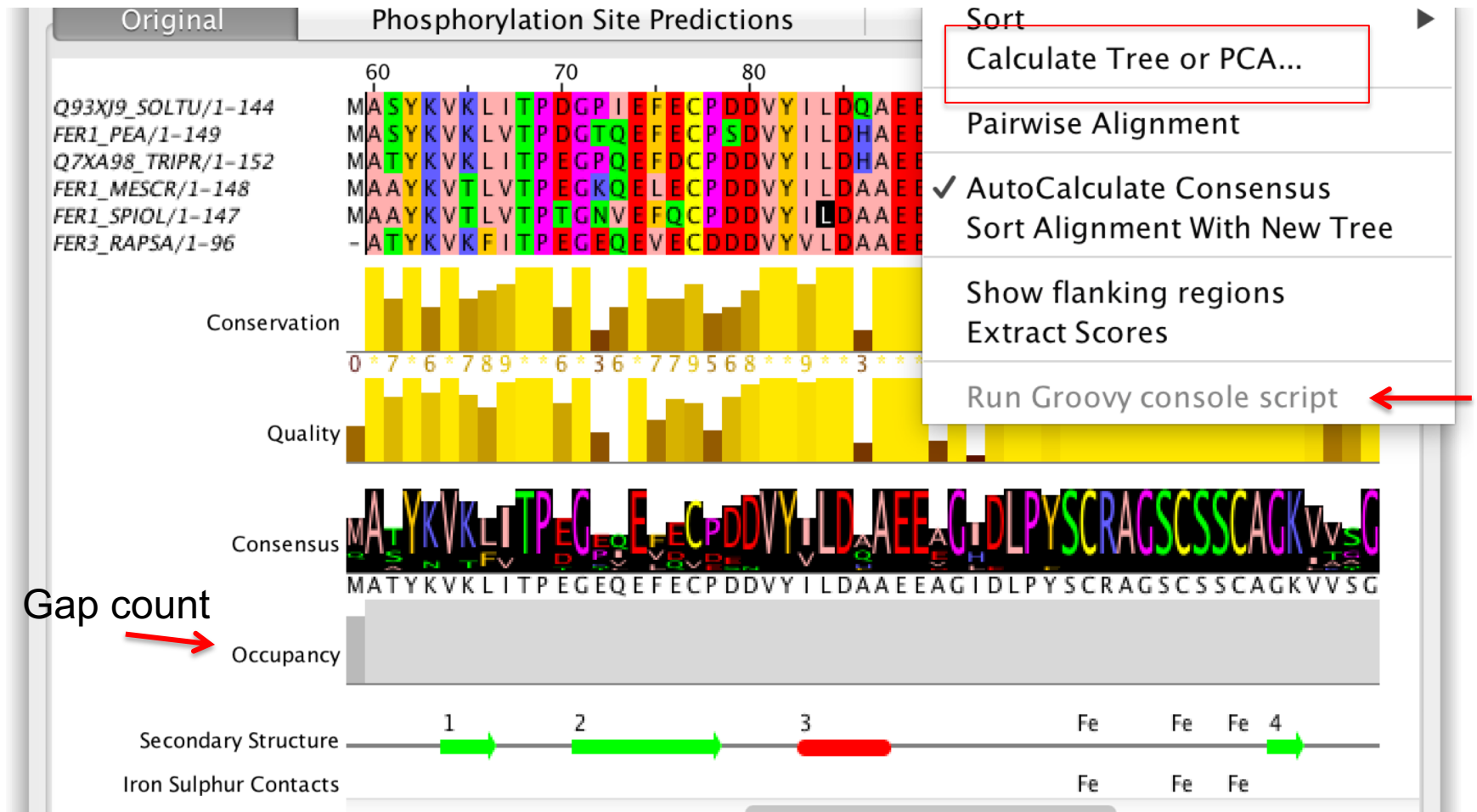
PCA, Trees & subfamily analysis (chapter 6)

- Principal components analysis (PCA)
- Phylogenetic trees
- Redundancy removal (homework)
- Tree based conservation analysis
- Subdividing alignment by mutation

Phylogenetic analysis and Jalview

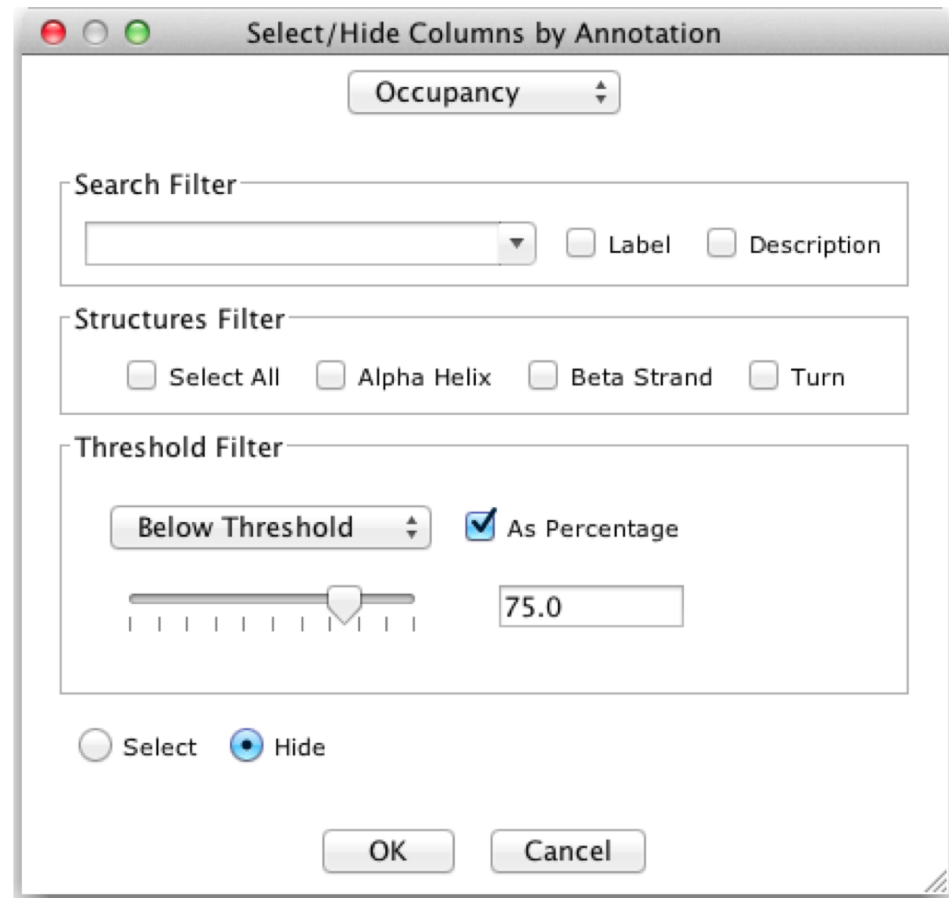
- **Built in tree methods**
 - UPGMA
 - Fast, simple, good for understanding alignment structure
 - Neighbour joining
 - Slower than UPGMA
 - Useful for a first approximation for evolutionary analysis
 - NJ does not work well for very divergent sequence sets
 - » Need to add in close relatives to get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

Jalview's Calculation menu



Select or hide ...

- By Alignment
 - Hide Insertions for a selection
- By annotation



PCA and Trees

Exercise 17 and 18

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

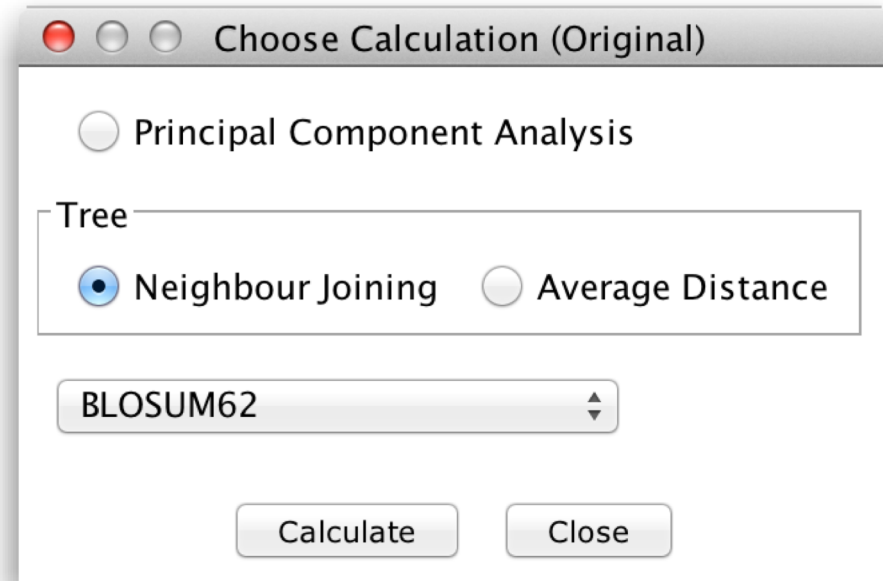
PCA and Trees

Exercise 17 and 18

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

Score models in Jalview

- PCA and Tree calculations
 - From Distance Matrices
- Import your own substitution matrices
 - [NCBI](#) & [AAIndex](#)
- Create your own score models
 - Via a **Groovy** script



Tree based conservation analysis

Exercise 20

- *ad hoc* '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

Exercise 20

- *ad hoc* '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
- Questions
 - How can you tell which differences are important ?
 - How can you navigate the sub-groups of a large alignment ?

Sub-groups and Sub-group Annotation

Exercise 22

- 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 22

- Task
 - Use the group consensus sequence logos to more easily compare tree subgroups
 - Use ‘Make groups for selection’ to subdivide groups by specific mutation
- Questions
 - How can you work out which group is associated with which annotation row ?

End of Session 3

Colouring, Trees and PCA

- Colouring sequences and alignments
- Calculating PCA and Trees
- Colouring subgroups to highlight patterns

<http://www.jalview.org/tutorial/training-materials/2018/May/Liverpool/>

ALIGNMENT RELIABILITY

Uses of alignments...

Are conservation scores trustworthy ?

- Good quality alignments
 - Conservation == probable molecular similarity
- Poor alignments
 - Conservation == random noise
- How do you measure MSA reliability ?
 - Try different methods. Vary parameters
 - Compare the results.

Score = 740 Length = 17
PID = 100%

Q93Z60_ARATH MASTA**LSSA**IVSTSFLR
 |||||
 FER2_ARATH MASTA**LSSA**IVSTSFLR

Score = 160 Length = 17
PID=29.41
FER1_MAIZE LGSPR**APAF**FFSSSLR
 ..| .. . |. | ||
Q93Z60 ARATH MASTA**LSSA**IVSTSFLR

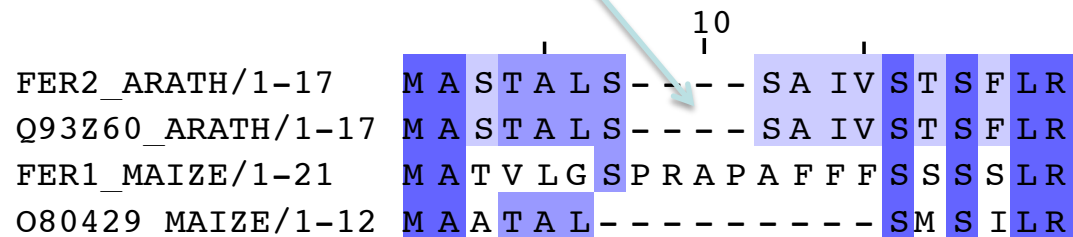
```
Score = 160 Length = 17
PID=29.41%
FER1_MAIZE  LGSPRAPAFFFSSSLR
                ..|   .. . |.| ||
FER2_ARATH  MASTALSSAIVSTSFLR
```

```
Score = 310 Length = 12
PID = 58.33%
O80429_MAIZE MAATALSMSILR
          ||.|||| .|.
Q93Z60_ARATH MASTALSSAIVS
```

```
Score = 310 Length = 12
PID=58.33%
O80429_MAIZE MAATALSMSILR
          ||.|||| .|.
FER2 ARATH MASTALSSAIVS
```

```
Score = 120 Length = 12
PID = 41.67%
```

080429_MAIZE MAATAL**MS**ILR
 . | . | | |
 FER1 MAIZE **APAFF**SSSSLR



- Calculate 'shift' between all pairwise alignments and the multiple sequence alignment
 - Higher shifts are less reliable

Consistent Pairwise Alignments

Score = 740 Length = 17
PID = 100%
Q93Z60_ARATH MASTA**LSSA**IVSTSFLR
 |||||
FER2_ARATH MASTA**LSSA**IVSTSFLR

Score = 160 Length = 17
PID=29.41
FER1_MAIZE LGSPR**APAF**FFSSSSLR
 ..| .. . |. | ||
Q93Z60_ARATH MASTA**LSSA**IVSTSFLR

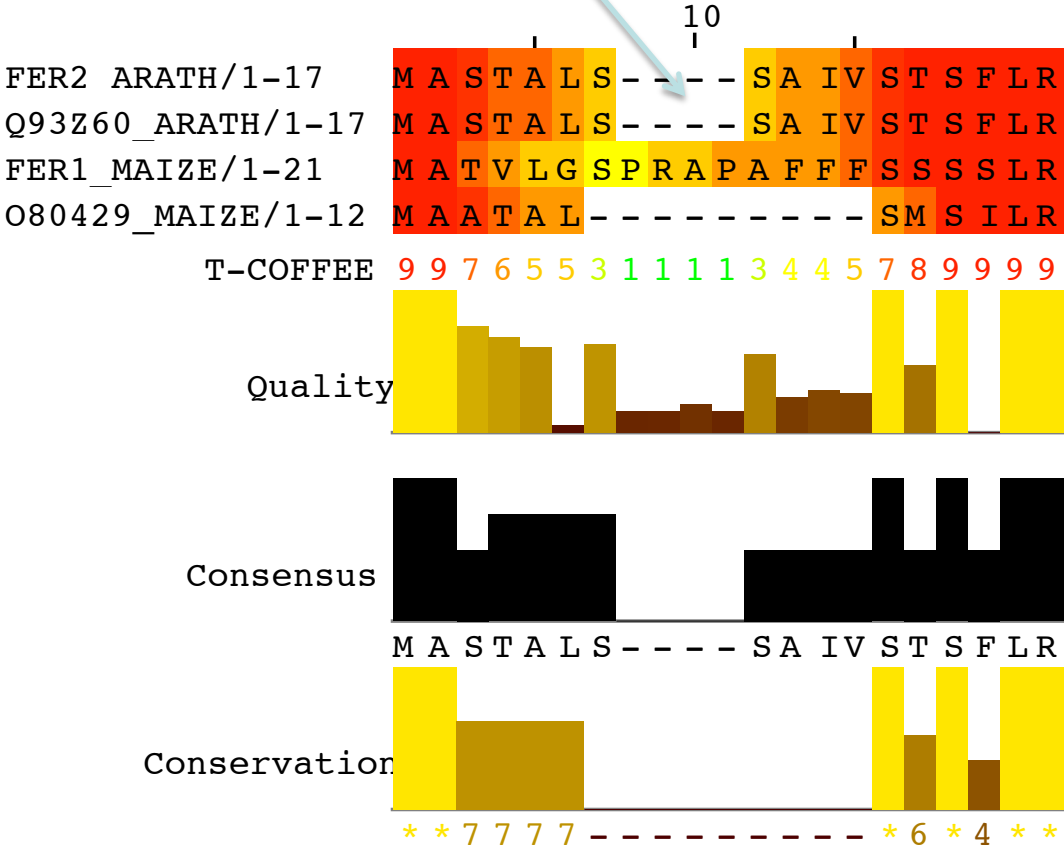
Score = 160 Length = 17
PID=29.41%
FER1_MAIZE LGSPR**APAF**FFSSSSLR
 ..| .. . |. | ||
FER2_ARATH MASTA**LSSA**IVSTSFLR

