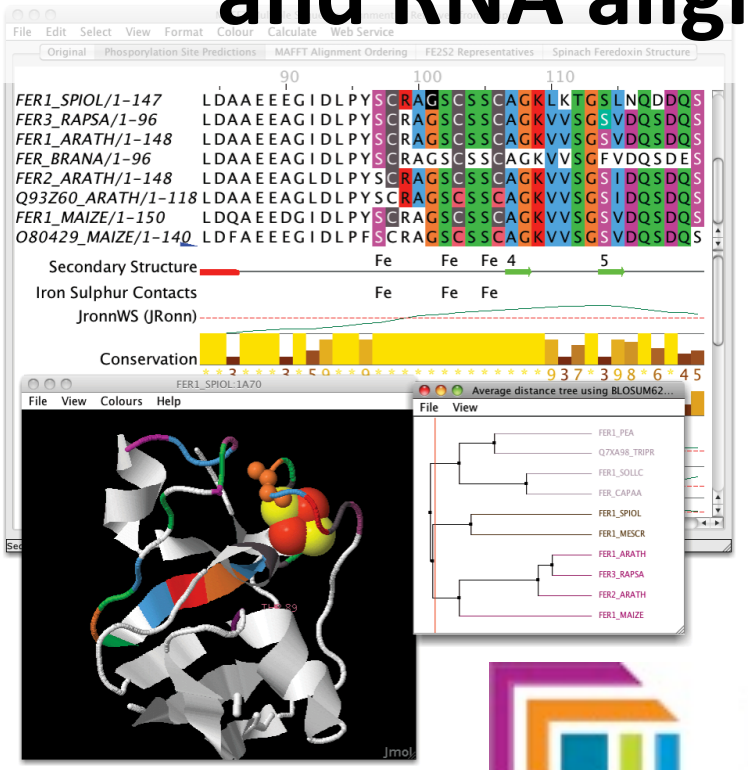
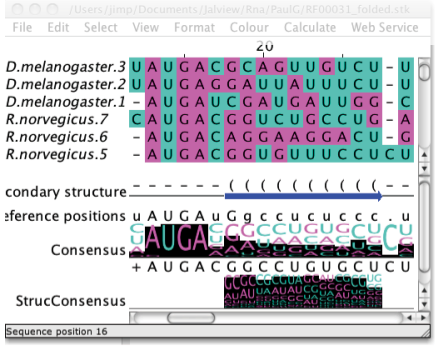
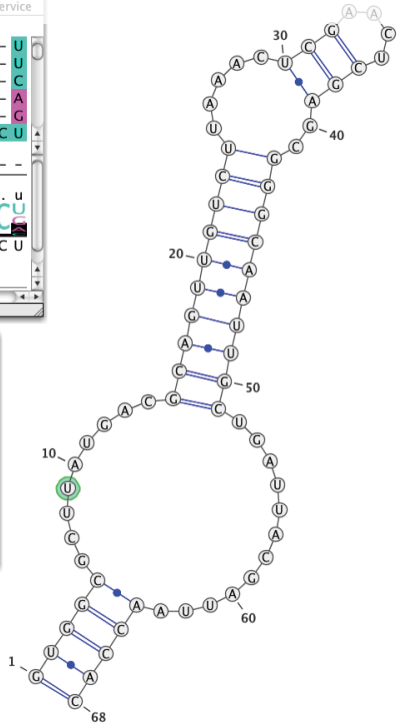
Creation, curation and analysis of Protein and RNA alignments with Jalview







Jalview

www.jalview.org

Jim Procter, Jalview Coordinator

j.procter@dundee.ac.uk

Scottish Phylogeny Discussion Group

James Hutton Institute 18th February 2013

University of Dundee

The Barton Group

School of Life Sciences, University of Dundee, Dow St, Dundee DD1 5EH, Scotland, UK

BBSRC bioscience for the future

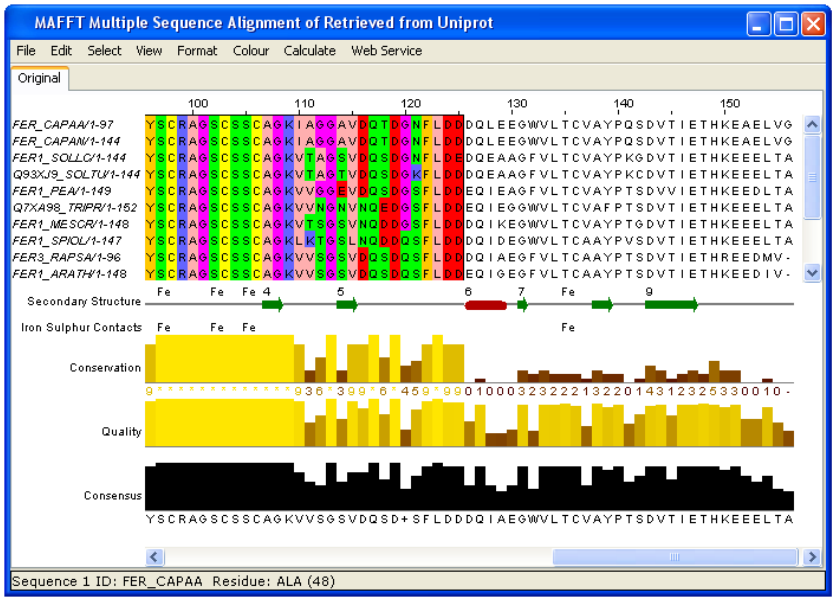
open source initiative

What is Jalview ?

- A java alignment viewer

  java ***al***ignment ***view***er

  .. but not just for viewing..
- Java ?
  - Platform independence
  - Standalone or web based tool



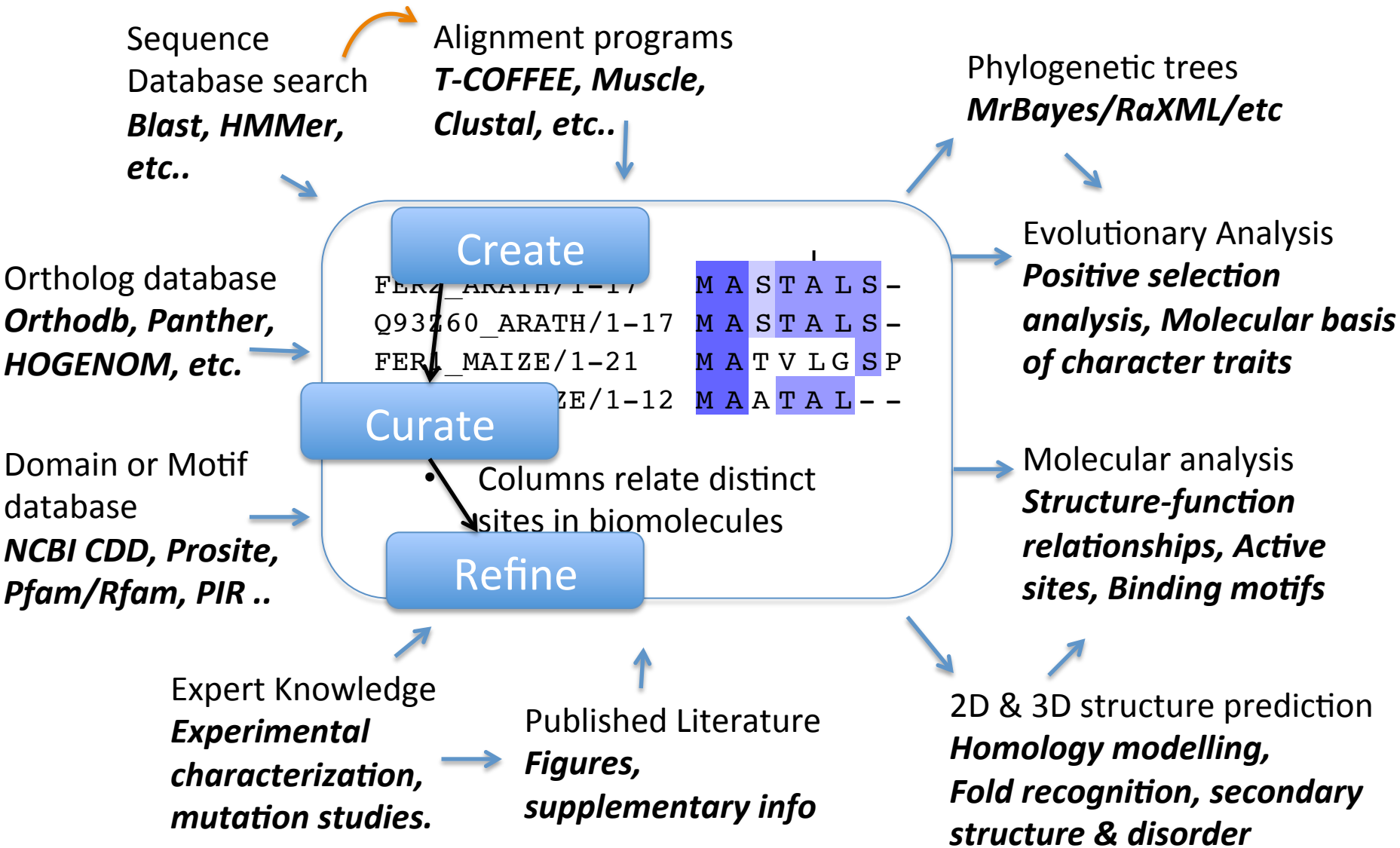
- Open source since '98
  - Coordinated in Dundee
  - Core development funded by …

UNIVERSITY OF DUNDEE

The Barton Group
School of Life Sciences, University of Dundee, Dow St, Dundee DD1 5EH, Scotland, UK

BBSRC bioscience for the future

open source initiative


Sequence Database search
Blast, HMMer, etc..
Alignment programs
T-COFFEE, Muscle, Clustal, etc..
Phylogenetic trees
MrBayes/RaXML/etc
Ortholog database
Orthodb, Panther, HOGENOM, etc.
Create
Q93Z60_ARATH/1-17 M A S T A L S -
FER1_MAIZE/1-21 M A T V L G S P
M A A T A L - -
Curate
Columns relate distinct sites in biomolecules
Refine
Evolutionary Analysis
Positive selection analysis, Molecular basis of character traits
Domain or Motif database
NCBI CDD, Prosite, Pfam/Rfam, PIR ..
Molecular analysis
Structure-function relationships, Active sites, Binding motifs
Expert Knowledge
Experimental characterization, mutation studies.
Published Literature
Figures, supplementary info
2D & 3D structure prediction
Homology modelling, Fold recognition, secondary structure & disorder

Curating Alignments

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

```
FER_CAPAA/28-82       AEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQLEEGWVLTCVAYPQ
FER_CAPAN/75-129      AEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQLEEGWVLTCVAYPQ
FER1_LYCES/75-129     AEEEGHDLPYSCRAGSCSSCAGKVTAGSVDQSDGNFLDEDQEAAGFVLTCVAYPK
Q93XJ9_SOLTU/75-129   AEEEGHDLPYSCRAGSCSSCAGKVTAGTVDQSDGKFLDDDQEAAGFVLTCVAYPK
FER1_PEA/80-134       AEEVGIDLPYSCRAGSCSSCAGKVVGGEVDQSDGSFLDDEQIEAGFVLTCVAYPT
Q7XA98_TRIPR/83-137   AEEVGIELPYSCRAGSCSSCAGKVVNGNVNQEDGSFLDDEQIEGGWVLTCVAFPT
FER1_MESCR/79-133     AEEAGIDLPYSCRAGSCSSCAGKVTSGSVNQDDGSFLDDDQIKEGWVLTCVAYPT
FER1_SPIOL/78-132     AEEEGIDLPYSCRAGSCSSCAGKLKTGSLNQDDQSFLDDDQIDEGWVLTCAAYPV
FER3_RAPSA/28-82      AEEAGIDLPYSCRAGSCSSCAGKVVSGSVDQSDQSFLDDDQIAEGFVLTCAAYPT
FER_ARATH/80-134      AEEAGIDLPYSCRAGSCSSCAGKVVSGSVDQSDQSFLDDEQIGEGFVLTCAAYPT
FER_BRANA/28-82       AEEAGIDLPYSCRAGSCSSCAGKVVSGFVDQSDESFLDDDQIAEGFVLTCAAYPT
FER2_ARATH/80-134     AEEAGLDLPYSCRAGSCSSCAGKVVSGSIDQSDQSFLDDEQMSEGYVLTCVAYPT
FER1_MAIZE/80-134     AEEDGIDLPYSCRAGSCSSCAGKVVSGSVDQSDQSYLDDGQIADGWVLTCHAYPT
O80429_MAIZE/72-126   AEEEGIDLPFSCRAGSCSSCAGKVVSGSVDQSDQSFLNDNQVADGWVLTCAAYPT
```

Alignments are central …

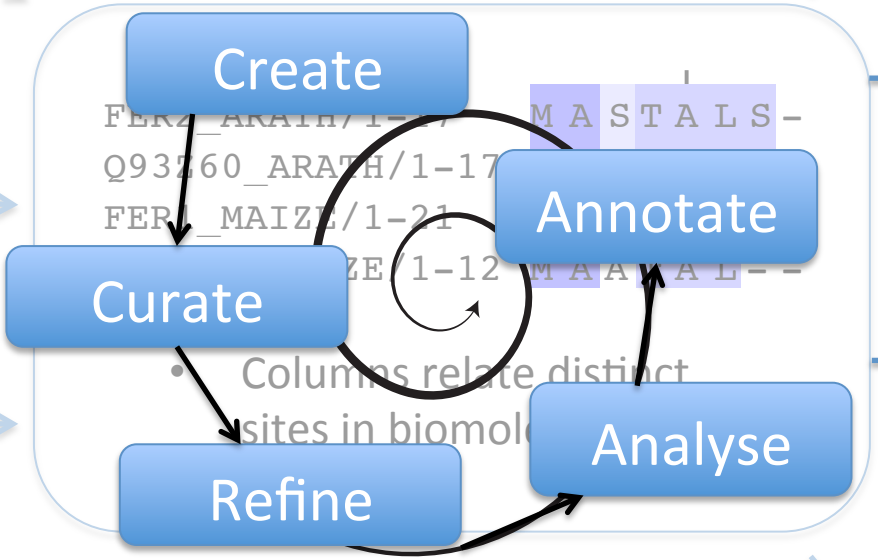
Sequence Database search
Blast, HMMer, etc..

Alignment programs
T-COFFEE, Muscle, Clustal, etc..

Phylogenetic trees
MrBayes/RaXML/etc

Ortholog database
Orthodb, Panther, HOGENOM, etc.

Domain or Motif database
NCBI CDD, Prosite, Pfam/Rfam, PIR ..



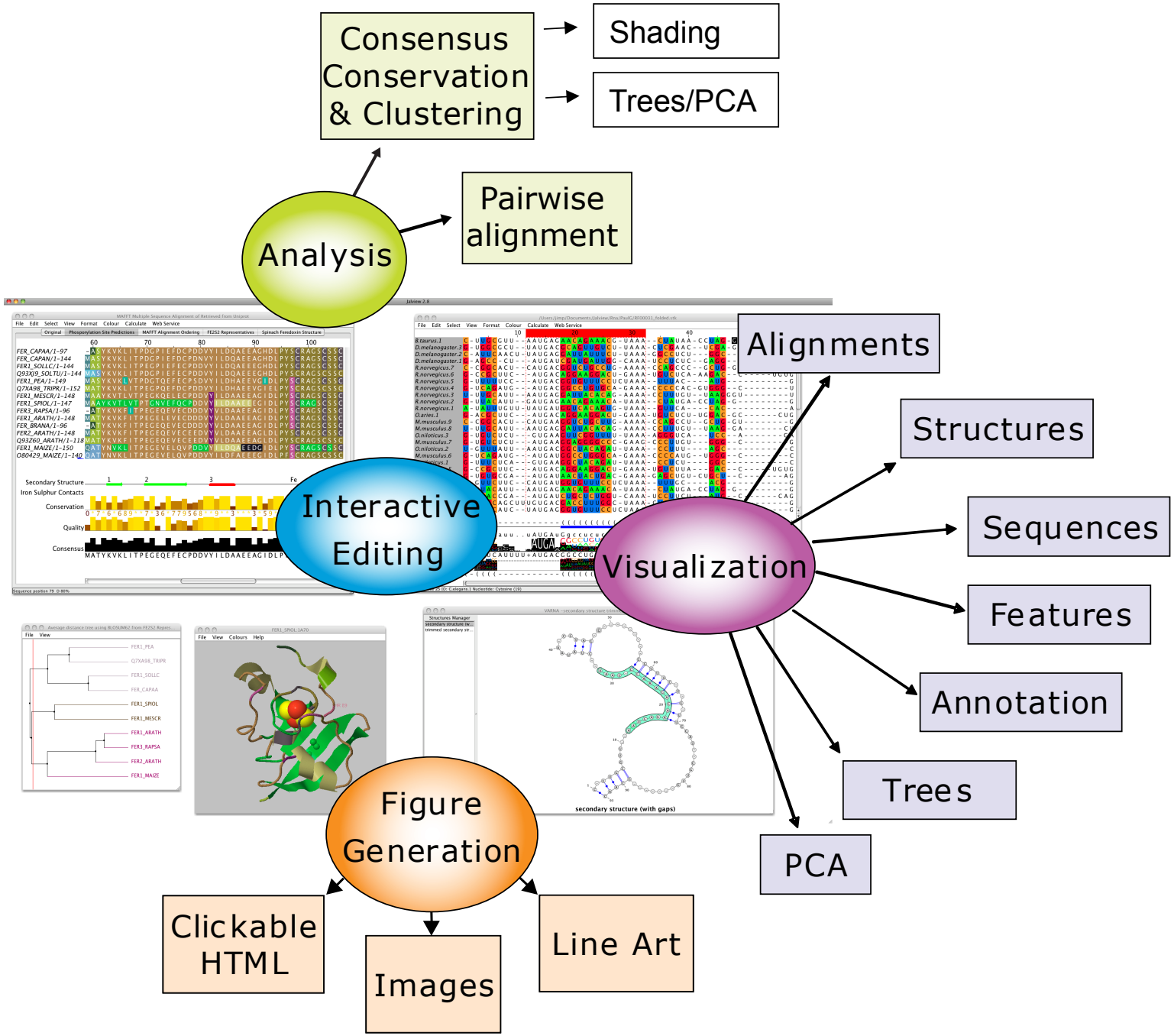
Evolutionary Analysis
Positive selection analysis, Molecular basis of character traits

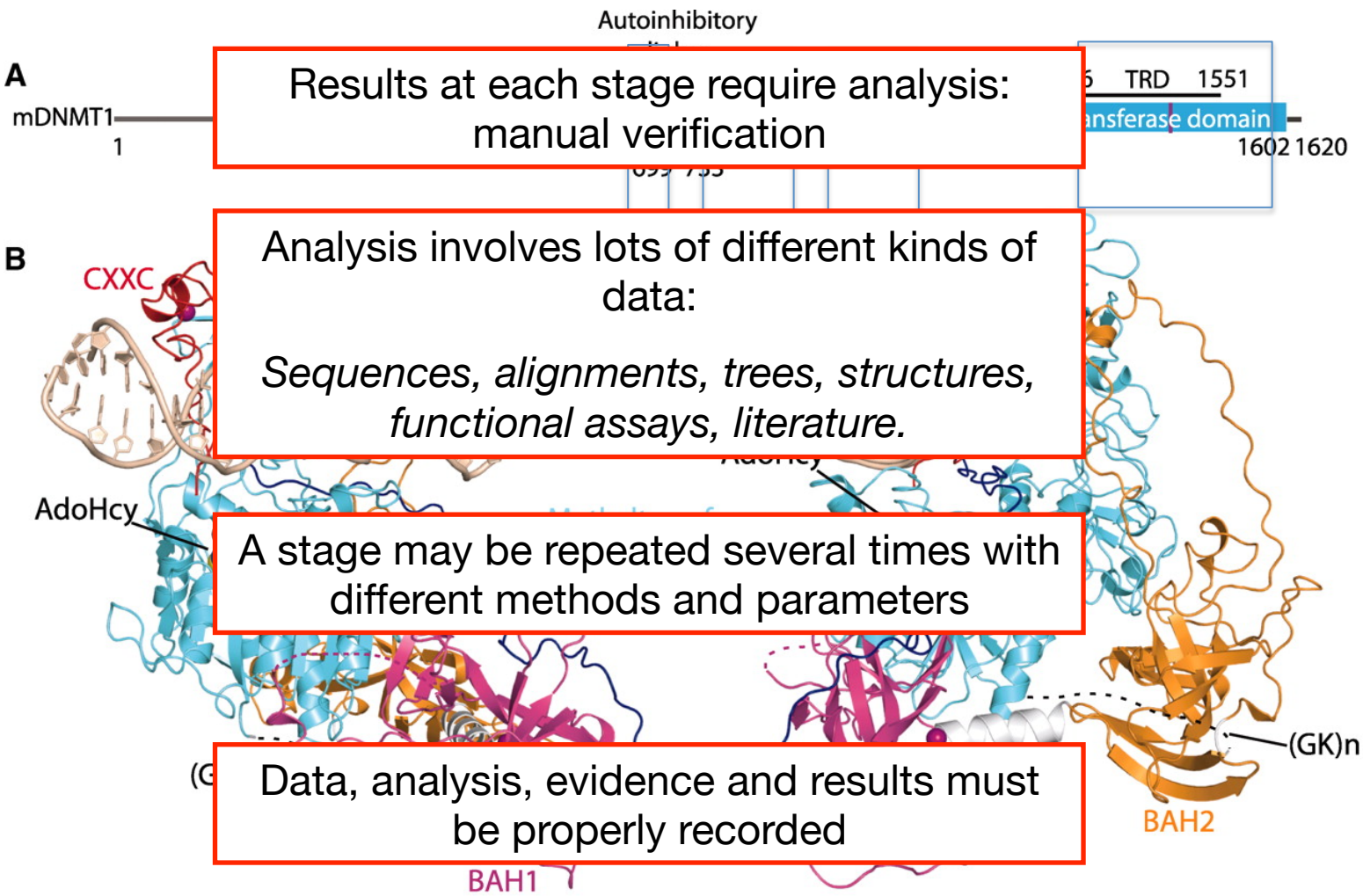
Molecular analysis
Structure-function relationships, Active sites, Binding motifs

Expert Knowledge
Experimental characterization, mutation studies.

Published Literature
Figures, supplementary info

2D & 3D structure prediction
Homology modelling, Fold recognition, secondary structure & disorder


Consensus Conservation & Clustering
Shading
Trees/PCA
Analysis
Pairwise alignment
Alignments
Structures
Sequences
Features
Annotation
Trees
PCA
Interactive Editing
Visualization
Figure Generation
Clickable HTML
Images
Line Art



Results at each stage require analysis: manual verification

Analysis involves lots of different kinds of data:

Sequences, alignments, trees, structures, functional assays, literature.

