Otterlace, Zmap, Blixem and Dotter user manual

5th January 2010

Blank

Contents

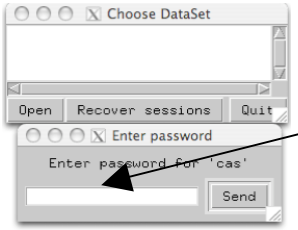
Blank

Otterlace

Otterlace is an interactive, graphical client, which uses a local acedb database with Zmap and perl/Tk tools to curate genomic annotation. Annotation is stored in an extended Ensembl schema (the "otter" database), which presents the annotator with contiguous regions of a chromosome. The acedb database provides local persistent storage, so that if the software or desktop machine crashes, reboots or is exited, the editing session can be recovered. Since all communication goes through the Sanger web server, annotators can work wherever there is a network connection.

Starting an Otterlace Session

Type: **otterlace &** in a terminal window. If you are using Mac OS X, double-click on the otterlace icon. You will be required to authorise your session by entering your password. If you experience any problems, email anacode@sanger.ac.uk

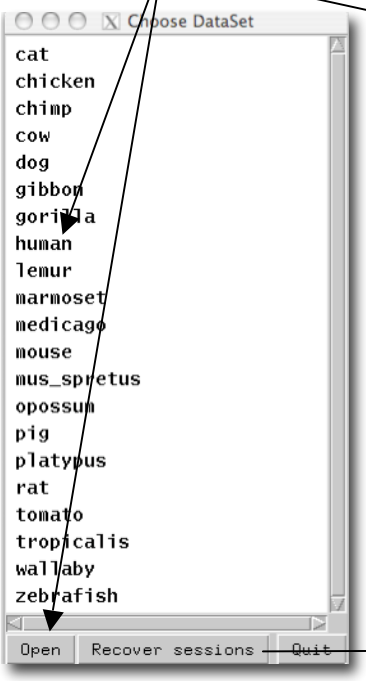


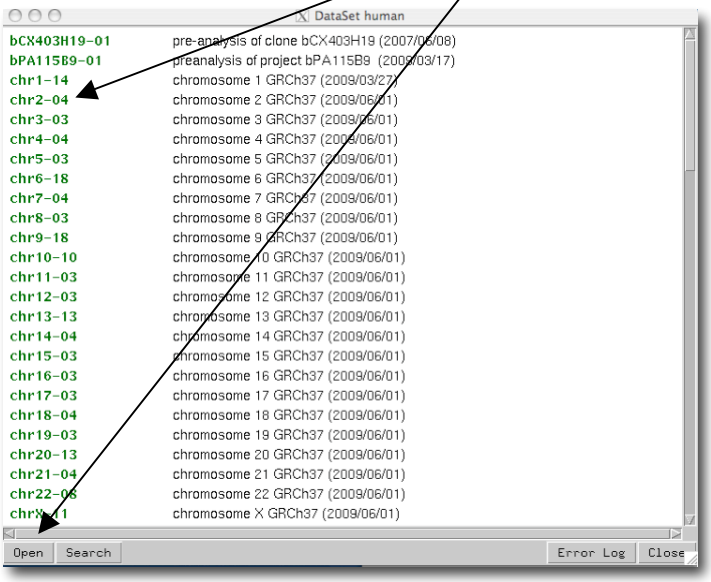
1) Enter your password in the box and click on **Send**.

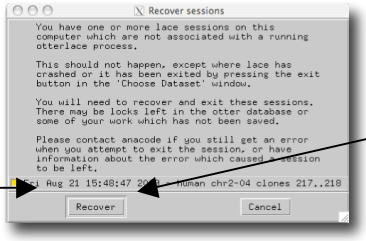
DataSet chooser

3) Select the dataset using left click. Then click on **Open**.

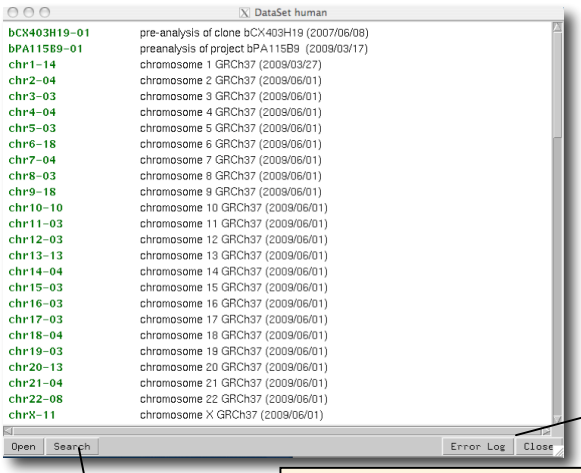
2) Select the species using left click and click on Open or just double click. This will open the **DataSet** window.