Score = 310 Length = 12
PID = 58.33%
O80429_MAIZE MAATA**LSMS**ILR
 ||. ||| |. |
Q93Z60_ARATH MASTA**LSSA**IVS

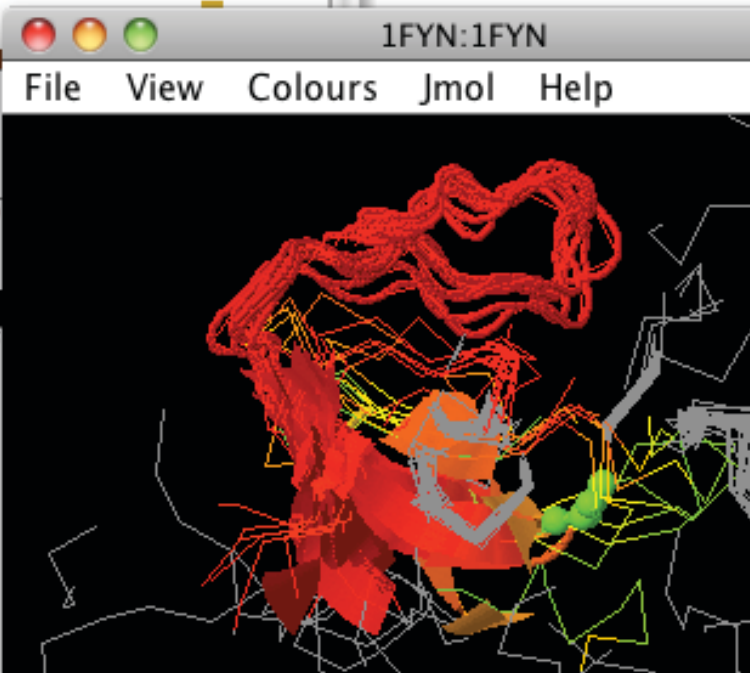
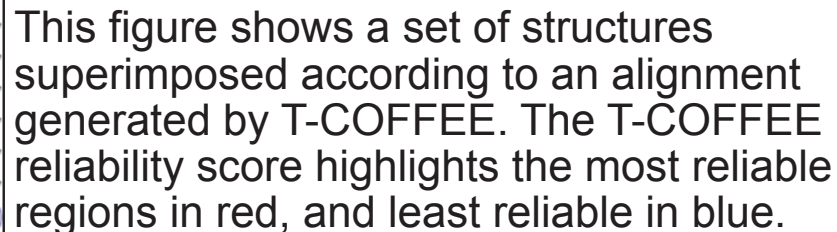
Score = 310 Length = 12
PID=58.33%
O80429_MAIZE MAATA**LSMS**ILR
 ||. ||| |. |
FER2_ARATH MASTA**LSSA**IVS

Major inconsistency: +4 Shift

Score = 120 Length = 12
PID = 41.67%
O80429_MAIZE MAATA**LSMS**ILR
 .| .| | ||
FER1_MAIZE **APAF**FFSSSSLR



T-COFFEE alignment reliability scores



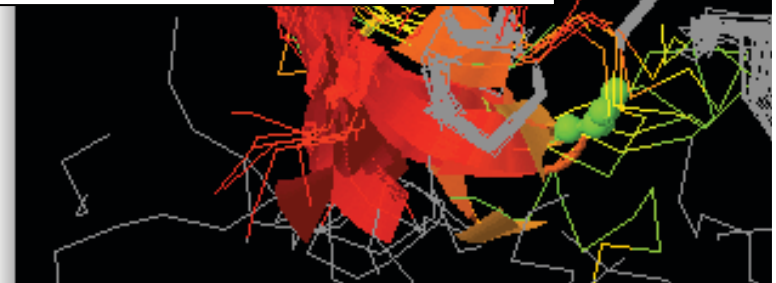
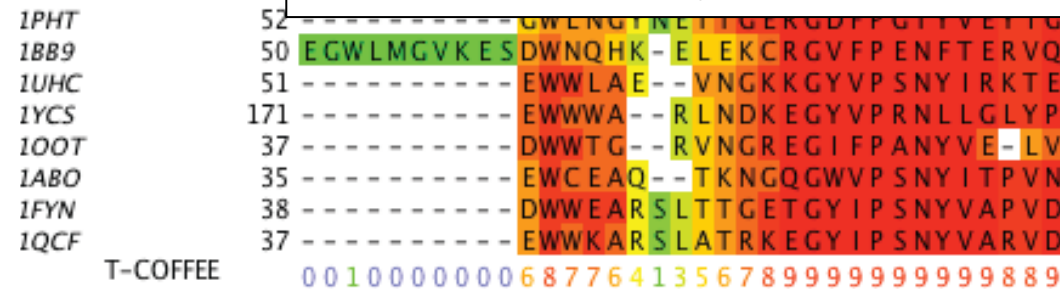
T-COFFEE alignment reliability scores



Easiest to do this is from command line, or from the T-COFFEE web site

- **CORE** – score an alignment
 - <http://tcoffee.org.cat/apps/tcoffee/do:core>
- **M-COFFEE** – combine results from many ‘popular aligners’
 - <http://tcoffee.org.cat/apps/tcoffee/do:mcoffee>
- Jalview can read the ‘score_ascii’ file for an alignment

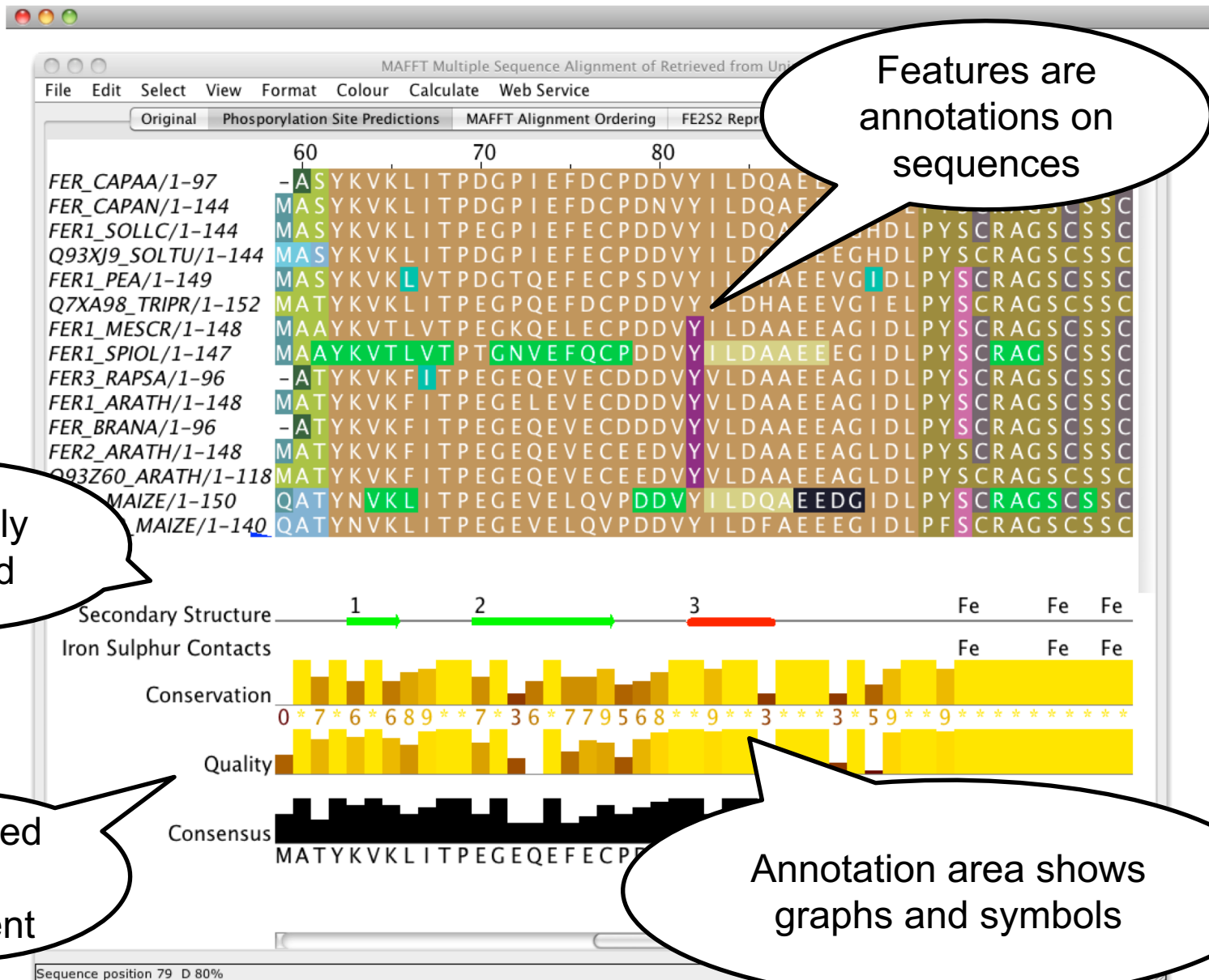
See supplementary exercises for instructions



Session 5:

Features, Annotation, Databases

- Understand the difference between Features and Annotation
- learn where features & annotation come from
- Customise feature display
- Use feature and annotation to query the alignment



Create New Feature...

I P A L Y G I A V S T S F L R I Q P M P M S V I I K A F S N - - G F L G L K I
I P A L Y G T A V S T S F M R R Q P V P M S V A T T T T T K A F P S G F G L K S
I A A I S G A T M S T A F A P K T - - P P M T A A I P T N V G - R A I F G I K S
A T T

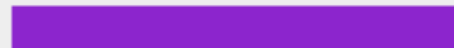
STA
STA
STA
SPR



Name Binding site

Group: Jalview

Colour:



Description

Deduced by homology

Cancel

OK

Jalview Test

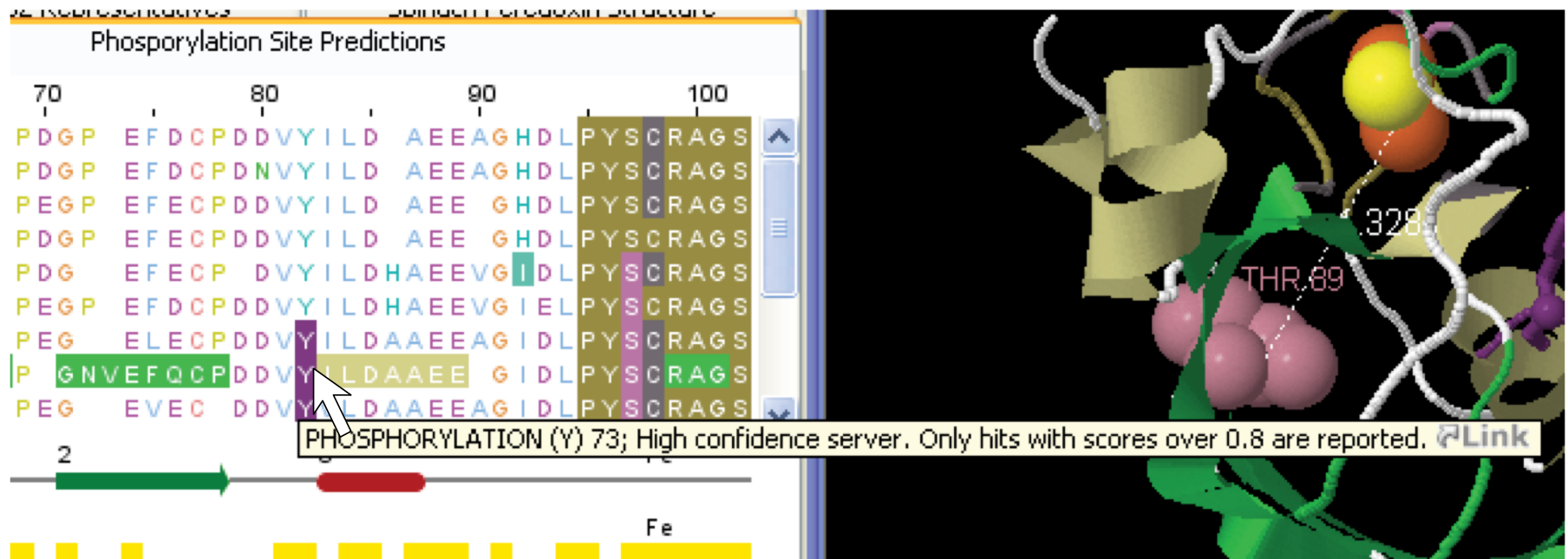
<http://www.jalview.org/tutorial/alignment.fa>