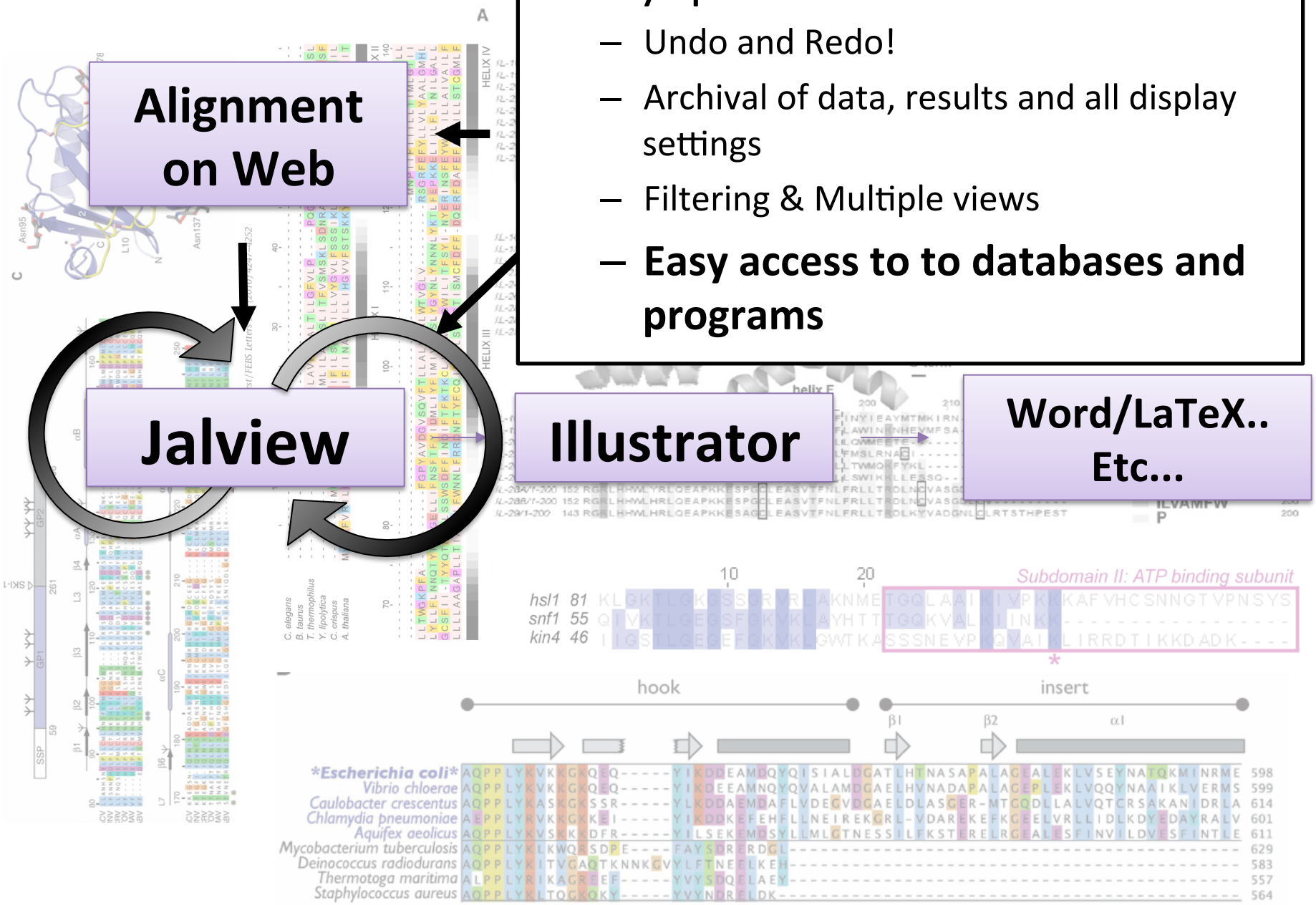
A stage may be repeated several times with different methods and parameters

Data, analysis, evidence and results must be properly recorded

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**



Jalview Flavours

Multi-windowed UI

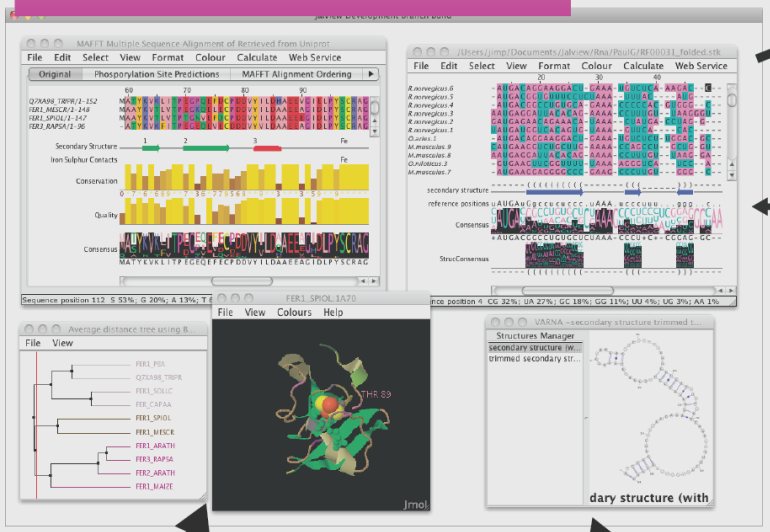
Visualization & Analysis

Common Data & Analysis
Editing, messaging and
File Input/Output

Lightweight UI

Integrate with web sites

The Jalview Desktop

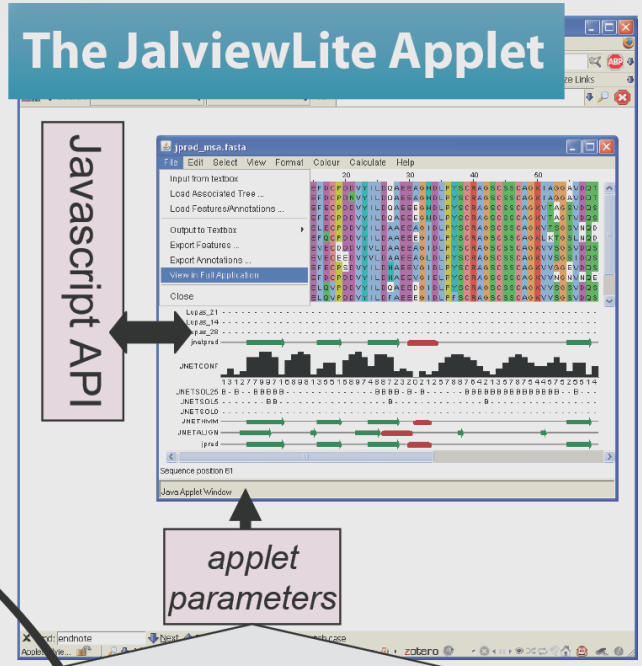


Figures
HTML
EPS
PNG

Local File system

import from web

The JalviewLite Applet



Javascript API

applet parameters

Alignment And Annotation

Bioinformatics Web Server

Analysis Services

Protein 2ndary structure

Jpred 3 Incorporating Jnet

JABAWS 2.0

Alignment Analysis

Protein Disorder

Multiple alignment

AACon

RONN

DisEMBL™

GLOBPLOT 2

ClustalW

PROBCONS

Mafft

TCoffee

MUSCLE

Biological Data Services

Pfam

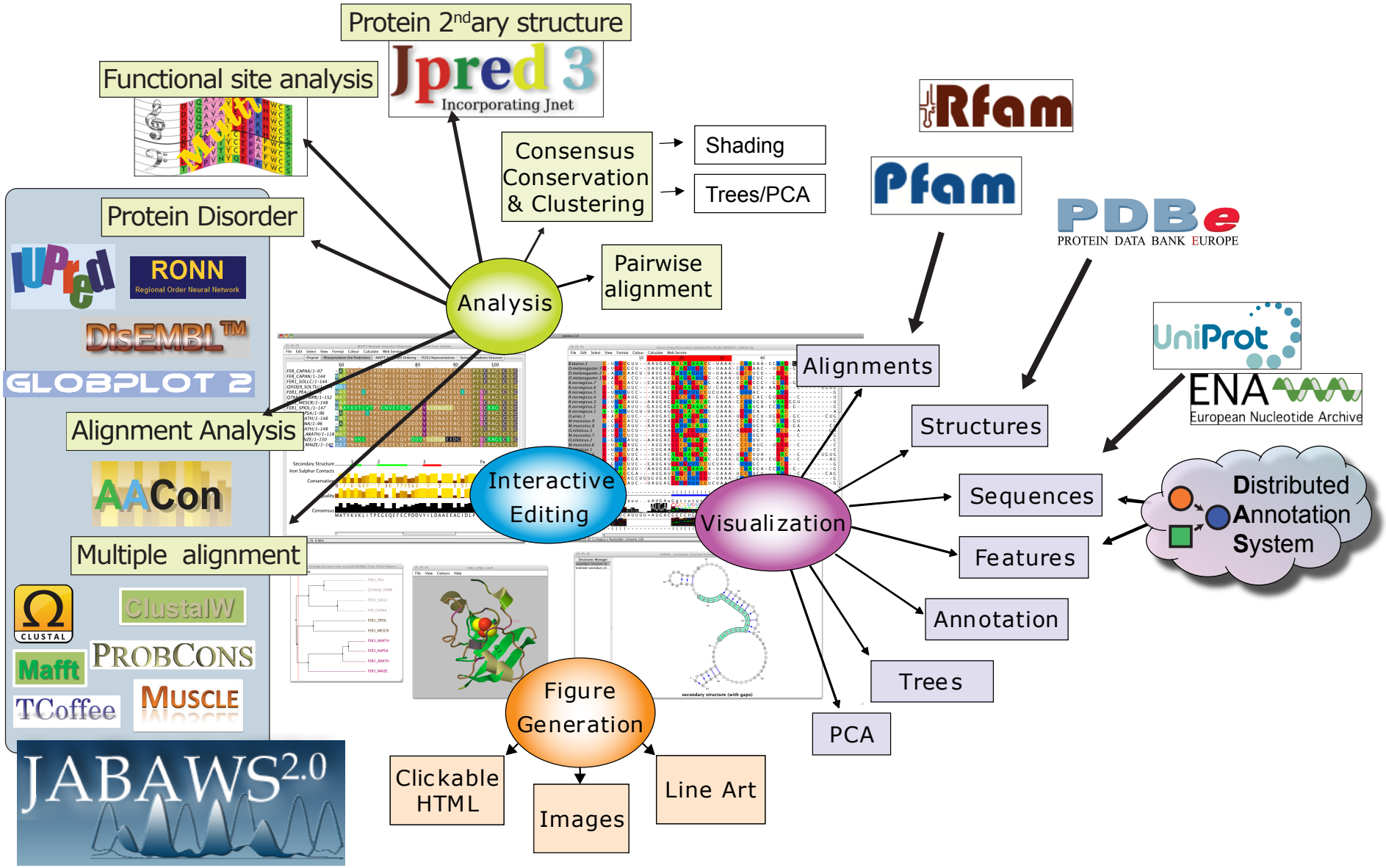
Rfam

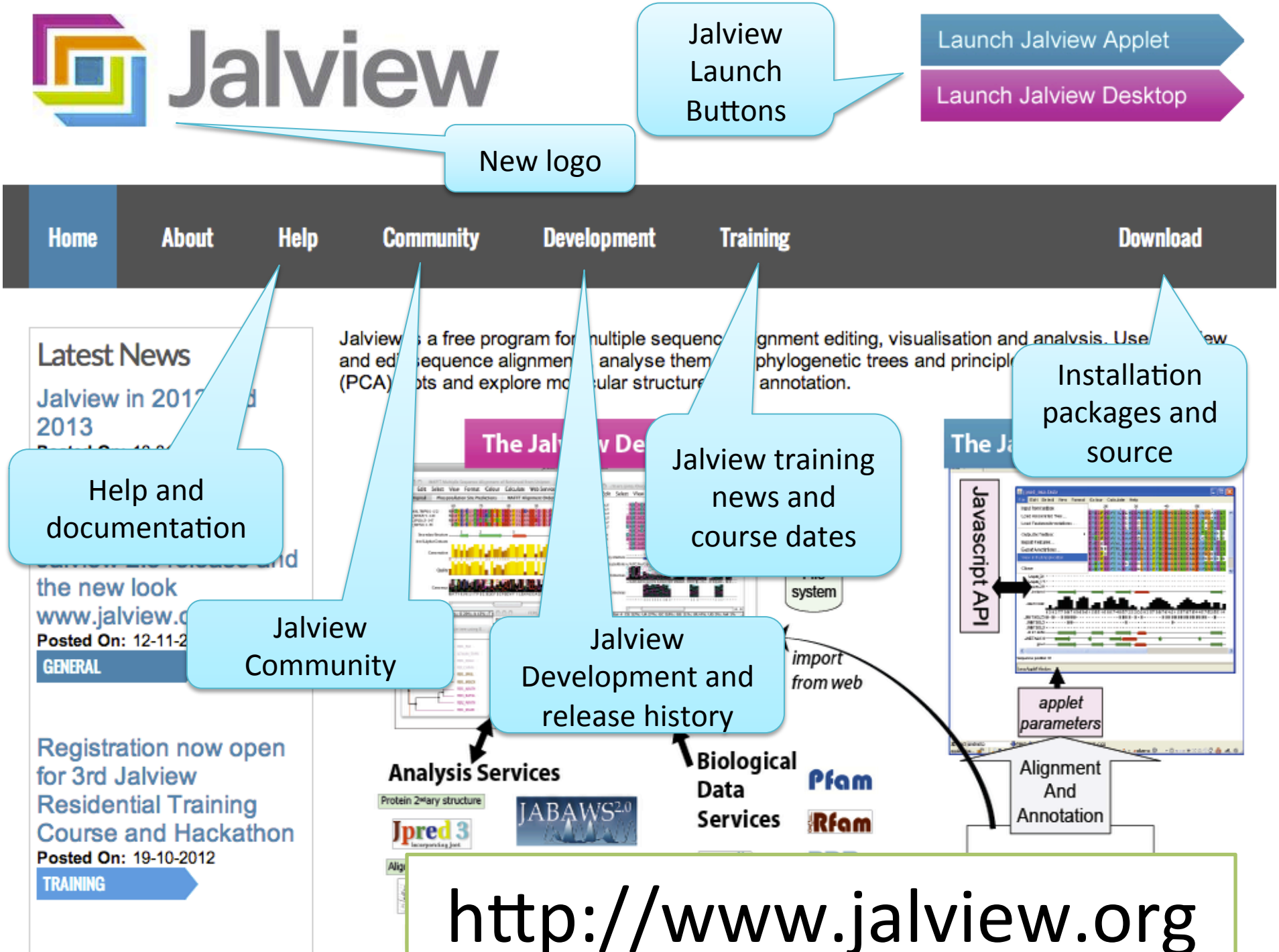
UniProt

PDBe PROTEIN DATA BANK EUROPE

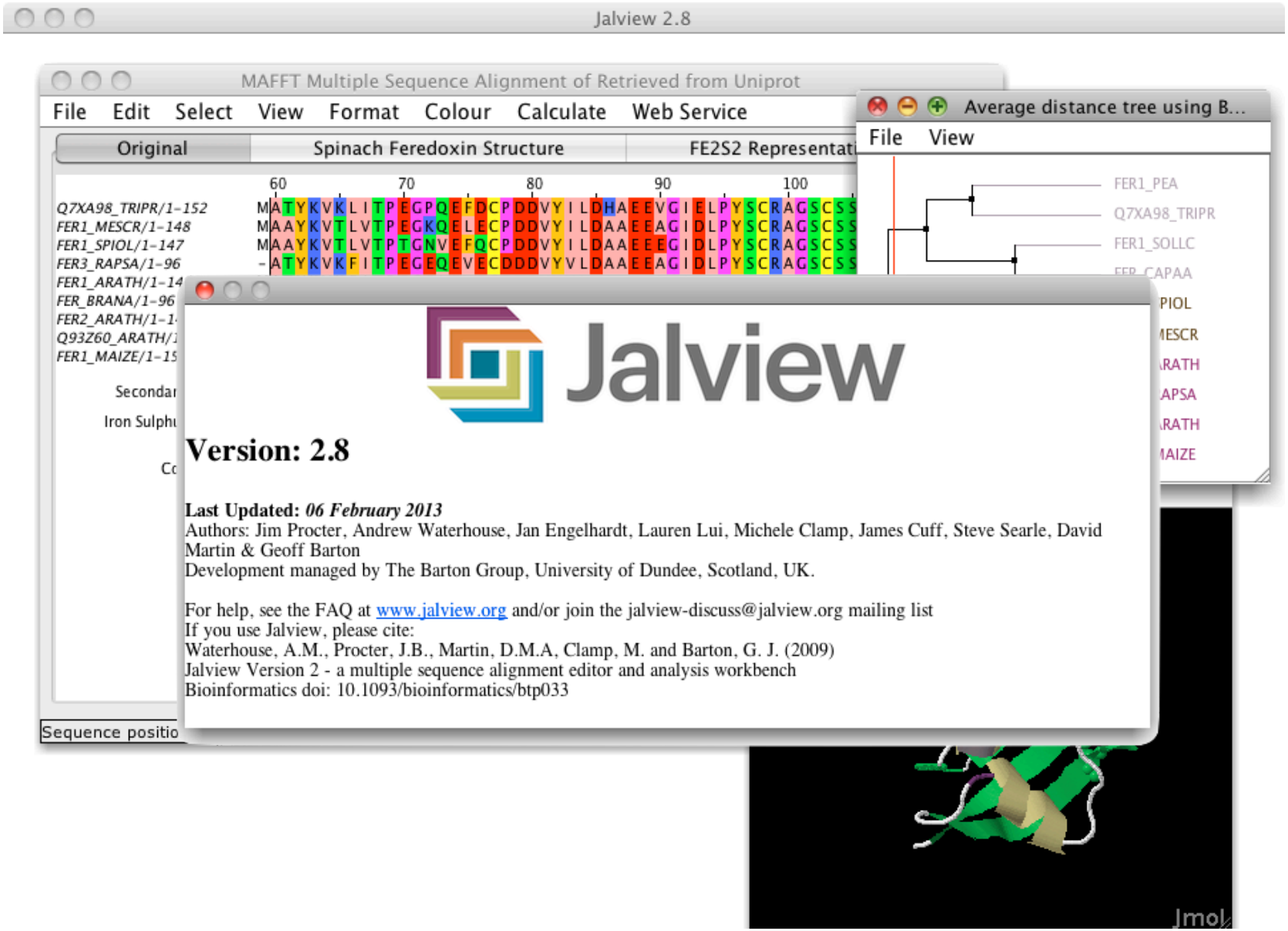
ENA European Nucleotide Archive

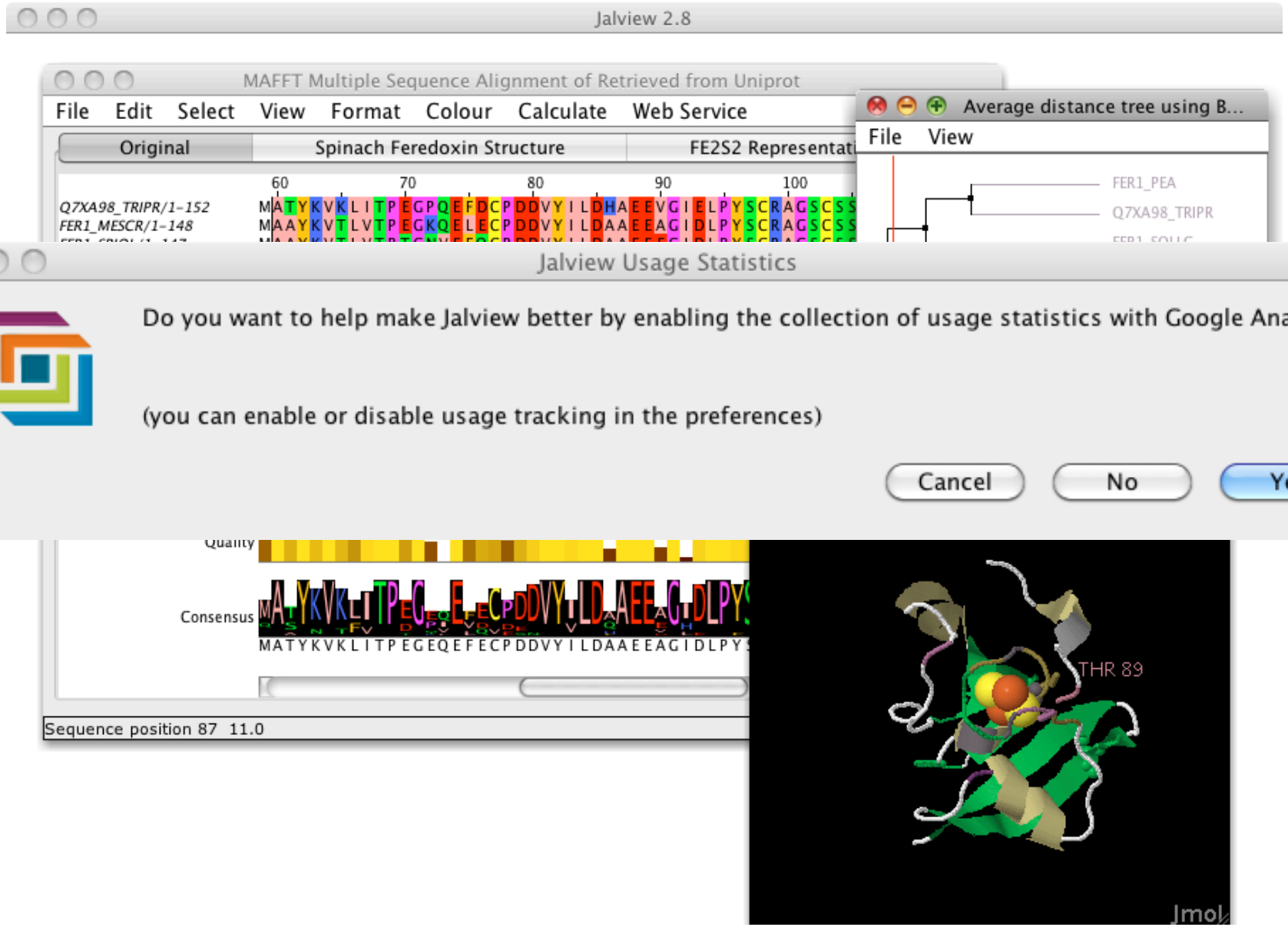
Distributed Annotation System


Protein 2ndary structure
Jpred 3
Incorporating Jnet
Functional site analysis
Protein Disorder
IUPred
RONN
Regional Order Neural Network
DisEMBL™
GLOBPLOT 2
Alignment Analysis
AACon
Multiple alignment
ClustalW
CLUSTAL
Mafft
PROBCONS
TCoffee
MUSCLE
JABAWS2.0
Consensus Conservation & Clustering
Shading
Trees/PCA
Pairwise alignment
Analysis
Interactive Editing
Visualization
Figure Generation
Clickable HTML
Images
Line Art
Alignments
Structures
Sequences
Features
Annotation
Trees
PCA
Rfam
Pfam
PDBe
PROTEIN DATA BANK EUROPE
UniProt
ENA
European Nucleotide Archive
Distributed Annotation System

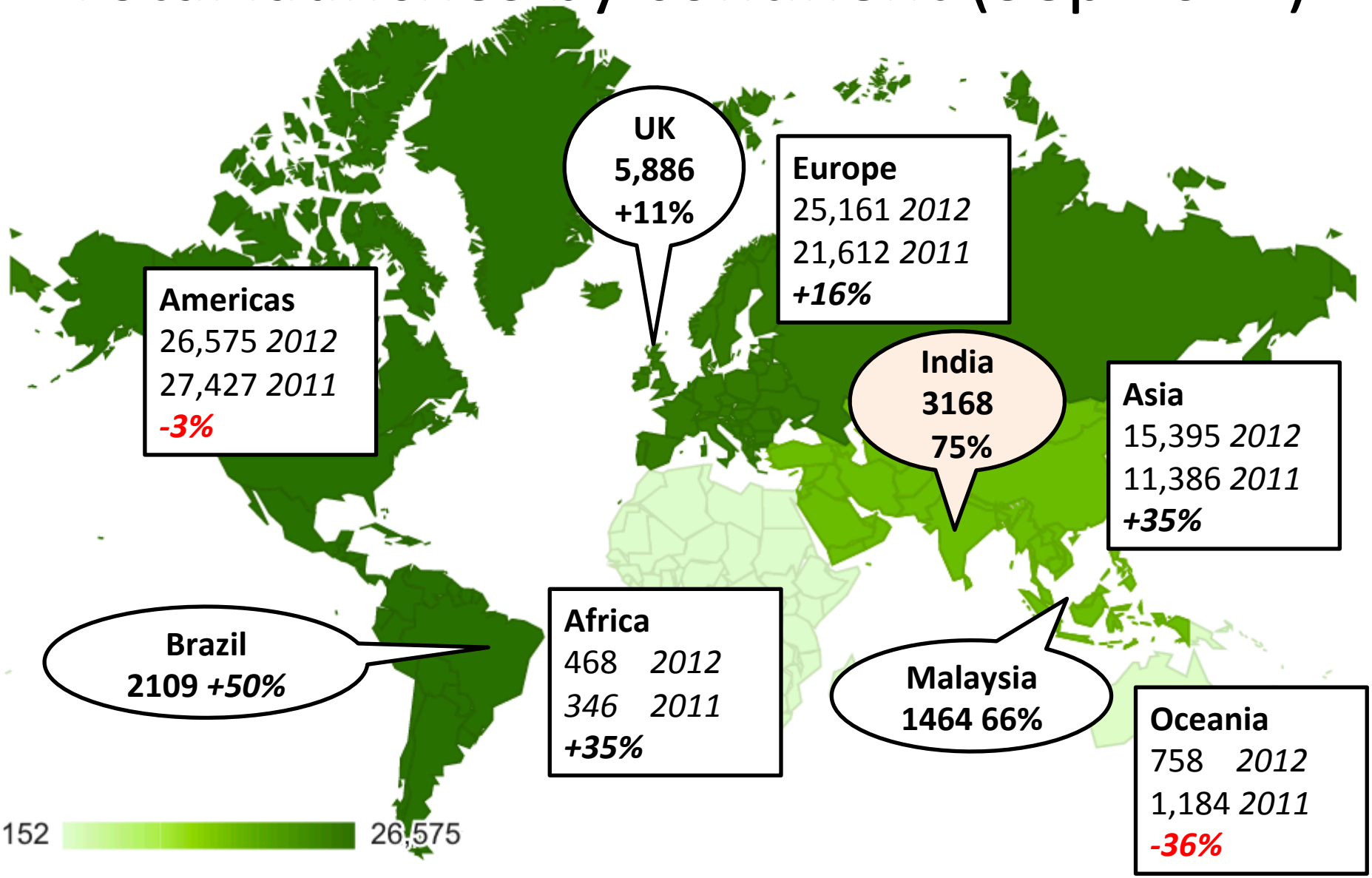

Jalview
Jalview Launch Buttons
Launch Jalview Applet
Launch Jalview Desktop
New logo
Home
About
Help
Community
Development
Training
Download
Latest News
Jalview in 2013
the new look
Posted On: 12-11-2
GENERAL
Registration now open for 3rd Jalview Residential Training Course and Hackathon
Posted On: 19-10-2012
TRAINING
Help and documentation
Jalview Community
Jalview Development and release history
Jalview training news and course dates
Installation packages and source
Javascript API
applet parameters
Alignment And Annotation
import from web
system
Analysis Services
Protein 2ary structure
Jpred 3
JABAWS2.0
Biological Data Services
Pfam
Rfam
http://www.jalview.org

