

DAS in Jalview 2: Past, present and future.



Jim Procter

jimp@compbio.dundee.ac.uk

Geoff Barton

jalview-discuss@jalview.org

<http://www.jalview.org>

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



10th March 2009

Jalview 2 and DAS

- What is Jalview
- DAS Capabilities in Jalview 2
- New features in Development
- The future...

Linked DNA/Protein Sequences

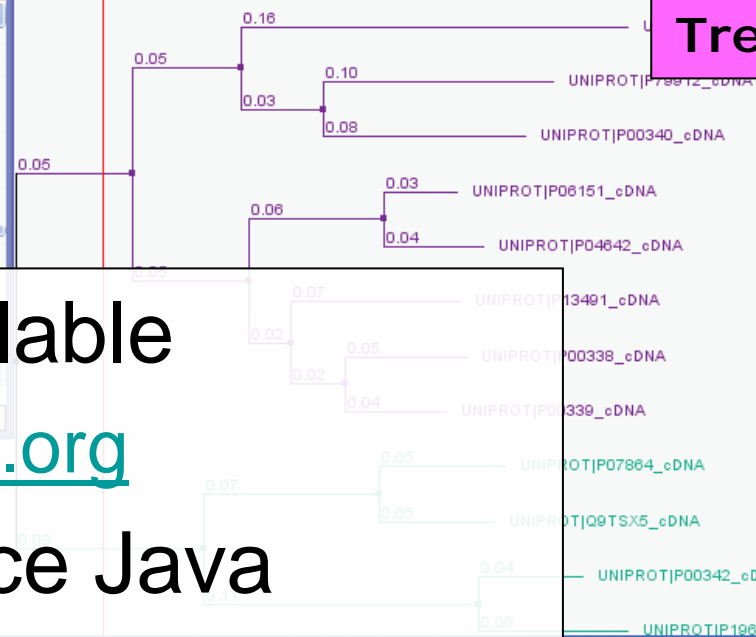
TCTCTGAAGA 637
TCCCTGAAGA 637
TCCCTGAAGA 637
TCCCTGAAGT 637
TCCCTGAAGT 637
AGTTTGCAA 640
UNIPROT|P19629_cDNA/1-999 606 TGGTGTAAACGTTGCTGCGTAACCTCTGAAGT 637
UNIPROT|P00340_cDNA/1-999 606 CGGAGTGAATGTTGCTGGTGTCTCCCTGAAGG 637
UNIPROT|P07864_cDNA/1-999 606 TGGGGTGAATGTTGCTGGTGTCTCTGAAGA 637
UNIPROT|Q9TSX5_cDNA/1-999 606 TGGAGTAAATGTTGCTGGTGTCTCTGAAGA 637
UNIPROT|P00342_cDNA/1-999 606 TGGTGTAAACGTTGCTGCGTAACCTCTGAAGT 637
UNIPROT|P19629_cDNA/1-999 606

Freely Available

www.jalview.org

- Open Source Java
- DAS Capable since 2006

Trees



Alignments

UNIPROT|P00339/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|P06151/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|P04642/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|Q9W7M6/1-333 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|P79912/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|P00340/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|P07864/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|Q9TSX5/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|P00342/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL
UNIPROT|P19629/1-332 DSSVPVWSGVNVAGVSLKLNHPDLGTDADKEQWKEVHKQVVDSSAYEVIKLGKGTWSAIGL

Features

Feature Settings		
<input checked="" type="checkbox"/> uniprot_aristotle	<input checked="" type="checkbox"/> nephos	<input type="checkbox"/> Phobius
<input checked="" type="checkbox"/> PDBsum_ligands	<input type="checkbox"/> netnes	<input type="checkbox"/> li10
<input type="checkbox"/> 91db	<input type="checkbox"/> 91dt	
Feature Type	Colour	Display
SO4 contact		<input checked="" type="checkbox"/>
non_covalent_binding_site		<input checked="" type="checkbox"/>
catalytic_residue		<input checked="" type="checkbox"/>
natural_variant_site		<input checked="" type="checkbox"/>

Annotation

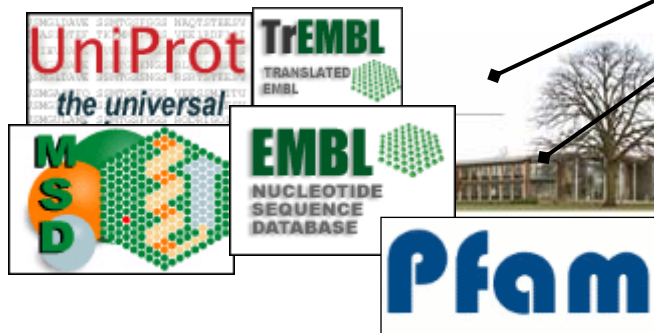
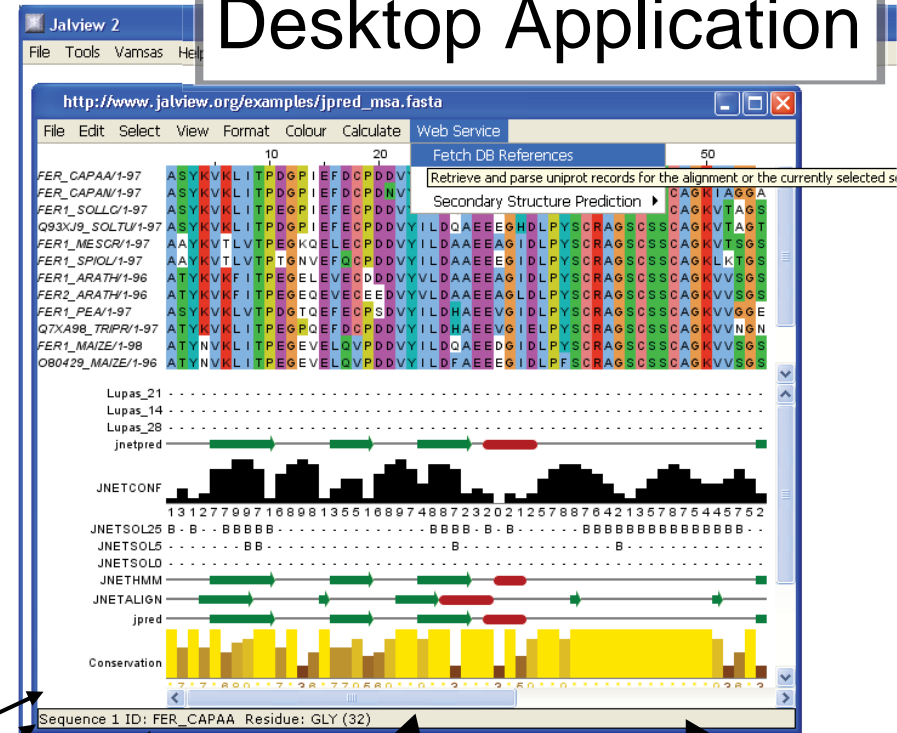


Jalview is available in two flavours...

Application can access:

- Web Services
 - Analysis
 - Databases
 - DAS
- Local System
 - Jalview Project Files
 - Figure Generation
 - EPS/PNG/HTML
 - Print System

Desktop Application



System
Clipboard

Local
File
system

HTTP
Distributed
Annotation
System



a multiple alignment editor

~~NEWS: 19th January 2009 - Jalview 2 Application Note published in Bioinformatics~~

Overview

Download

Applet Version

Screenshots

FAO

Documentation

Release history

Source Code

Links

News Mailing List

Discussion Mailing List

Please send problems
and bug reports to the discussion
list.