File Edit Select View Annotations Format Colour Calculate Web Service

	10	20	30	40
FER_CAPAA/1-97	-----			
FER_CAPAN/1-144	-MASVSATMI STSFMPRKPAVTSLKPI P-NVG-EALFGLK			
FER1_SOLLC/1-144	-MASISGTMISTSF LPRKPAVTSLKAI S-NVG-EALFGLK			
Q93XJ9_SOLTU/1-144	-MASISGTMISTSF LPRKPVVTSLKAI S-NVG-EALFGLK			
FER1_PEA/1-149	TTPALYGTAVSTSFLRTQPMPSVTTTKAFSN--GFLGLK			
Q7XA98_TRIPR/1-152	TTPALYGTAVSTSFMRRQPVPMSVATTTTKAFPSGFGLK			
FER1_MESCR/1-148	TTAALSGATMST Binding site 4 15; Deduced by homology /G-RALFGLK			
FER1_SPIOL/1-147	AATTTTMMGMATTFV PKPQAPPMMAALPSNTG-RSLFGLK			
FER3_RAPSA/1-96	-----			
FER1_ARATH/1-148	ASTALSSAIVGTSFIRRSPAPISLRSLPSANT-QSLFGLK			
FER_BRANA/1-96	-----			
FER2_ARATH/1-148	ASTALSSAIVSTSFLRRQQTPISLRSLPFANT-QSLFGLK			
Q93Z60_ARATH/1-118	ASTALSSAIVSTSFLRRQQTPISLRSLPFANT-QSLFGLK			
FER1_MAIZE/1-150	GSPRAPAFFSSSSSLRAAPAPTAVALPAKVG---IMGRS			
O80429_MAIZE/1-140	----MAATALSMSILRAPP-PCFSSPLRLRVAVAKPLAAP			

Sequence 8 ID: FER1_SPIOL Residue: THR (13)

Sequence Features are visualised over the alignment



Sources of sequence features

- Sequence databases
- 3D Structure associations
- GFF2/3 (*Generalised Feature Format*)
- Jalview format feature files
- Certain 'rich' alignment text file formats
 - Stockholm
 - AMSA
- Manual input of features in Jalview

[illegible]

Annotated Protein Domain Alignments

EMDB > EMD-5119

Cathrin D8 coat

Source organism: *Escherichia coli*

Fitted atomic model: [1.50 Å](#)

Primary publication: [DOI: 10.1016/j.jmb.2012.05.010](#)

ED: Molecular model: [Riley A., Cheng Y., et al. \(2012\)](#)

WTUW 432,473

EMD: [1.50 Å](#)

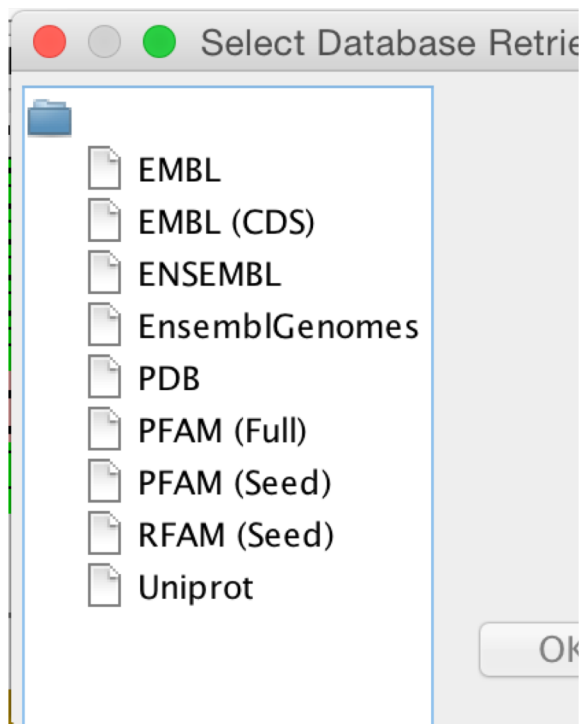
PDB **e**

PROTEIN DATA BANK EUROPE

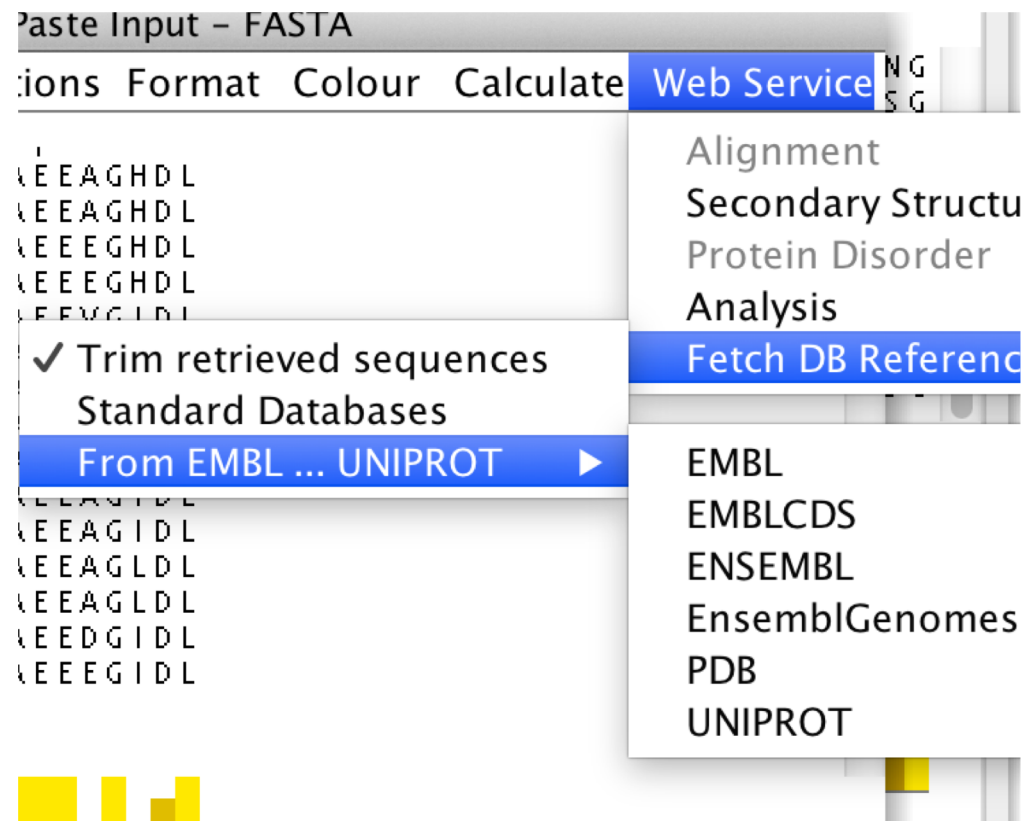
[illegible][illegible][illegible]

Annotated RNA Domain Alignments

Retrieve from database



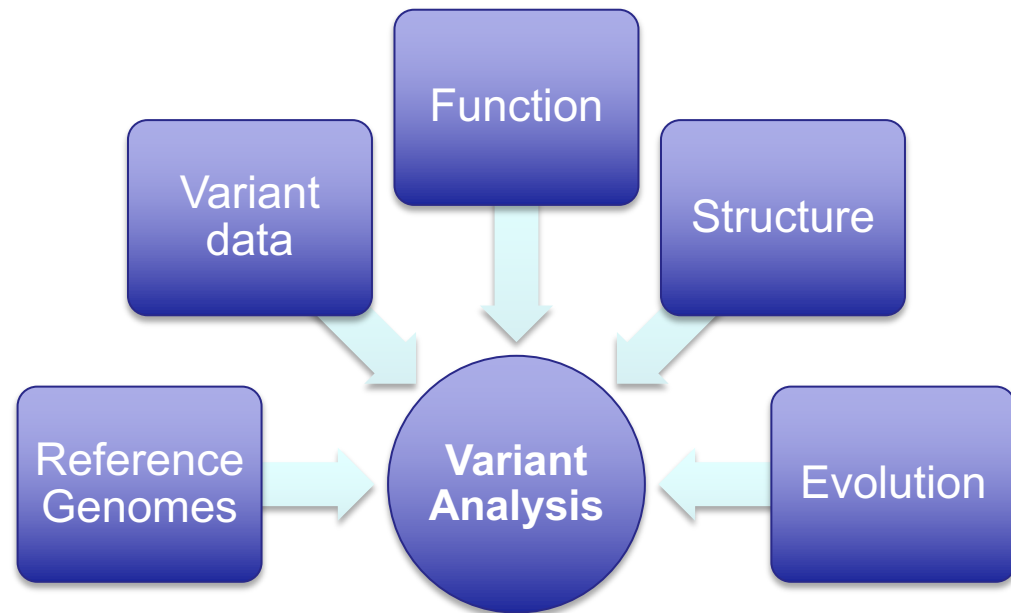
Match sequence to database record



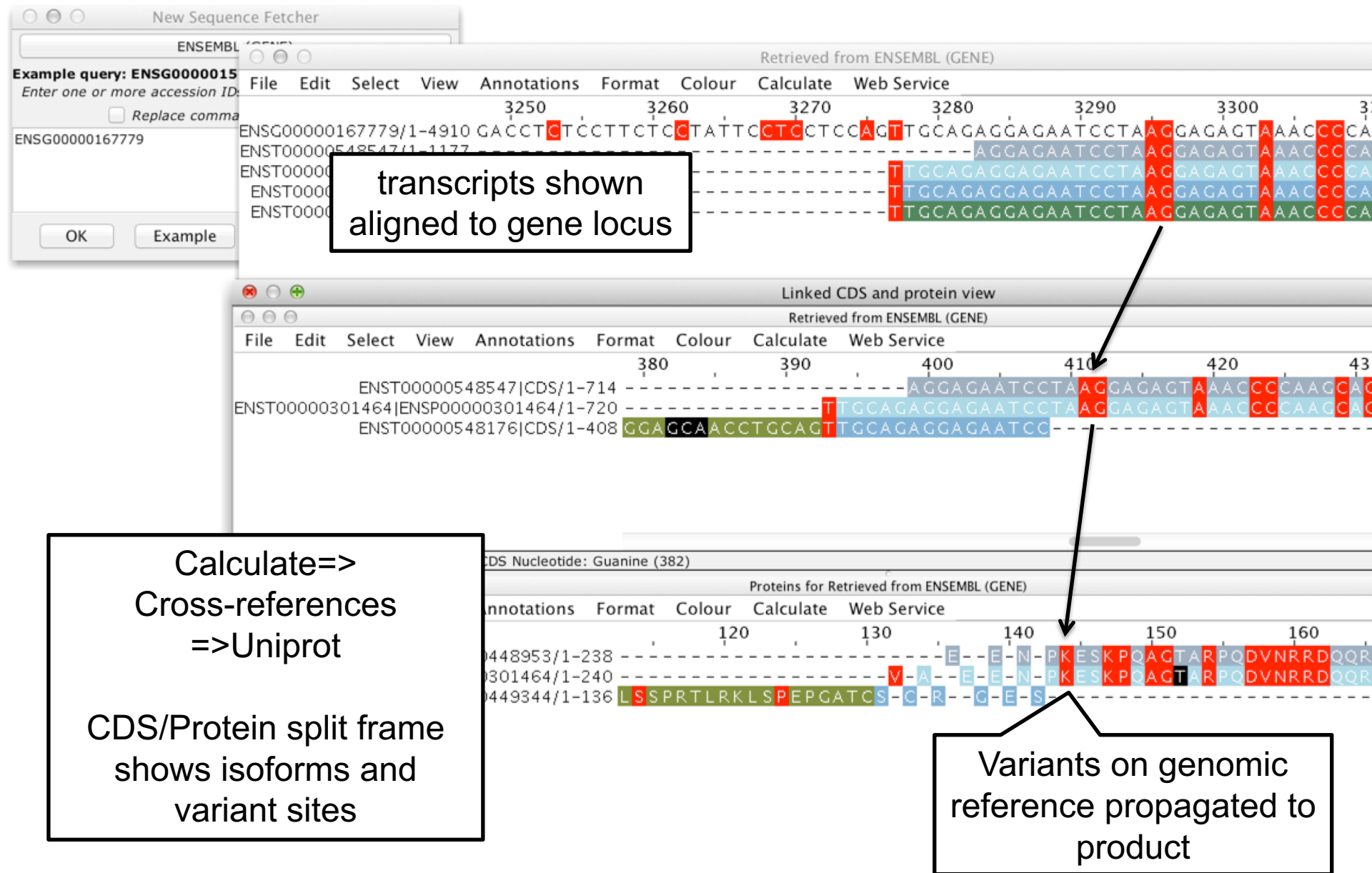


– new in Jalview 2.10

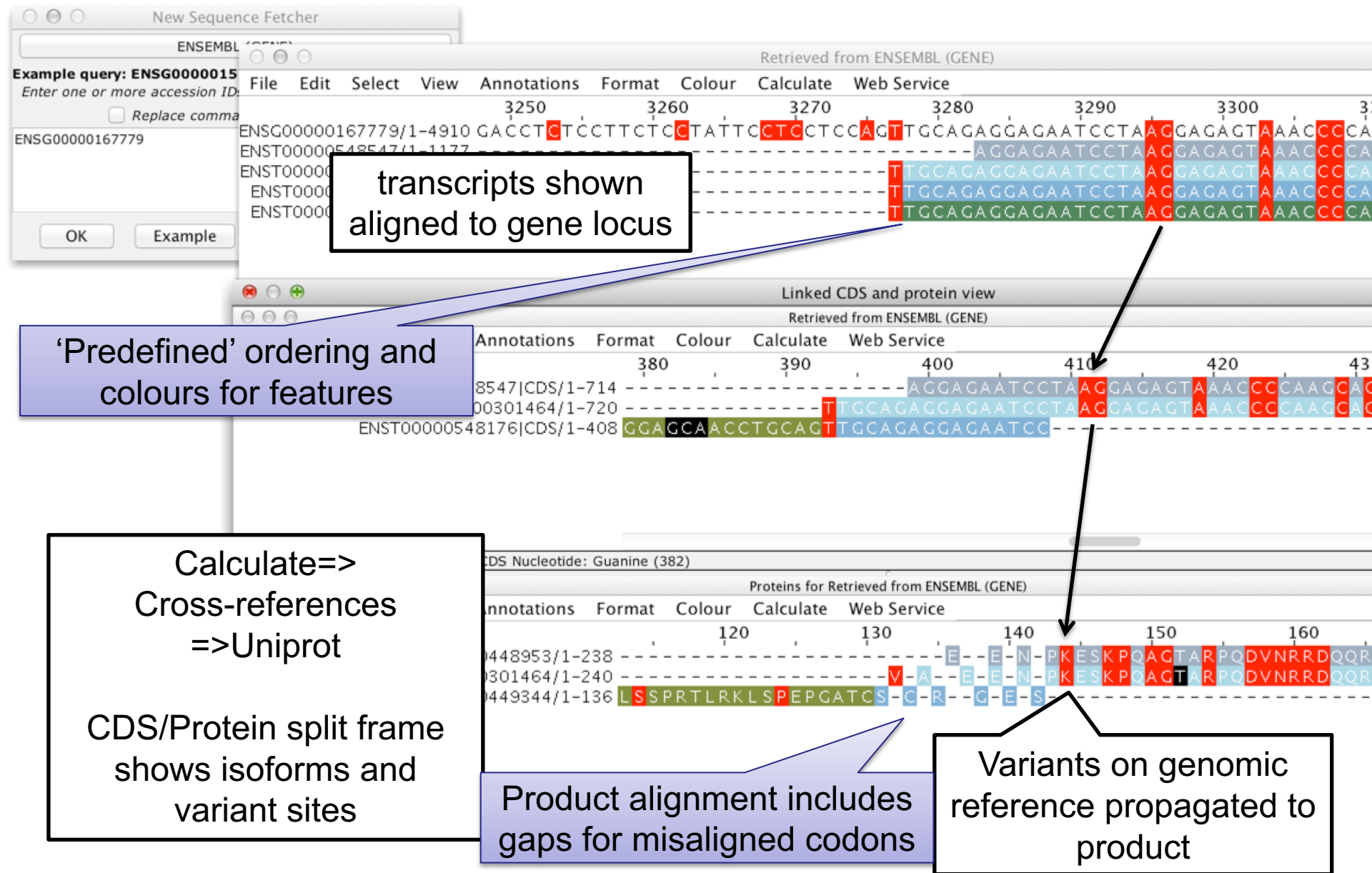
- Principal source for
 - Reference genomes
 - Transcripts
 - Protein products
- Sequence variants