Launching the jalview desktop



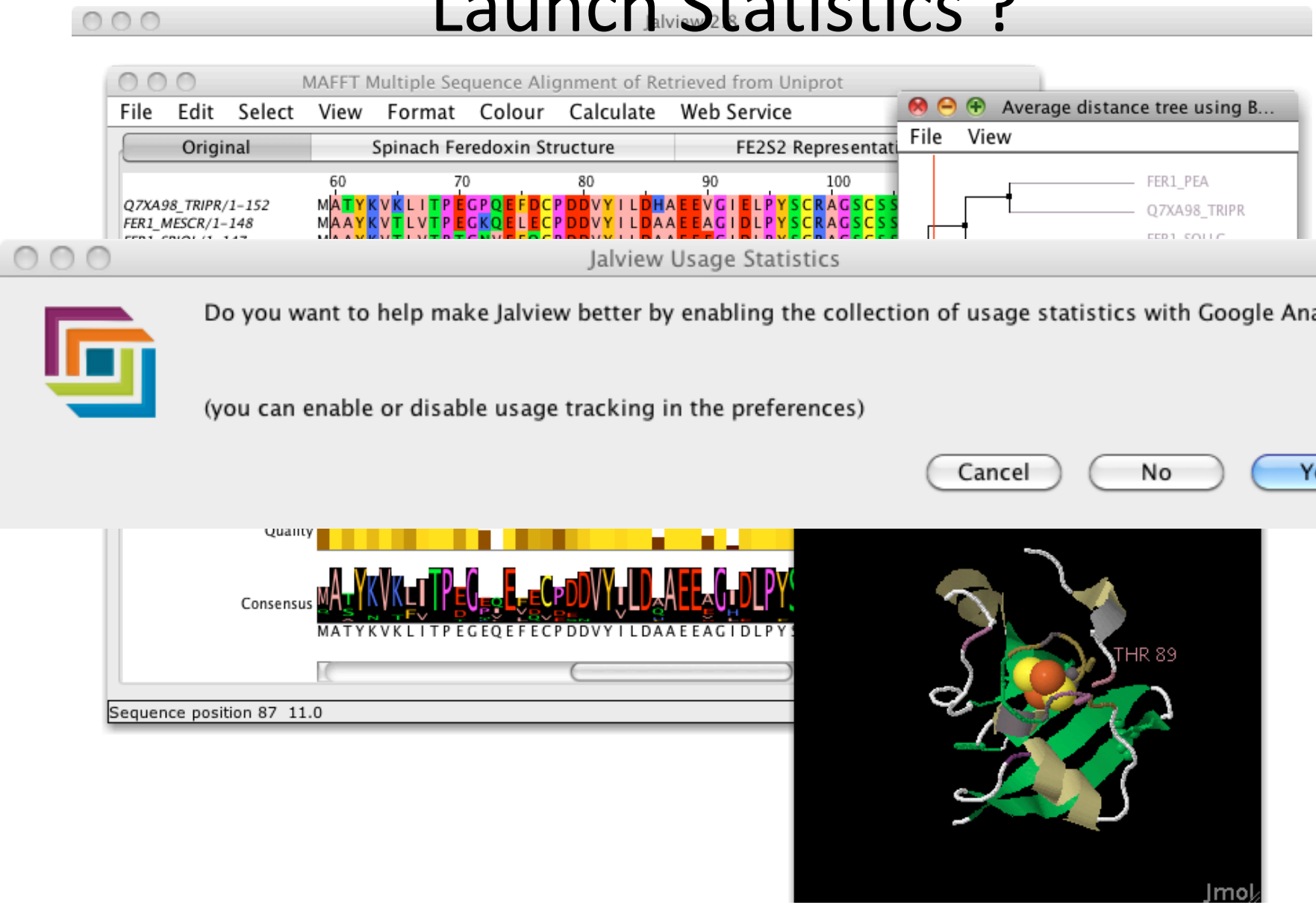

Jalview 2.8
MAFFT Multiple Sequence Alignment of Retrieved from Uniprot
File Edit Select View Format Colour Calculate Web Service
Original
Spinach Feredoxin Structure
FE2S2 Representati
Q7XA98_TRIPR/1-152
FER1_MESCR/1-148
Average distance tree using B...
File View
FER1_PEA
Q7XA98_TRIPR
Jalview Usage 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
Consensus
Sequence position 87 11.0
THR 89
Jmol

Total launches by continent (Sep 2012)



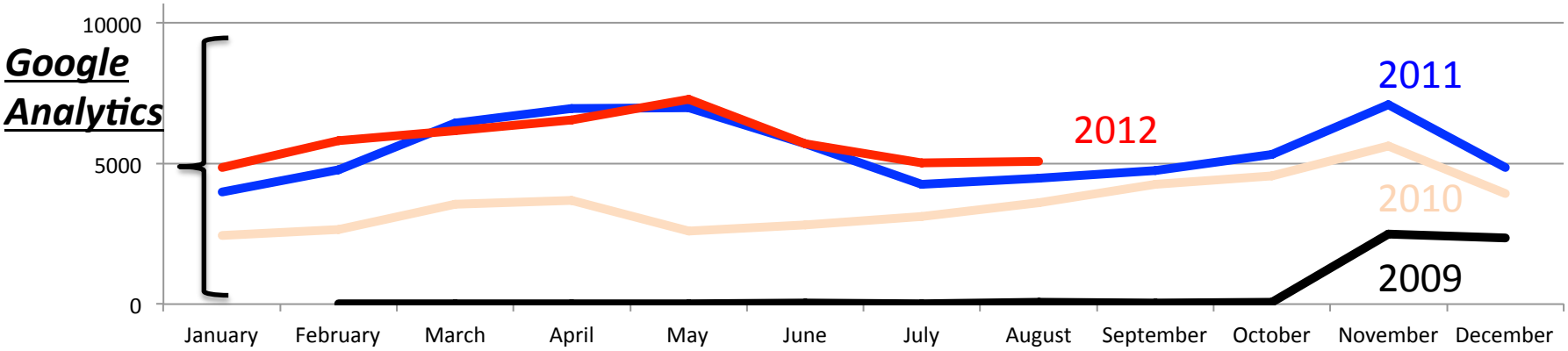
Jalview News

How accurate are the Google Analytics Launch Statistics ?

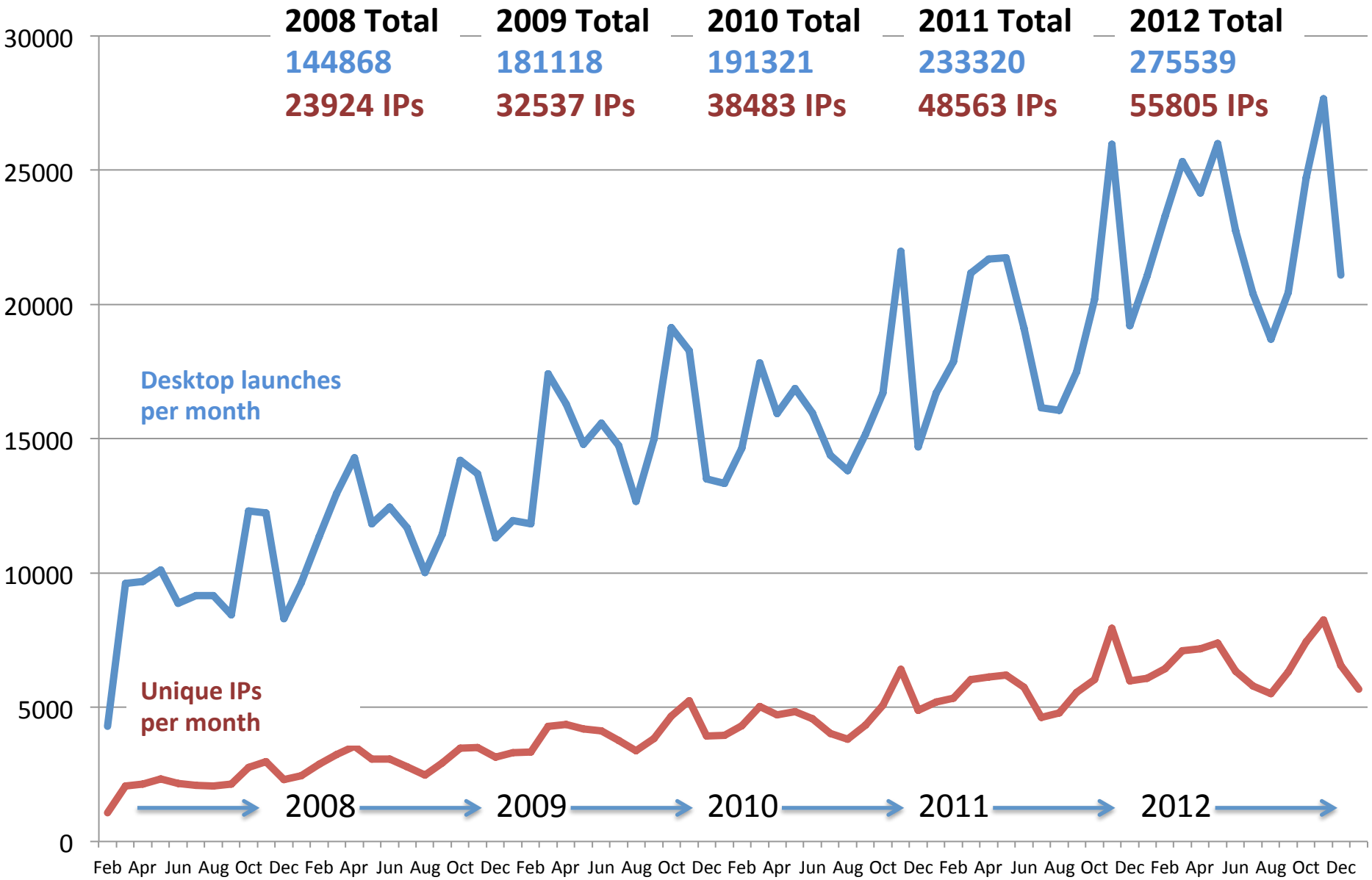


Year on Year increase in Desktop Launches as measured by Google Analytics vs our web server logs

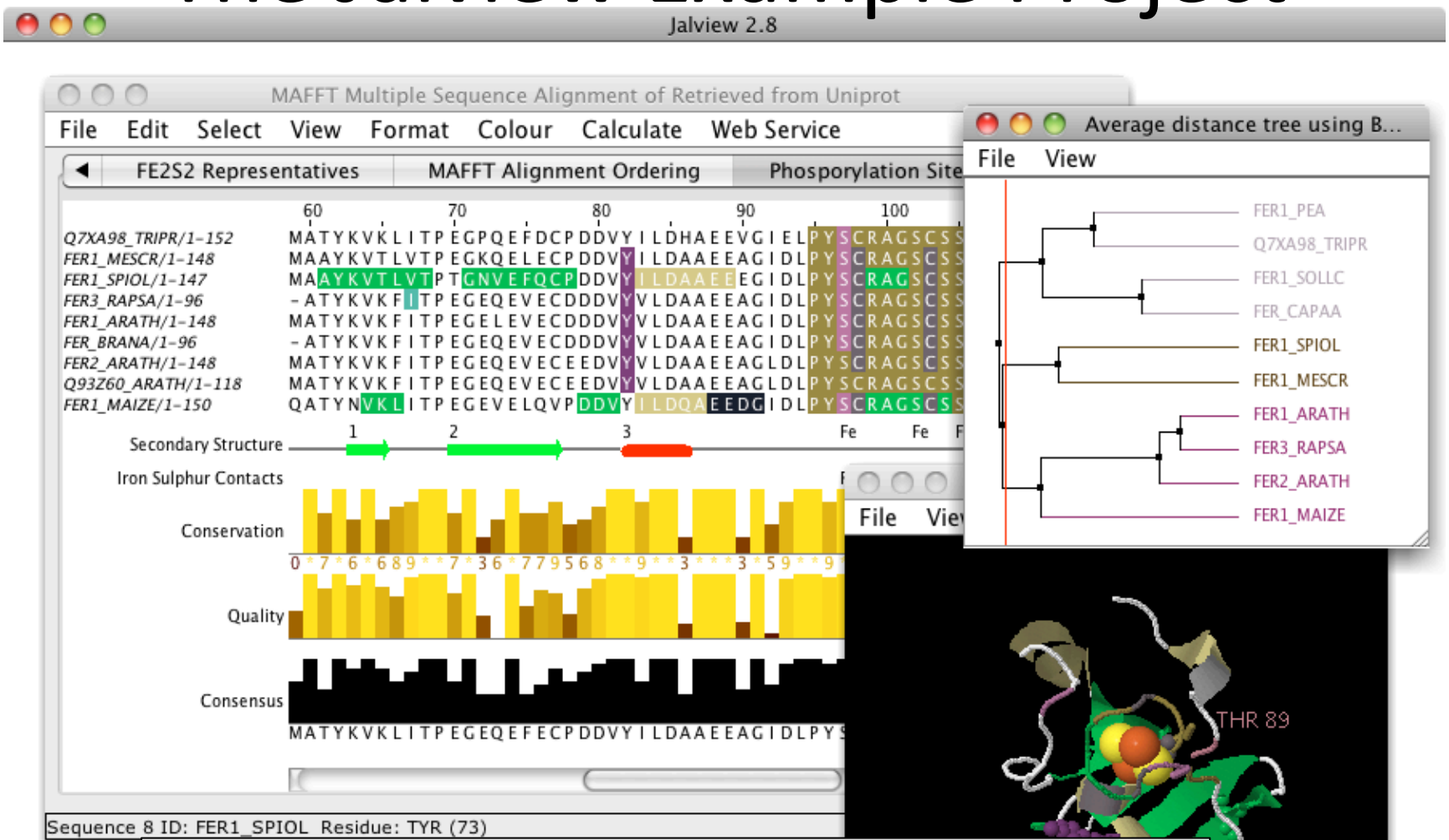
1 in 4 users say 'no' to Google Analytics



Jalview Desktop monthly usage from 2007-2012



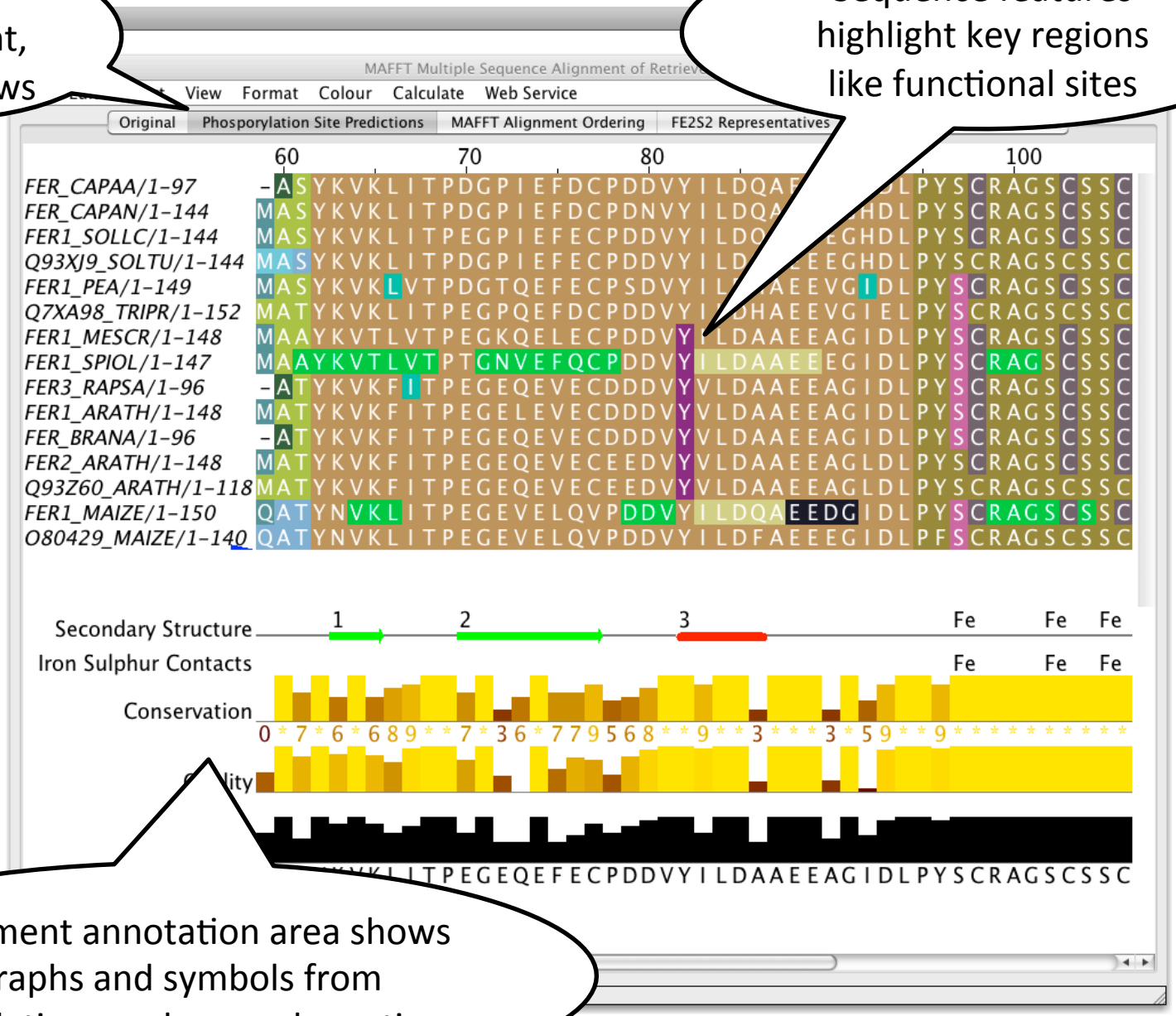
The Jalview Example Project



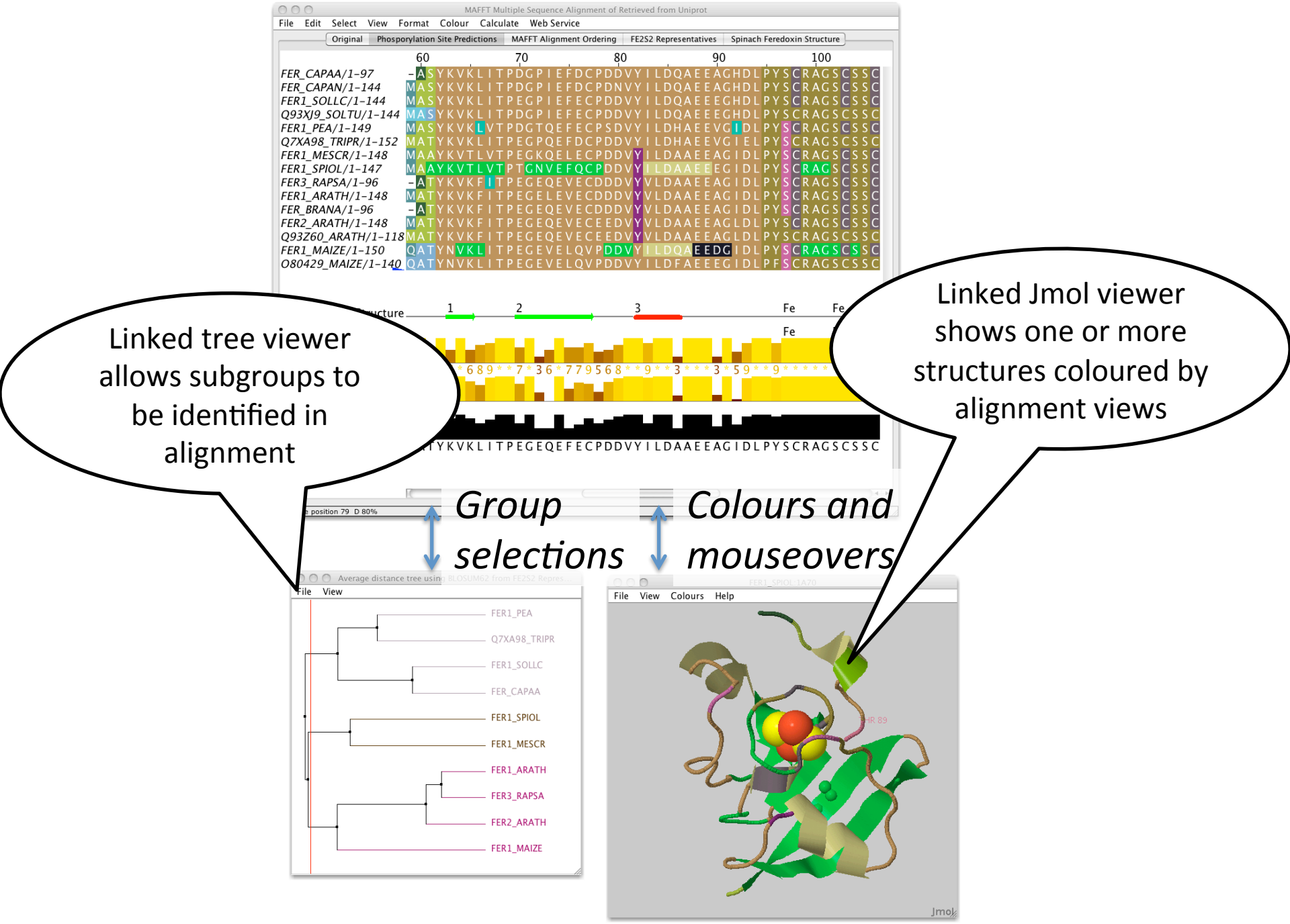
- Demonstrates most of Jalview's key features

One alignment, many views

Sequence features highlight key regions like functional sites



Alignment annotation area shows graphs and symbols from calculations and manual curation


MAFFT Multiple Sequence Alignment of Retrieved from Uniprot
File Edit Select View Format Colour Calculate Web Service
Original
Phosporylation Site Predictions
MAFFT Alignment Ordering
FE2S2 Representatives
Spinach Feredoxin Structure
FER_CAPAA/1-97
FER_CAPAN/1-144
FER1_SOLLC/1-144
Q93XJ9_SOLTU/1-144
FER1_PEA/1-149
Q7XA98_TRIPR/1-152
FER1_MESCR/1-148
FER1_SPIOL/1-147
FER3_RAPSA/1-96
FER1_ARATH/1-148
FER_BRANA/1-96
FER2_ARATH/1-148
Q93Z60_ARATH/1-118
FER1_MAIZE/1-150
O80429_MAIZE/1-140
Linked tree viewer allows subgroups to be identified in alignment
Linked Jmol viewer shows one or more structures coloured by alignment views
Group selections
Colours and mouseovers
File View
FER1_PEA
Q7XA98_TRIPR
FER1_SOLLC
FER_CAPAA
FER1_SPIOL
FER1_MESCR
FER1_ARATH
FER3_RAPSA
FER2_ARATH
FER1_MAIZE
File View Colours Help
Jmol

Typical jalview workflow

1. Import sequence or alignment
   - Drag/drop, Paste, or URL
2. Decorate sequences with references & annotation from external databases
3. Create alignment
4. Use built in shading, conservation analysis, tree and PCA capabilities to explore
   - Also use annotation and structure data if available
5. Select regions for refinement or further analysis
6. Import trees, annotation, etc. created with other programs to explore further
7. Prepare annotated views for publication

