Introduction to Jalview
a guided tutorial
EMBL-EBI, Jan 21st, 2013.

Jim Procter & Geoff Barton
College of Life Sciences
University of Dundee, UK.
jprocter@dundee.ac.uk
Day 2 - Tuesday

09:00-10:30   Advanced Jalview and Jalview Hackathon

10:30-11:00   Coffee/Tea Break

11:00-12:30   Advanced Jalview and Jalview Hackathon continued

12:30-13:30   Lunch

13:30-15:00   Jalview Clinic and Jalview Hackathon

15:00-15:30   Coffee/Tea Break

15:30-15:50   Jalview Clinic and Jalview Hackathon continued

15:50-16:15   Wrap-up

16:15   Departure
Course materials

Everything is (or will) be online at
http://www.jalview.org/tutorial/training-materials/2013/EBI/Jan/

- These slides
- Jalview v2.8 Tutorial (v1.4.1)

- A copy of the tutorial is also on Penelope
Getting and working with sequence features and annotation

• Sequence Databases

• Sequence feature sources
  – DAS Sequence feature retrieval
  – GFF and Jalview feature files

• Visualizing features
  – Highlighting annotated regions
  – Shading and reordering based on scores and labels
Distributed Annotation System

Interactive Editing

Visualization

- Alignments
- Structures
- Sequences
- Features
- Annotation
- Trees
- PCA

Rfam

Pfam

PDBe

UniProt

ENA

Distributed Annotation System
Jalview and Sequence Databases
Sec 2.9.1 Ex. 29

- Can retrieve new sequences or match against existing records using IDs

- Task
  - Recover the Uniprot annotation for the ferredoxin sequences using their IDs
    - make sure you are using the latest 2.8 Build
  
  - Verify retrieval by examining sequence annotation
Jalview and Sequence Databases
Sec 2.9.1 Ex. 29

• Can retrieve new sequences or match against existing records using IDs

• Task
  – Recover the Uniprot annotation for the ferredoxin sequences using their IDs
  – Verify retrieval by examining sequence annotation

• Question
  – What happens if only a subsequence is present in the alignment?
  – Does database annotation get shared between alignments?
Sequence Features
Sequence Features
Section 2.8.1-3 & Ex 27

- Annotate the whole or part of a sequence
- Database refs are special case.

• Tasks
  - Visualise, create, modify, import and export features.
Sequence Features
Section 2.6.1-3 & Ex 27

- Annotate the whole or part of a sequence
- Database refs are special case.

• Tasks
  - Visualise, create, modify, import and export features.

• Questions
  - What are the different types of file formats available for import and export
  - What services allow you to discover annotation for sequence?
The Distributed Annotation System
Section 2.9.2, Exercise 30

– Web servers that jalview can use to discover annotation for a sequence

• Task
  – Browse available DAS sources for protein sequences
  – Retrieve annotation for the ferredoxin alignment.
The Distributed Annotation System
Section 2.9.2, Exercise 30

– Web servers that jalview can use to discover annotation for a sequence

● Task
  – Browse available DAS sources for protein sequences
  – Retrieve annotation for the ferredoxin alignment.

● Question
  – What does the ‘optimise order’ button do?
Working with sequence features
Ex 31 and 32 (Sec 2.9.4)

• Task
  – Shading features using labels and scores
  – Sorting alignment using feature scores
Shading, thresholding, colour by label.
Working with sequence features
Ex 31 and 32 (Sec 2.9.4)

• Task
  – Shading features using labels and scores
  – Sorting alignment using feature scores

• Questions
  – What types of features are best displayed with a ‘label’ colourscheme?
  – How would you display only the highest or lowest scoring features?
Protein secondary structure prediction

Section 2.6 onwards in the Manual
Page 64
Protein Secondary Structure Prediction
Sect 2.6

• Jalview interfaces with the Jpred protein secondary structure predictor

• Prediction is based on
  – Neural net which can recognise helical, coil or beta strand using amino acid patterns
  – Amino acid profile for a sequence
    • Multiple sequence alignment
    • Profile from sequence database search
      – Position Specific Substitution Matrix
Exercise 26

• Tasks
  – Perform a variety of Jnet predictions
    • Note the effect of hidden regions
    • Learn about sequence associated annotation
  – Save your results for the next exercise

• Questions
  – …
Exercise 26

• Tasks
  – Perform a variety of Jnet predictions
    • Note the effect of hidden regions
    • Learn about sequence associated annotation
  – Save your results for the next exercise

• 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 ?
Alignment Annotation and sequence features.

Annotation is shown below alignment.

Feature settings & annotation display settings.