Retrieve Ensembl ENSG0000016779



Retrieve Ensembl
ENSG0000016779



'Split Frame' cDNA & Protein MSAs

Opened with any linked cDNA/Protein data

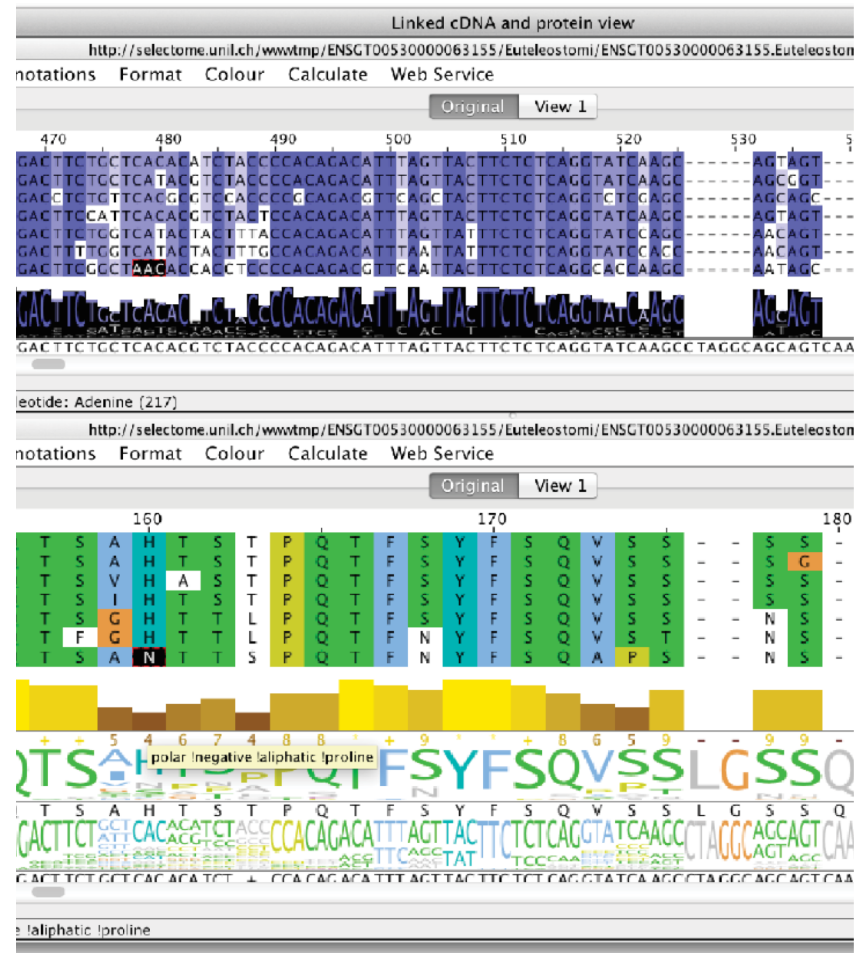
e.g.

- ENA records
- Flat file
- cDNA translation

cDNA/Protein Alignment Reconstruction:

- Gaps inserted according to 'reference alignment'

Layout and display are configurable.



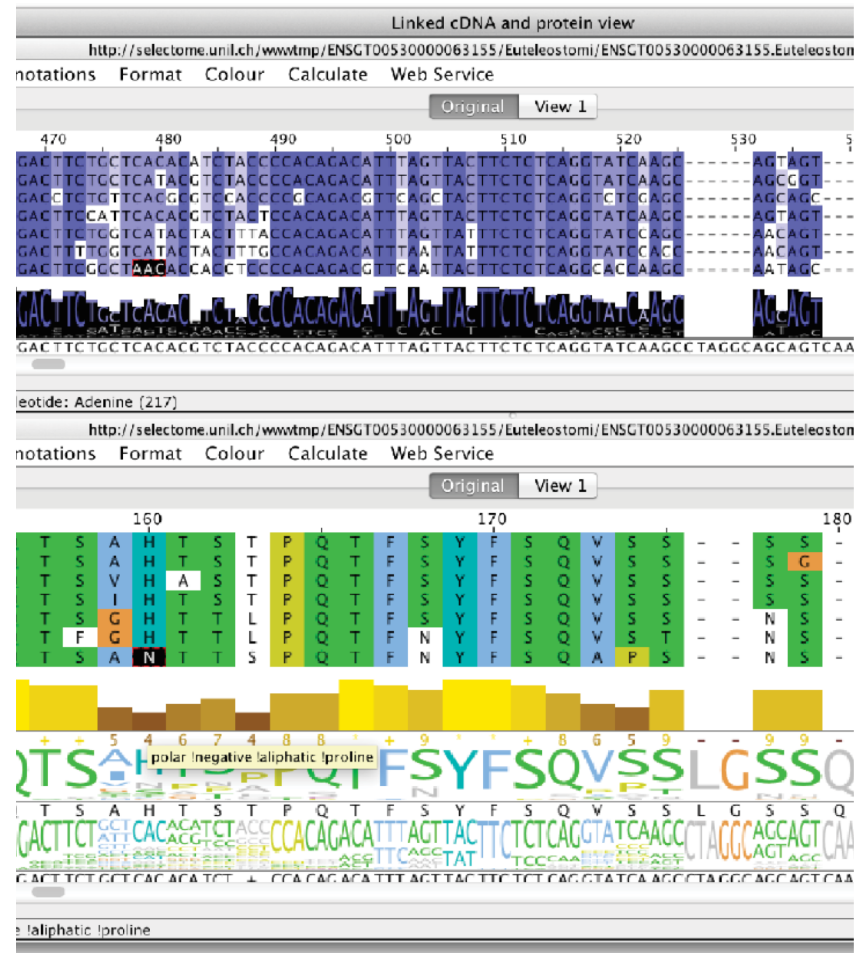
'Split Frame' cDNA & Protein MSAs

Linked alignments share:

- Selections
- Gap editing
- Multiple Views
- Tree sub-groups

Allow new Capabilities

- Codon conservation



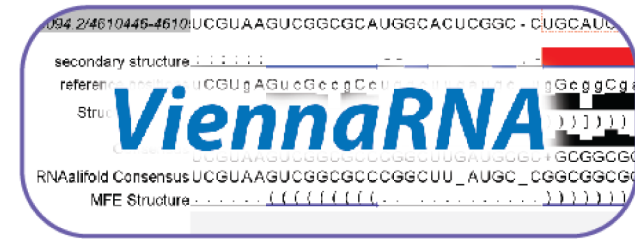
End of Session 5:

Features and Annotation

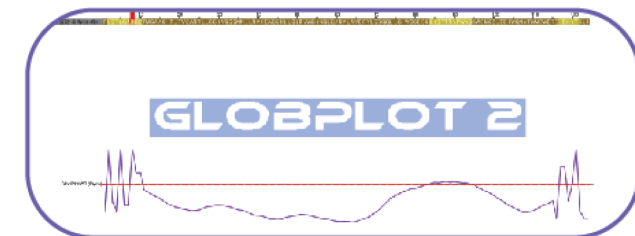
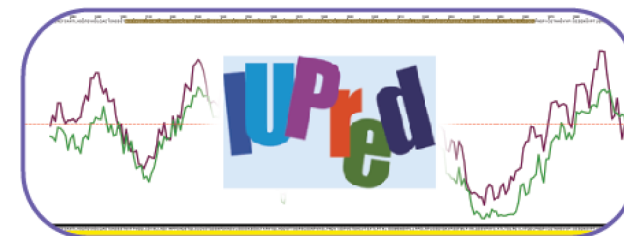
- Understand the difference between Features and Annotation
- learn where features & annotation come from
- Customise feature display
- Use feature and annotation to query the alignment
- CDS/Protein split-view
 - Ensembl
 - 'Show Cross-references'

SECONDARY STRUCTURE PREDICTION

RNAAlifold
Not today!



PROTEIN DISORDER PREDICTION



<http://www.compbio.dundee.ac.uk/jpred>

Jpred 4

Incorporating Jnet



A Protein Secondary Structure Prediction Server

[Home](#)[REST API](#)[About](#)[News](#)[F.A.Q.](#)[Help & Tutorials](#)[Monitoring](#)[Contact](#)[Publications](#)

Input sequence(?)

MQQWPIEGIKKFETLSYLPPLTVEDLLKQIEYLLRSKWVPCLEFSKVGFFVYRENHRSPGYDGRYW
TMWKLPMFGCTDATQVLKEEEAKKAYPDAFVRIIGFDNVRQVQLISFIAYKPPGC

[Advanced options \(click to show/hide\)](#)

[Make Prediction](#)

[Reset Form](#)

Primary citation: Drozdetskiy A, Cole C, Procter J & Barton GJ. Nucl. Acids Res.
(first published online April 16, 2015) doi: 10.1093/nar/gkv332 [[link](#)]
More citations: [link](#).



Protein Secondary Structure Prediction

Section 8.1

- Neural network trained on amino acid profiles
 - Predicts **Helix**, **Sheet**, or **Coil** based on sliding window
 - Also predicts **coiled coils** and **surface accessibilities**
- Server can take
 - **Single Sequence**
 - Service find homologs with PSI-Blast
 - **Alignment**
 - Service uses MSA to calculate profile for prediction

Exercise 27

- Tasks
 - Perform a variety of Jpred4 predictions
 - Note the effect of hidden regions
 - Learn about sequence associated annotation
 - Learn about reference sequence views
- Questions
 - ...

Exercise 27

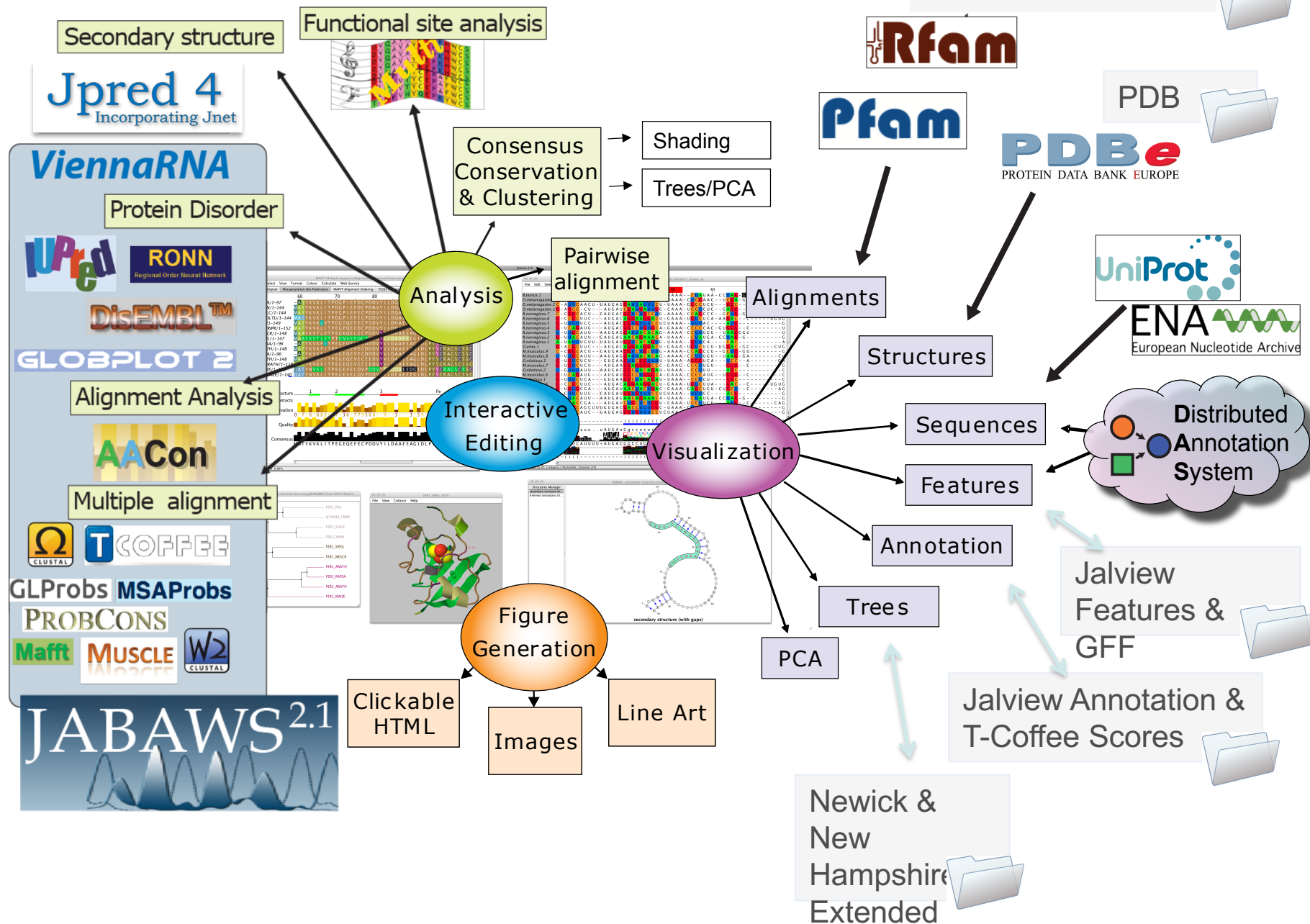
- Tasks
 - Perform a variety of Jpred4 predictions
 - Note the effect of hidden regions
 - Learn about sequence associated annotation
 - Learn about reference sequence views
- Questions
 - What other data does Jnet provide ?
 - Which is better – a PSI blast prediction or an MSA based prediction ?
 - What happens when you have hidden regions ?