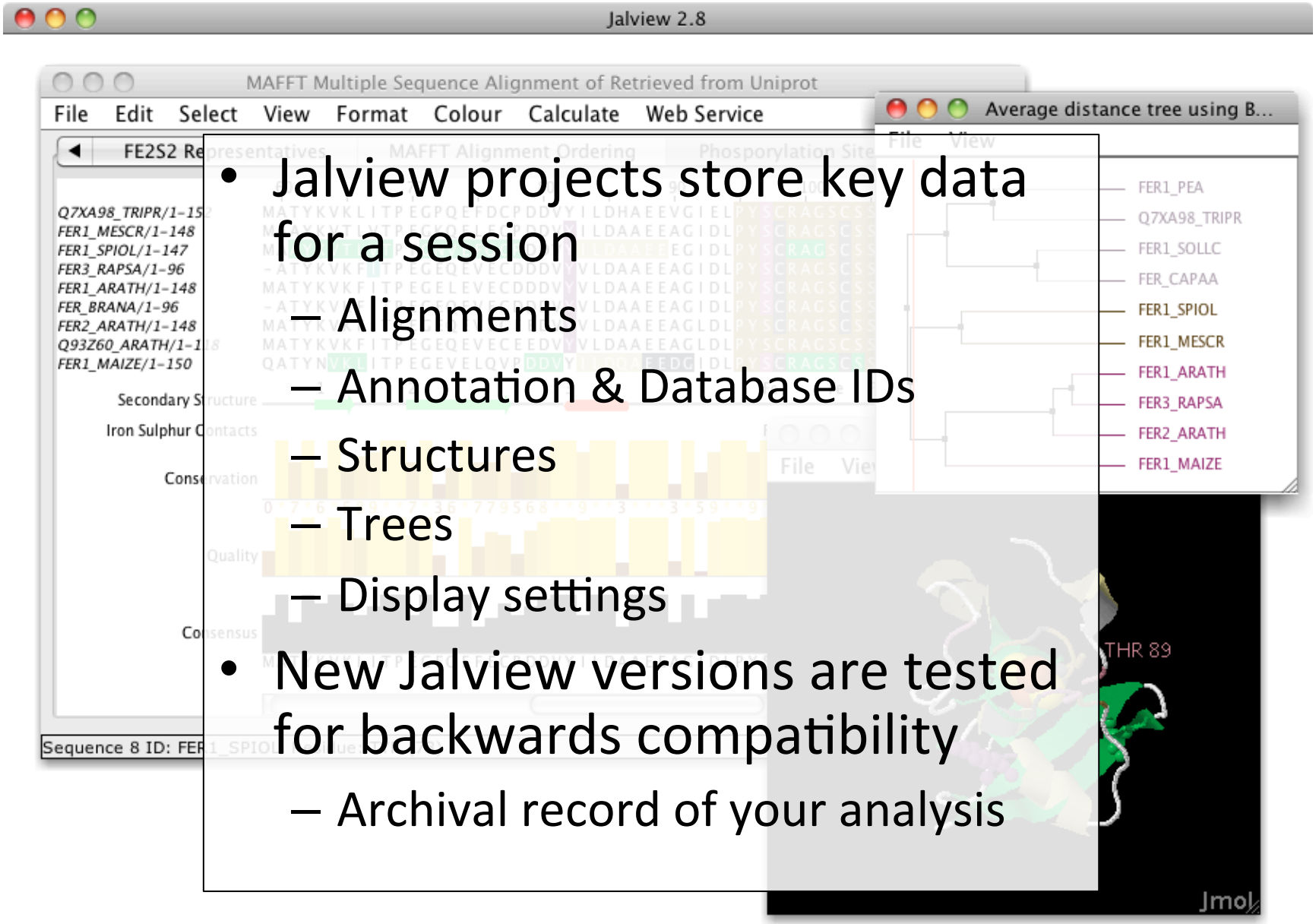
Phylogenetic analysis and Jalview

- **Built in tree methods**
  - UPGMA
    - Fast, simple, but not reliable for phylogenetic inferrence
  - Neighbour joining
    - Slower than UPGMA
    - Perhaps useful for a first approximation
      - Jalview's implementation is not the most efficient
- **Import trees for subgroup analysis**
  - Load any number of Newick/NH Extended files onto an alignment from another program
    - Bootstraps are displayed

Typical jalview workflow

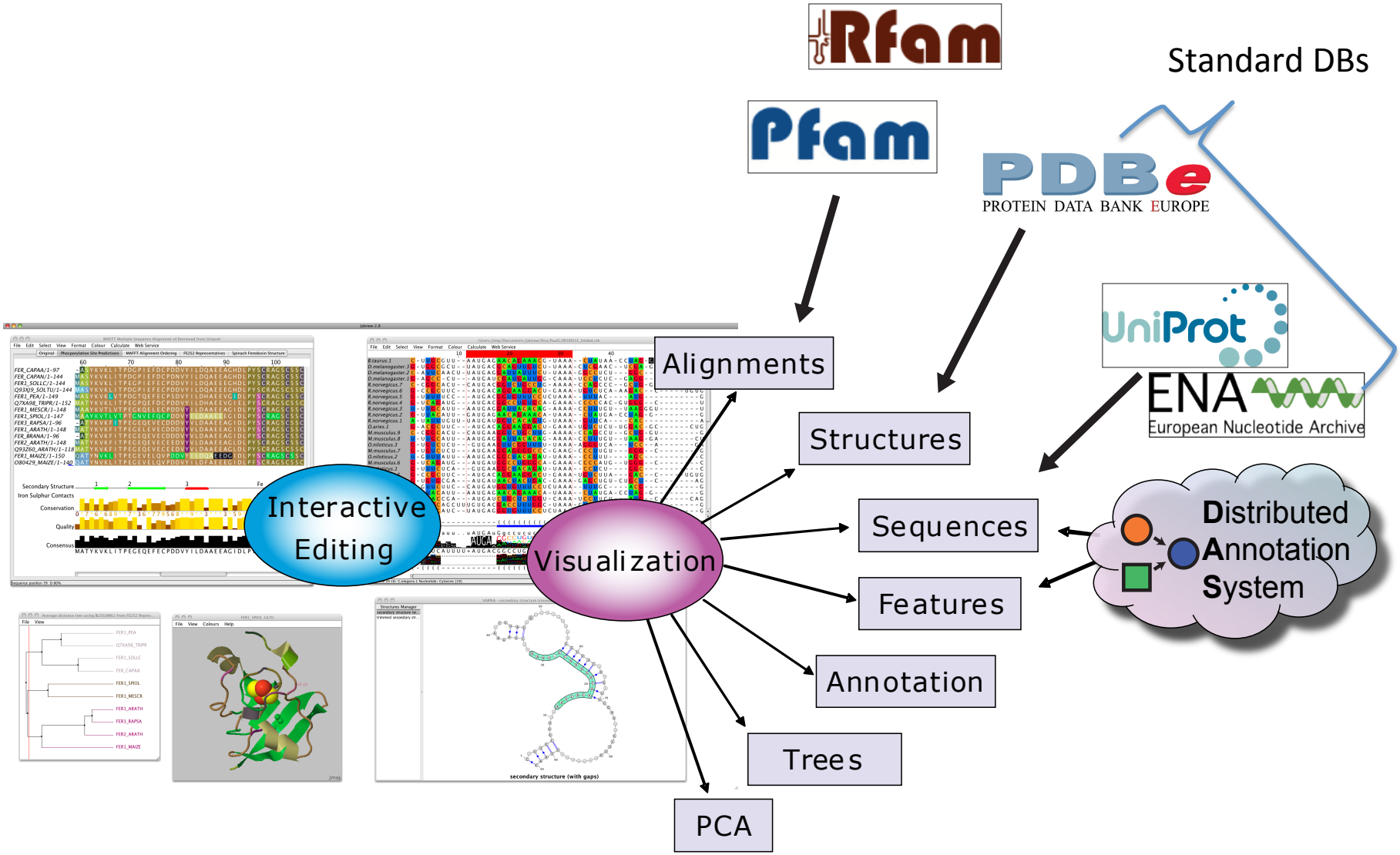
Save as
Jalview project
at strategic points

1. Import sequence or alignment
   - Drag/drop, Paste, or URL
2. Decorate sequences with references & annotation from external databases
3. Create alignment
4. Use built in shading, conservation analysis, tree and PCA capabilities to explore
   - Also use annotation and structure data if available
5. Select regions for refinement or further analysis
6. Import trees, annotation, etc. created with other programs to explore further
7. Prepare annotated views for publication

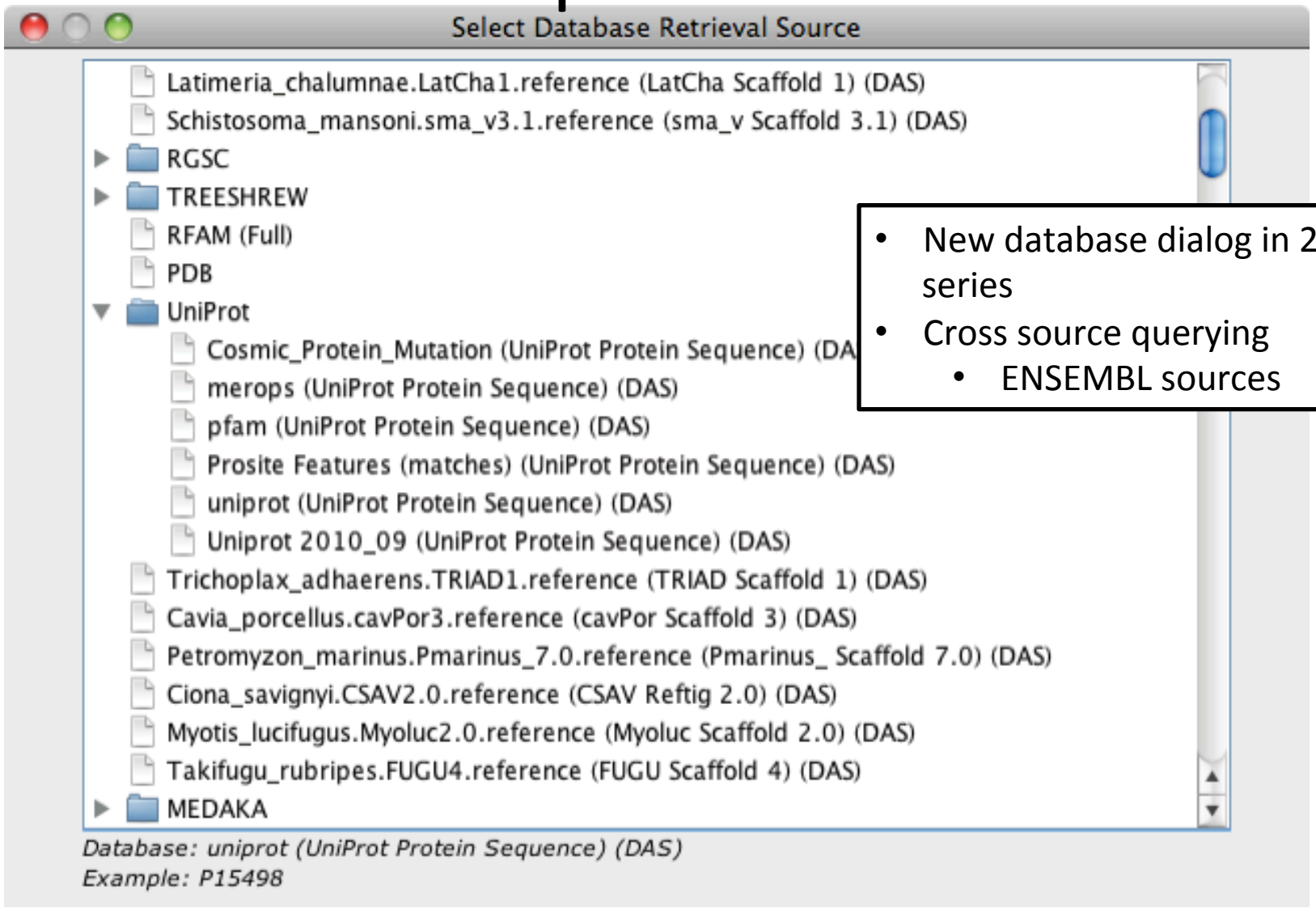


- Jalview projects store key data for a session
  - Alignments
  - Annotation & Database IDs
  - Structures
  - Trees
  - Display settings
- New Jalview versions are tested for backwards compatibility
  - Archival record of your analysis

Retrieval from External Databases



DAS allows Jalview access to Over 270 Sequence Databases…



- New database dialog in 2.8.x series
- Cross source querying
  - ENSEMBL sources

jdas

JDAS - JAXB DAS Client Library

Project Home | Downloads | Wiki | Issues | Source | Administer

Summary | People

Tip: Project owners, see our Getting Started guide for steps to configure your project.

Project Information

Recommend this on Google

Starred by 1 user
Project feeds

Code license
Apache License 2.0

Labels
DAS, JDAS

Members
rajido, gustavoadolfo.salazar, leylajael, warren.jonathan, andy.jenkinson, macias.jr, foreveremain
1 committer

Your role
Owner

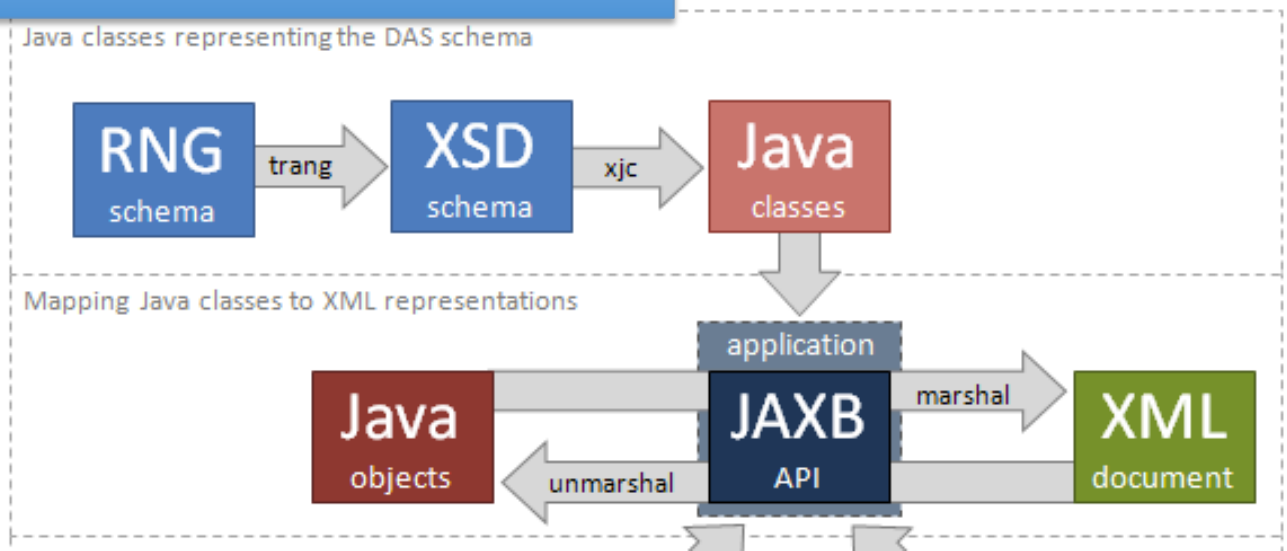
Featured

Downloads
jdas-1.0.1.jar
jdas-1.0.2.jar
Show all »

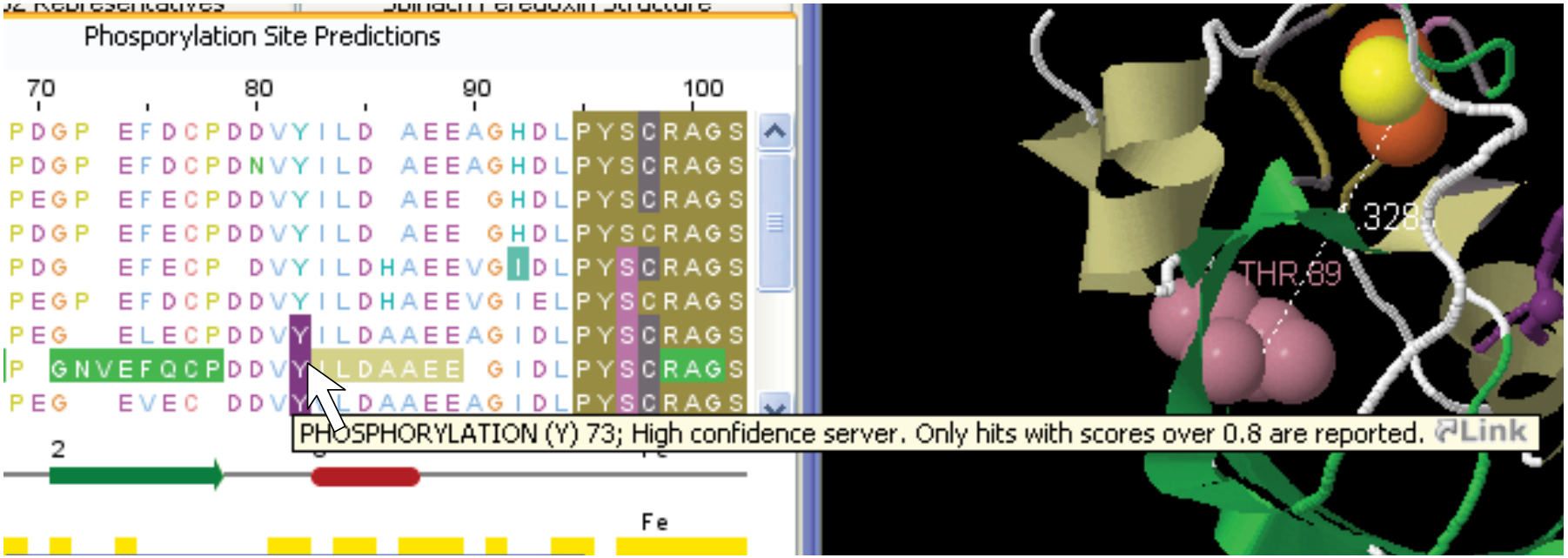
JDAS is a Java DAS client library which simplifies DAS queries and maps Java classes to DAS XML representations. JDAS includes JAVA classes generated from DAS RNG schemas. These Java classes are used through the JAXB API to store and retrieve data in memory without the need to implement a specific set of XML loading and saving routines. JAXB provides two main features: the ability to marshal Java objects into XML and the inverse, to unmarshal XML into Java objects. JDAS additionally implements adaptor classes which are used to refactor generated classes with a complex structure. Finally JDAS implements client classes to facilitate querying DAS information.

Developed by Rafael Jiminez, Jonathan Warren and other Java developers in DAS Communty

Table Of Contents



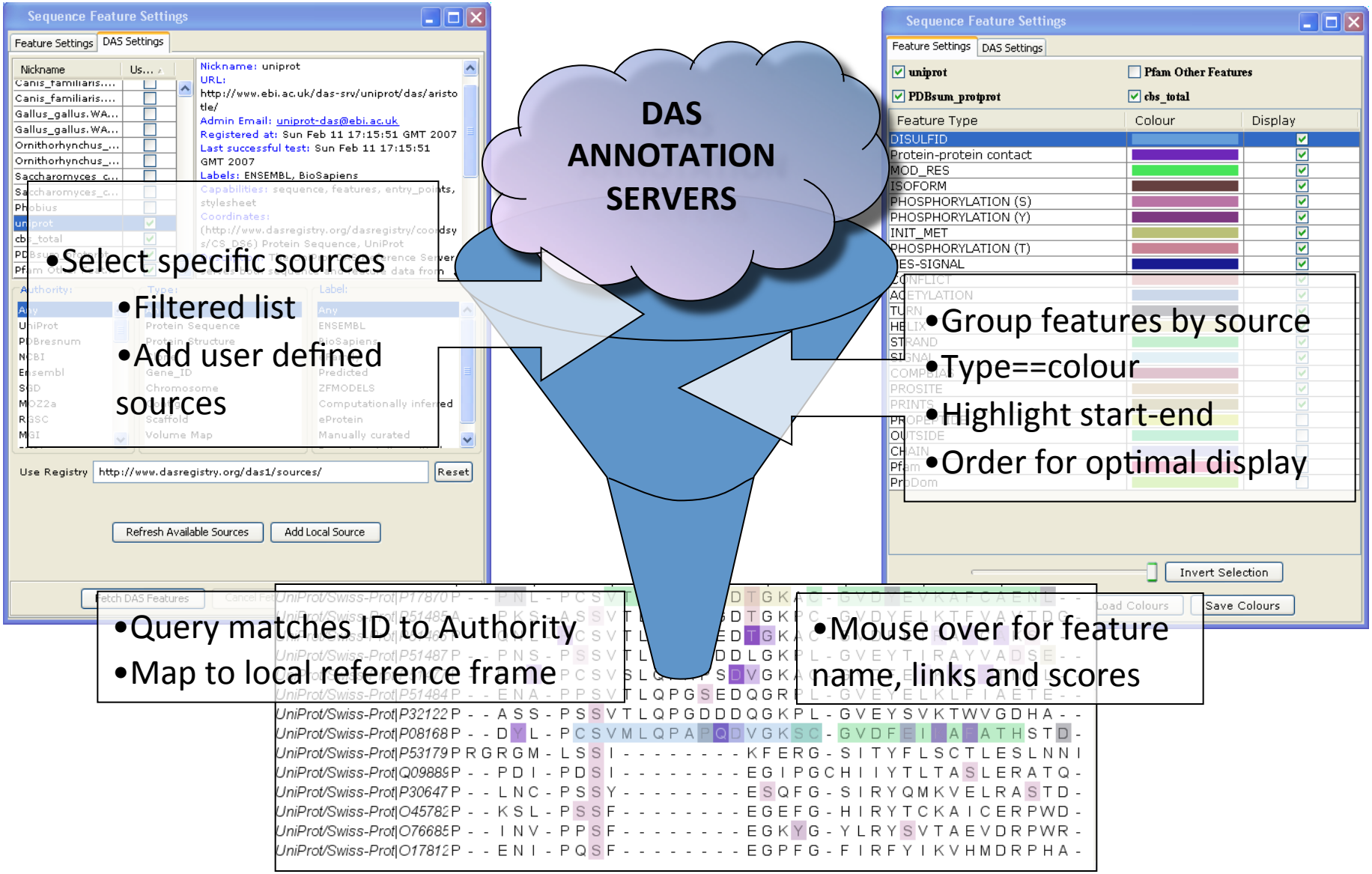
Sequence Features



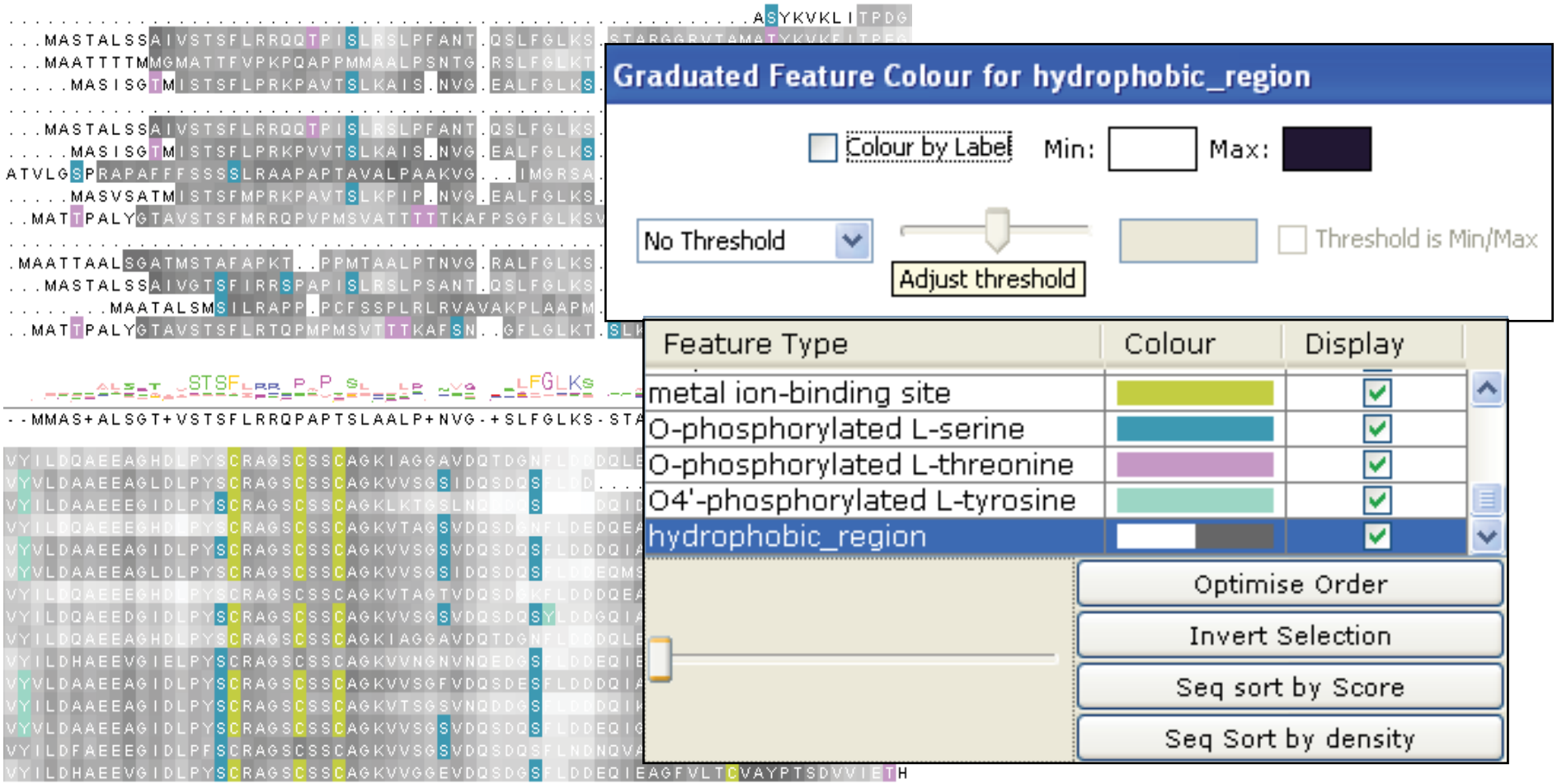
Sources of sequence feature data

- Jalview sequence annotation files
- DAS sources
- GFF files
- Certain ‘rich’ alignment formats
  - Stockholm
  - AMSA