Jalview 2.4 Manual and Tutorial

Jalview is a multiple alignment editor written in Java. It is used widely in a variety of web pages (e.g. the EBI Clustalw server and the Pfam protein domain database) but is available

JahView 2 development was funded from 2004 to 2007 by the BBSRC as part of the "VAMSAS" Project. This project is coordinated by Geoff Barton at the University of Dundee with partners at EBI and BioSS and consultancy (blessing :-) from Michele Clamp; the originator of

jalview

devel

Authors: [David Martin](#); [Geoff Barton](#)
Version 1+ [Michele Clamp](#); [James Cuff](#); [Stephen Searle](#); [Geoff Barton](#)

Link to development version
on release history page

Thanks to Andreas Prlic for code and suggestions for DAS feature capabilities and Benjamin Schuster-Böckler for his Stockholm parsing code, both from the Wellcome Trust Sanger Institute, Cambridge.

If you use Jalview in your work, please quote this publication:

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

Please send all bug reports and help requests to the [jalview-discuss mailing list](mailto:jalview-discuss@lists.sourceforge.net)



Jalview version history



Jalview

a multiple alignment editor

Home

Overview

Download

Applet Version

Screenshots

FAQ

Documentation

Release history

Source Code

Links

News Mailing List

Discussion Mailing List

Please send problems
and bug reports to the discussion
list.



The original version of Jalview which was designed to run under Java Version 1.8 will remain available for users who are not able to obtain or install Java 1.4+ on their system. You can download it [here](#).

The [download page](#) and [source code page](#) on this website will always display the latest stable version of Jalview.

See the [Jalview Version Archive](#) for links to [older versions](#).

Link to development version

Release	New Features	Issues Resolved
	User Interface <ul style="list-style-type: none">• Linked highlighting of codon and amino acid from translation and protein products• Linked highlighting of structure associated with residue mapping to codon position• Sequence Fetcher provides example accession numbers and 'clear' button• MemoryMonitor added as an option under Desktop's Tools menu• Extract score function to parse whitespace separated numeric data in description line• Column labels in alignment annotation can be centred.• Tooltip for sequence associated annotation give name of sequence Web Services and URL fetching	<ul style="list-style-type: none">• selected region output includes visible annotations (for certain formats)• edit label/displaychar contains existing label/char for editing• update PDBEntries when DBRefEntries change (vamsas)• shorter peptide product names from EMBL records• Newick string generator makes compact representations• bootstrap values parsed correctly for tree files with comments

<http://www.jalview.org/versions.html>



Jalview

a multiple alignment editor

- Home
- Overview
- Download
- Applet Version
- Screenshots
- FAQ
- Documentation
- Release history
- Source Code
- Links
- News Mailing List
- Discussion Mailing List

*Please send problems
and bug reports to the discussion
list.*



Jalview Version Archive

Full information on the changes for each jalview 2 release is given on the [Jalview 2 release history page](#) .
Below is a table of links to different versions of Jalview 2.

The latest stable release is always given on the [Jalview download page](#) and [JalviewLite applets page](#) , as well as that version's [online documentation](#) .

Development Versions

Build Type	Webstart	InstallAnywhere	Applet	Source
Latest CVS Build This version comes with no guarantees - back up your data!	Webstart CVS Build	InstallAnywhere CVS Build	Applet CVS Build	Source Tarball for CVS Build
Latest Build of Current Release Branch The latest and greatest version of the current Jalview release (Jalview 2.4). Any bug fixes will be tested in this version before they are incorporated in a new release.	Webstart Release Build	InstallAnywhere Release Build	Applet Release Build	Source Tarball for Release Build

Previous Jalview Releases

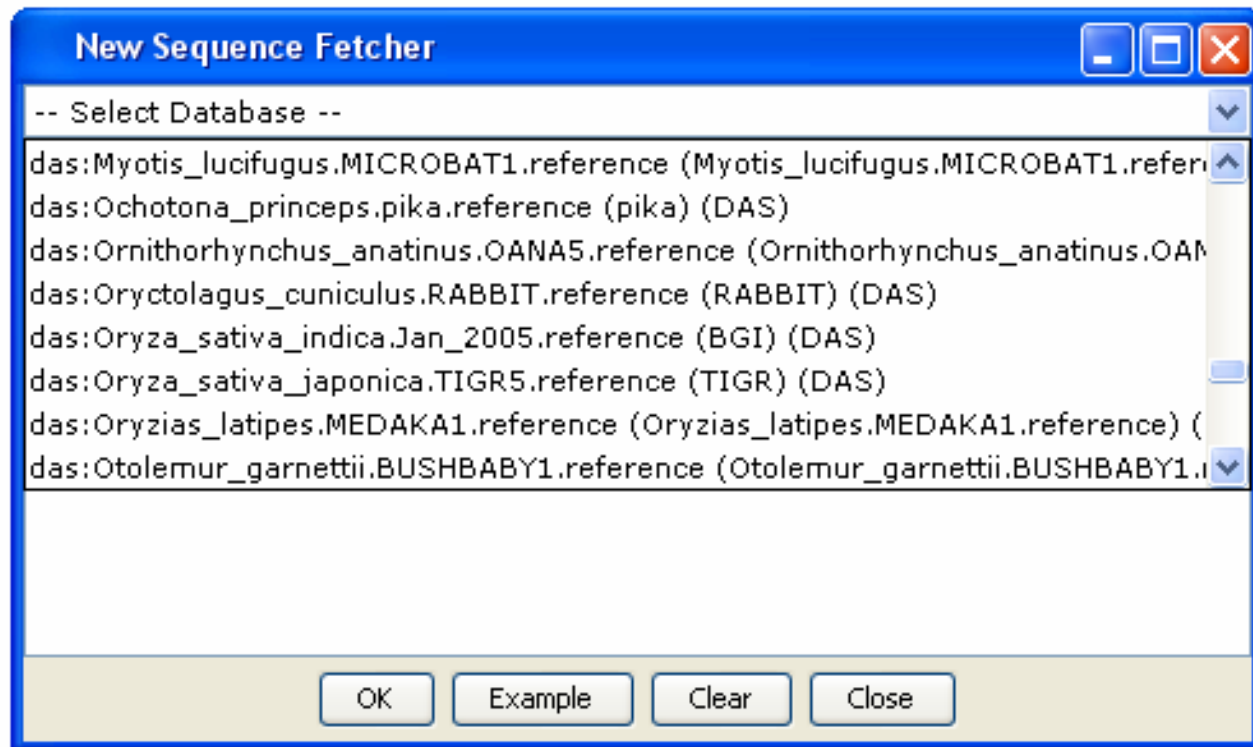
Jalview 2.3	Webstart Jalview 2.3	InstallAnywhere Jalview 2.3	Jalview 2.3 Applet	Source Tarball for Jalview 2.3
-----------------------------	--------------------------------------	---	------------------------------------	--

Jalview 2.4

- Released in September 2008
- Bioinformatics App Note – January 2009
- Final BBSRC supported Jalview 2 release
- Unveiled for EBI-ENFIN Workshop
 - [Protein function prediction tools](#)
 - Manual and exercises available on Jalview website
- 2.4 DAS features described last year...