Exercise 28 – Disorder predictions

- Run a variety of disorder predictions on Interleukin7
- look for available structure data to evaluate predictions

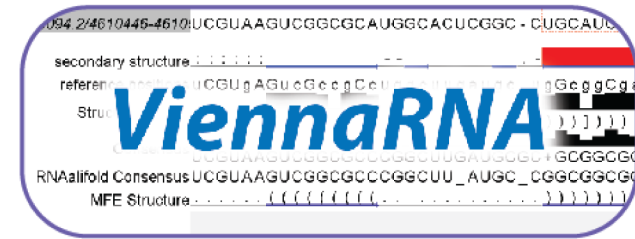
DAY 1 WRAPUP

The Jalview Desktop

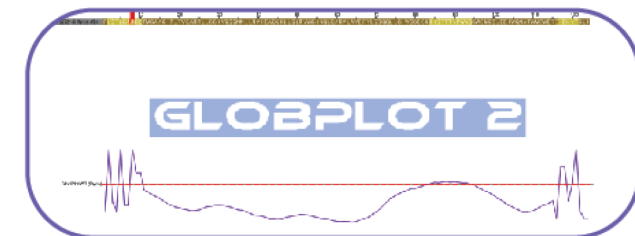
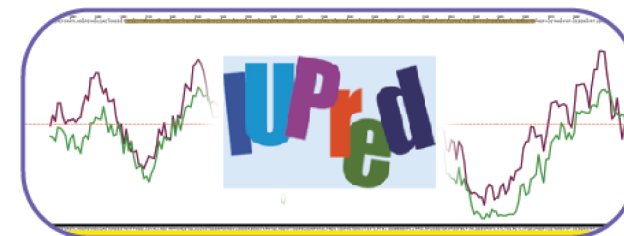


SECONDARY STRUCTURE PREDICTION

RNAAlifold
Not today!



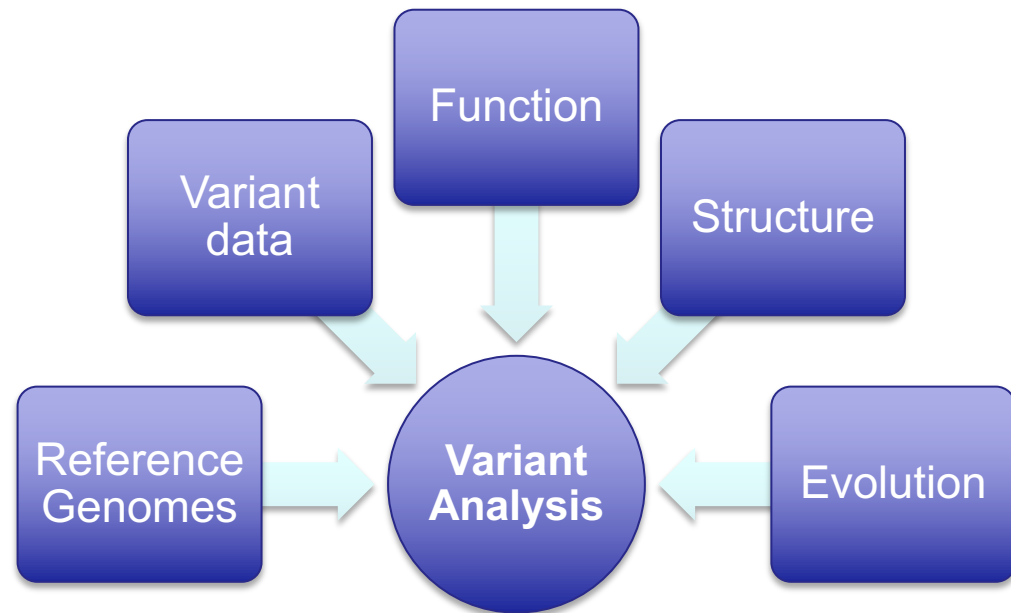
PROTEIN DISORDER PREDICTION

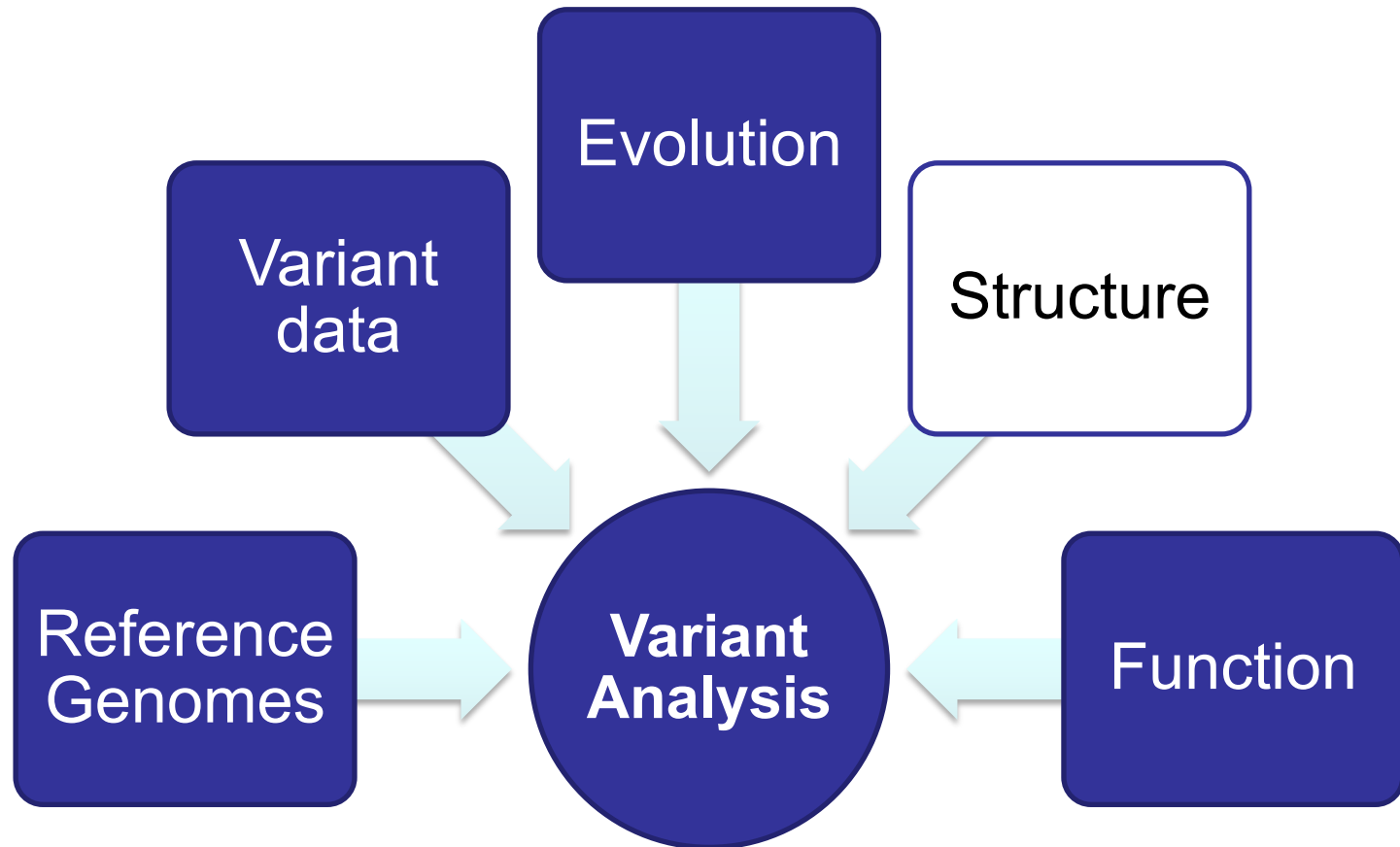




– new in Jalview 2.10

- Principal source for
 - Reference genomes
 - Transcripts
 - Protein products
- Sequence variants





New Sequence Fetcher

ENSEMBL

Example query: ENSG00000157764

Enter one or more accession IDs separated by a semi-colon

☐ Replace commas with semicolons

ENSG00000113924

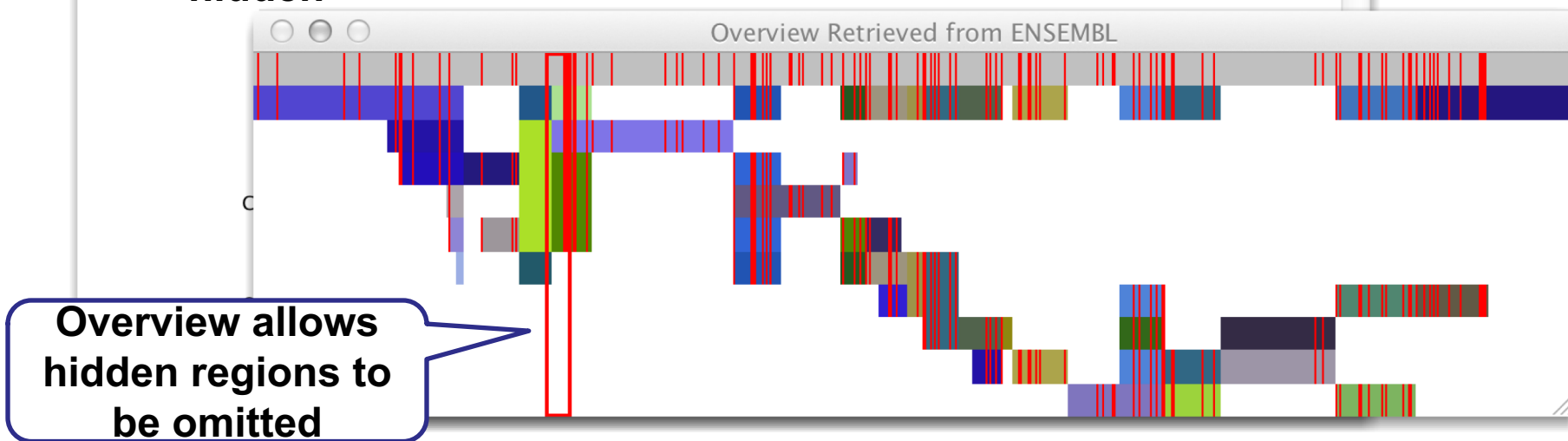
**Ensembl REST API's resolver
discovers loci associated with
entered query**

Retrieved from ENSEMBL

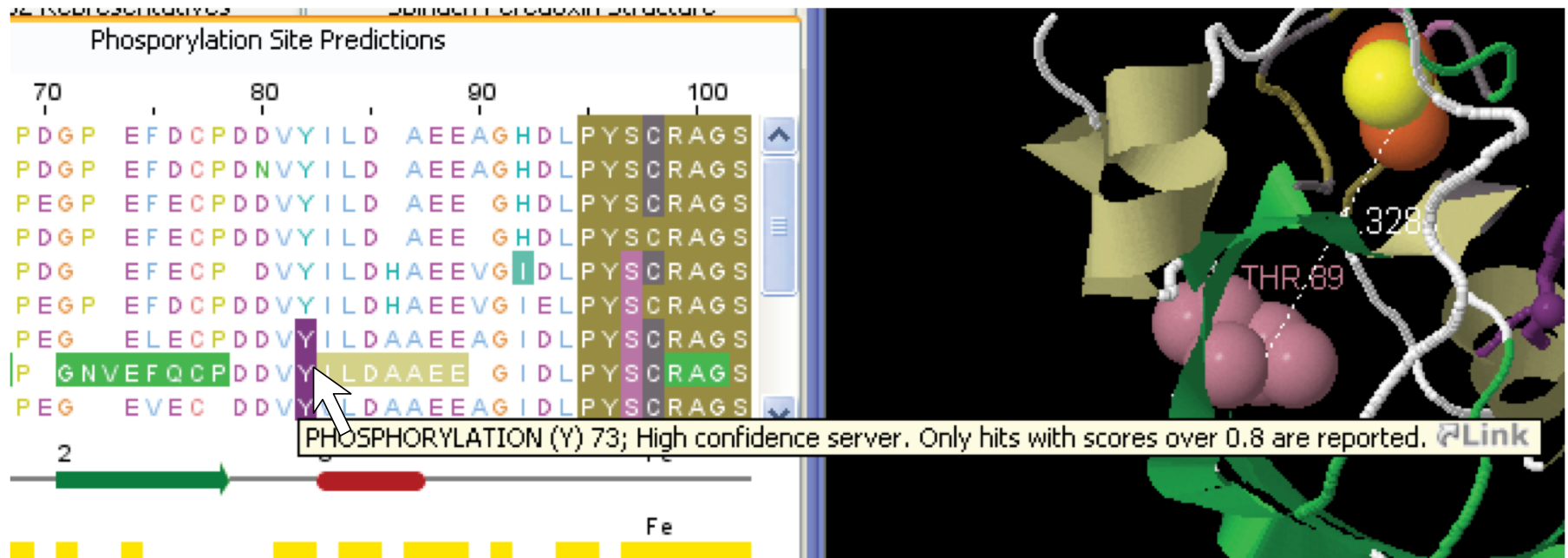
	File	Edit	Select	View	Annotations	Format	Colour	Calculate	Web Service
OK	ENSG00000113924/1-54399				6777	7589	7599	7609	7619
	ENST00000283871/1-2005								
	ENST00000480862/1-656								
	ENST00000488183/1-572								
	ENST00000466528/1-444								
	ENST00000485313/1-528								
	ENST00000476082/1-462								
	ENST00000492108/1-631								
	ENST00000475447/1-565								
	ENST00000494453/1-683								