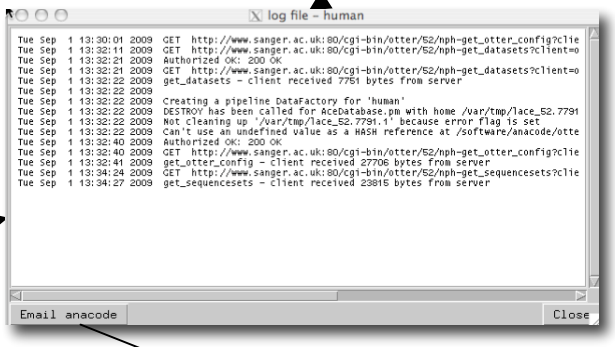




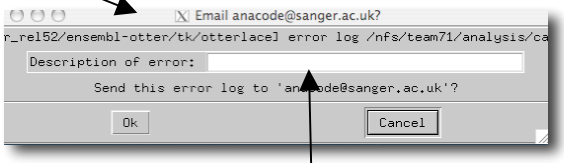
This allows you to recover sessions that have crashed, or when lace has been exited by pressing the **Quit** button in the **Choose Dataset** window. This window will appear automatically when opening a new otterlace session and previous sessions are still present.

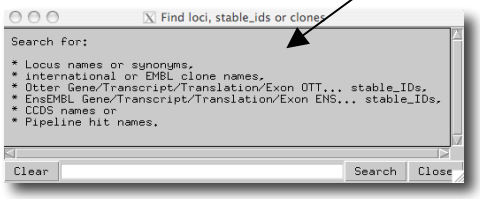


Otterlace software and/or database problems are shown in the **Error Log**.

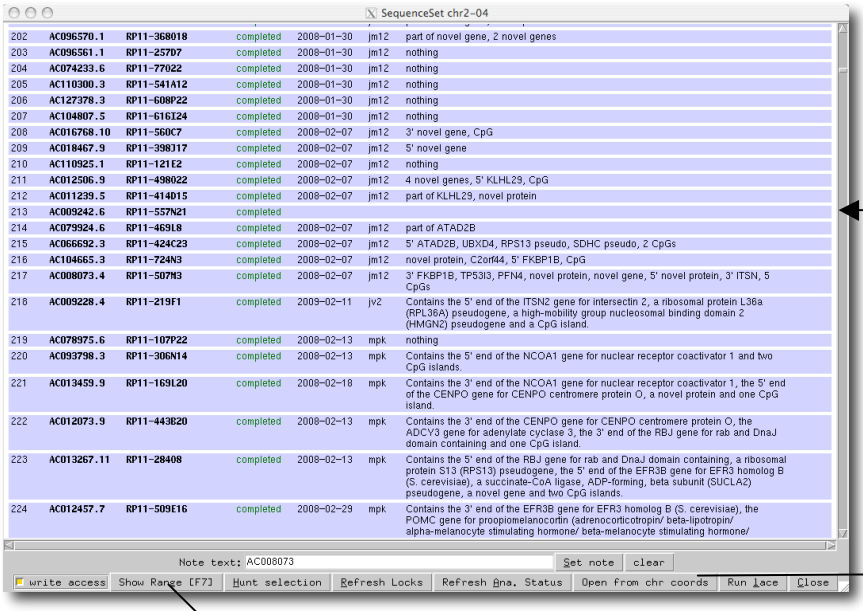


The **Search** feature allows you to search the Dataset for any feature such as Otter ID, gene name etc.

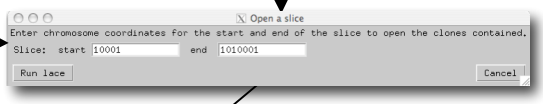


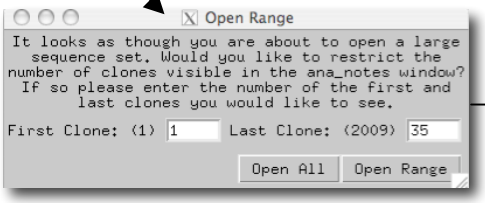


An option to email anacode with the errors is provided to facilitate a diagnosis. Always include a "useful" description in the email!