DAS Sequence Retrieval

- Extended existing Sequence fetcher



Sequence Feature Settings

Feature Settings

DAS Settings

☒ EMBL ☒ uniprot ☒ PDBsum_ligands

Feature Type	Colour	Display
cleaved_initiator_methionine		<input checked="" type="checkbox"/>
polypeptide_region		<input checked="" type="checkbox"/>
natural_variant_site		<input type="checkbox"/>
glycosylated residue		<input checked="" type="checkbox"/>
PO4 contact		<input checked="" type="checkbox"/>
OXY contact		<input checked="" type="checkbox"/>
HEM contact		<input checked="" type="checkbox"/>
metal_binding		<input checked="" type="checkbox"/>
turn		<input checked="" type="checkbox"/>
alpha_helix		<input checked="" type="checkbox"/>
mature_protein_region		<input checked="" type="checkbox"/>

Invert Selection

Optimise Order

OK

Cancel

Load Colours

Save Colours

- DAS features retrieved for Uniprot ID
- Locally mapped to EMBL cDNA annotated region

```

EMBL|V00488/1-1138 487 a c c c c t g t c c c c t g a g c a c c t g c a c g c g c a a a g c t t c g g t t g a c c c g g 540
EMBL|V00488/1-1138 541 t c a a c t t c a a g g t g a g c g g c g g g c c g g g a g c g a t c t g g g t c g a g g g g c g a g a t g 594
EMBL|V00488/1-1138 595 g c g c c t t c c t c t c a g g g c a g a g g a t c a c g c g g g t t g c g g g a g g t g t a g c g c a g g 648
EMBL|V00488/1-1138 649 c g g c g g c g c g g c t t g g g c c g c a c t g a c c c t c t t c t c t g c a c a g c t c c t a a g c c a 702
EMBL|V00488/1-1138 703 c t g c c t g c t g g t g a c c c t g g c c g c c a c c t c c c c g c c g a g t t c a c c c c t g e g g t 756
EMBL|V00488/1-1138 757 g c a c g c t t c c c t g g a c a a g t t c c t g g c t t c t g t g a g c a c c g t g c t g a c c t c c a a 810
EMBL|V00488/1-1138 811 a t a c c g t t a a g c t g g a g c c t c g g t a g c c g t t c c t c c t g c c c g c t g g g c c t c c c a 864
EMBL|V00488/1-1138 865 a c g g g c c c t c c t c c c c t c c t t g c a c c g g c c c t t c c t g g t c t t t g a a t a a a g t c t 918

```

Sequence 1 ID: EMBL|V00488 Nucleotide: Adenine (25)








- Mouseover for more info
 - Feature description
 - URL Links

Sequence Feature Settings

Feature Settings

DAS Settings

☐ uniprot aristotle☐ Uniprot☒ chs_total☒ PDBsum_protprot

Feature Type	Colour	Display
PHOSPHORYLATION (Y)		<input checked="" type="checkbox"/>
PHOSPHORYLATION (S)		<input checked="" type="checkbox"/>
Protein-protein contact		<input checked="" type="checkbox"/>
PHOSPHORYLATION (T)		<input checked="" type="checkbox"/>
NES-SIGNAL		<input checked="" type="checkbox"/>
ACETYLATION		<input checked="" type="checkbox"/>
PROPEPTIDE		<input checked="" type="checkbox"/>

TFEFP - - PNL - P+SVTLQPGPEDTGKPC - GVDYE I
 YFTIP - - QNL - CSVTLQPGPEDTGKAC - GVDFE I
 TFHFP - - PNS - PSSVTLLQEGDLSKPL - GVEYT I
 VLEFP - - DFL - PCSVSLQAPSDVGKAC - GVDFE I
 TFALP - - ENA - PPSVTLLQPGSEDQGRPL - GVEYEL
 TFHFP - - ASS - PSSVTLLQPGDDQDGKPL - GVEYSV
 LLTFP - - DYL - PCSVMLQAPQDVGKSC - GVDFE I

Protein-protein contact 137; 1AYR: chain C to B 137 189 190 191 192 193 194 195 196 197 344 [?Link](#)

Protein-protein contact 137; 1CF1: chain B to C 137 186 189 190 191 192 193 194 195 197 235 334 339 340 341 342 343 344 345 [?Link](#)

Protein-protein contact 137; 1CF1: chain D to A 137 138 186 189 190 191 192 193 194 195 197 235 334 339 340 342 343 344 345 [?Link](#)

Protein-protein contact 137; 1AYR: chain A to D 137 189 190 191 192 193 194 195 196 197 199 [?Link](#)

PROPEPTIDE 1 286; Furin-type cleavage site prediction (arginine/lysine residues). [?Link](#)

PHOSPHORYLATION (T) 137; High confidence server. Only hits with scores over 0.8 are reported. [?Link](#)

Invert Selection

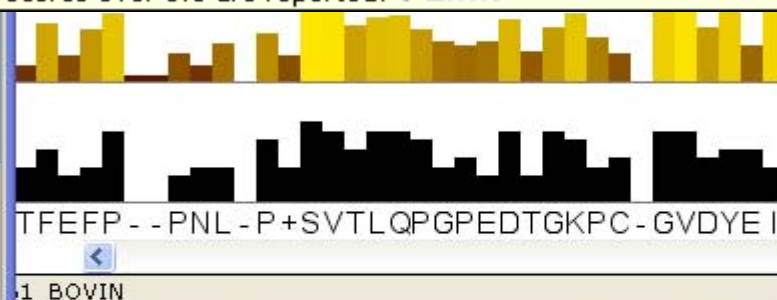
Optimise Order

OK

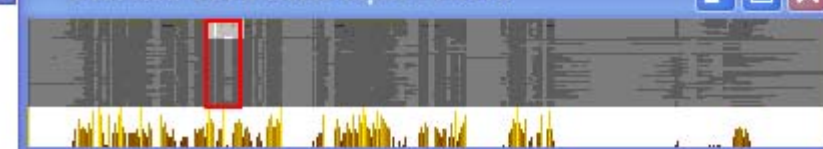
Cancel

Load Colours

Save Colours



Overview Cut & Paste input - FASTA



Jalview 2.4 DAS Functionality

DAS ANNOTATION SERVERS

Sequence Feature Settings (Left Window)

Feature Settings | DAS Settings

Nickname | Use S... | <http://www.ebi.ac.uk/das-srv/proteinanda>

uniprot ar... | ☒ | Admin Email: tim.massingham@ebi.ac

cbs_total | ☒ | Registered at: Sun Feb 24 20:22:42 Gf

CSA - ext... | ☒ | Last successful test: Sun Feb 24 20:22

PDBsum... | ☒ | Labels:

Filter on:

- Authority
- Coordinate System
- Labels

Use Registry <http://www.dasregistry.org/das1/sources/> [Reset]

[Refresh Available Sources] [Add Local Source]

Sequence Feature Settings (Right Window)

Feature Settings | DAS Settings

☐ uniprot aristotle | ☐ Uniprot

☒ cbs_total | ☒ PDBsum_protprot

Feature Type	Colour	Display
PHOSPHORYLATION (Y)	[Purple]	<input checked="" type="checkbox"/>
PHOSPHORYLATION (S)	[Purple]	<input checked="" type="checkbox"/>
Protein-protein contact	[Purple]	<input checked="" type="checkbox"/>
PHOSPHORYLATION (T)	[Purple]	<input checked="" type="checkbox"/>
MES-SIGNAL	[Blue]	<input checked="" type="checkbox"/>
ACETYLATION	[Blue]	<input checked="" type="checkbox"/>
PROPEPTIDE	[Yellow]	<input checked="" type="checkbox"/>
OUTSIDE	[Green]	<input checked="" type="checkbox"/>