**Introns
automatically
hidden**

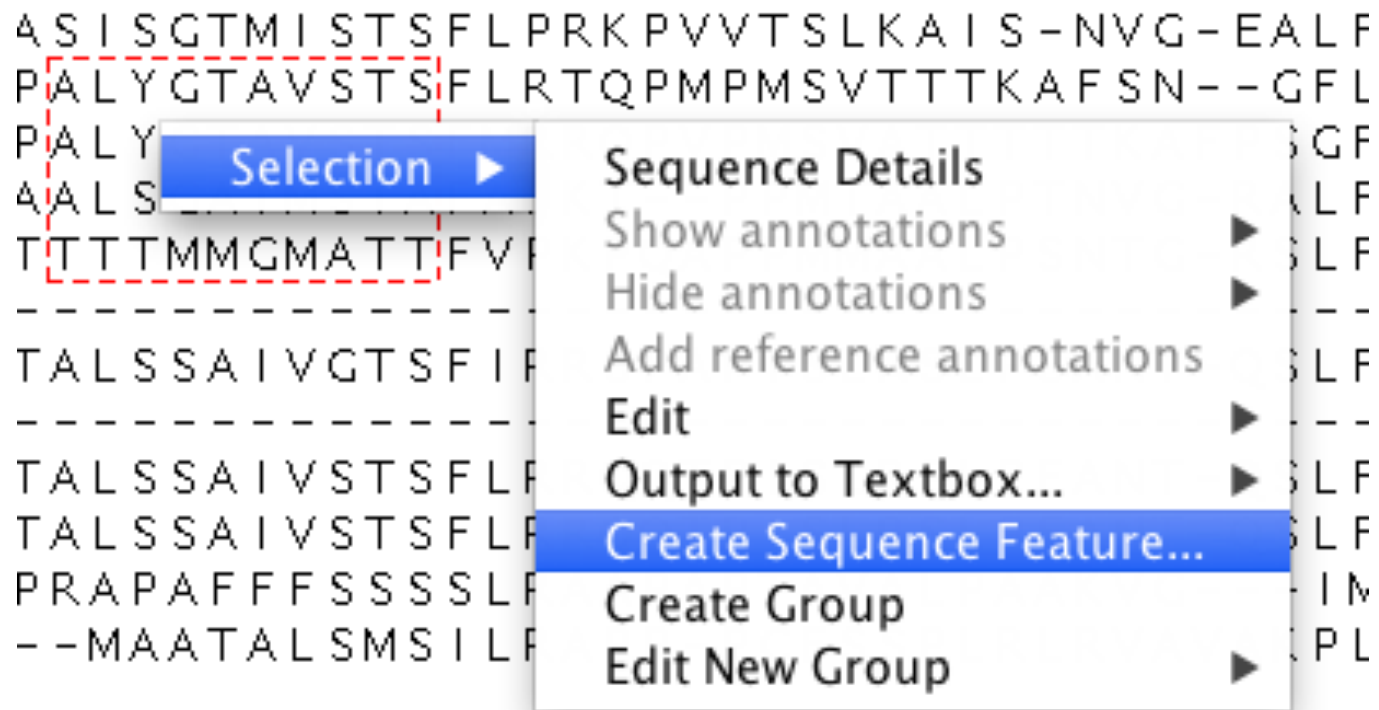
**Calculate->Show Cross-
References to open
CDS/Protein alignment**



Sequence Features have: Name, Description, Group



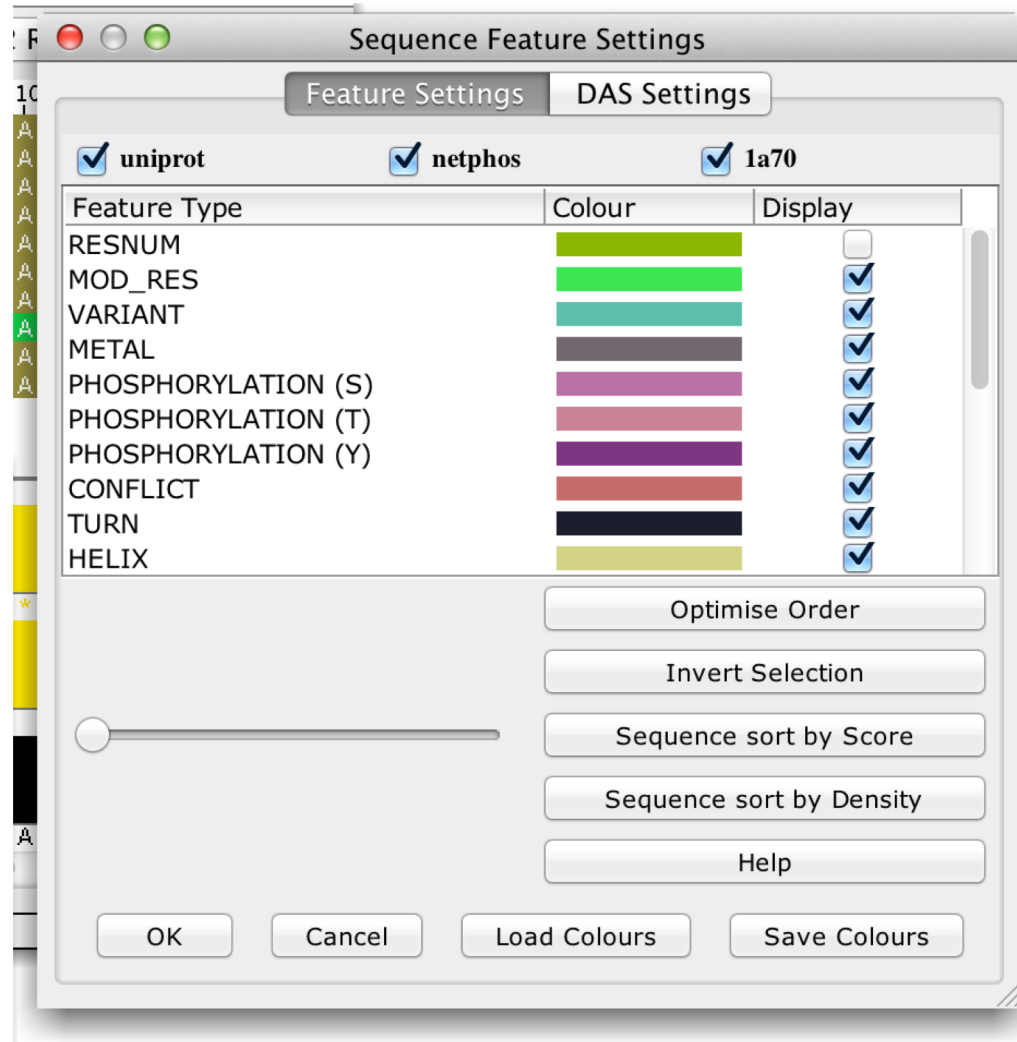
Adding Features Manually



Customising Feature Display

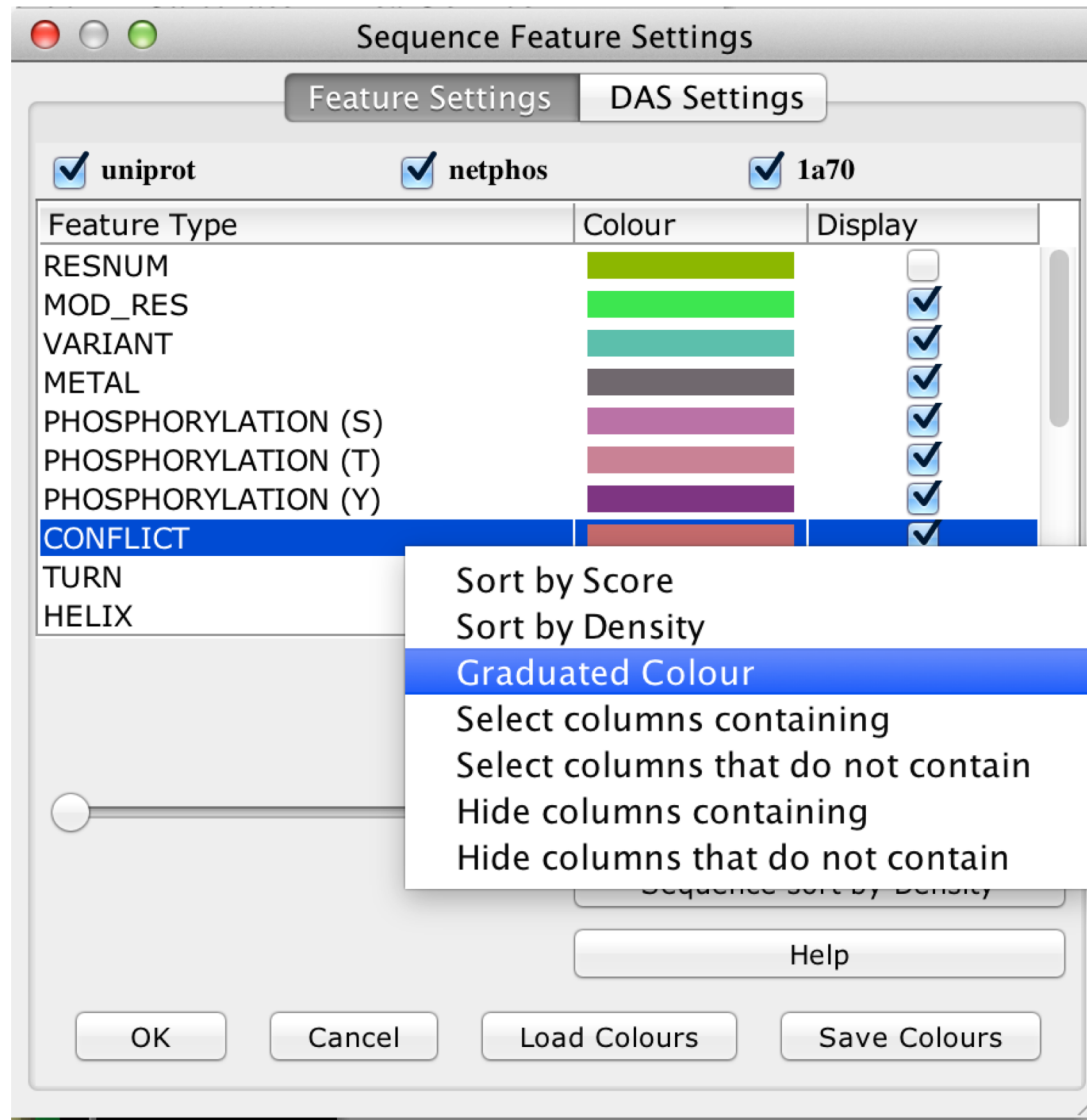
View => Feature Settings...

- Drag features to reorder on alignment
 - Or make transparent
- Change colour
- Enable/disable by name and group



Customising Feature Display

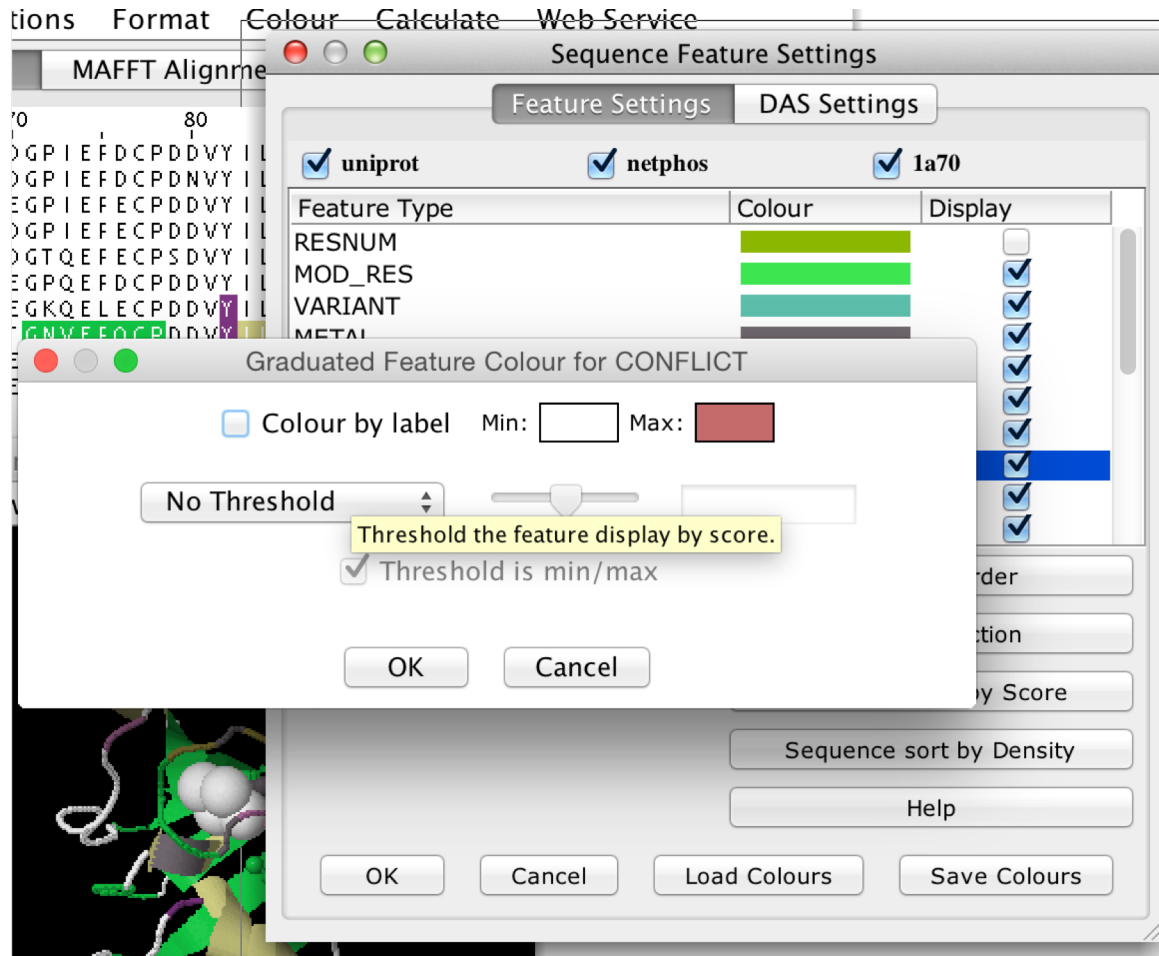
Right-click on feature => Popup menu



**Select
or Hide
columns**

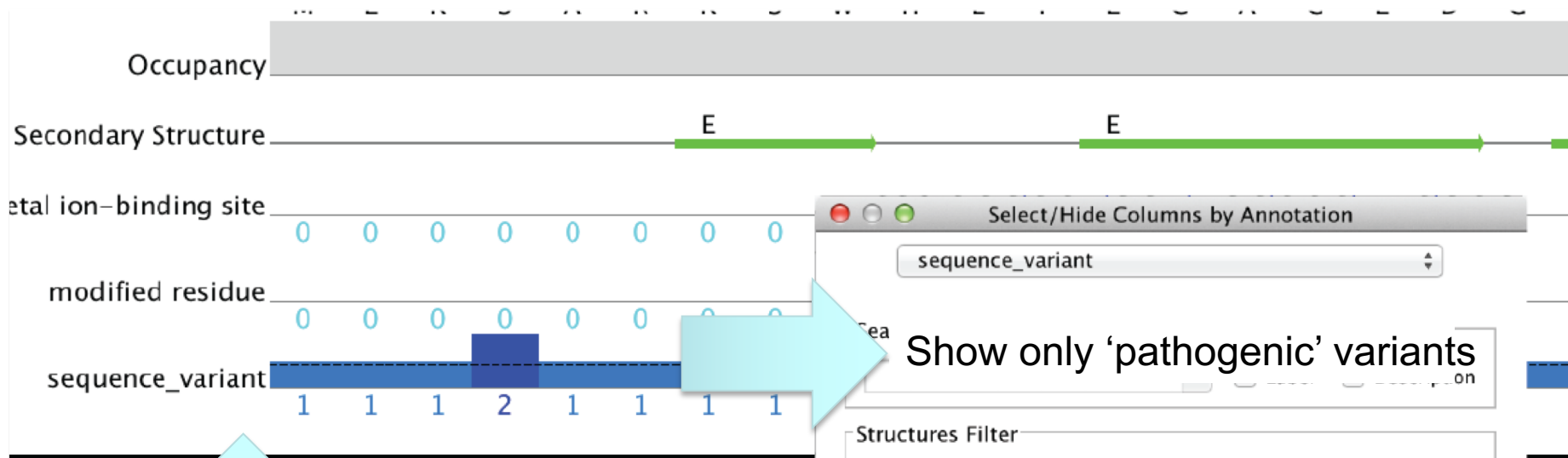
Customising Feature Display

Graduated Colour Scheme



Sources of sequence features

- Sequence databases
- 3D Structure associations
- **GFF2/3 (*Generalised Feature Format*)**
- **Jalview format feature files**
 - Like GFF
 - can create GFF in Excel
 - Allows scores to be added