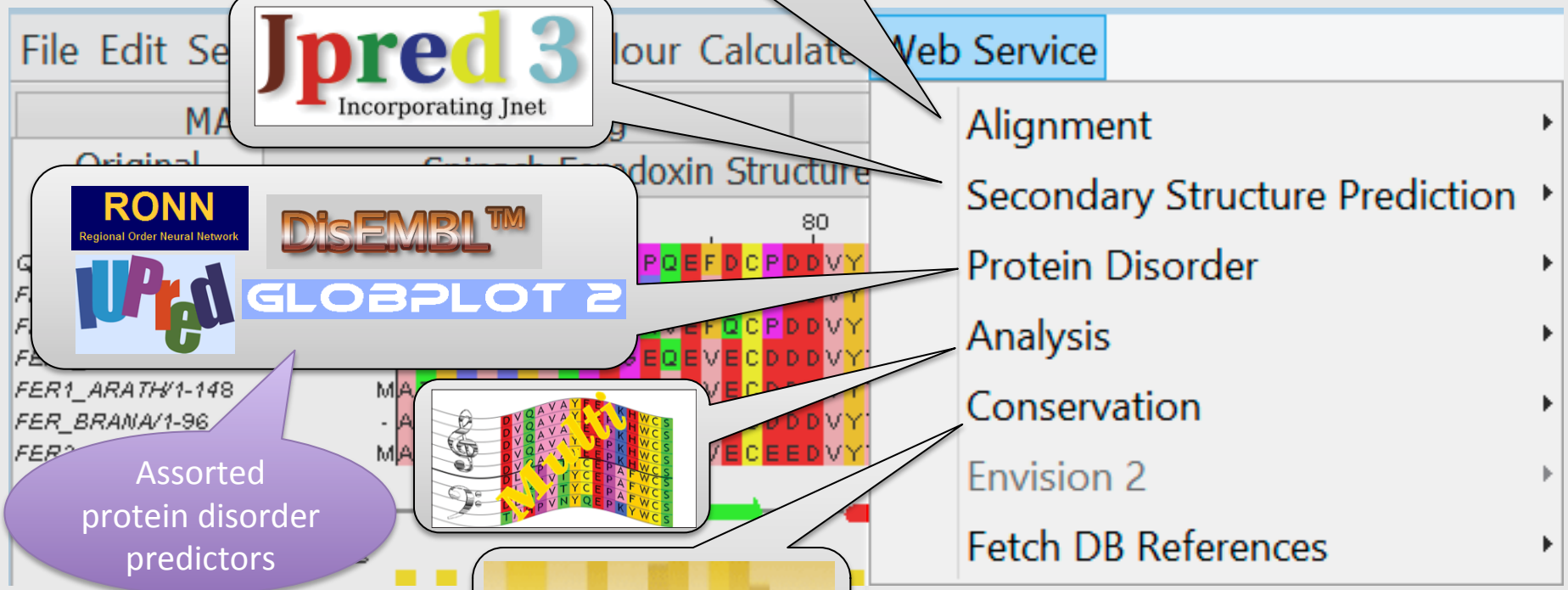
Sequence Features Dialog box



Shading, thresholding, colour by label.



Clustal Omega quick alignment of millions of sequences



Assorted protein disorder predictors

Protein conservation calculations

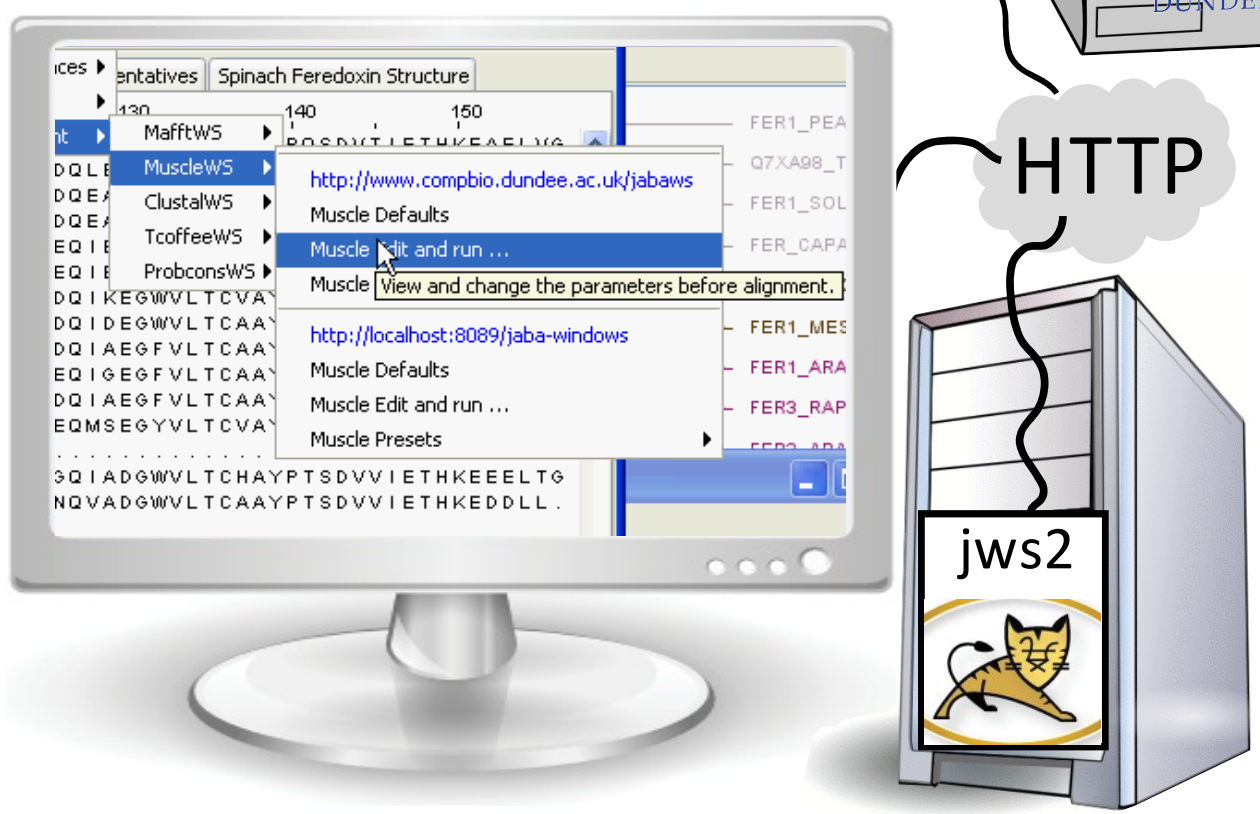
JABAWS 2 services in Jalview 2.8

Alignment Web Services: JABAWS:MSA

Replaces original Jalview 2 services:

- Extensible framework for wrapping command line programs
- Can be installed on user's own machine/cluster

See Troshin et al. application note in *Bioinformatics* for more details.



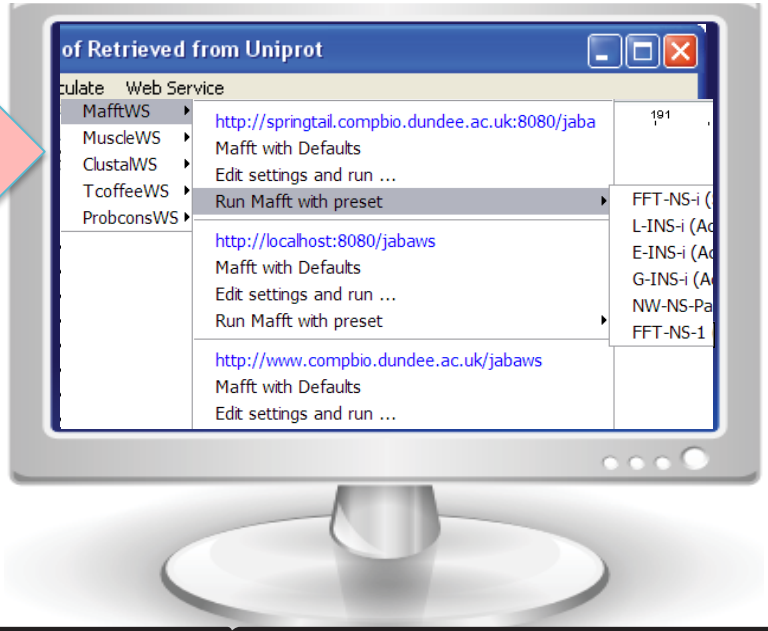
Peter Troshin

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.

vmware

Native JABAWS installs on a range of platforms

Apache Tomcat

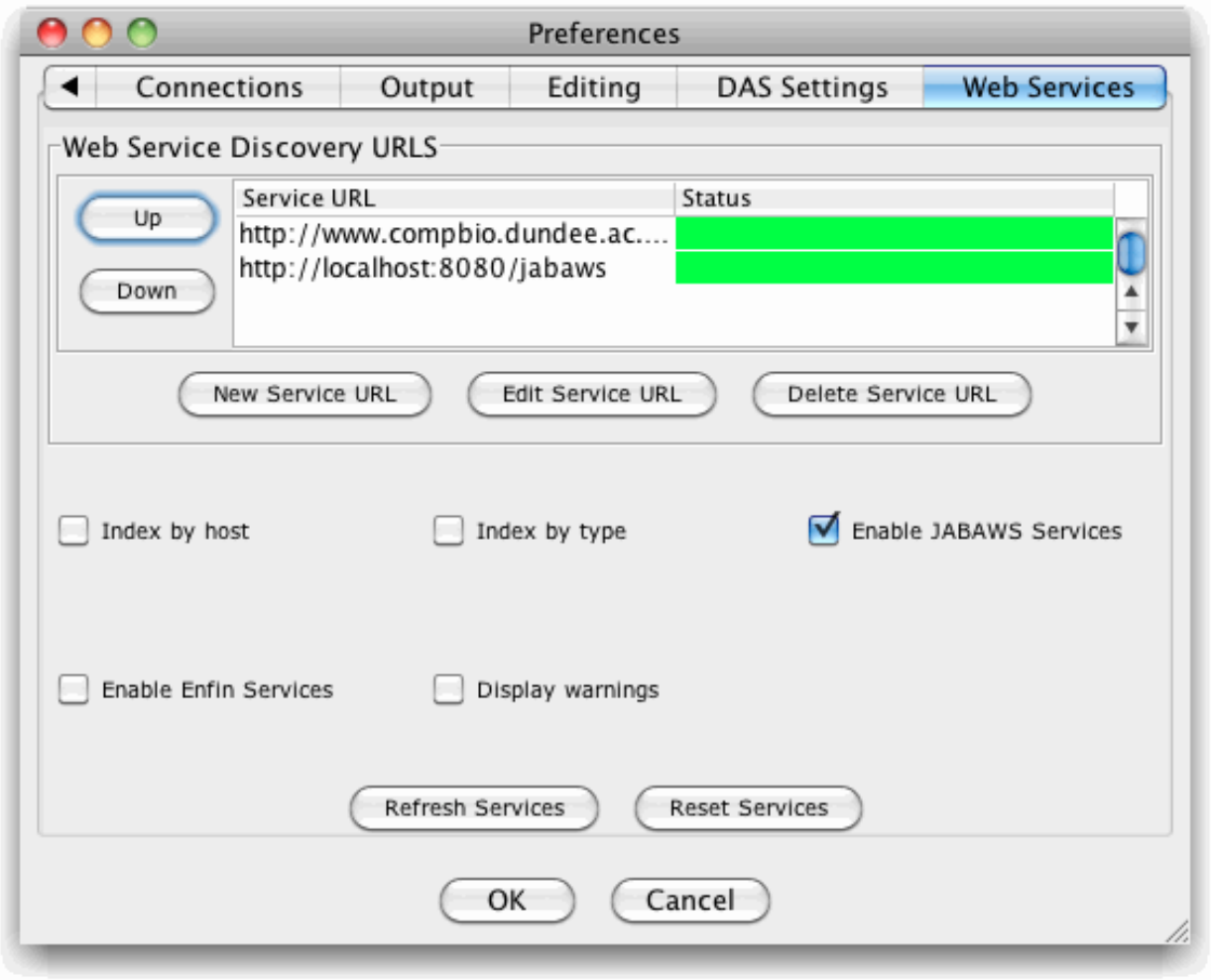
DRMAA
Distributed Resource Management Application API — www.drmaa.org

JABAWS Amazon Machine Image on EC2

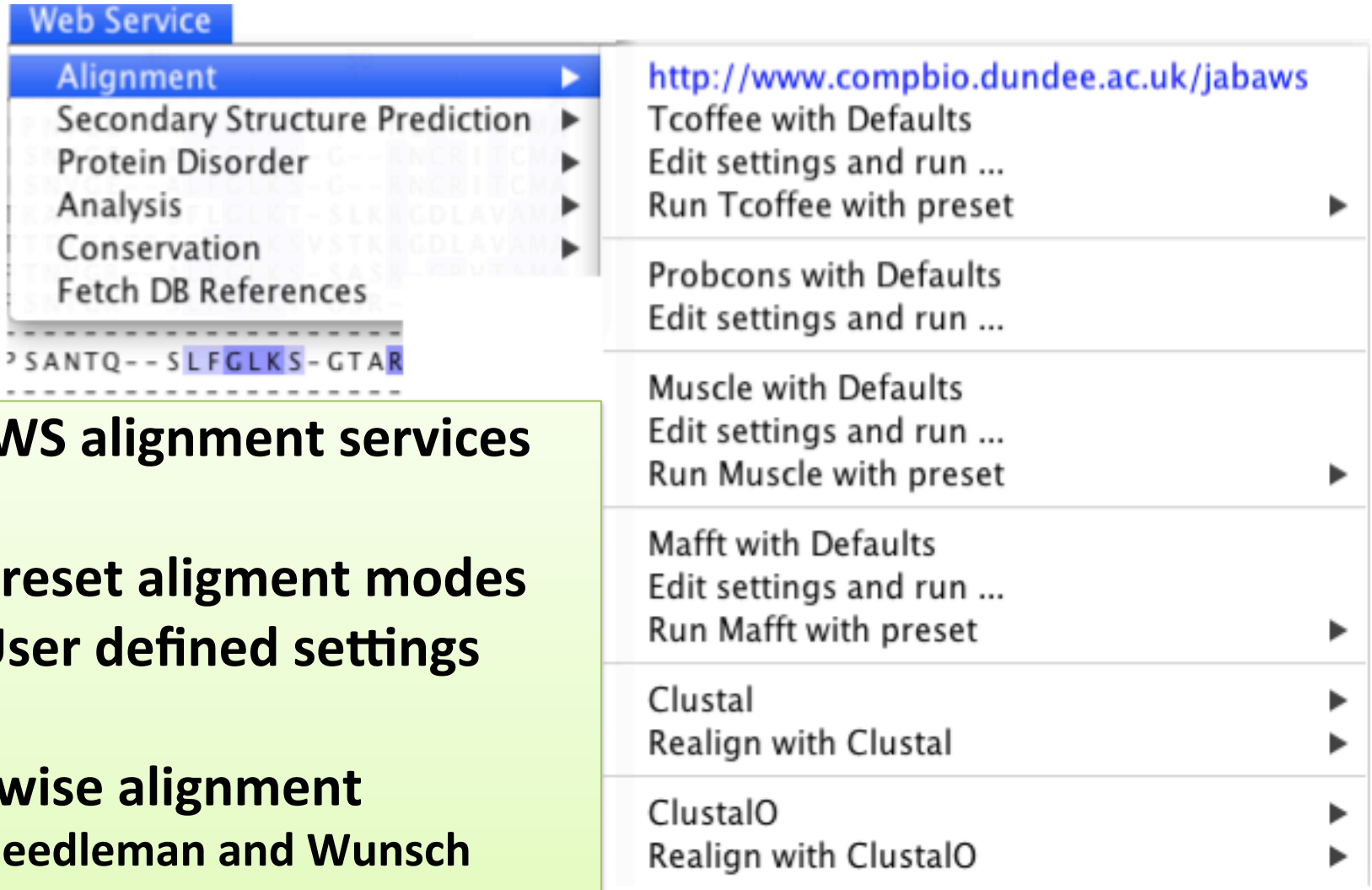
powered by amazon web services

the cloud market

Jalview's JABAWS Configuration Panel

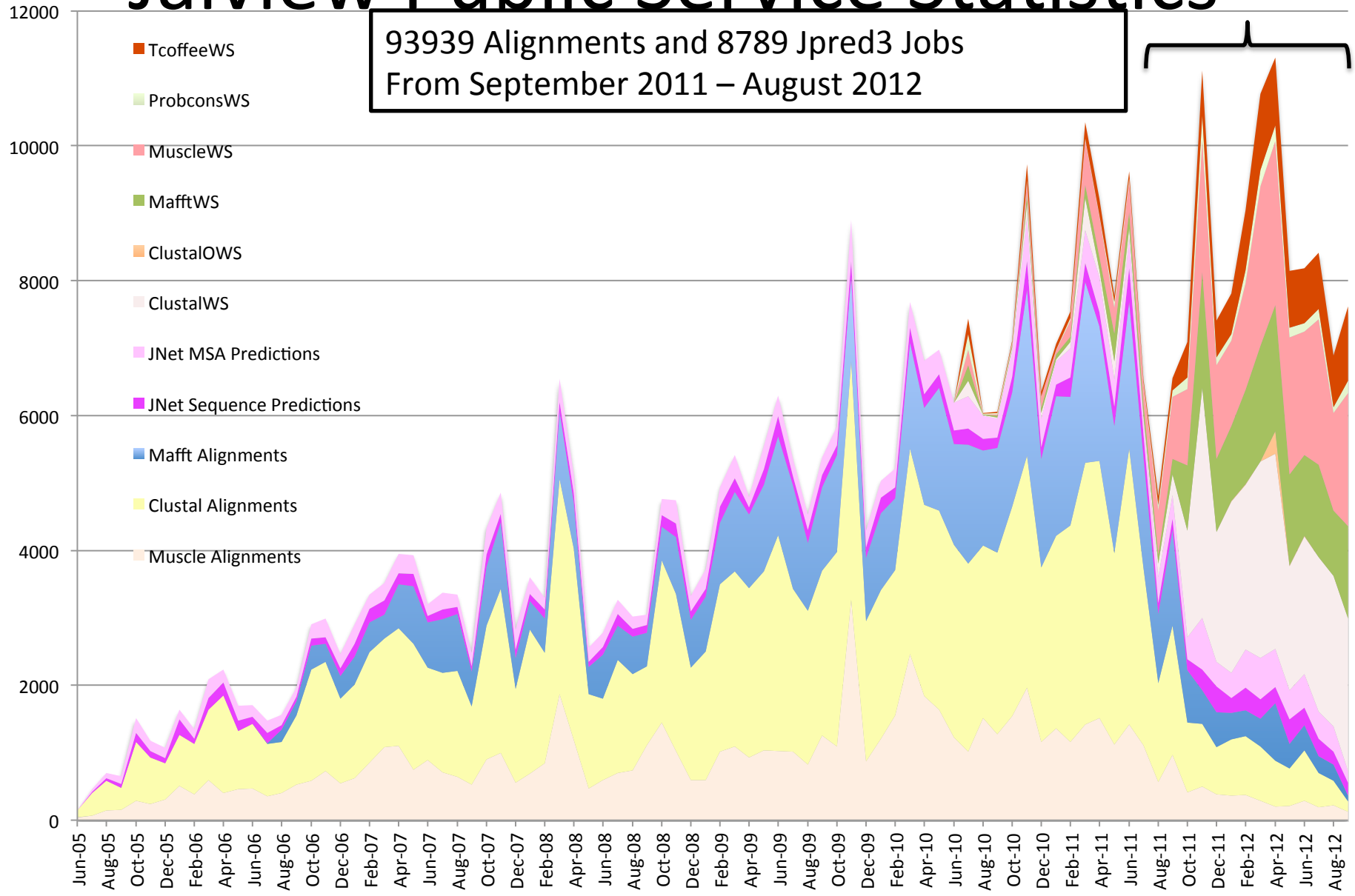


Jalview's Alignment Methods



- **JABWS alignment services**
  - **Preset aligment modes**
  - **User defined settings**
- **Pairwise alignment**
  - **Needleman and Wunsch**
    - Mostly used internally

Jalview Public Service Statistics



Common types of alignment algorthm

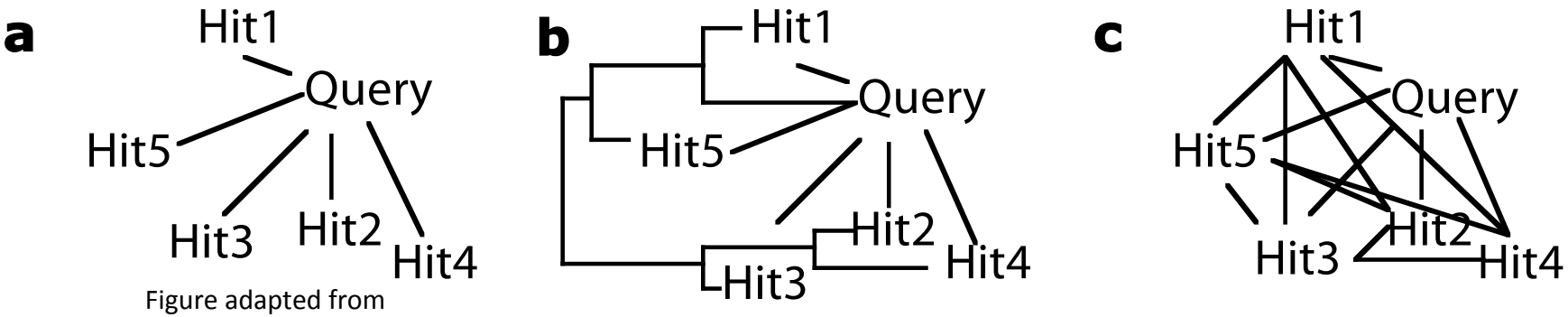


Figure adapted from
Procter et al. (2010) *Nature Methods* **7** S16 - S25

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**, are hybrids – employ sampling strategies to speed up tree building & refinement.