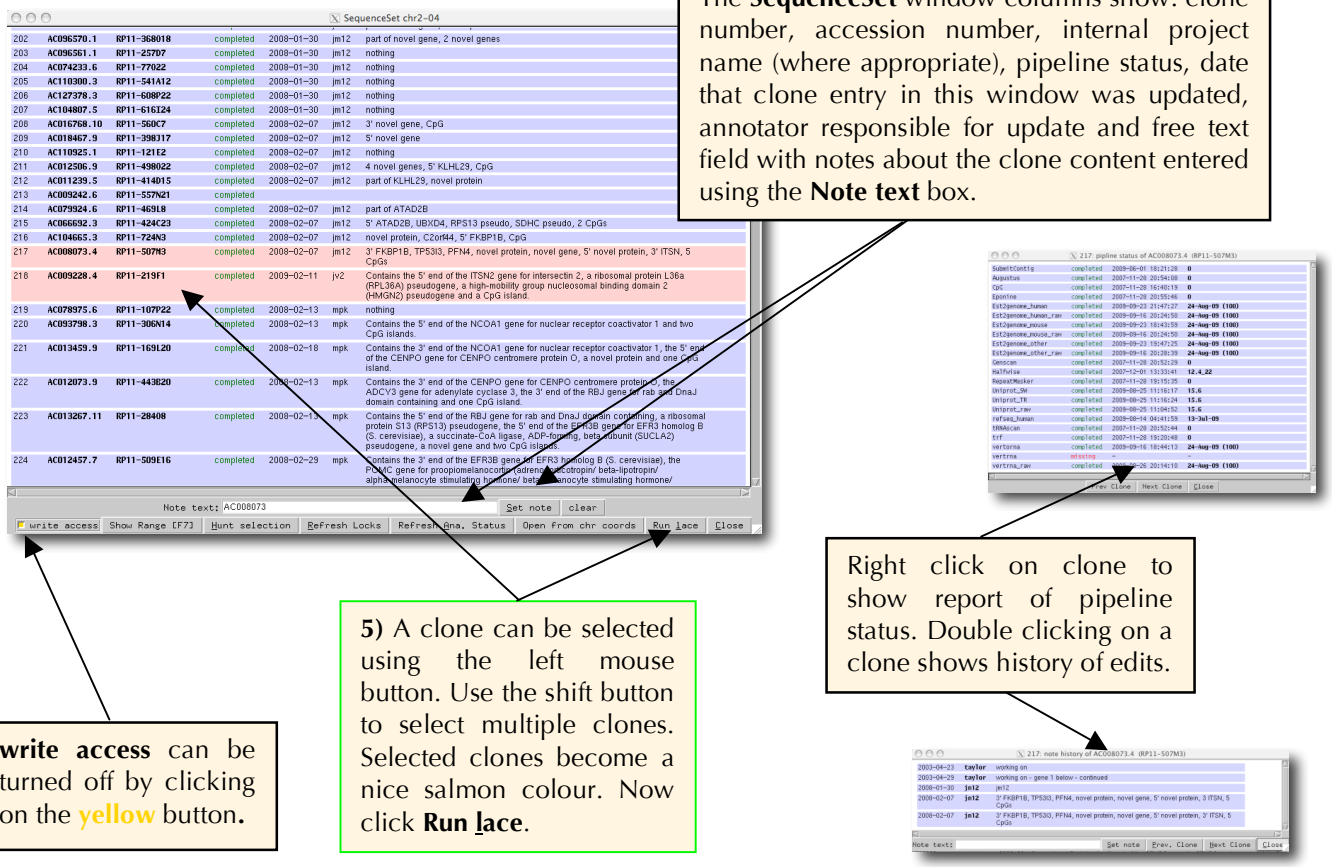


4) The **SequenceSet** window appears, (also known as **Ana_notes)**. It shows remarks that can be added using the entry field at the bottom to help track annotation progress. This window also allows you to either open the whole contig range in one scrollable window or open a selected range of your choice.





These options allow you to open specific regions and are designed to make opening clones quicker.



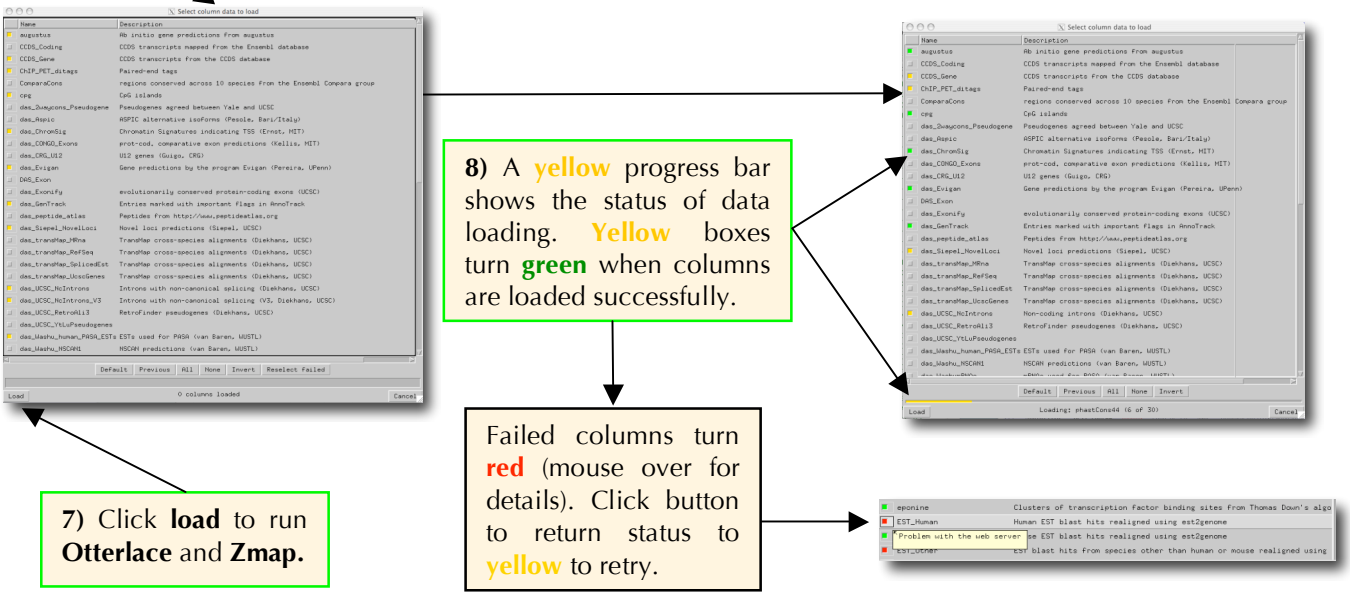
The **SequenceSet** window columns show: clone number, accession number, internal project name (where appropriate), pipeline status, date that clone entry in this window was updated, annotator responsible for update and free text field with notes about the clone content entered using the **Note text** box.