Feature count tracks
Via Groovy Script

```

* for each feature in visibleFeatures
* Argument 'features' will be a list of SequenceFeature
*/
def getCounts =
{ features ->
    int[] obs = new int[visibleFeatures.size]
    for (sf in features)
    {
        /*
        * Here we inspect the type of the sequence feature.
        * You can also test sf.description, sf.score, sf.featureGroup,
        * sf.strand, sf.phase, sf.begin, sf.end
        * or sf.getValue(attributeName) for GFF 'column 9' properties
        */
        String clinSig = (String) sf.getValue(GffConstants.CLINICAL_SIGNIFICANCE);
        if (sf.type.equals("sequence_variant") && (clinSig==null || clinSig.equals("")))
        {
            int pos = 0
            for (type in visibleFeatures)
            {
                if (type.equals(sf.type))
                {
                    obs[pos]++
                }
                pos++
            }
        }
    }
    obs
}

```

Select/Hide Columns by Annotation

sequence_variant

Show only 'pathogenic' variants

Structures Filter

☐ Select All ☐ Alpha Helix ☐ Beta Strand ☐ Turn

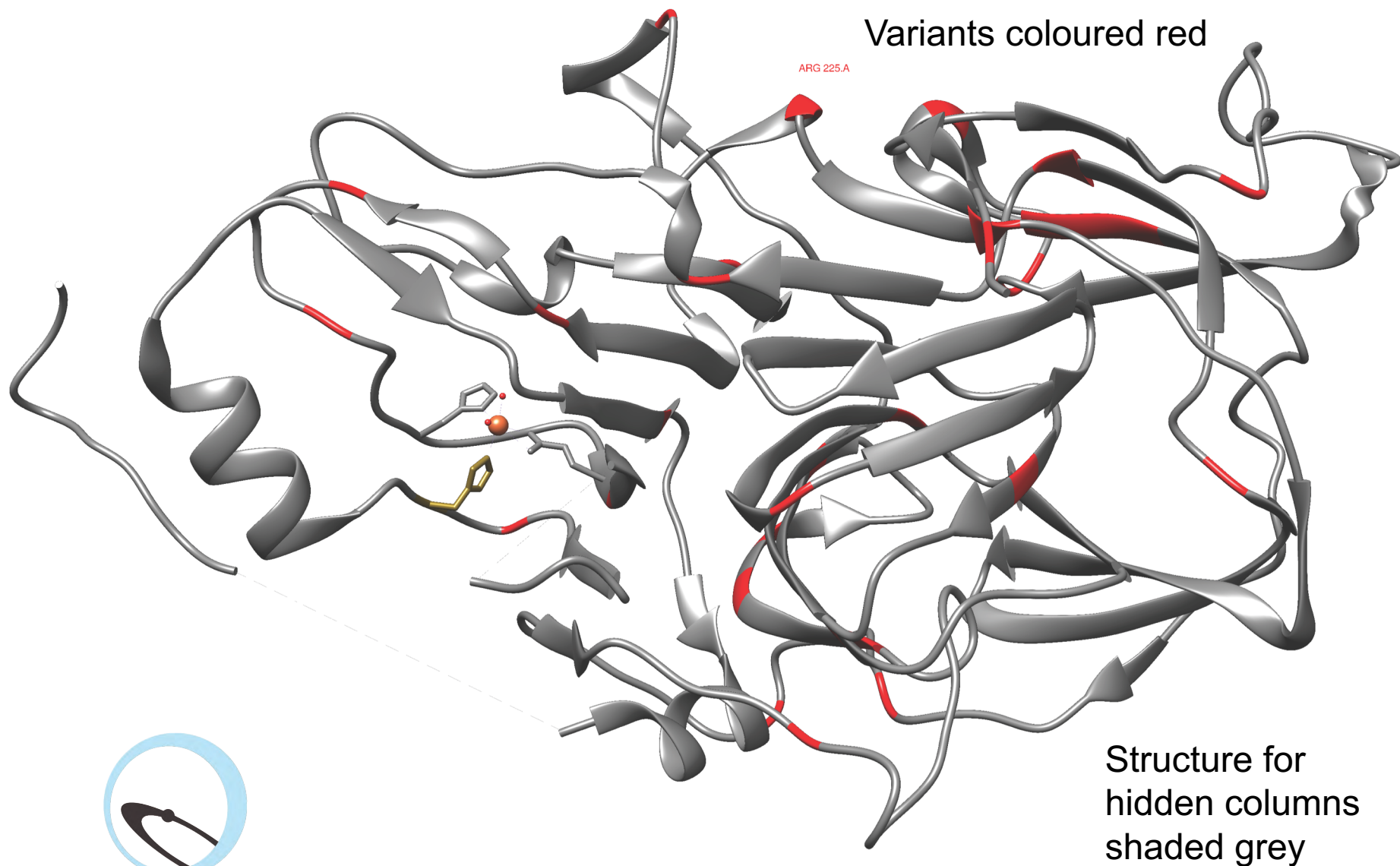
Threshold Filter

Below Threshold ☐ As Percentage

0.853

☐ Select ☒ Hide

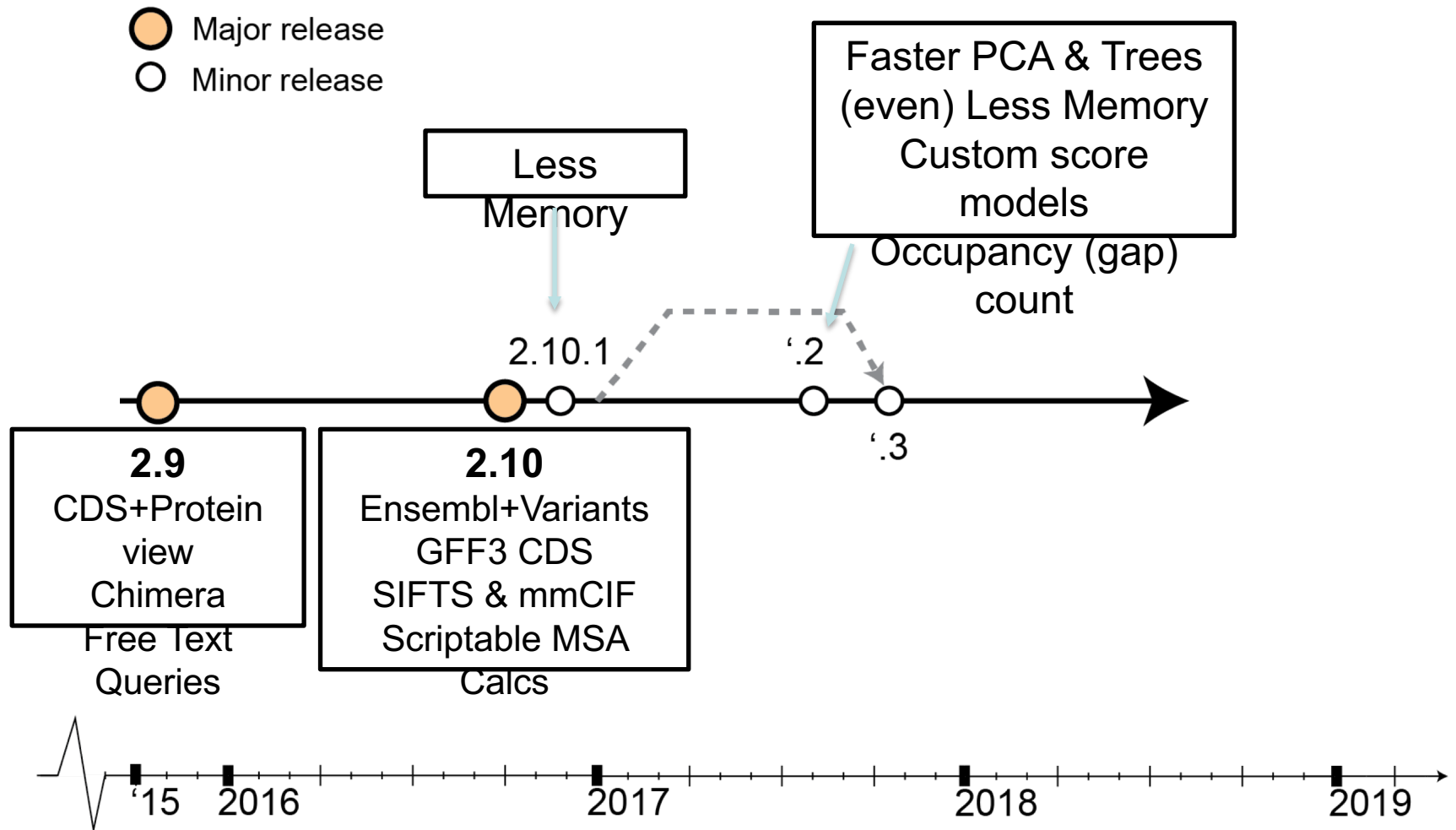
OK Cancel



UCSF CHIMERA

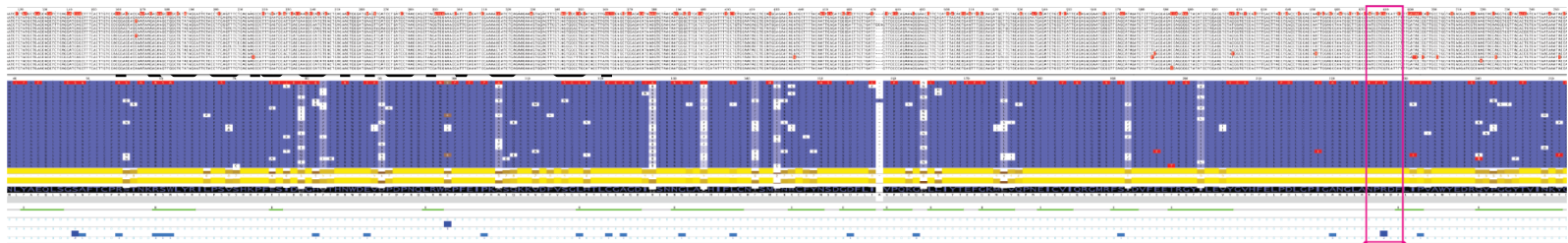
Human Homogentisate 1-2, dioxygenase
PDB ID 1eyp. Sites with pathogenic variants
in red

Key new features in Jalview releases



2.10.3: Optimisations for variant data

- Homogentisate diogxygenase locus is **54kbp**
- **10,000 variants** from Ensembl
- Jalview must enable:
 - **Fast import**