[Invert Selection] [Optimise Order]

[OK] [Cancel] [Load Colours] [Save Colours]

Callout Boxes:

- Discover PDB/Uniprot/EMBL CS
- Match IDs
- Correct reference frame
- Query with raw ID or accession
- Mouseover for labels and scores
- Menu to access links

UniProt/Swiss-Prot|P17870|ITPTFEIP... - PCSVTLQGFEDLQKAC - GVDYEL

UniProt/Swiss-Prot|P51485|FPFWFEVA... - ASSVTLQBARGDQKRC - GVDYEL

UniProt/Swiss-Prot|P51466|HPFYFTIP... - QNL - PCSVTLQGFEDLQKAC - GVDYEL

UniProt/Swiss-Prot|P51467|HPFTTFHP... - PNS - PSSVTLQQEGDDLQKPL - GVEYTI

UniProt/Swiss-Prot|P51477|FPFVLEFP... - DFL - PCSVSLQAPSDVQKAC - GVDFEI

UniProt/Swiss-Prot|P32122|EPFTEHFP... - ENA - PPSVTLQPGSEDDQGRPL - GVEYEL

UniProt/Swiss-Prot|P08168|YPFLLTFP... - DYI - PCSVMLQAPQDVQKSC - GVDFEI

UniProt/Swiss-Prot|P53179|FPFRIRIPRGRGM - LSSI - - - - - KFERG - SITYFL

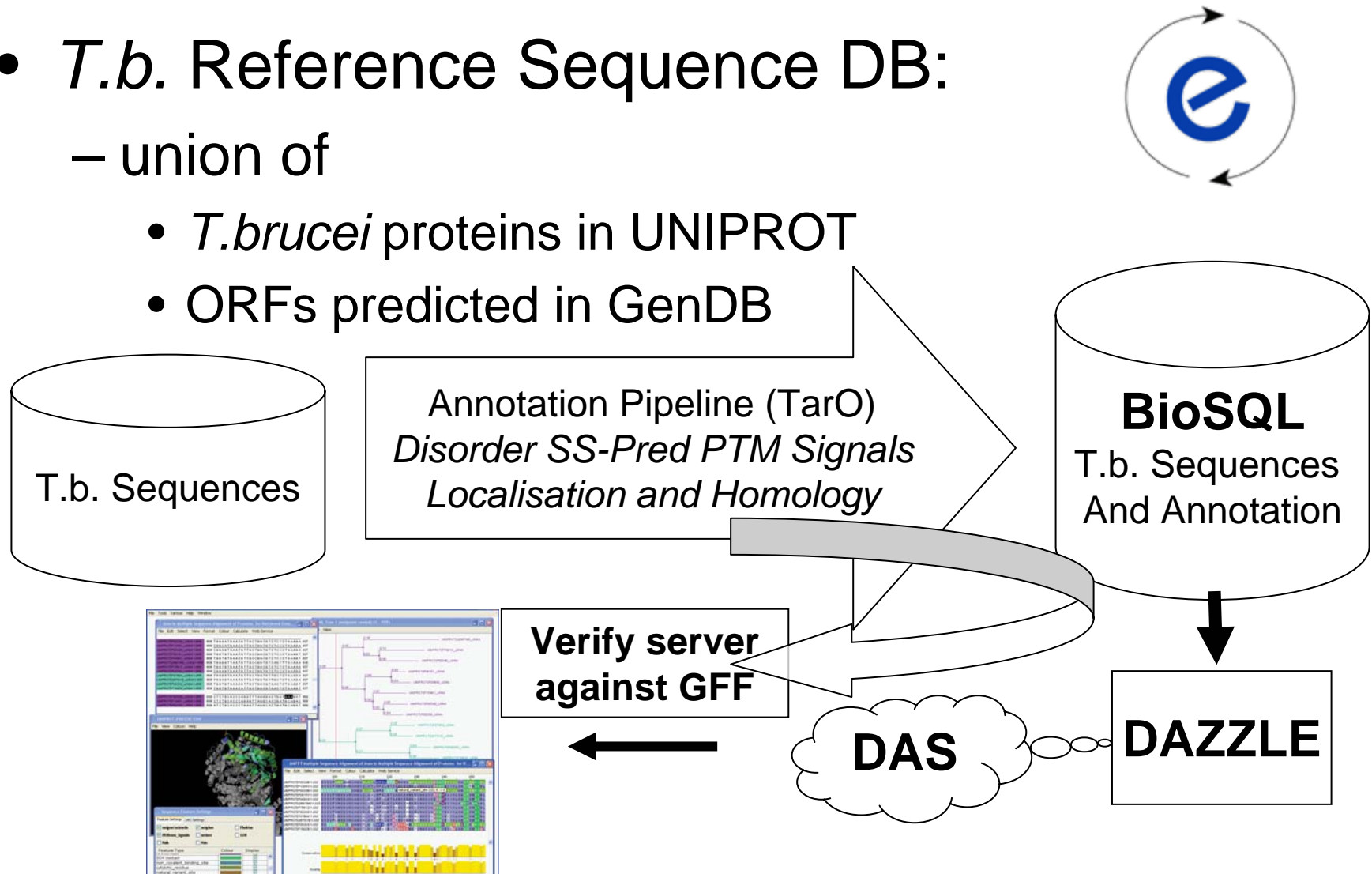
Jalview Development Version

- Unstable build of applet and Application
 - Distinct from 2.4 Maintenance Build
- Includes New Functionality from
 - Community
 - New Unconserved Residue Display
 - Olivier Martin, Uni. Lausanne
 - Jim's Personal Need
 - Side-effect of Wet/Dry Experimental Collaboration
 - ENFIN
 - » European Network of excellence for Functional INtegration

My Day Job:

T. Brucei in silico annotation

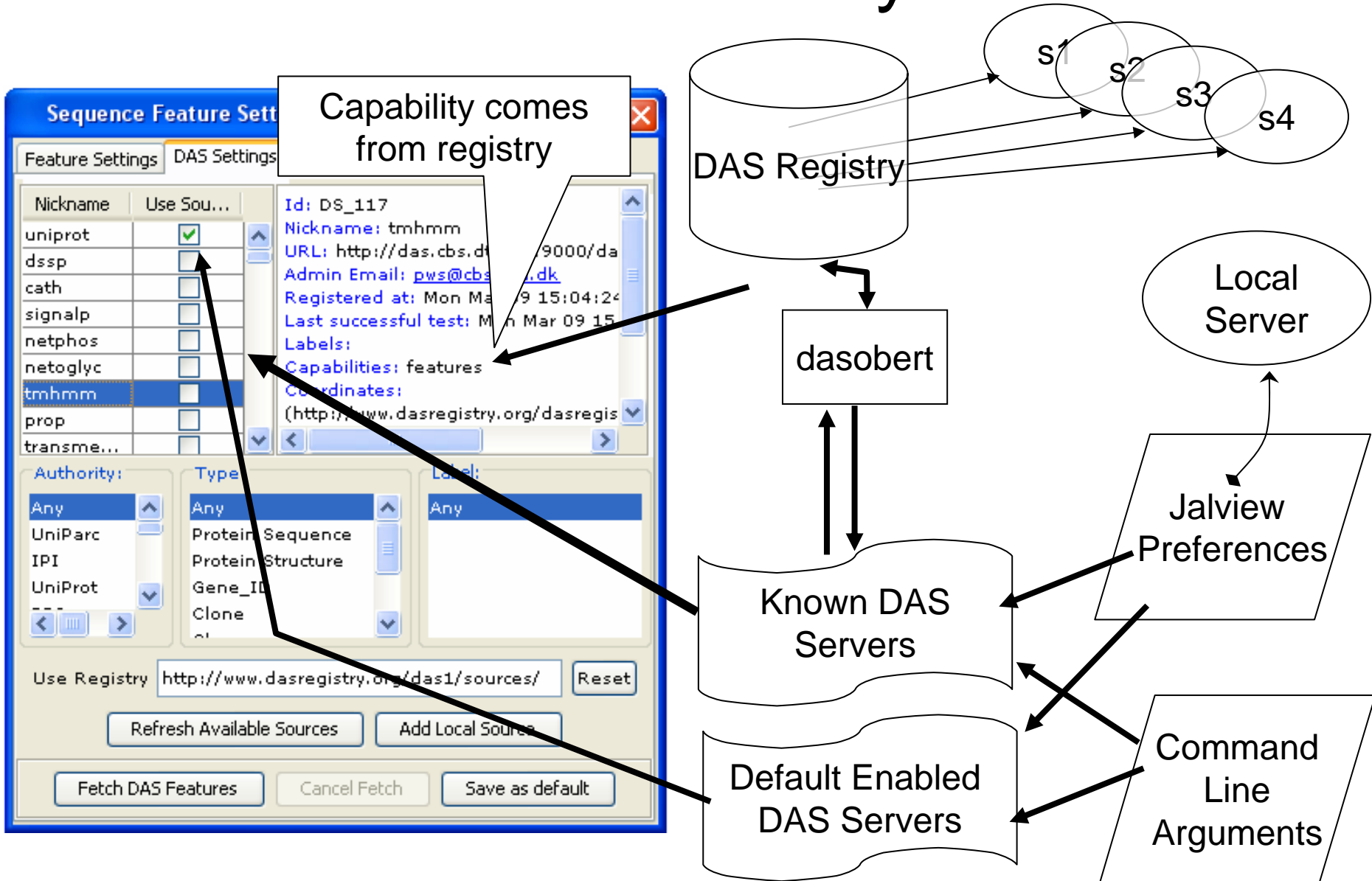
- *T.b.* Reference Sequence DB:
 - union of
 - *T.brucei* proteins in UNIPROT
 - ORFs predicted in GenDB



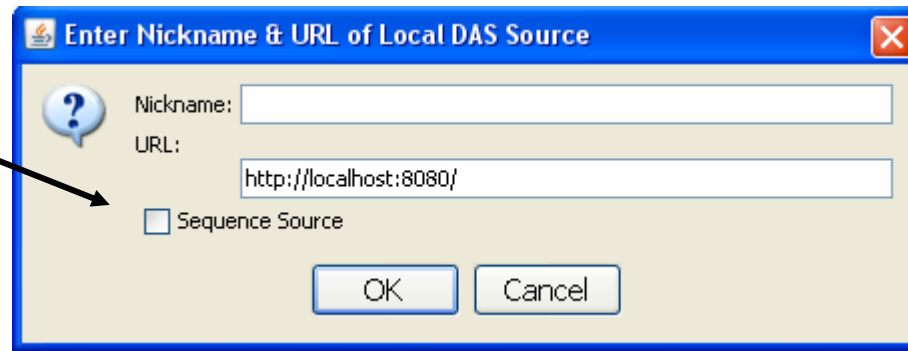
What's new ?

- Sequence Retrieval
 - Support for local sequence sources
- DAS Feature processing semantics
 - Database References
 - Protein Names
- DAS Feature Scores
 - Import/Display
- Feature analysis
 - New sorting functions
 - feature density and (average) score
 - Thresholding and colouring

DAS Source Discovery in Jalview



Local Sequence Source Switch



- Alternative: Could use Sources* command ?
 - Not guaranteed to be faithfully supported by all servers

* Thanks to Andrew J. For the correction!

DAS Feature type parsing: More friendly sequence retrieval

- protein_name
 - Set protein description line (if not set)
- dbxref / dbref
 - Create new accession cross references
 - Enables validation/annotation in other coord. systems

Primary and Secondary References (with hrefs)

atb2268/1-818

hypothetical protein, conserved [1363639 - 1366095] MW:88633.76

TBDB atb2268

GeneDB Tb927.5.4580

GeneDB Tb05.26C7.220

[Link](#)

[Link](#)

Note on source consistency

Further Text->Annotation should be done server

Note on source consistency:
Further Text->Annotation
should be done server side?

Feature Creep and Usability : Validation of sequences against DAS Reference Sources

Reference Sequence Validation



1. Query DB with words from sequence ID
2. Match any returned sequence to local version
3. Add new database reference with mapping between local and external coordinate system

DAS Reference Source Querying can take a looong time!

- If Jalview tried to verify against all sequence sources

Query 50+ sources with average of 3 IDs on each of N sequences

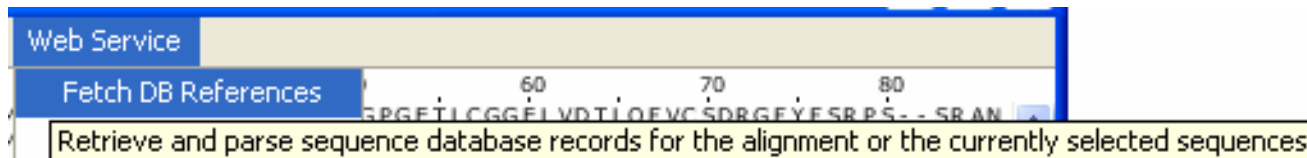
Two Seconds per source per query $\Rightarrow N \times 6$ seconds for each source

Pragmatic Solution:

\Rightarrow Limit to currently selected sources or a specifically chosen source

Leads to an Ugly UI hack

- 2.4 Fetch DB Refs Menu



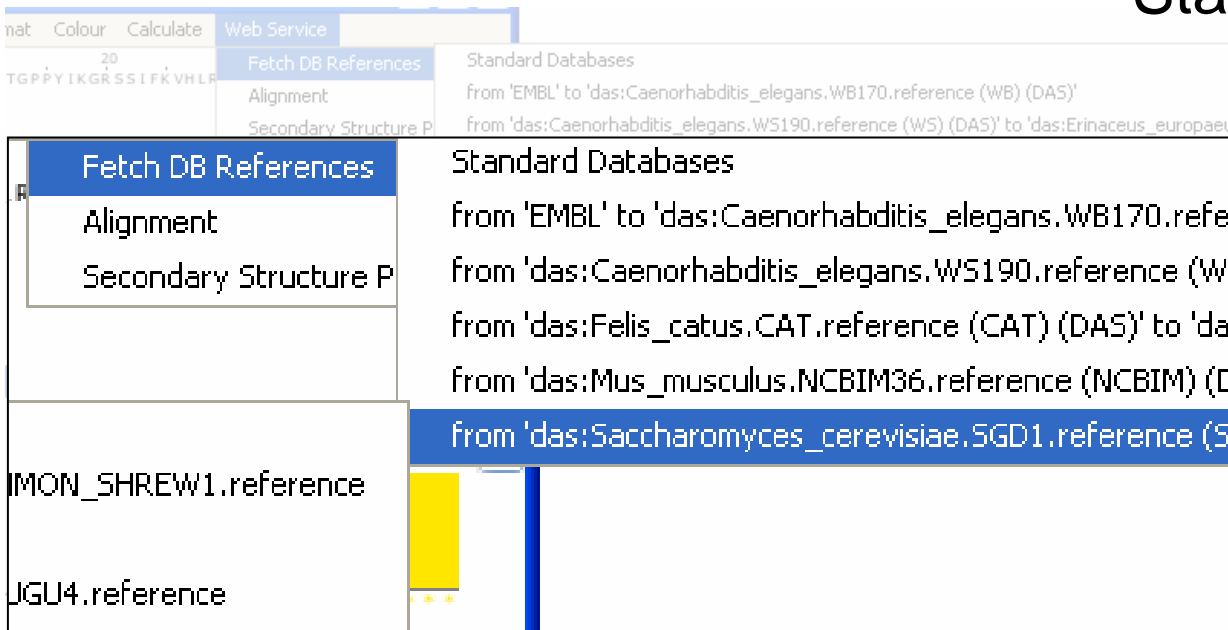
Development Version

– Standard Databases

– Currently selected sources

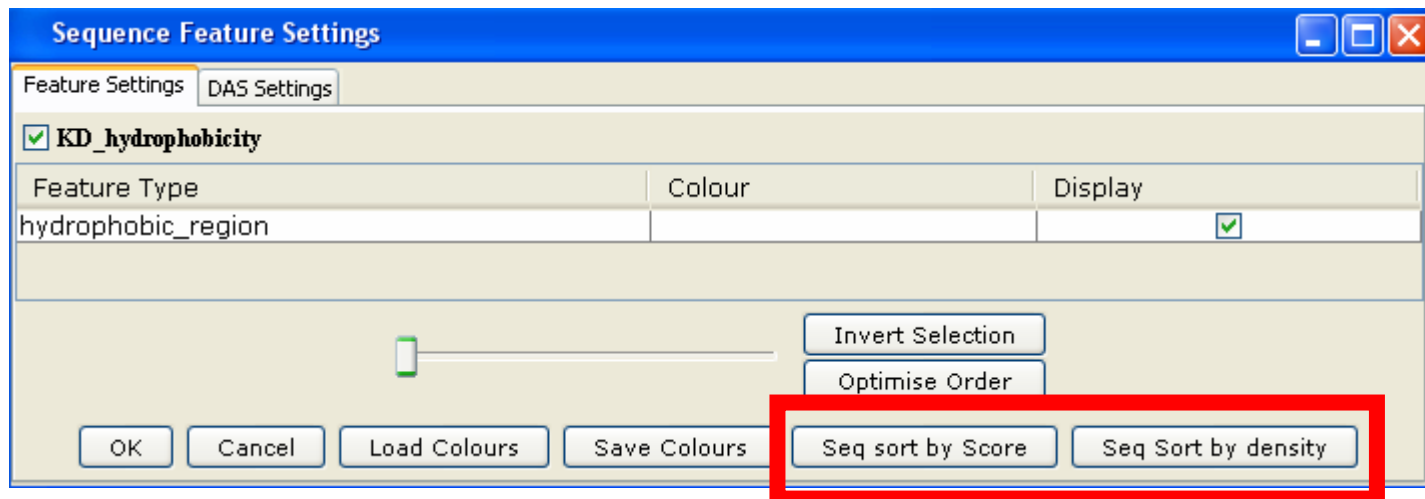
OR

Verify against individual source



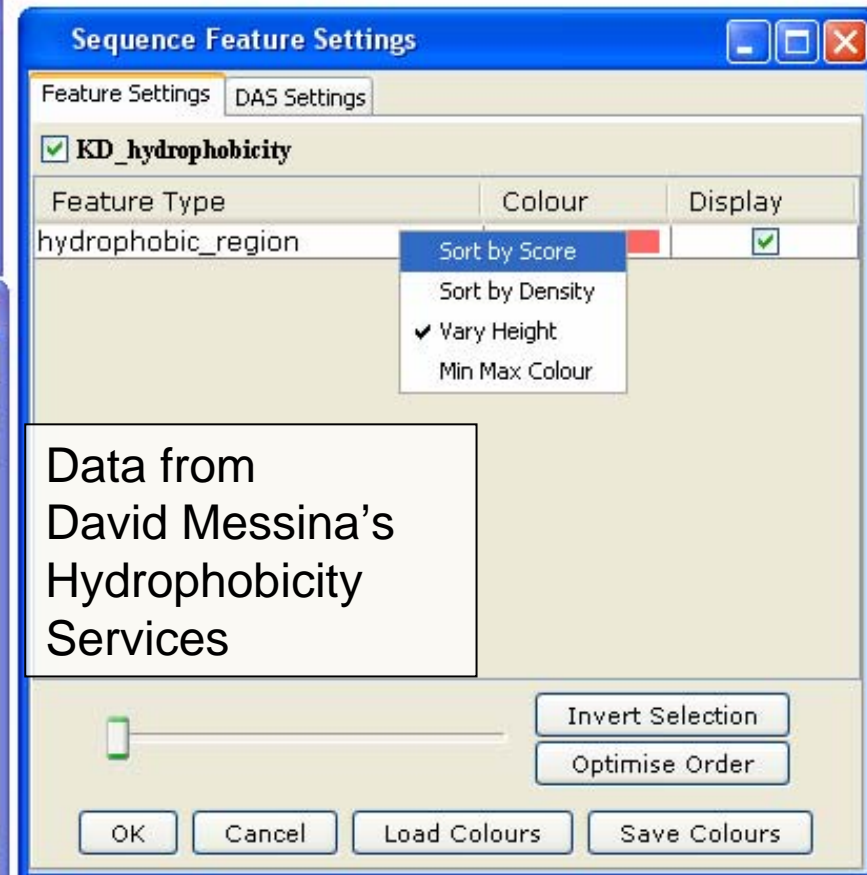
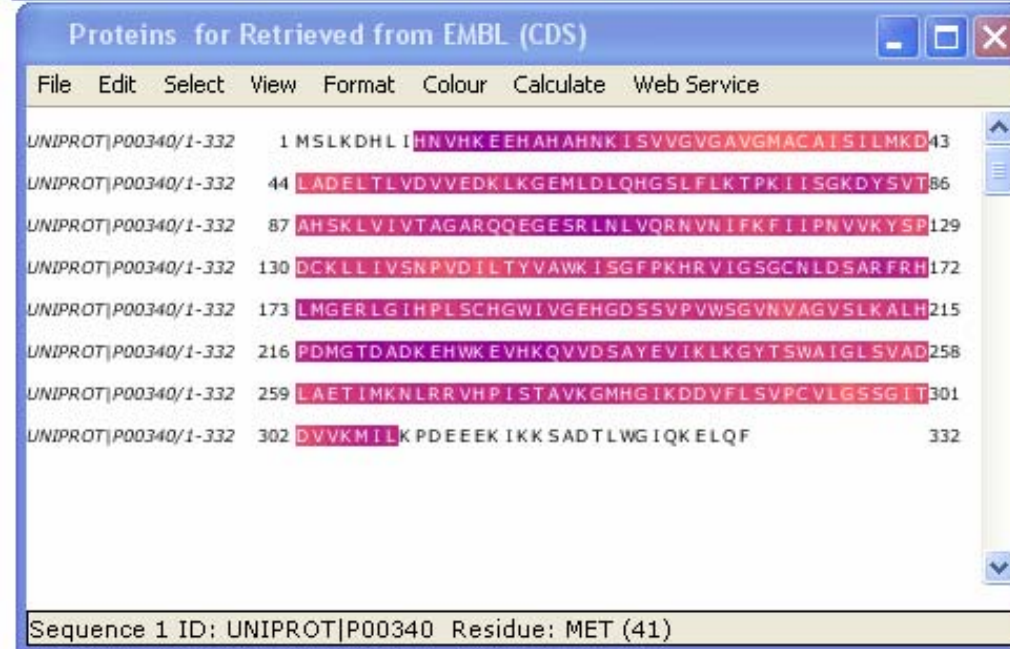
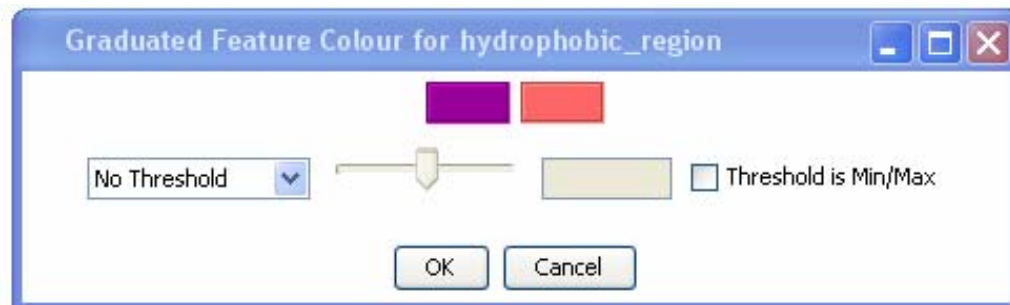
Sorting Alignments with Features