Right click on clone to show report of pipeline status. Double clicking on a clone shows history of edits.

write access can be turned off by clicking on the yellow button.

5) A clone can be selected using the left mouse button. Use the shift button to select multiple clones. Selected clones become a nice salmon colour. Now click **Run lace**.

6) The **Select column data to load** box appears next, which allows you to select the analysis and features you wish to see in Zmap. Fewer selected columns will mean a shorter period of time required to open your clones. Selected columns have a yellow box next to them.



8) A yellow progress bar shows the status of data loading. Yellow boxes turn **green** when columns are loaded successfully.

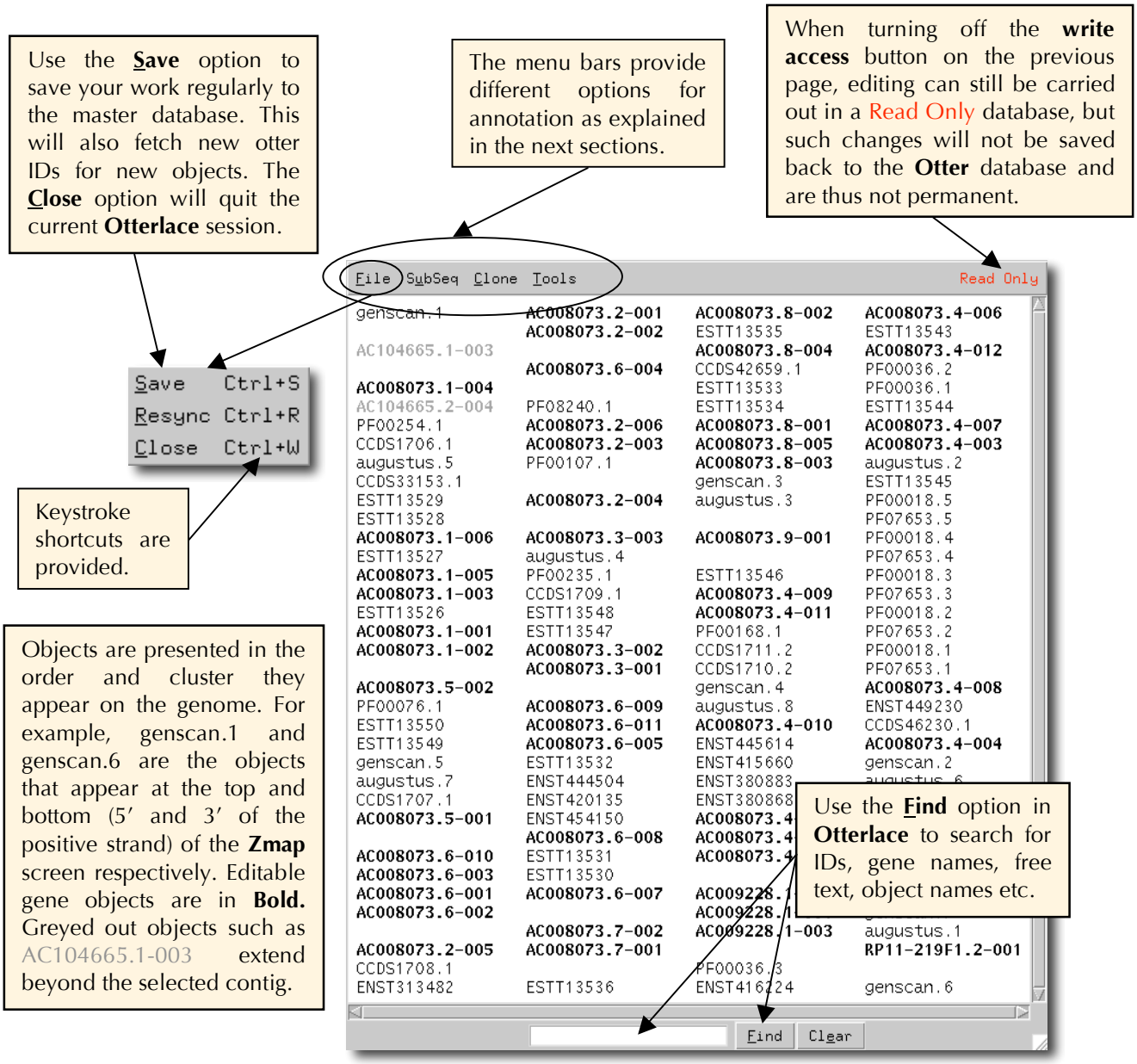
Failed columns turn **red** (mouse over for details). Click button to return status to yellow to retry.