Alignment Job Parameter Settings

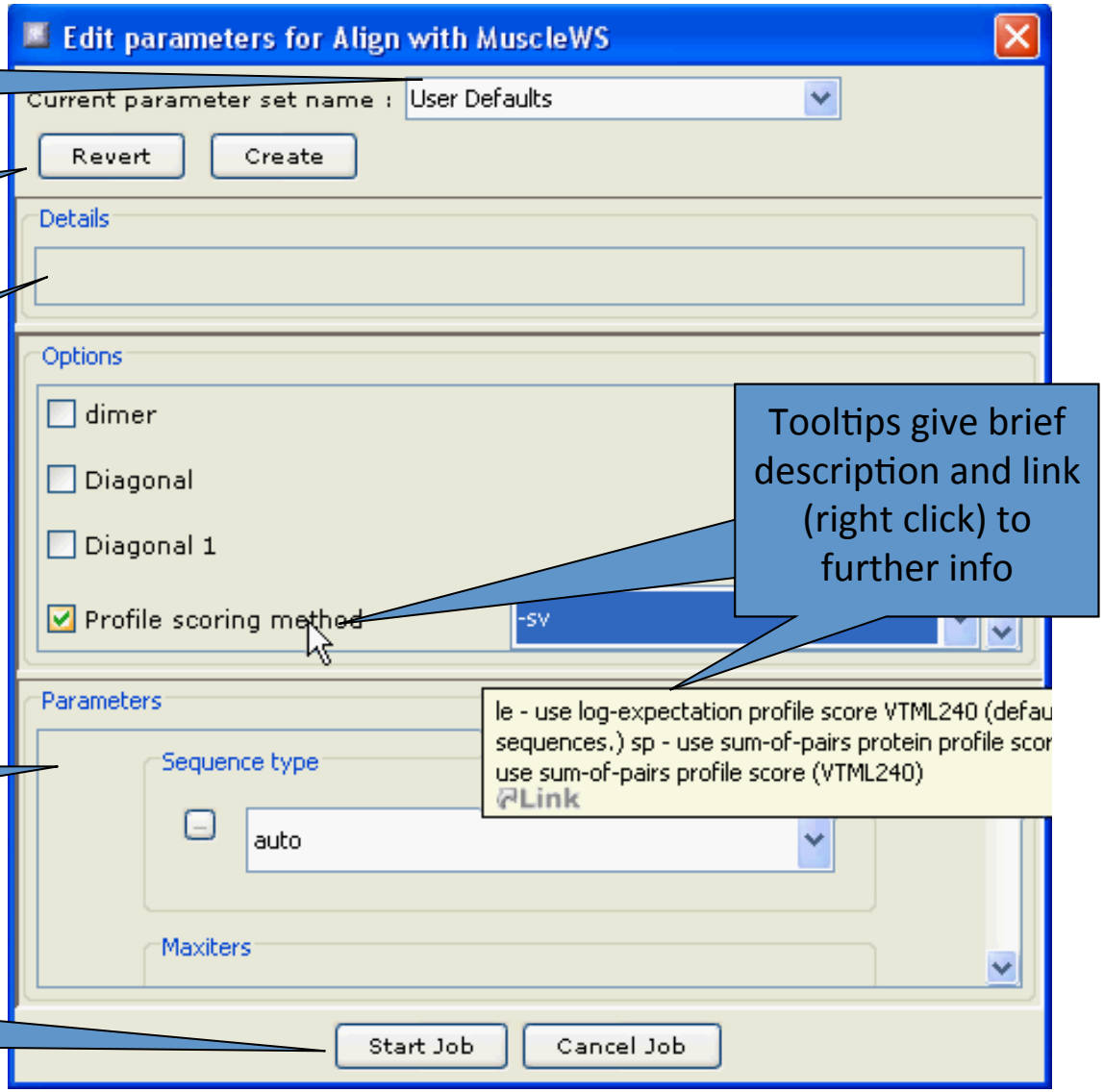
Browse or edit to change name of set

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

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

‘Realignment’ – adding sequences to an existing alignment

Clustal Realign options:

- Jalview leaves gaps in when sending sequences to JABAWS
- ClustalW
  - Fixes sequences with gaps, aligns other sequences to profile
- ClustalO
  - Creates an HMM from sequences with gaps
  - Aligns all sequences to HMM

Clustal Omega quick alignment of millions of sequences

File Edit Select View Format Colour Calculate Web Service

MAFFT Alignment Ordering

Original | Spinach Feredoxin Structure

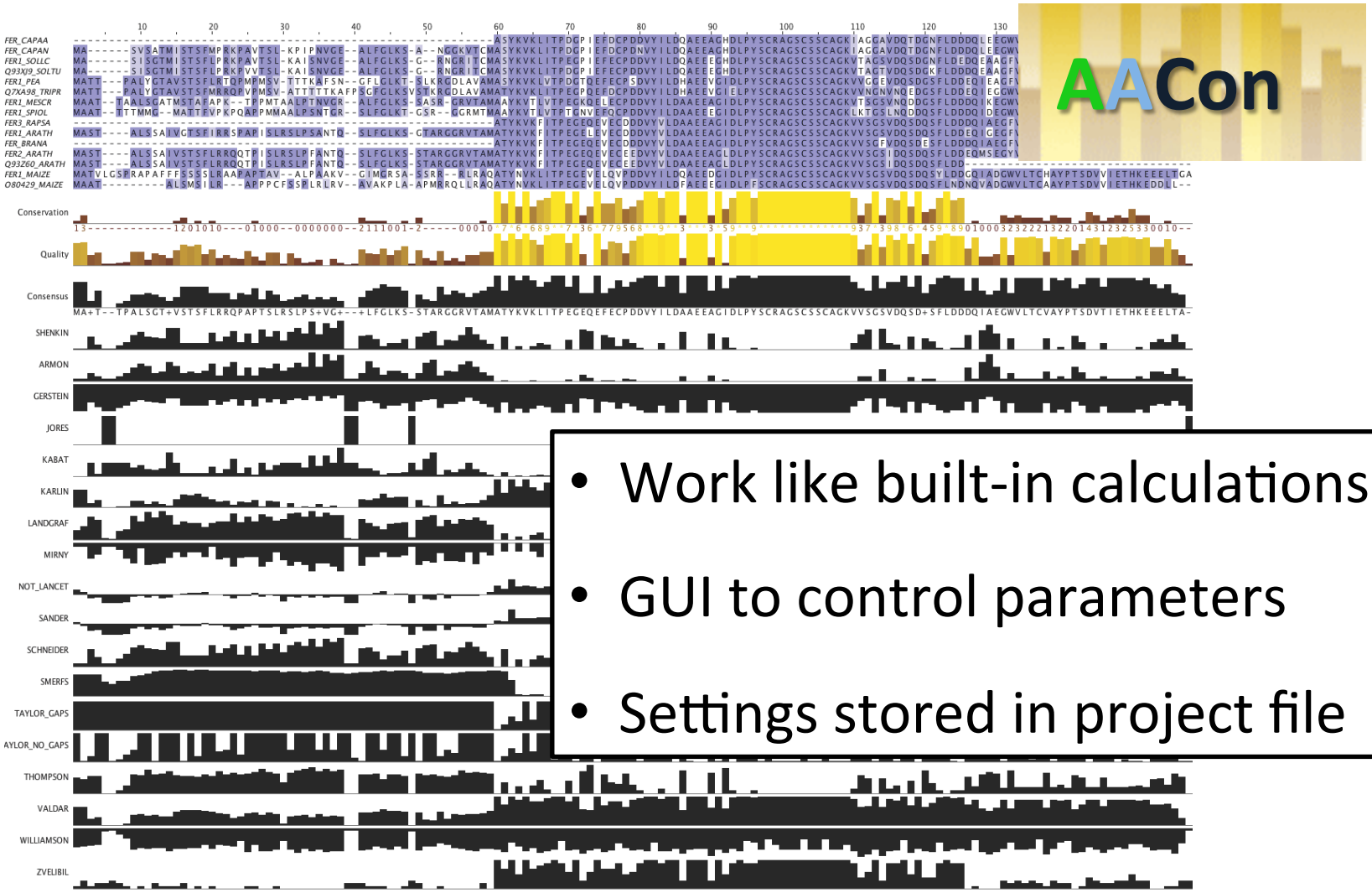
```
                    60        70        80
Q7XA98_TRIPR/1-152  MATYKVKLITPEGPQEFDCPDDVY
FER1_MESCR/1-148    MAAYKVTLVTPEGKQELECPDDVY
FER1_SPIOL/1-147    MAAYKVTLVTPTGNVEFQCPDDVY
FER3_RAPSA/1-96     -ATYKVKFITPEGEQEVECDDDVY
FER1_ARATH/1-148    MATYKVKFITPEGELEVECDDDVY
FER_BRANA/1-96      -ATYKVKFITPEGEQEVECDDDVY
FER2_ARATH/1-148    MATYKVKFITPEGEQEVECEEDVY
```

Secondary Structure

Iron Sulphur Contacts

- Alignment
- Secondary Structure Prediction
- Protein Disorder
- Analysis
- Conservation
- Envision 2
- Fetch DB References

18 new alignment conservation calculations provided as web services



- Work like built-in calculations
- GUI to control parameters
- Settings stored in project file

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.

Consistent Pairwise Alignments

```
Score = 740 Length = 17
PID = 100%
Q93Z60_ARATH MASTALSSAIVSTSFLR
             ||||||||||||||||
  FER2_ARATH MASTALSSAIVSTSFLR

Score = 160 Length = 17
PID=29.41
  FER1_MAIZE LGSPRAPAFFFSSSSLR
             ..|   .. . |.| ||
Q93Z60_ARATH MASTALSSAIVSTSFLR

  Score = 160 Length = 17
  PID=29.41%
  FER1_MAIZE LGSPRAPAFFFSSSSLR
             ..|   .. . |.| ||
  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
```

Major inconsistency: +4 Shift

```
Score = 120 Length = 12
PID = 41.67%
O80429_MAIZE MAATALSMSILR
             .|  .| | ||
  FER1_MAIZE APAFFFSSSSLR
```

```
                                  10
                         |        |         |
FER2_ARATH/1-17          MASTALS----SAIVSTSFLR
Q93Z60_ARATH/1-17        MASTALS----SAIVSTSFLR
FER1_MAIZE/1-21          MATVLGSPRAPAFFFSSSSLR
O80429_MAIZE/1-12        MAATAL---------SMSILR
```

- Calculate ‘shift’ between all pairwise aligments and the multiple sequence alignment
  - Higher shifts are less reliable

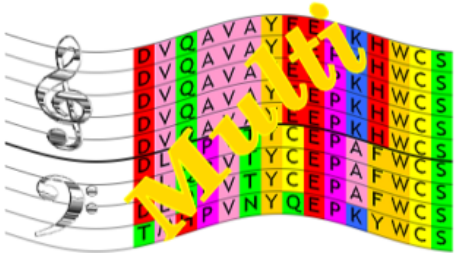
T-COFFEE alignment reliability scores

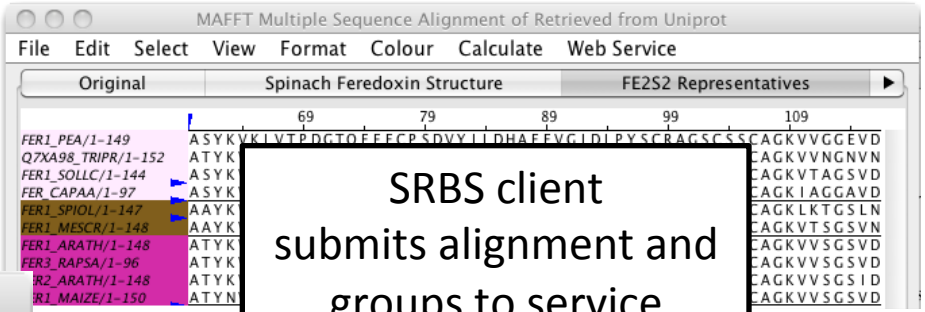
Easiest to do this is from command line, or from the T-COFFEE web site

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

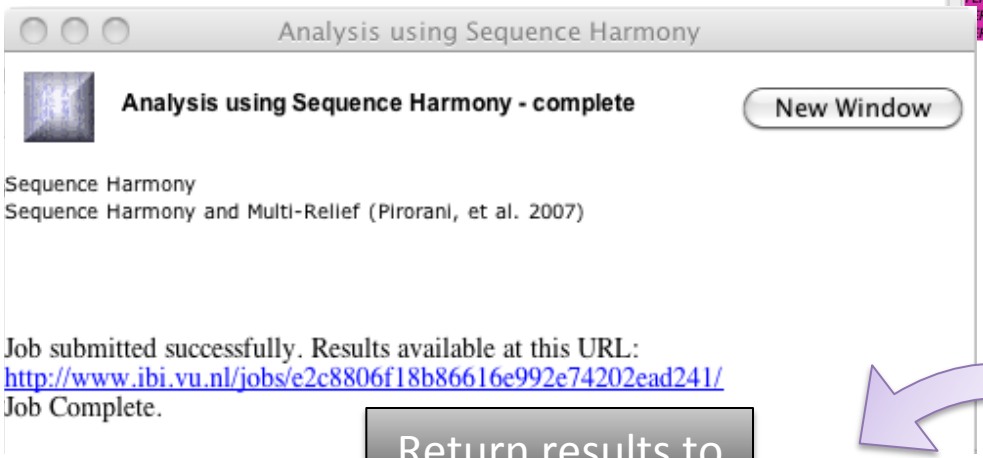
Hope to add functionality to JABAWS in future

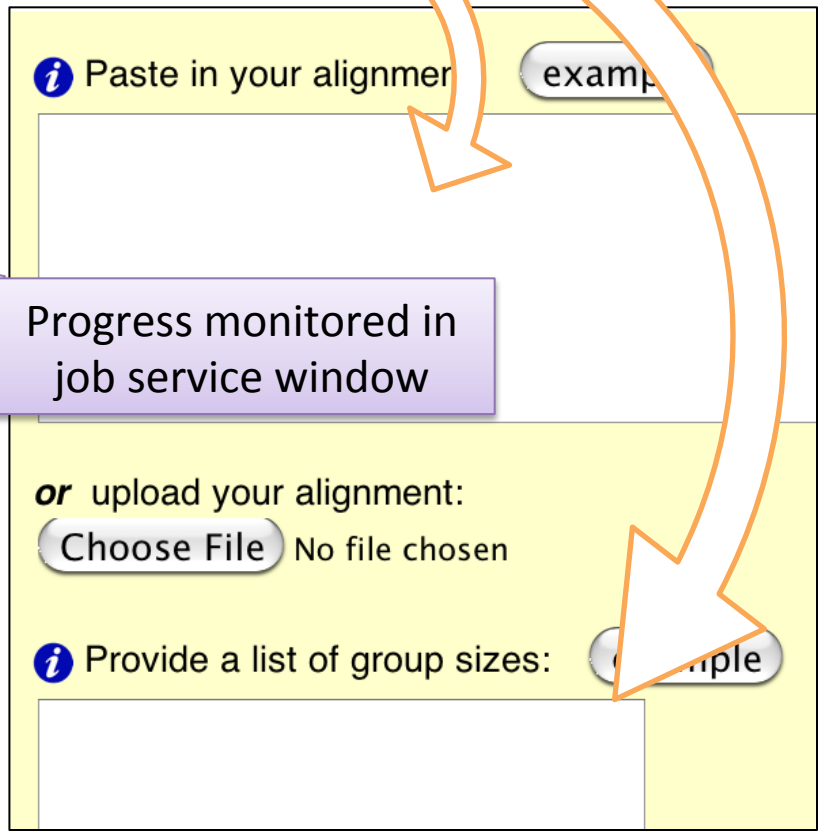
Multi Harmony – from Jaap Heringa's Group at Amsterdam Free University





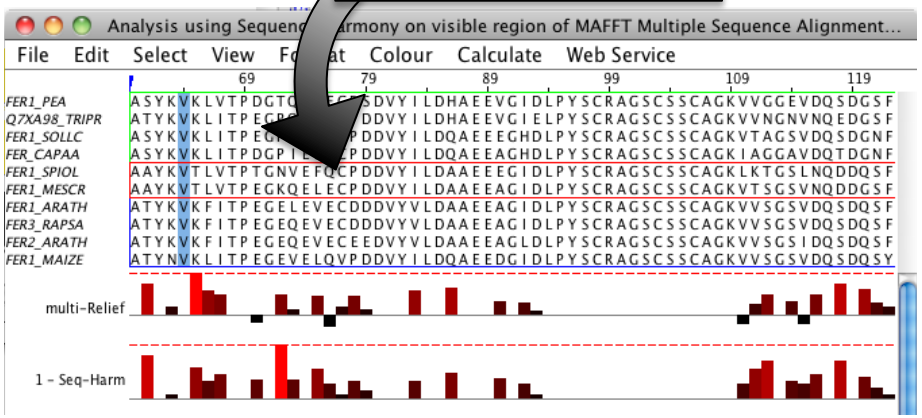
SRBS client submits alignment and groups to service

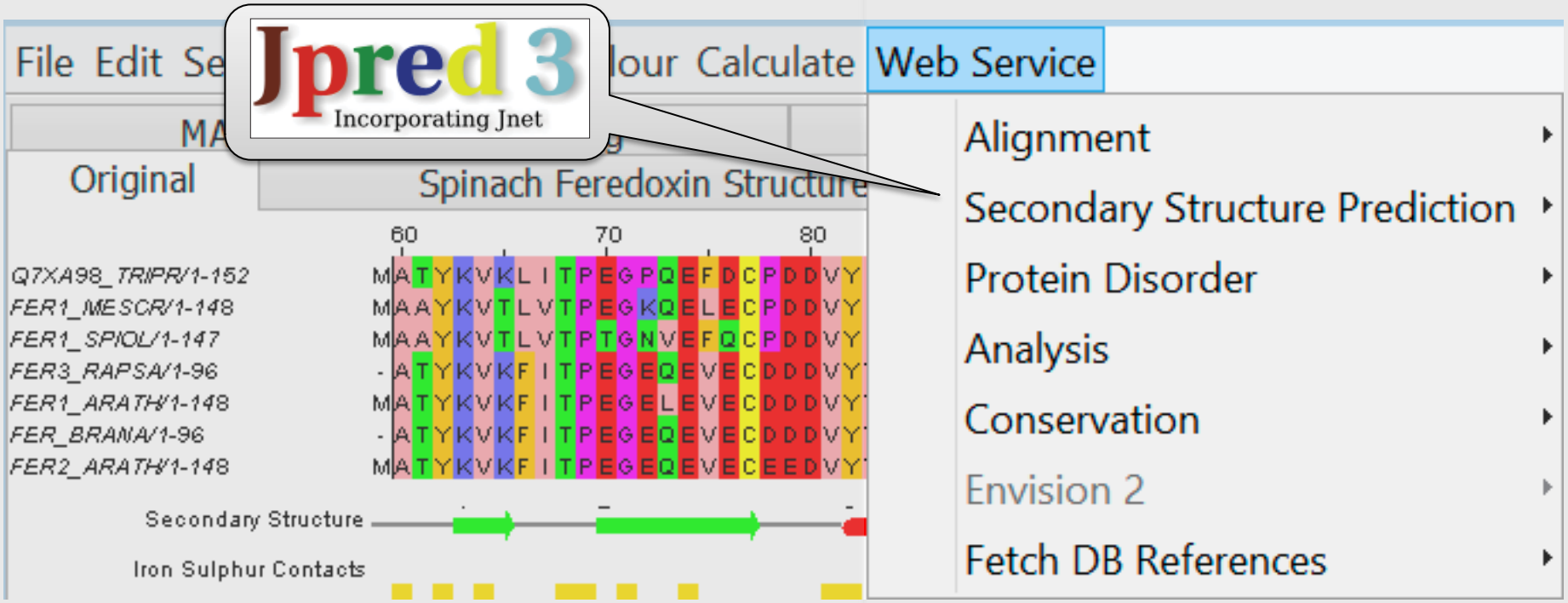




Progress monitored in job service window

Return results to user



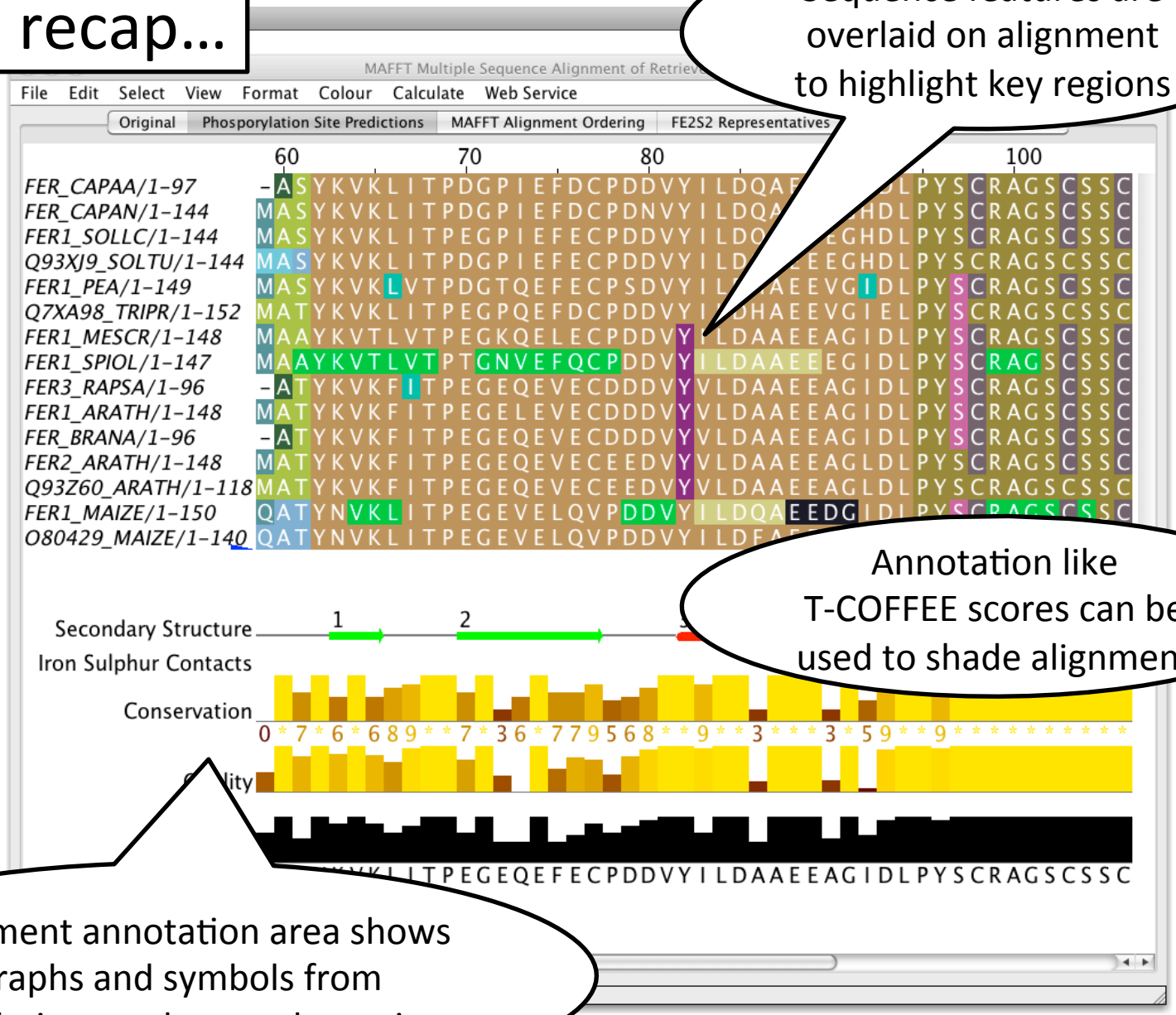

File Edit Se
Jpred 3
Incorporating Jnet
lour Calculate
Web Service
Original
Spinach Feredoxin Structure
60
70
80
Q7XA98_TRIPR/1-152
FER1_MESCR/1-148
FER1_SPIOL/1-147
FER3_RAPSA/1-96
FER1_ARATH/1-148
FER_BRANA/1-96
FER2_ARATH/1-148
Secondary Structure
Iron Sulphur Contacts
Alignment
Secondary Structure Prediction
Protein Disorder
Analysis
Conservation
Envision 2
Fetch DB References

Protein Secondary Structure Prediction

- Neural network trained on amino acid profiles
  - Predicts **H**elix, sh**E**et, or **C**oil 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

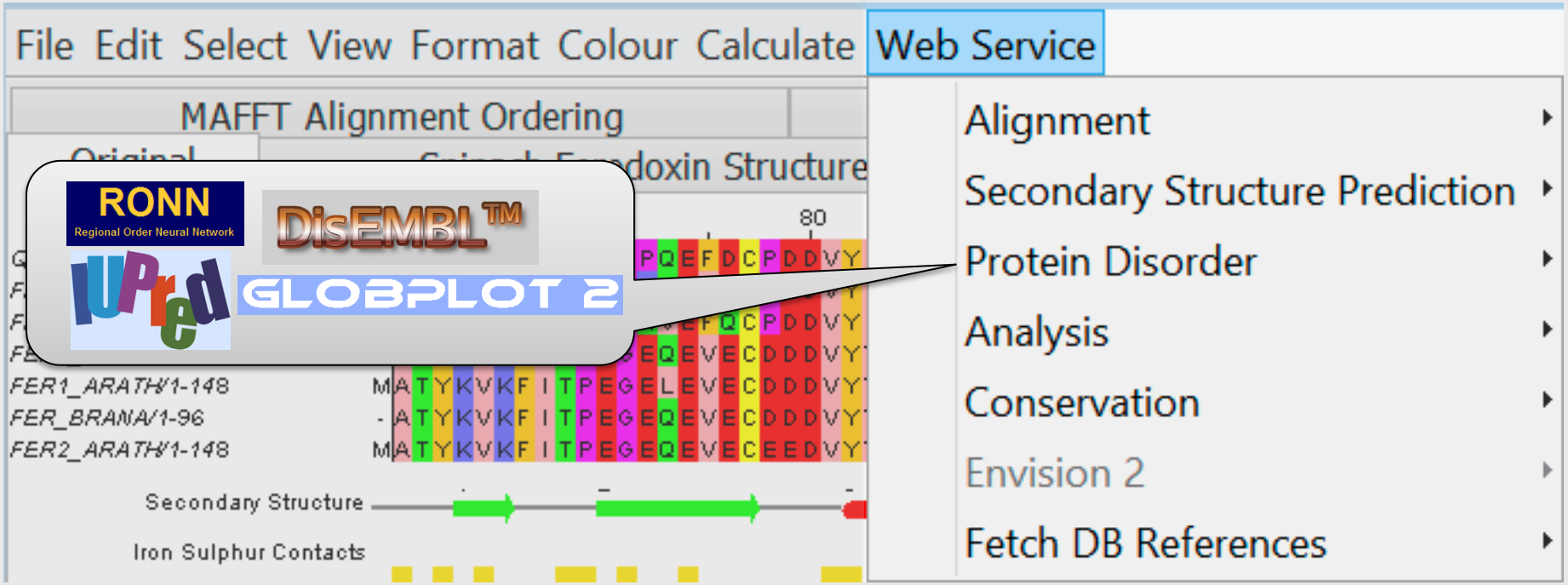
Just to recap…

Sequence features are overlaid on alignment to highlight key regions



Annotation like T-COFFEE scores can be used to shade alignment

Alignment annotation area shows graphs and symbols from calculations and manual curation