- Two New Buttons
 - Only visible when dialog is made extra wide (currently)
 - Reorder current selection
 - Total Number or average score
 - Only counts displayed features in selected region



Sorting and Colouring on a specific type

- New Pop-up Menu for each Feature Type
 - Set Min-Max colourscheme
 - Sort on specific type



Data from
David Messina's
Hydrophobicity
Services

Jalview Alumni 1997-2008



Michele Clamp
Broad Institute, USA



Geoff Barton
Dundee, Scotland.

James Cuff
Harvard, USA



Andrew Waterhouse
Riken, Japan



David Martin



Steve Searle
Sanger, UK

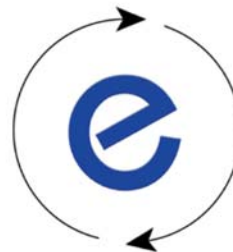


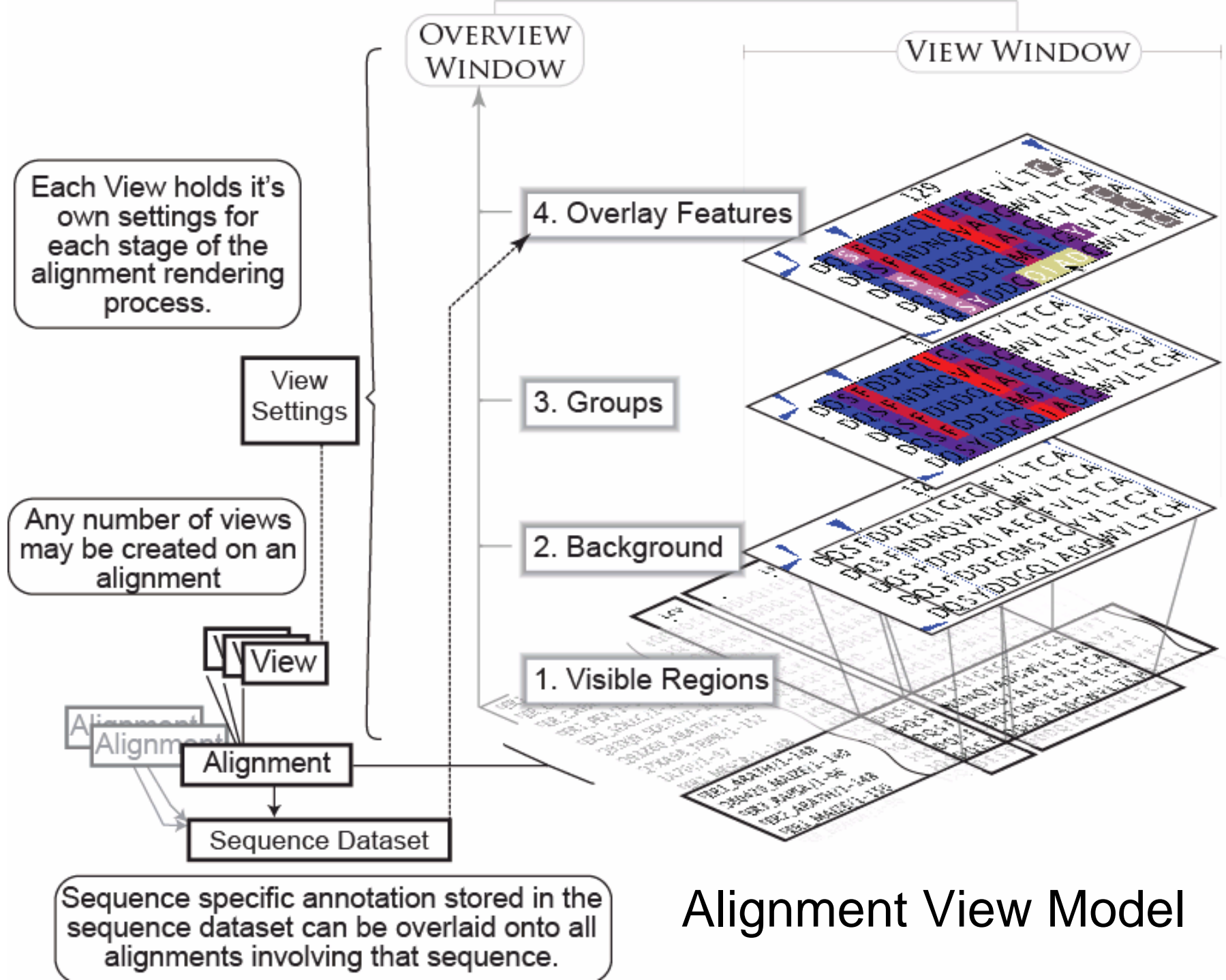
Jim Procter



Acknowledgements

- Jalview 2
 - Andrew Waterhouse
 - David Martin
 - Geoff Barton
- *T.b. Proteome Work*
 - Ian Overton (TarO)
 - David Martin (TBDB)
 - Joao Rodrigues
 - Luis Izquierdo
 - Mike Ferguson