7) Click **load** to run **Otterlace** and **Zmap.**

Further description information for the **column data** can be found here:
http://scratchy.internal.sanger.ac.uk/wiki/index.php/Otterlace_filter_descriptions

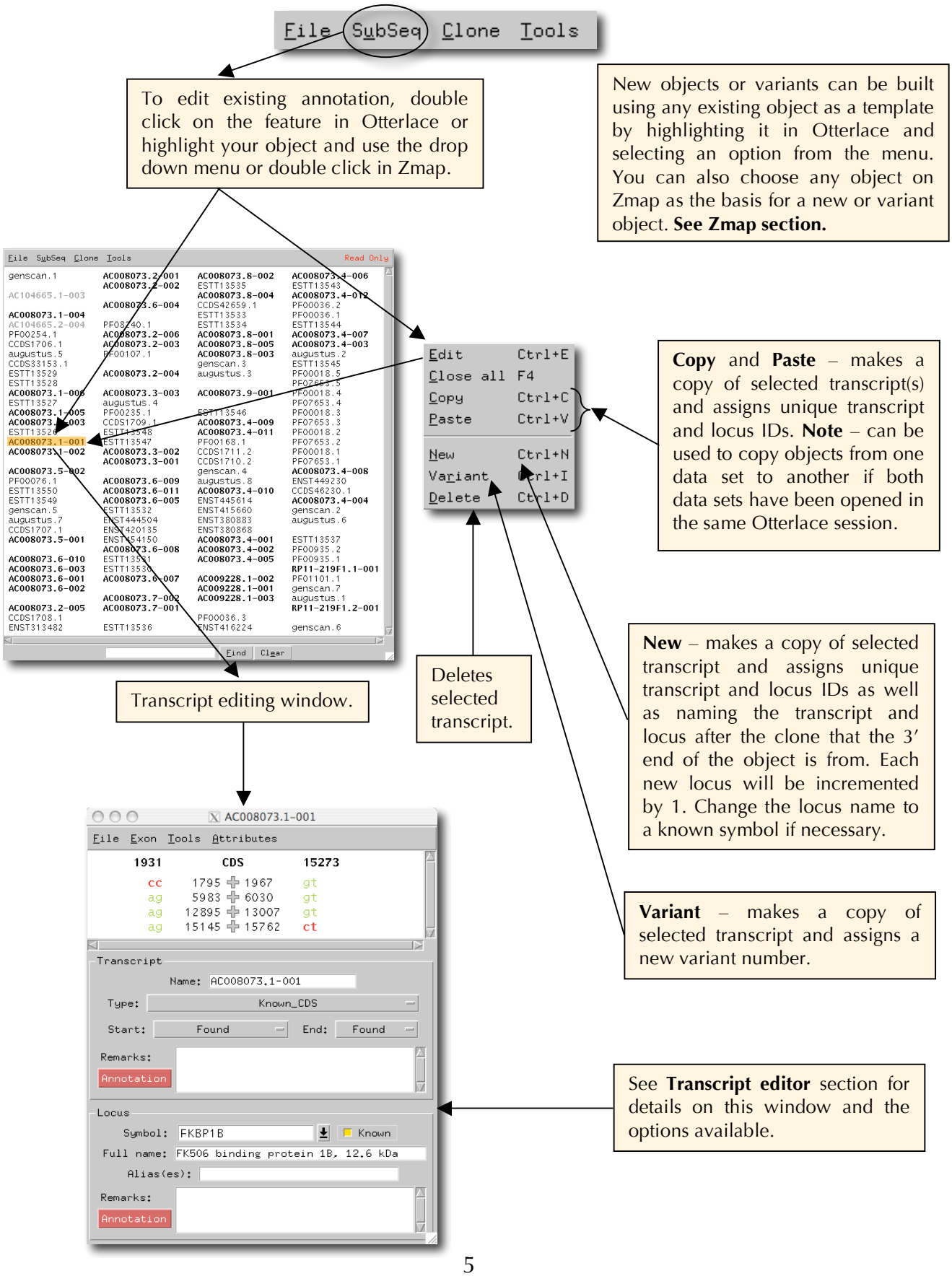
Transcript chooser section

File menu: Manage the Otterlace editing session.