File Edit Select View Format Colour Calculate Web Service
MAFFT Alignment Ordering
RONN
Regional Order Neural Network
DisEMBL™
IUPred
GLOBPLOT 2
FER1_ARATH/1-148
FER_BRANA/1-96
FER2_ARATH/1-148
Secondary Structure
Iron Sulphur Contacts
Alignment
Secondary Structure Prediction
Protein Disorder
Analysis
Conservation
Envision 2
Fetch DB References

Protein Disorder prediction

- Complementary problem to secondary structure prediction
  - Recognise structured & unstructured domains
  - Predict holes in density maps (REM450)
  - Detect flexible loops ('HOTLOOPS')

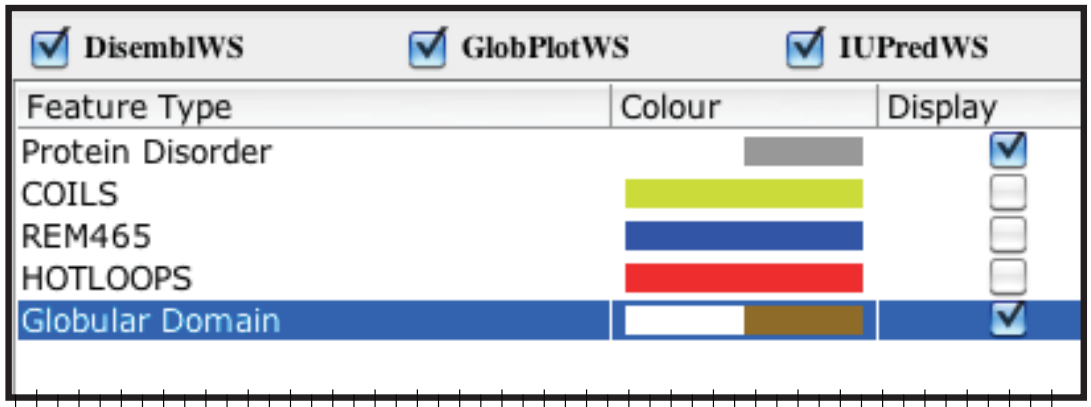
- Programs provided by JABAWS 2 employ
  - Machine learning methods (**DisEMBL**)
  - Similarity to disordered sequences (**RONN**)
  - Empirical amino acid statistics (**IUPred**, **GlobPlot**)

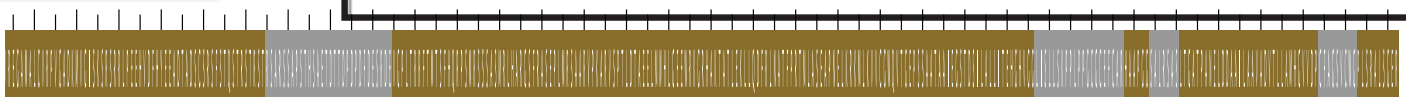
Disorder Predictions from JABAWS

JABAWS Analysis Service

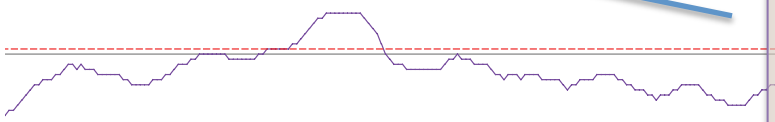
Jalview JABAWS 2.0 Client

Process results into both **annotation** and **features**





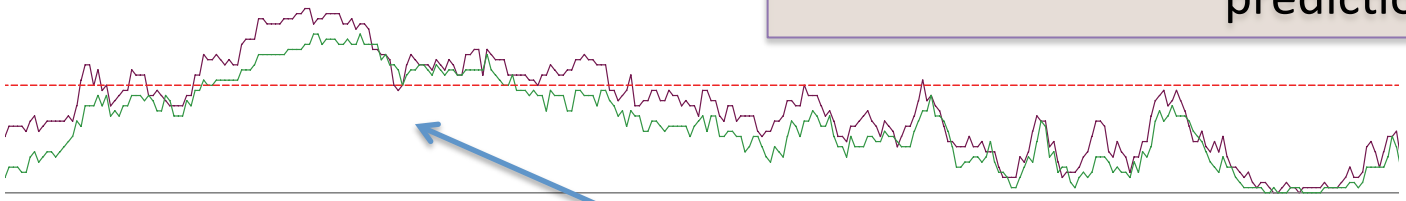
GlobPlotWS (Dydx)



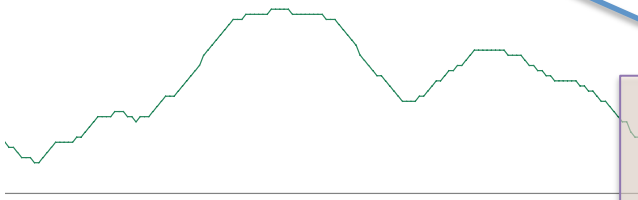
Features highlight disordered region or structured domain predictions

IUPredWS (Short)

IUPredWS (Long)

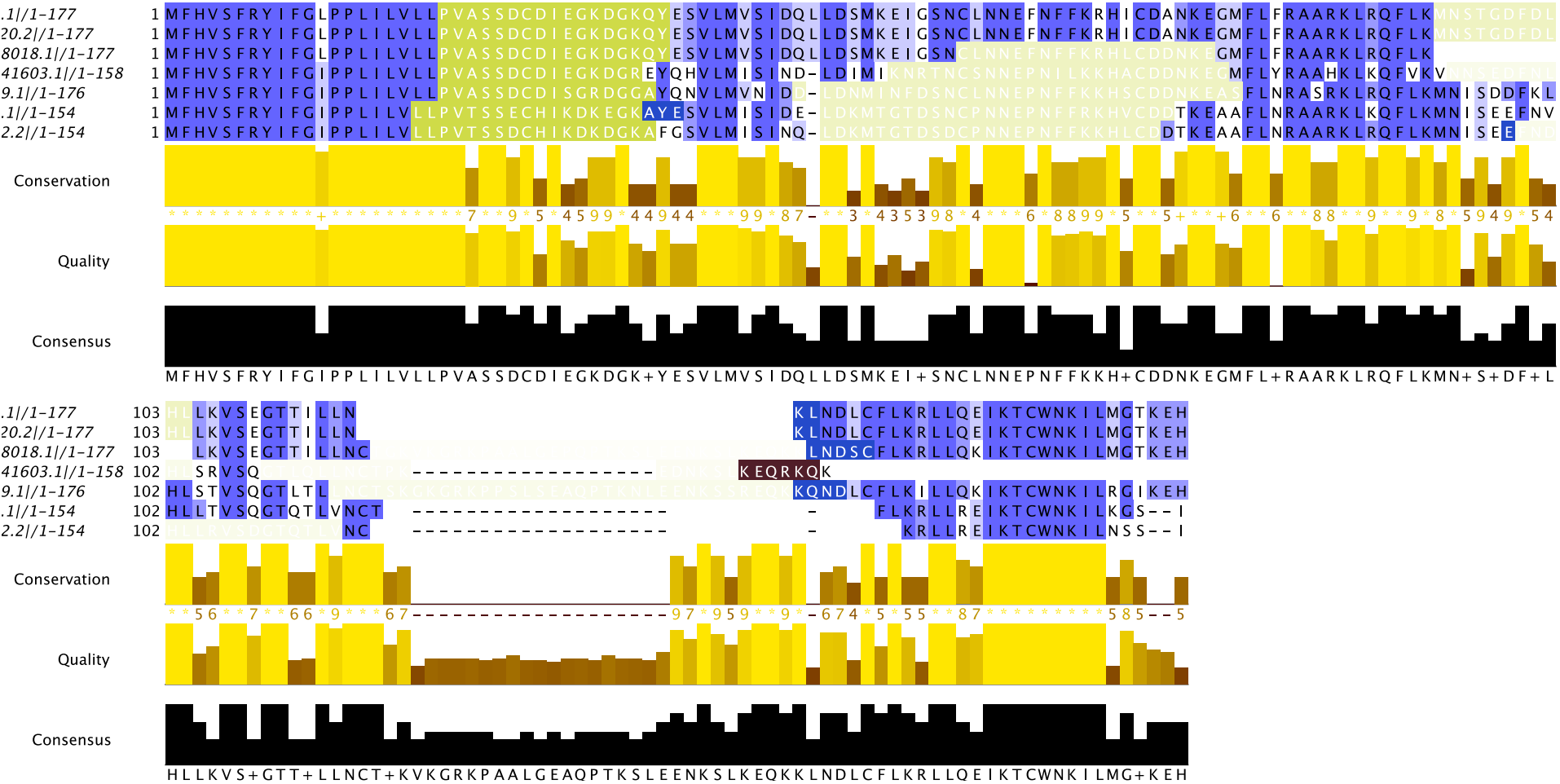


JronnWS (JRonn)



Use Threshold & Per-sequence option on 'Colour by Annotation' dialog to shade alignment using raw scores

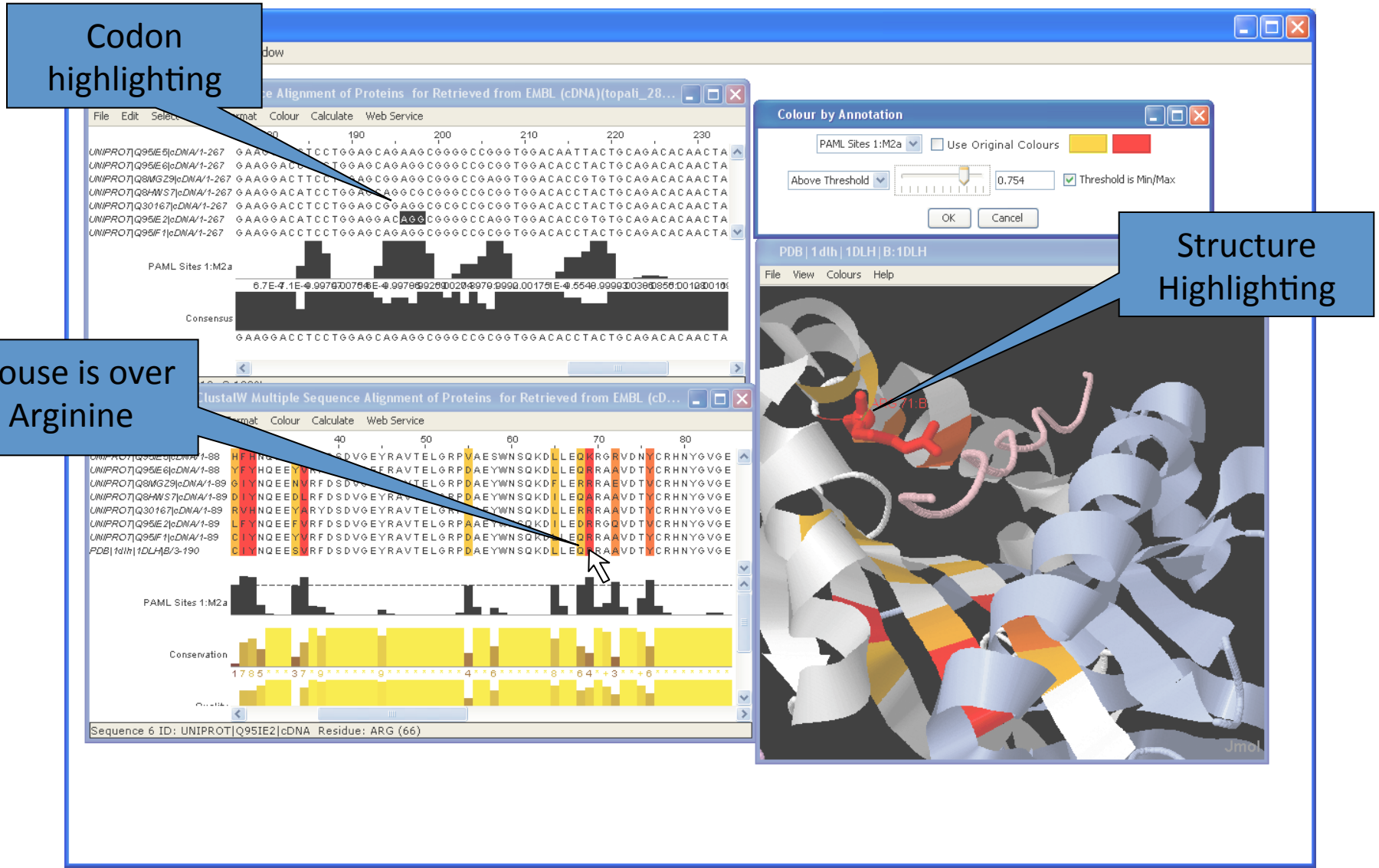
Disorder in Interleukin 7 Orthologs



Jalview and Nucleotide data

- Basic by current standards
- **Built in cDNA->Amino acid translation**
  - Works with Aligned cDNA
  - Preserves alignment annotation
  - Does **not** backtranslate
- **European nucleotide archive records**
  - Parse cDNA annotation in contigs
- **Display of WUSS or VIENNA RNA secondary structure notation**
  - Stockholm file import
  - LocaRNA Clustal files

Translation of annotated cDNA alignment



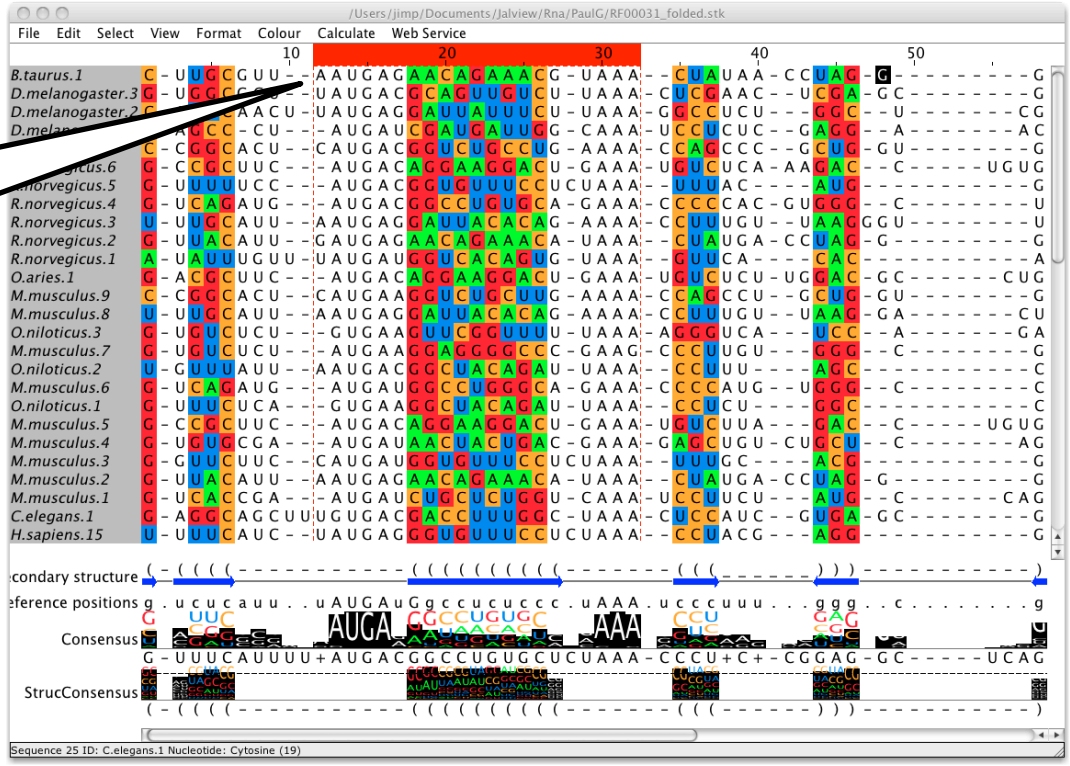
Jalview 2.8 and RNA 2nd-ary Structure

Structure Consensus Logo: Shows base pair distribution at each paired position in a given RNA secondary structure.

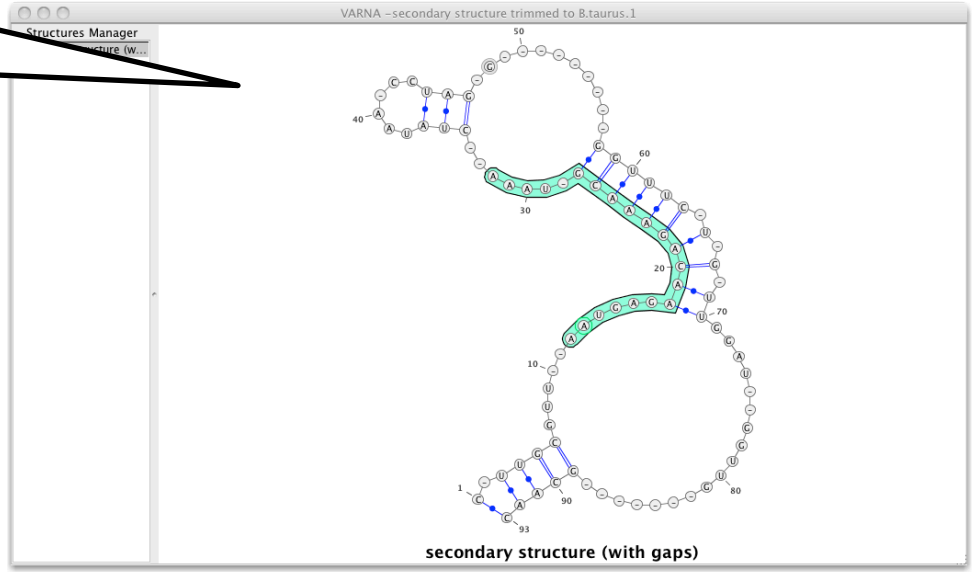
Linked VARNA RNA Secondary Structure viewer and editor.

RALEE style colouring highlights distinct stems and helices

Selections and mouse positions shared between alignment view and VARNA



VARNA has a wide range of 2D RNA plots and supports interactive annotation

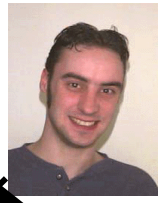


VARNA: Interactive drawing and editing of the RNA secondary structure
Kévin Darty, Alain Denise and Yann Ponty
Bioinformatics (2009) **25** 1974-1975

A brief history of Jalview

Jalview Version 1
1997

Michele Clamp
Director of Informatics and Scientific Applications, Harvard.



James Cuff
High Performance Computing, Harvard.

Steve Searle
Now at Sanger, UK

2004
Jalview 1 published.

David Martin

Jalview Version 2
2005

Andrew Waterhouse
U. Basel.