Sequence features are positional annotations mapped on to aligned positions.
Creating, editing and using annotation. Exercise 28 (Sec. 2.8.4)

- **Tasks**
  - Manually annotate some columns using the interactive editing functions
  - Learn about jalview annotation files
    - How to change the appearance of quantitative data.
    - Understand how to create sequence associated annotation

- **Questions**
  - What other things can be defined in jalview annotation files?
- Session 3
  - Sequence DB refs and Sequence Features
  - Protein secondary structure prediction
  - Alignment annotation

- Session 4
  - Working with structures
    - Viewing 3D structures
    - Viewing RNA
Protein Structure and Jalview
Section 2.1

• Jalview includes the Jmol molecular graphics viewer
  – Structures can be coloured by their aligned sequences
  – Position of mouse highlighted in sequence or structure
Structure shaded by sequence

Mouse is over Arginine

Codon highlighting
Associating structures with sequences

• Local PDB file
  – Attach PDB file to sequence manually
  – drag and drop to match files to sequences by ID

• Structures in the PDB database
  – Provide PDB id (and chain) for sequence
  – Discover references via sequence database
Protein Structures in Jalview
Sec 2.1. Exercise 14

• Task
  – Discover PDB structures for ferredoxin sequence(s)
    • Note use
      Fetch Database Refs->UNIPROT->Uniprot
  – Save and load structures and manipulate colouring
Protein Structures in Jalview
Sec 2.1. Exercise 14

• Task
  – Discover PDB structures for ferredoxin sequence(s)
  – Save and load structures and manipulate colouring

• Questions
  – How does Jalview match up sequence data to structural data
many structures can be shown in a single view

Structures can be superimposed using the visible region of alignment
Superposing Structures using Alignments
Sec 2.1.4 – Exercise 15

• Task
  – Align structures using the ferredoxin alignment

If ‘View all N structures’ doesn’t align structures:
  • Use Jmol->Align menu

  – Experiment with views to control what part of the alignment is used to superimpose the structures
Superposing Structures using Alignments
Sec 2.1.4 – Exercise 15

• Task
  – Align structures using the ferredoxin alignment
  – Experiment with views to control what part of the alignment is used to superimpose the structures

• Questions
  – What colourscheme would highlight the conserved parts of the structures?
  – Which view gave the ‘best’ structure superposition?
    • How did you decide this?
Colouring structures using many multiple alignments
Sect 2.1.5. Exercise 16
DNA and Protein in Jalview

• Discussed in Section 2.10 of manual

• From DNA to Protein
  – Calculations => Translate cDNA
  – View protein annotation on exons using EMBL records

• From protein to DNA
  – Recover DNA for proteins using EMBL cross references
Protein Feature visualization on DNA
Section 2.10, exercise 33

• Task
  – Retrieve a DNA contig and visualize features from UNIPROT at their coding positions.

• Question
  – What fields in an EMBL sequence record can Jalview use?
Semantic Processing: Database Reference Tracing
‘get me the sequences from database blah for the selected sequences’

1. Is this reference a cross reference?
2. Is there already a sequence associated with this reference?
   If not: Retrieve it.
3. Copy associated sequence to new 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.

Try it now:
1. Import RF00162 from Rfam (Full)
2. Select ‘Colour by RNA Helices’

RALEE style colouring highlights distinct stems and helices
VARNA has a wide range of 2D RNA plots and supports interactive annotation. Selections and mouse positions shared between alignment view and VARNA.

Try it now:
1. Open VARNA with 
2. Explore difference between trimmed and untrimmed
3. Bonus points – add and link a Jmol structure view

Bacillus_amyloliquef.9 corresponds to PDB 3NPB

VARNA: Interactive drawing and editing of the RNA secondary structure
Kévin Darty, Alain Denise and Yann Ponty
WRAP UP
One alignment, many views

Sequence features on sequences shown to highlighted key regions

Alignment annotation area shows graphs and symbols from calculations and manual curation
Linked tree viewer allows subgroups to be identified in alignment.

Linked Jmol viewer shows one or more structures coloured by alignment.

Group selections and mouseovers are also available.
18 new alignment conservation calculations provided as web services

- Work like built-in calculations
- GUI to control parameters
- Settings stored in project file
This figure shows a set of structures superimposed according to an alignment generated by T-COFFEE. The T-COFFEE reliability score highlights the most reliable regions in red, and least reliable in blue.
Disorder Predictions from JABAWS

JABAWS Analysis Service
Jalview JABAWS 2.0 Client

Process results into both annotation and features

GlobPlotWS (Dydx)
IUPredWS (Short)
IUPredWS (Long)
JronnWS (J Ronn)

Features highlight disordered region or structured domain predictions

Use Threshold & Per-sequence option on ‘Colour by Annotation’ dialog to shade alignment using raw scores
Disorder in Interleukin 7 Orthologs
The Jalview developers
Michele Clamp
Harvard, USA.
James Cuff
Harvard, USA.
Steve Searle
Sanger, UK.
Andrew Waterhouse
Basel, Switzerland.

RNA Features
Lauren Lui
UC Santa Cruz, USA.
Jan Engelhardt
Univ. Leipzig, Germany.
Yann Ponty (VARNA)

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…