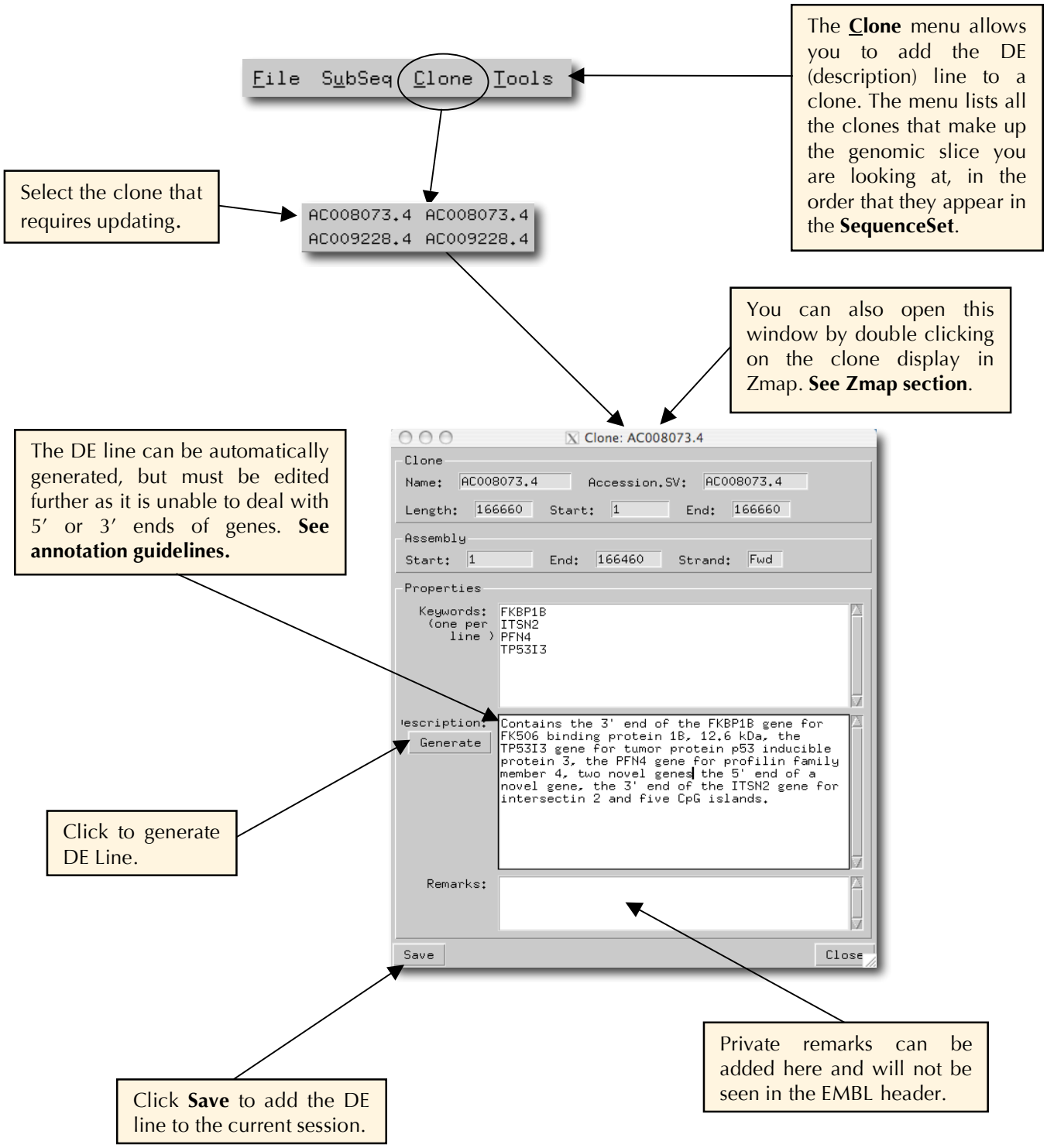


The locus and its associated transcripts and exons are attributed stable, versioned database IDs (e.g. OTTHUMG00000017411), generated and tracked within the **Otter** database. Whenever a gene locus is edited the version number will increase and the date of the change will be saved, allowing the user to find out when the annotation was last updated. It should be noted that versioning occurs within the database and such changes are not externally visible. Clearly, it is vital that current Otter IDs are not deleted, only modified, unless the object is no longer valid.

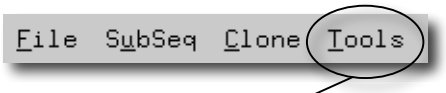
Subseq menu: Editing operations on the transcripts listed in the window.

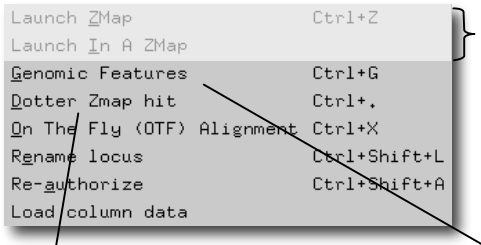


Clone menu: Edit properties of each of the clones (one or many) opened in the otterlace editing session.

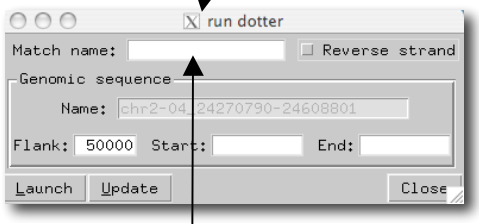


Tools menu: Useful things to run on the genomic sequence being annotated.