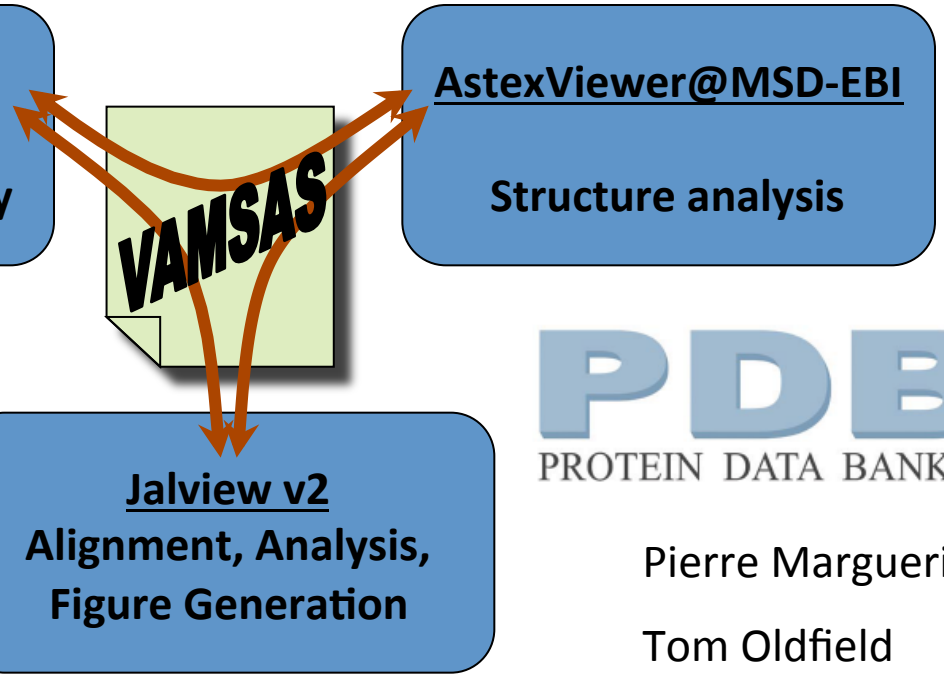
Jim Procter

VAMSAS eScience project

BBSRC
bioscience for the future

VAMSAS

Visualization and Analysis of Molecular Sequences, Alignments and Structures



Iain Milne

Dominik Lindner

Frank Wright

David Marshall

Pierre Marguerite

Tom Oldfield

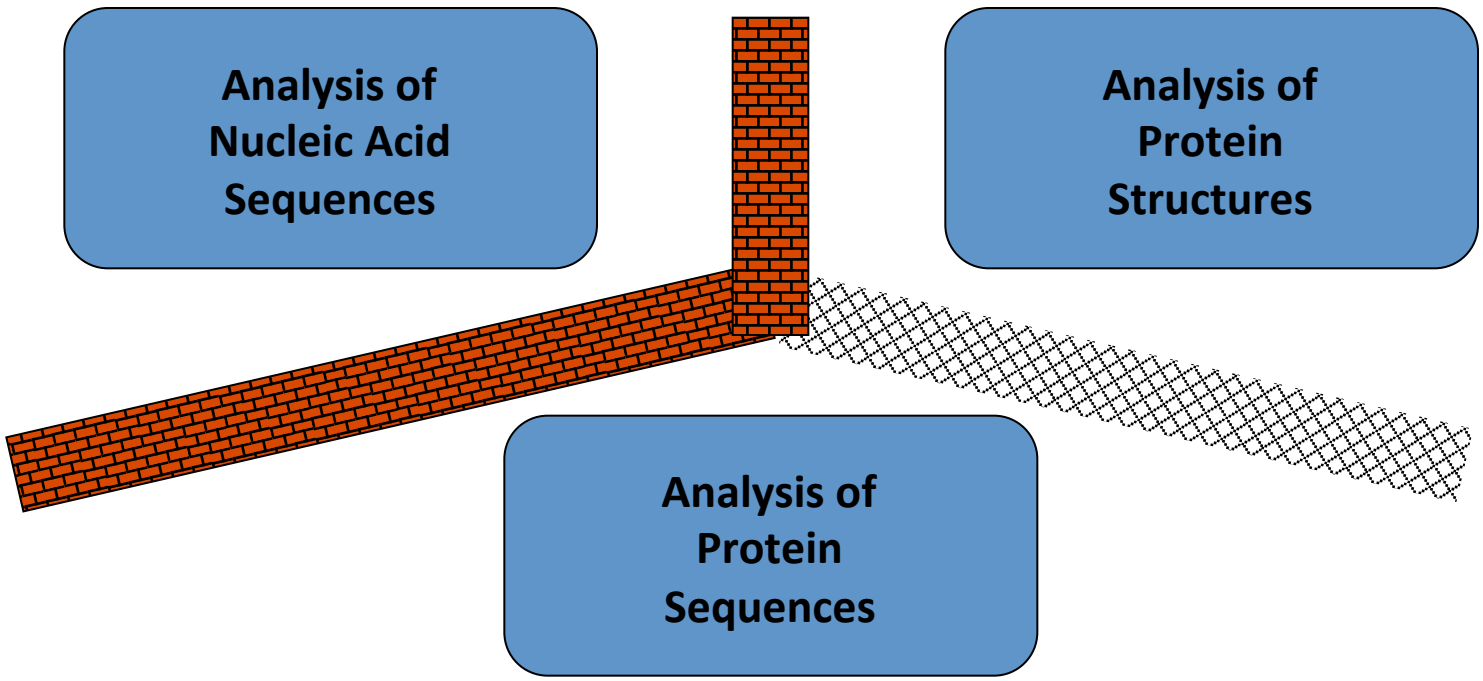
Andrew Waterhouse

Jim Procter

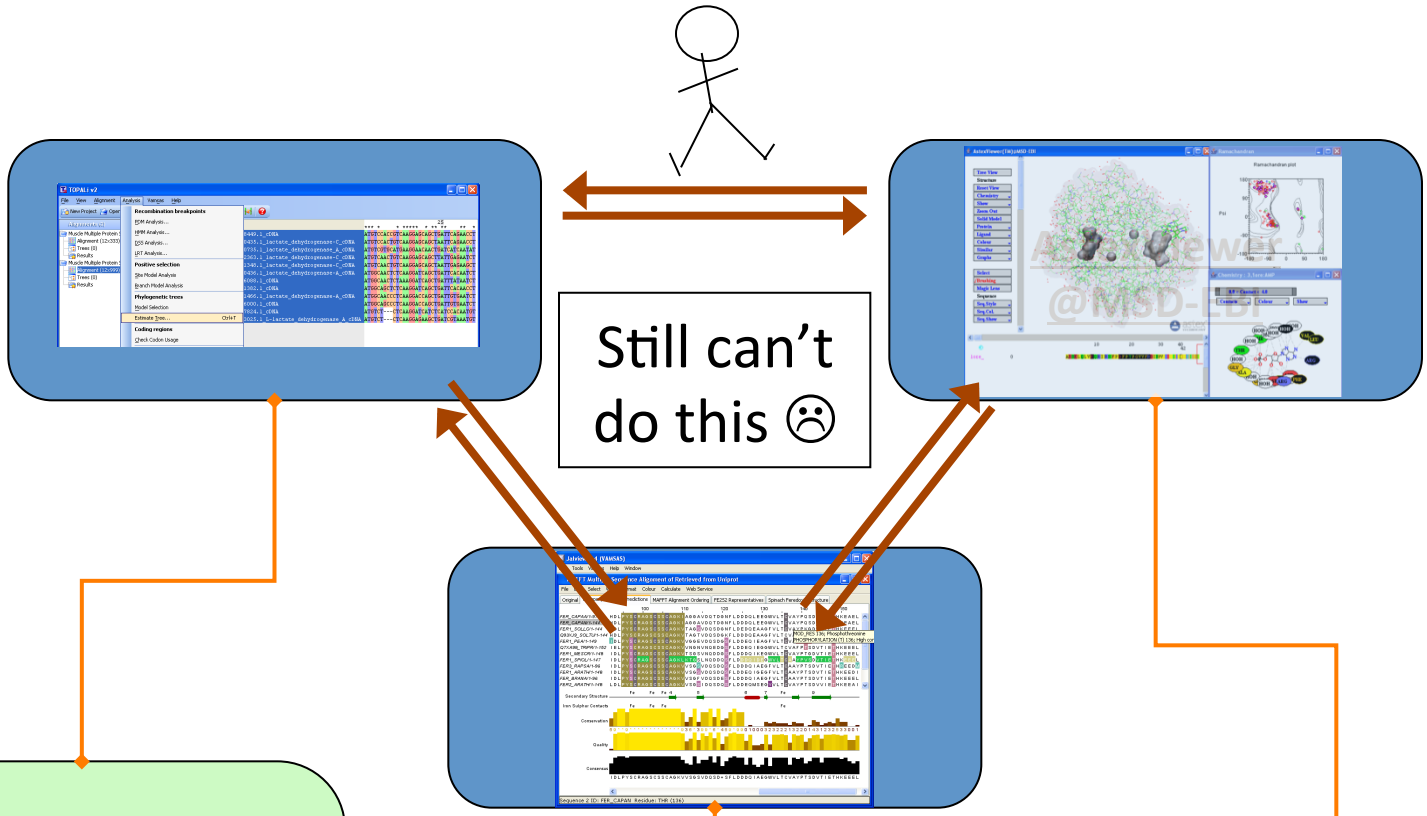
David Martin, Geoff Barton

VAMSAS

Visualization and Analysis of Molecular Sequences, Alignments and Structures



Aim: Enable user to move between different VAMSAS Applications



Model Selection
Phylogeny
Ancestral Sequences
Positive Selection
Recombination

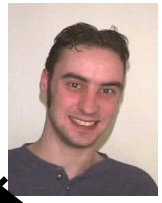
Databases, Annotation
Alignment
2-ary Structure Prediction

Structure Databases
Structural Clustering
Uniprot/MSD Mapping

A brief history of Jalview

Jalview Version 1
1997

Michele Clamp
Interim Director of Research Computing, Harvard.



James Cuff
Cycle Computing, USA.

Steve Searle
Now at Sanger, UK

2004
Jalview 1 published.

Jalview Version 2
2005

Andrew Waterhouse
U. Basel.

Jim Procter

2.1

VAMSAS eScience project

2.2

2.3

2.4

David Martin

Jalview 2 Paper
2009

Unsupported graphical tools cause headaches..

User Headaches:

- Can't do what you want:
  - Read/write format 'X'
  - Run my analysis 'Y'
  - Load my 10Gb dataset
- Slow or difficult to use
- Doesn't work properly

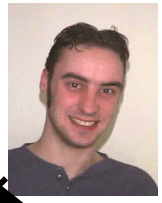
Core Developer needed for:

- Staying 'state of the art':
  - New standards
  - New analysis methods
  - New data
- Maintenance
  - Dependencies
  - New OS Versions
  - New hardware
- Hard to do 'research' and maintain software

A brief history of Jalview

Jalview Version 1
1997

Michele Clamp
Interim Director of Research Computing, Harvard.



James Cuff
Cycle Computing, USA.

Steve Searle
Now at Sanger, UK

2004
Jalview 1 published.

Jalview Version 2
2005

Andrew Waterhouse
U. Basel.

Jim Procter

Peter Troshin

2.1

VAMSAS eScience project

2.2

2.3

2.4 2.5 2.6 .1 2.7 **2.8**

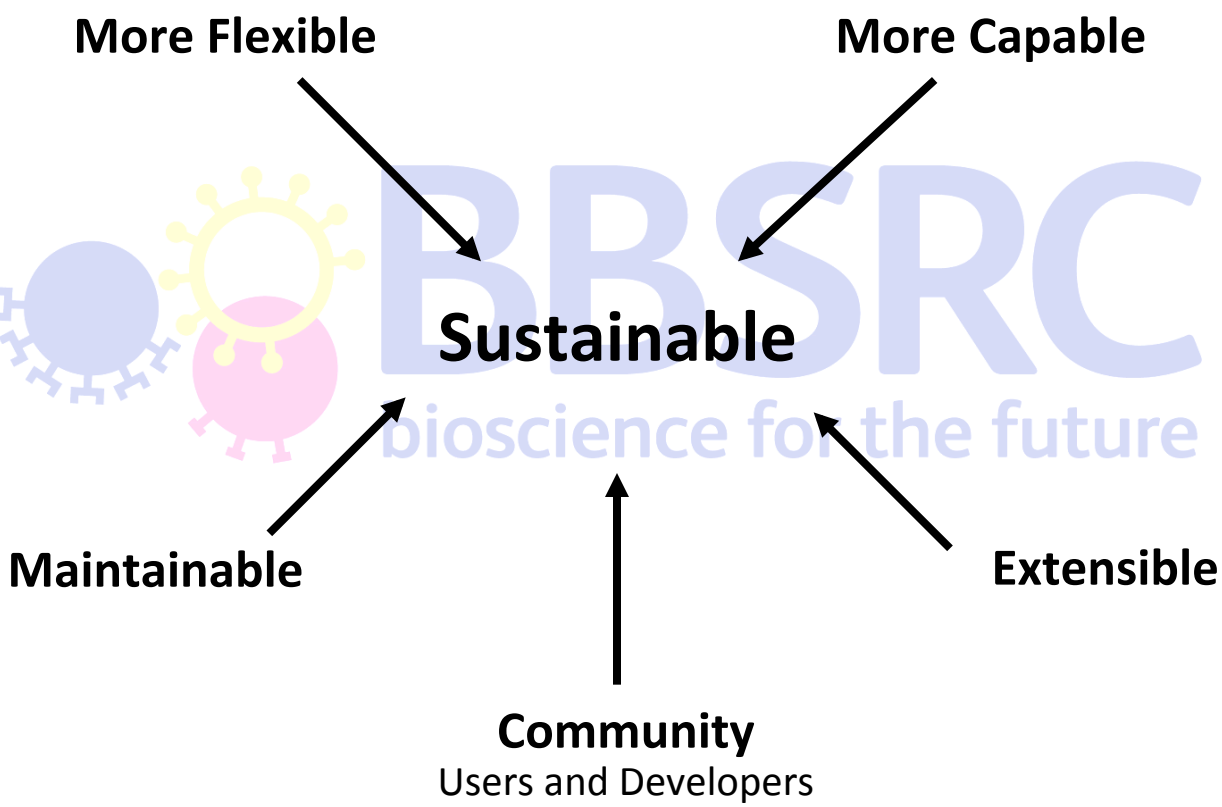
David Martin

Jalview 2 Paper
2009

New BBSRC Funding
2009-2014

1 Oct 2009 — 30th Sep 2014

The Jalview 5 Year Plan



5 Year Tools and Resources Development Fund Grant from the UK Biotechnology and Biological Sciences Research Council

1 Oct 2009 — 30th Sep 2014

The Jalview 5 Year Plan

More Flexible

GUI refactor & plugins

More Capable

Large Datasets

Sustainable

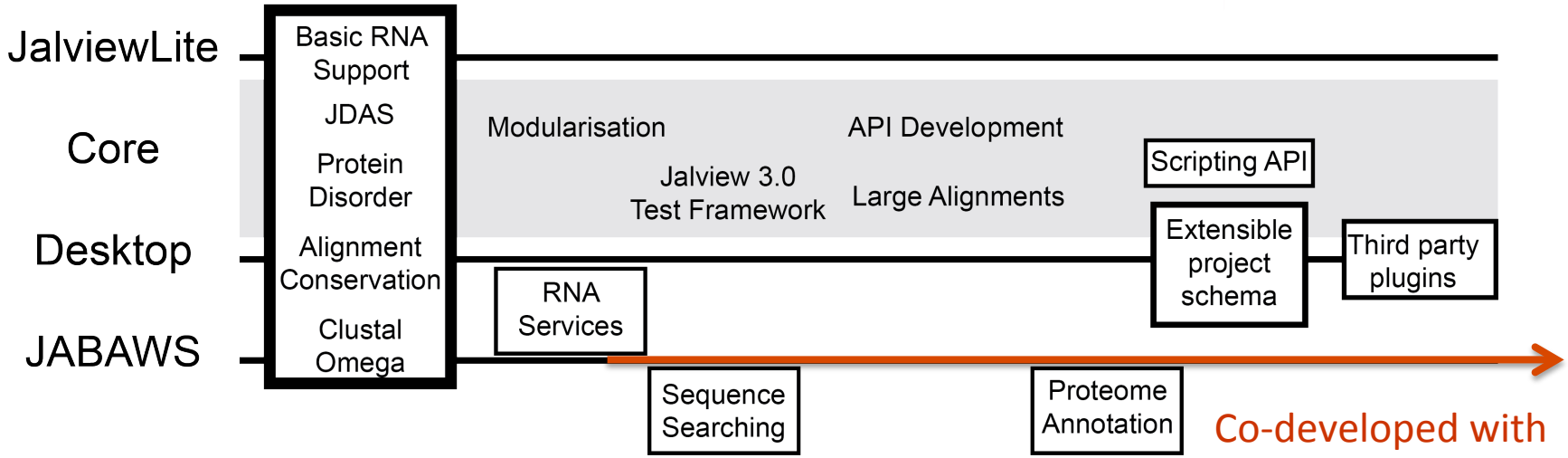
Maintainable

Software Engineering

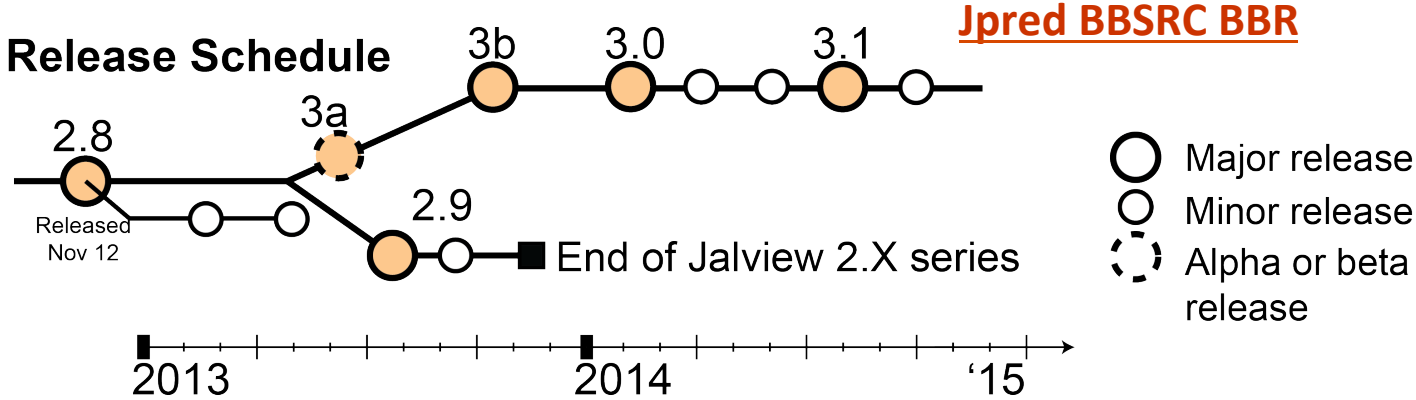
Community

Extensible

Analysis services



Co-developed with **Sasha Sherstnev** on **Jpred BBSRC BBR**



1 Oct 2009

30th Sep 2014

The Jalview 5 Year Plan

More Flexible

More Capable

GUI refactor & plugins

Large Datasets

Community Outreach

- Website redesign
- Training
  - Users and Developers
- Internationalisation
- True Open Source Development
  - Issue tracker and open repository

Maintain

tensible

Software Engineering

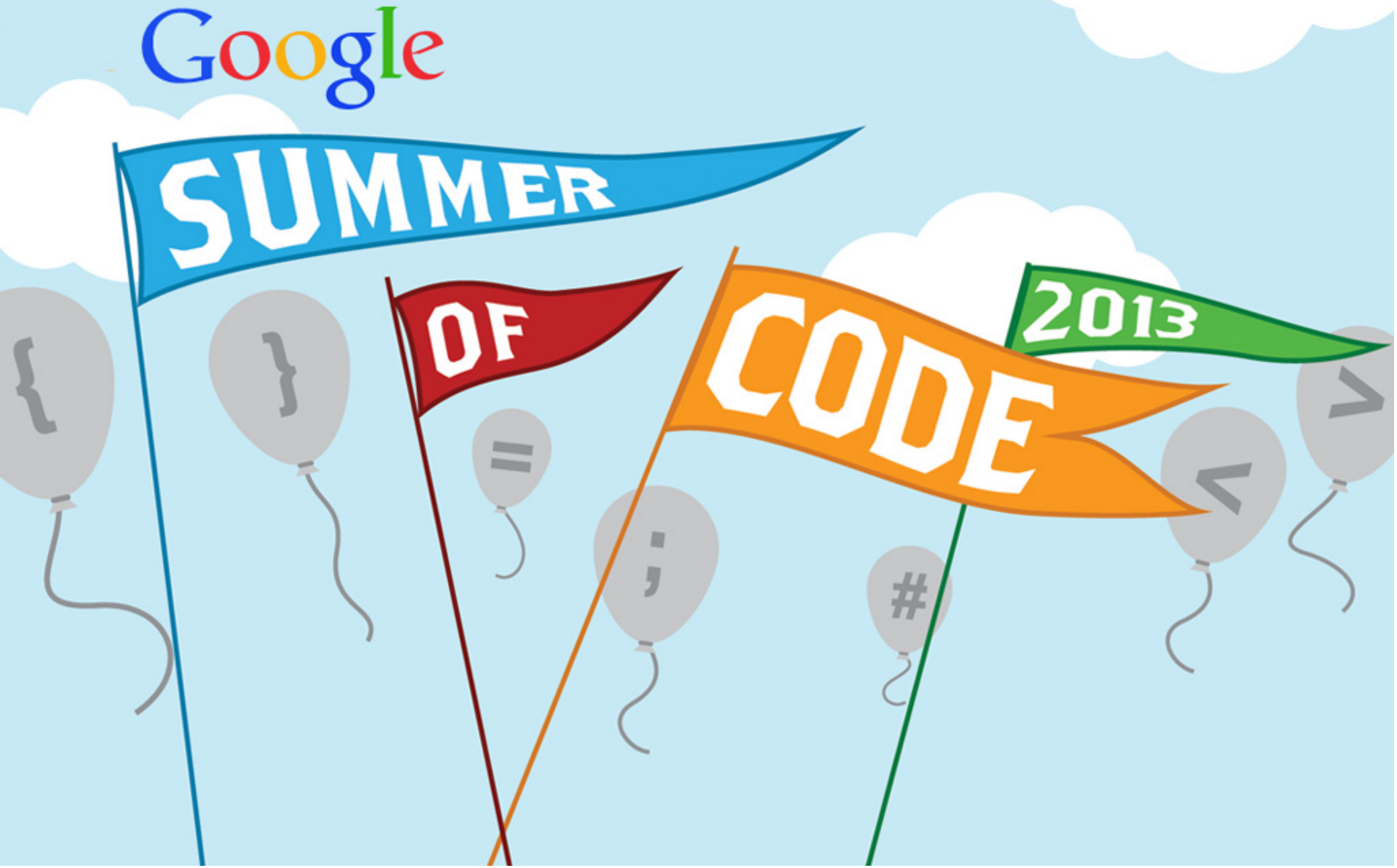
Analysis services

Community

Users and Developers


Practical Jalview
guided tutorial and Jalview clinic
Practical Jalvie
guided tutorial and Jalview clini

www.google-melange.com



Lauren Lui, UC Santa Cruz.

http://jalview-rnasupport.blogspot.com/

NESCent

Purine/pyrimidine colourscheme

Rfam

alignment fetcher

WUSS annotation parser (from RALEE)

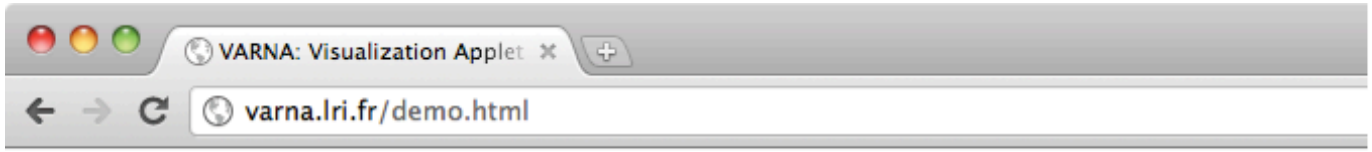
Colouring to highlight helical structure

SUMMER OF CODE // 2010

code.google.com/soc

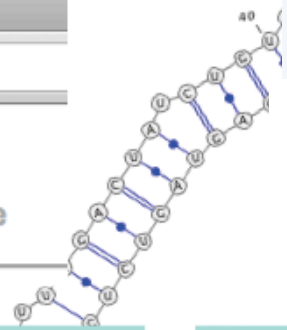
Google

Jan Engelhardt (Uni. Leipzig)

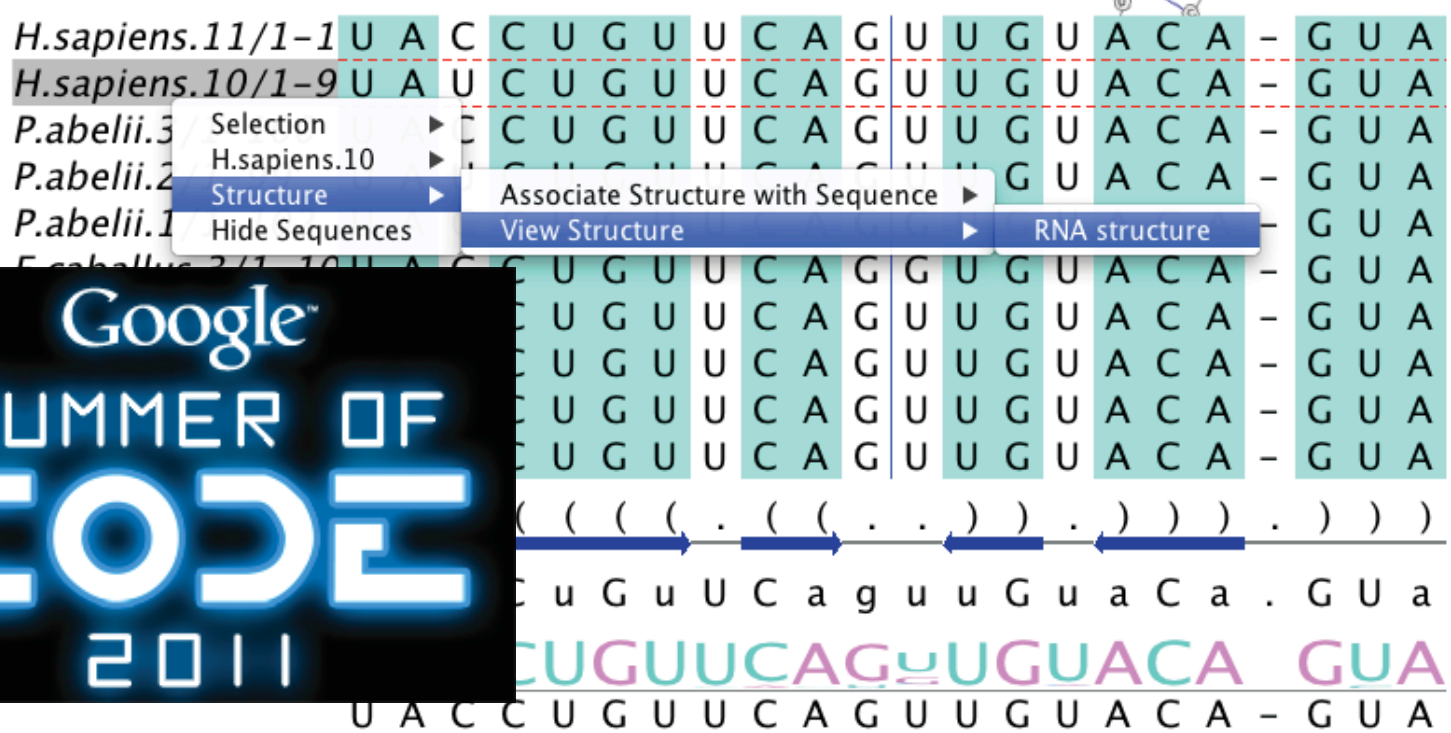


VARNA: Visualization Applet for RNA

A Java lightweight component and applet for drawing the RNA secondary structure



NESCent



Google Summer of Code 2013 : Why should I participate?

NESCent

Students gain:

- skills
- real world experience
- sample code
- contacts

Organizations gain:

- new contributions & contributors
- global exposure

Google and NESCent Summer of Code 2013

New phylogeny support in Jalview ?

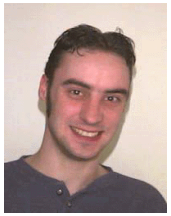
- cDNA/protein editing
  - Display AA translation of cDNA alignment
  - ORF support
- Bacterial genome alignment manipulation
  - New editing
- JABAWS
  - Phylogeny services
  - Extend alignment services to
    - Return guide trees and uncertainty scores
    - Support ‘add’ and profile alignment

Jalview

www.jalview.org

The Jalview developers

Michele Clamp
Harvard, USA.



James Cuff
Cycle Computing

Steve Searle
Sanger, UK

Andrew Waterhouse
Basel, Switzerland.

RNA Features

Lauren Lui
UC Santa Cruz, USA.

Jan Engelhardt
Univ. Leipzig, Germany.

Yann Ponty (**VARNA**)
École Polytechnique, Fr.

SUMMER OF CODE // 2010
code.google.com/soc
Google

Google SUMMER OF CODE 2011

T-COFFEE Scores

Paolo di Tomasso
Notredame Group, CRG, Spain.

Geoff Barton (**Money**)
David Martin (**Teaching**)
Peter Troshin (**JABAWS**)
Barry Strachan (**logo**)
Tom Walsh (**Apache**)
Ryan Maclaughlan (**CSS**)
Andrew Millar (**Drupal**)

All the Jalview users, and …

BBSRC
bioscience for the future