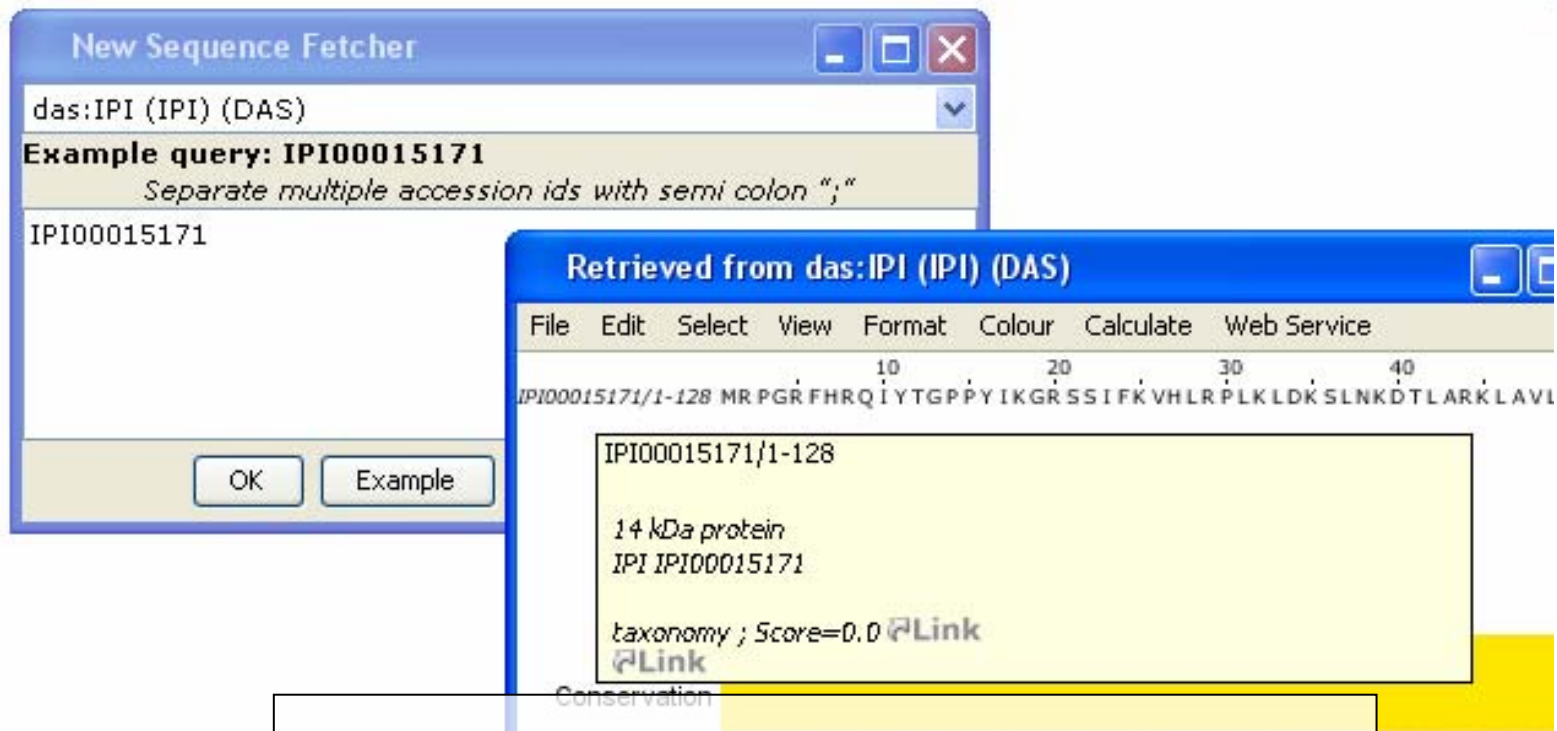


Alignment View Model

Cracks in the DAS 1 Pavement

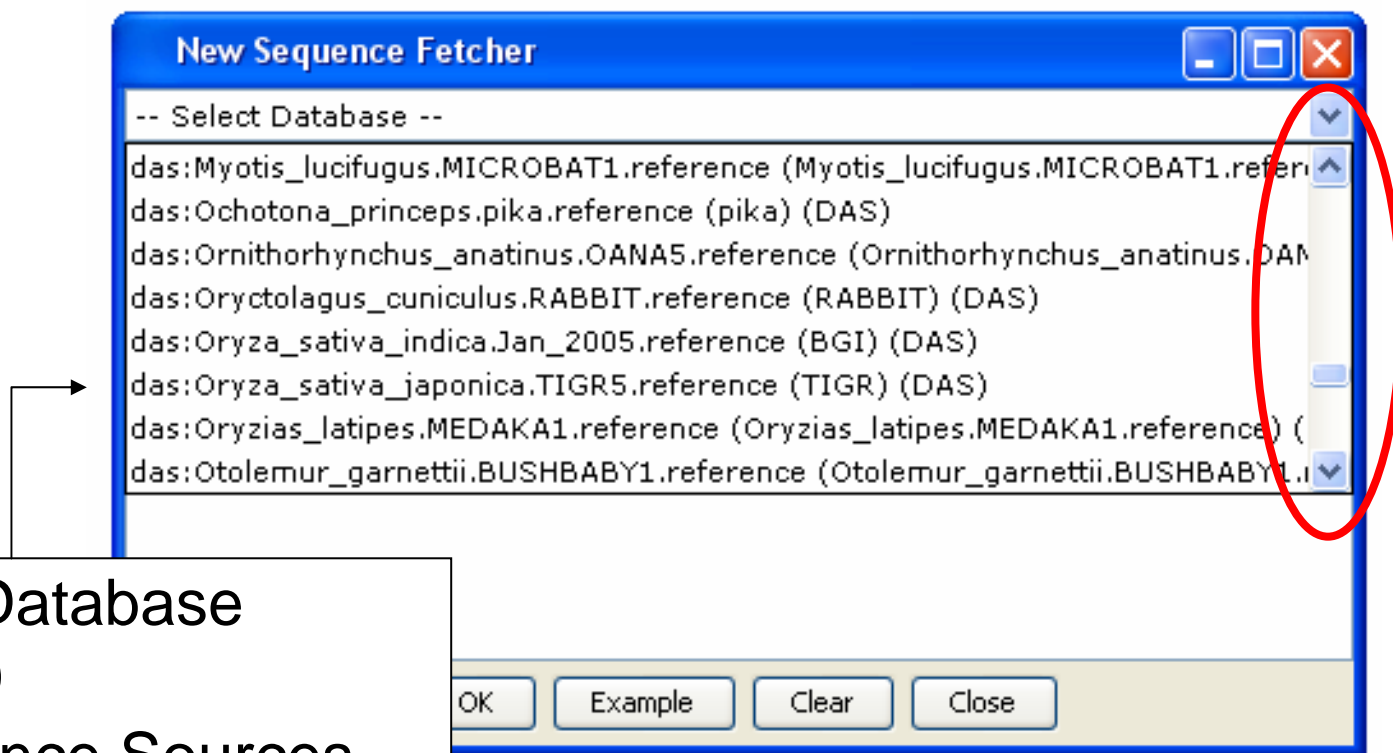
- Consistent use of Ontologies
- Consistency in Feature Metadata
 - Linkouts usually have a name
- DAS Example ID
 - Useful for users as well as testing
 - Possible extension of DAS capabilities response ?
 - Server provides an example ID

Utility of Example Query IDs



Choose an interesting Query ID ?
Provide Cross References where possible ?

Source Navigation: Usability Issues



- Direct EBI Database Services (4)
- DAS Sequence Sources from Registry (50)
- Local Sources