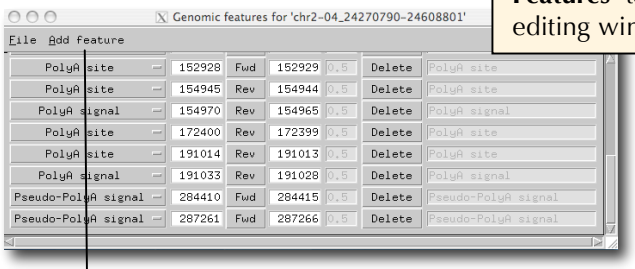




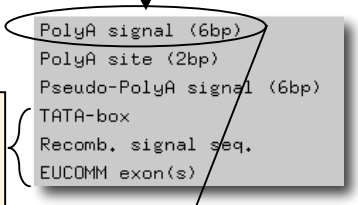
Use this to relaunch Zmap if it was accidentally closed. For Zmap options, see **Zmap** section.



Dotter alignment of any selected homology in paste buffer to object. See section on **Dotter.**



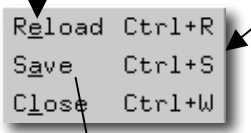
Select **Genomic Features** to bring up editing window.

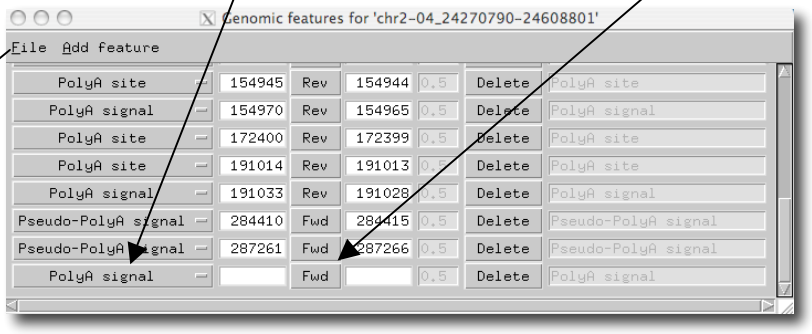


Some features are project specific and will be defined when working on that project.

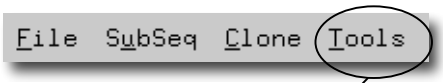
From the **Add feature** menu select the type of feature you want to add. This will then appear in the main box. For polyA features only one of the coordinates needs to be entered as the other is calculated automatically. If necessary, click to toggle the direction (Fwd/Rev). **Select strand before entering coordinates**.

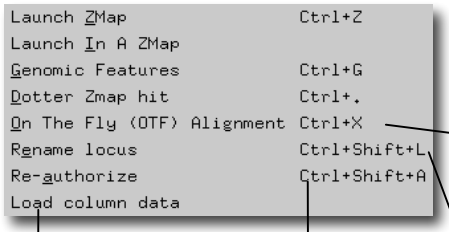
Reload reverts features back to the last save.



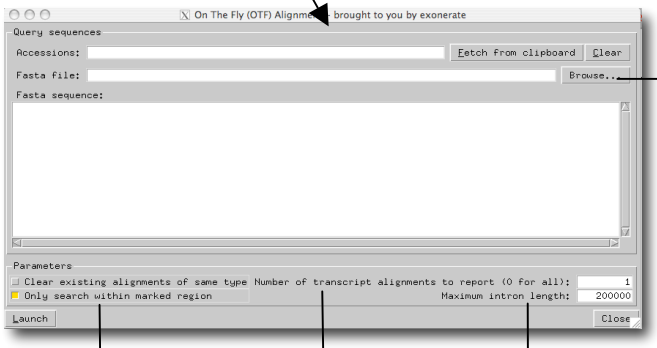


Once the coordinates have been entered, select **Save** from the main window to see the features in Zmap.





On The Fly (OTF) alignment uses exonerate to align sequences to Zmap. These can be single sequences, multiple sequences highlighted in Zmap, missing accession numbers or a fasta file of one or many sequences. Data can be entered in all three of the fields in the OTF window at the same time to search on accession(s), from a file and a seqtext area. Results are dynamically loaded onto Zmap to the right or left of the clone lines **(see later Zmap section)**, depending on orientation.

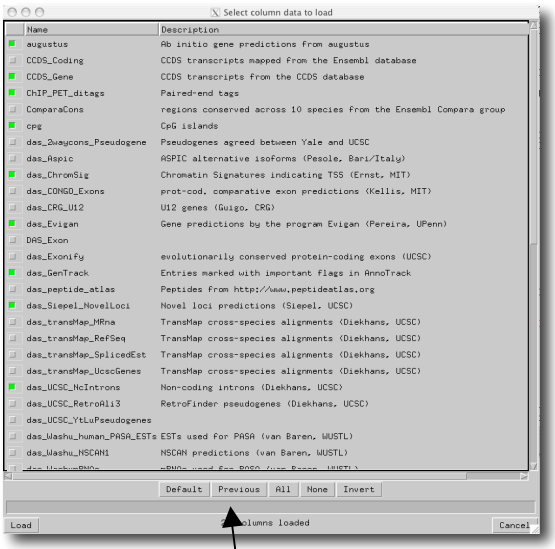


Re-authorize allows you re-establish connection to the database if login expires. This can occur if session has been running for a few days.