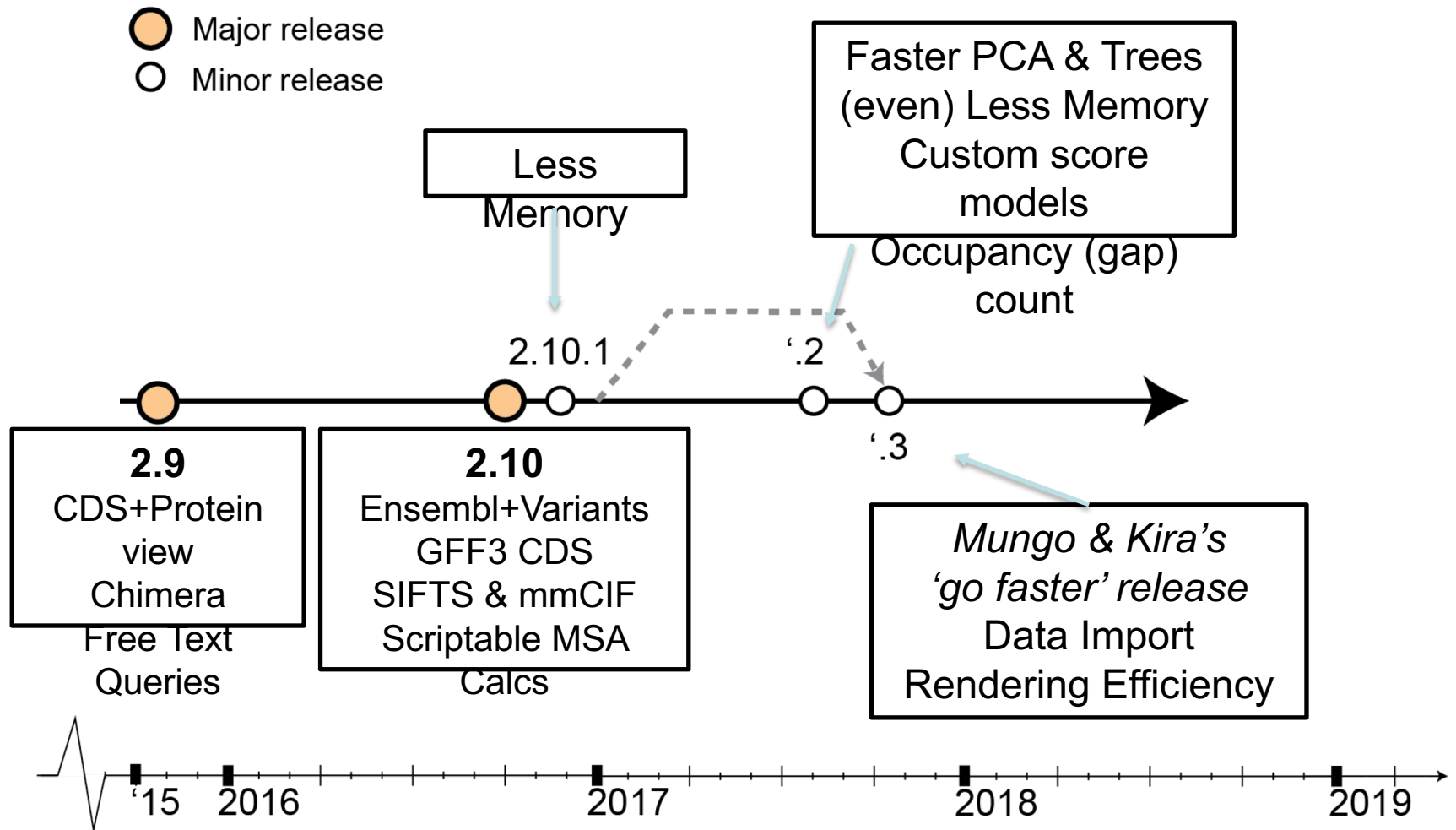


sequence_variant 674; G,C,A; likely pathogenic
sequence_variant 674; HGMD_MUTATION
sequence_variant 674; HGMD_MUTATION

Variants from locus translated to product

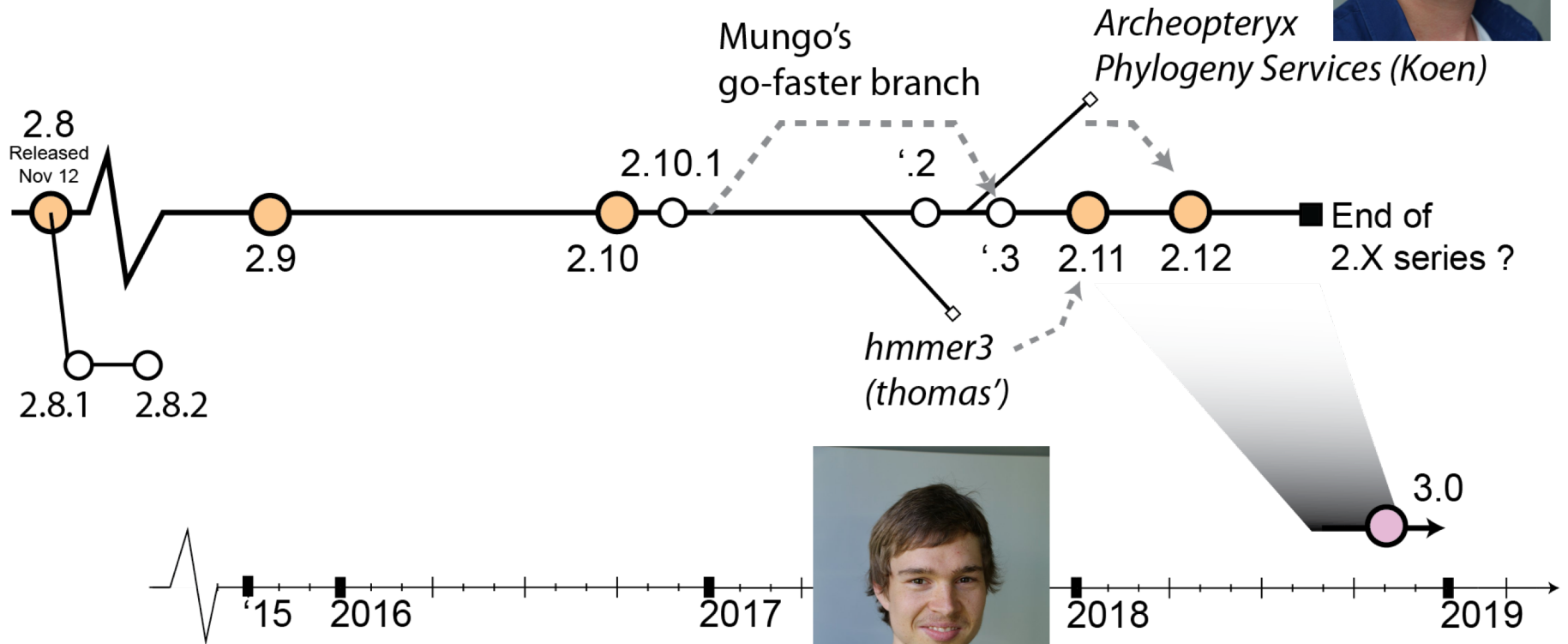


Key new features in Jalview releases



- Major release
- Minor release

Spike (Experimental)



A HMMER for Jalview 2.11

HBB_HUMAN/1-146 L T P E E K S A V T A L W G K V
 HBA_HUMAN/1-141 L S P A D K T N V K A A W G K V
 MYG_PHYCA/1-153 L S E G E W Q L V L H V W A K V
 GLB5_PETMA/1-149 L S A A E K T K I R S A W A P V

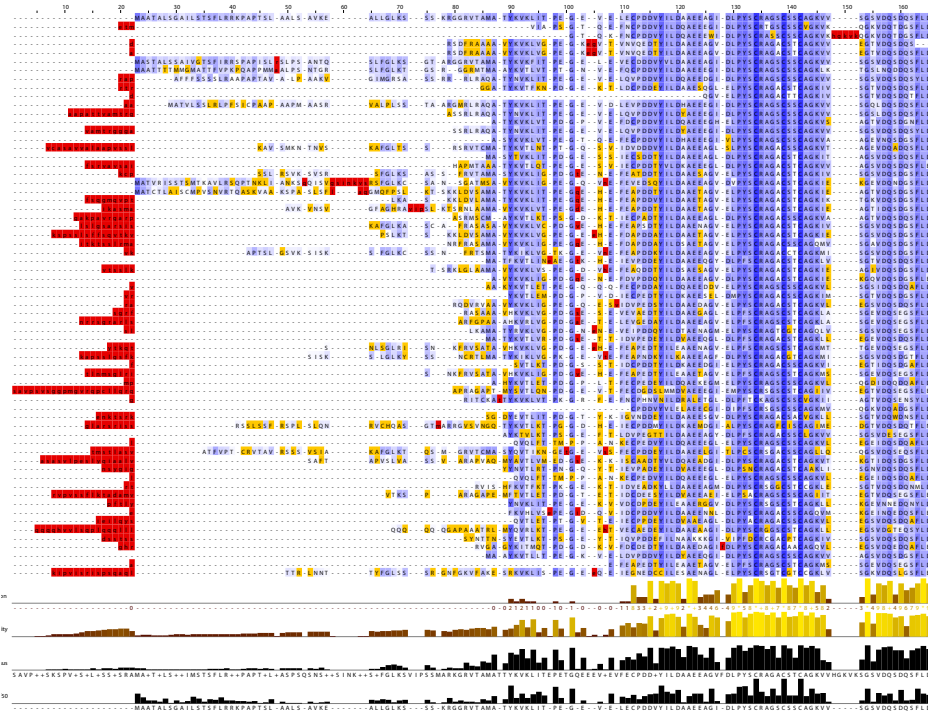
hmmbuild

G4_HMM/8-164 L S E A E K T K V K A V W A K V
 HBB_HUMAN/1-146 L T P E E K S A V T A L W G K V
 Group_HMM/0-7 - - - - K T L V K A V W - - -
 HBA_HUMAN/1-141 L S P A D K T N V K A A W G K V
 MYG_PHYCA/1-153 L S E G E W Q L V L H V W A K V
 GLB5_PETMA/1-149 L S A A E K T K I R S A W A P V

G4_HMM



Group_HMM



Deciding E-value



hmmsearch



2.11 - VCF import



VCF is a text file format (most likely stored in a compressed manner). It contains meta-information lines, a header line, and then data lines each containing information about a position in the genome.

There is an option whether to contain genotype information on samples for each

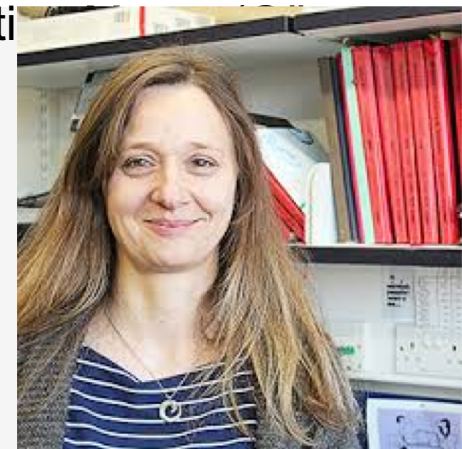
Example:

```
##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO
A00001 NA00002 NA00003
```

Challenges

1. Map **chromosomal coordinates** to aligned sequences
2. Visualise and Filter on **variant attributes**

Alpha testers: Beti (lab)



FORMAT

N

Feb '18: Jalview meets Forester



Powerful phylogeny viewer

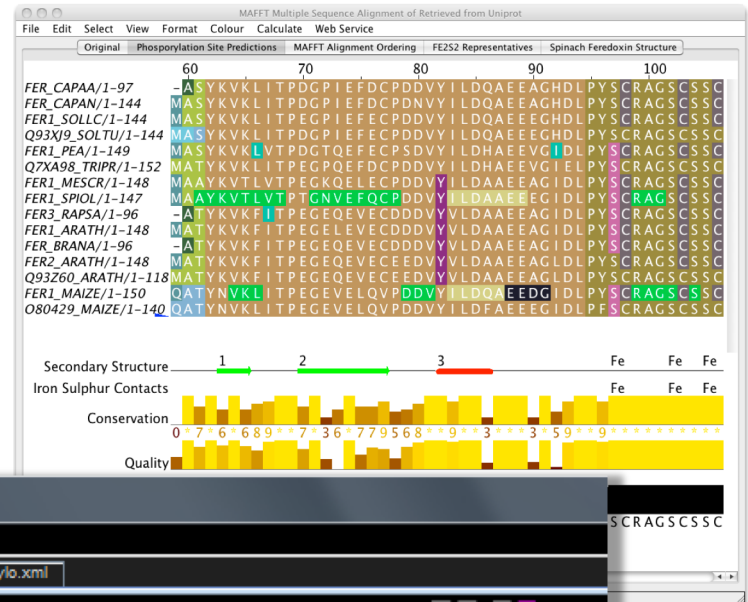
- replace Jalview's tree display

File format support

- NEXUS, phyloxml

Web services

- iqtree, mrbayes



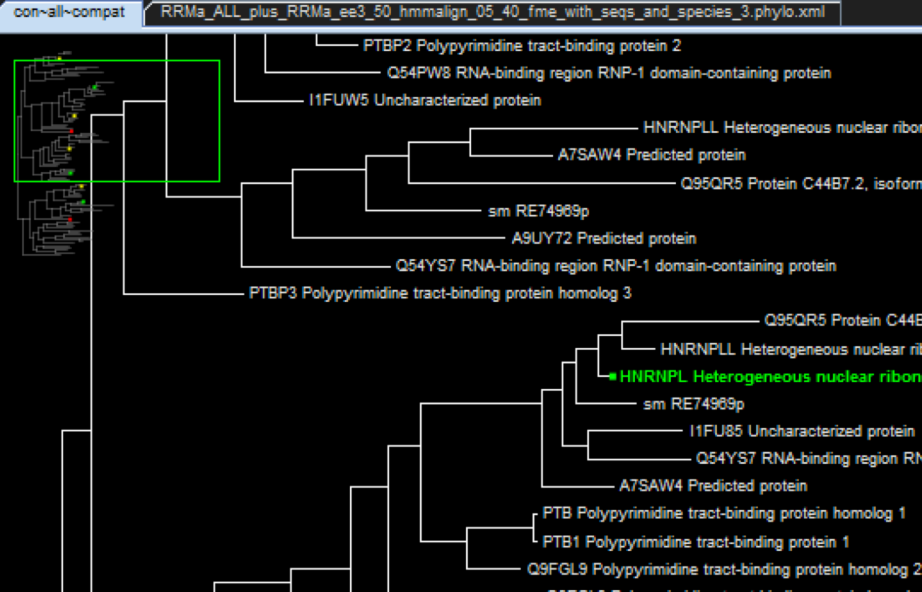
Archaeopteryx 0.9901 beta (141118)

File Inference Analysis Tools View Font Size Options Type Help

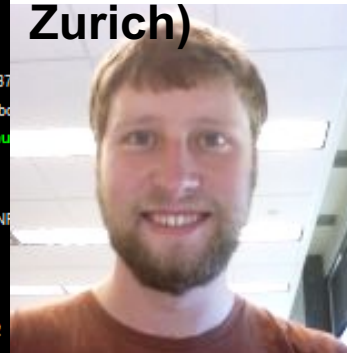
- ☒ Phylogram
- ☒ Dyna Hide
- ☒ Rollover
- ☒ Show Internal Data
- ☐ Colorize by Sequence
- ☐ Colorize by Taxonomy
- ☐ Colorize by Annotation
- ☐ Visual Styles/Branch Colors
- ☐ Branch Widths

Display Data:

- ☐ Node Name
- ☐ Taxonomy Code
- ☐ Taxonomy Scientific
- ☐ Taxonomy Common
- ☒ Seq Name
- ☒ Gene Name
- ☐ Seq Symbol
- ☐ Seq Accession
- ☐ Seq Annotations



Alpha testers: Spencer Bliven & Maria Anisimova (U. Zurich)



Homework: 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 ?
 - What happens when you drag an exported HTML file back on to Jalview ?

Thanks !

- Feedback forms
 - See your email
- Any more questions, problems, calls for help ?
- Interested in trying new features ?
- Interested in helping develop them ?

Join
[jalview-discuss](#)



Sign up at
issues.jalview.org