Limit search to the marked region in Zmap

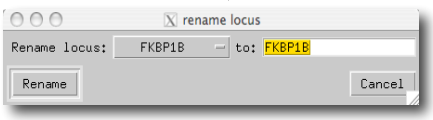
Set this box to "1" to search for the best match or set to "0" to search for all matches.

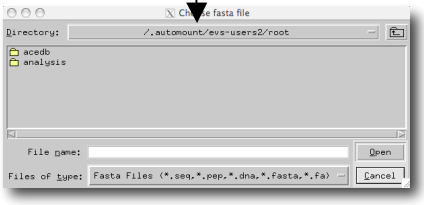
Use this window to increase the window of search for genes with large introns.



Renames a locus to a new locus name.

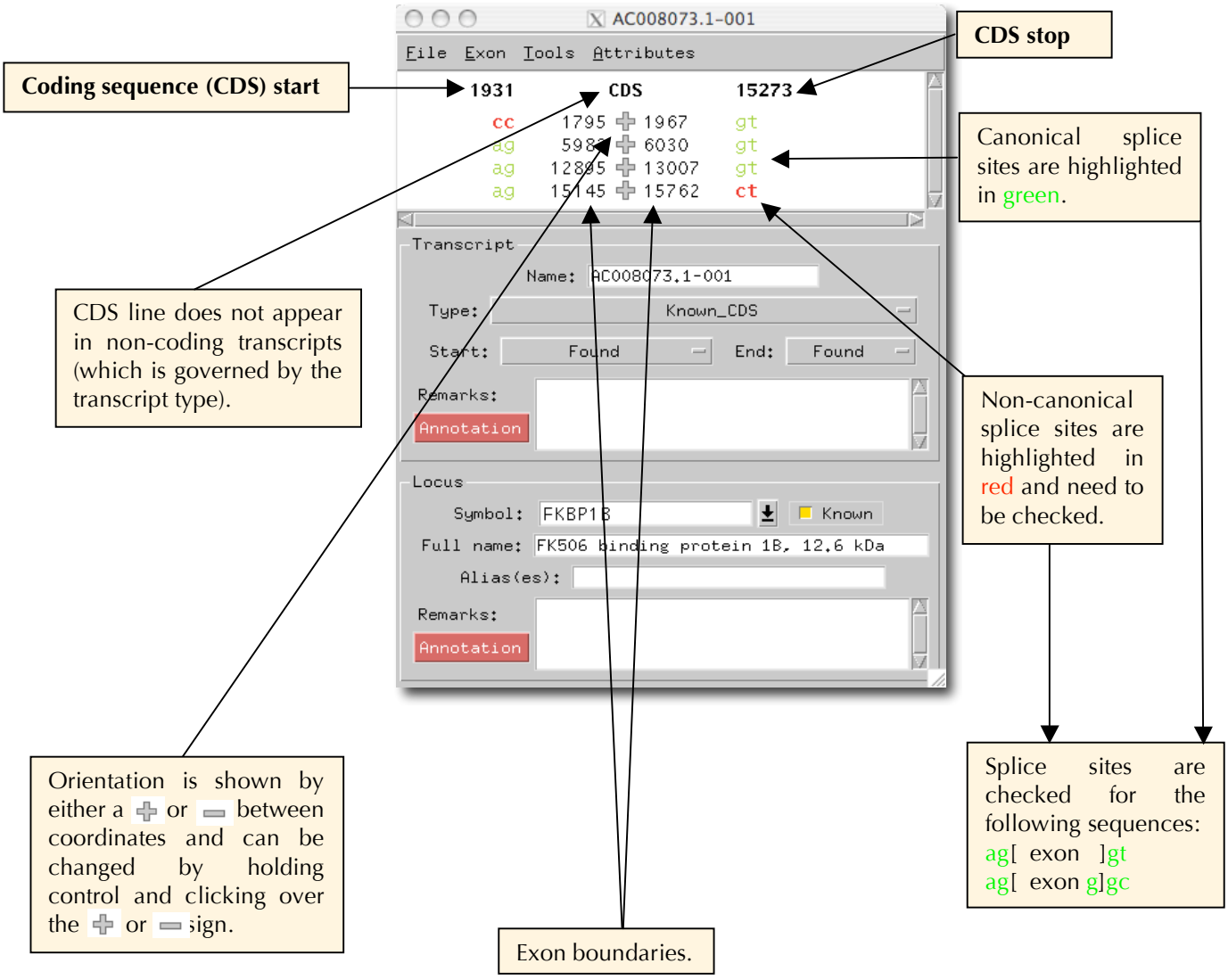
Browse local directory for sequence files.





Load column data gives you the option to load in further column data to a session that is already open.

Transcript editor section



Changing the coordinates can be done a number of ways:
a) Copy coordinates from Blixem (**see section on Blixem**) or select a block of your choice (exon, homology, …) in Zmap and paste coordinates in white space to create new exon(s).
b) Or select existing exon(s) and paste to create copies to edit.
c) Paste over existing coordinate to replace old with new.
d) Or select coordinate and use up and down cursor key to change value.
e) Or select coordinate and delete numbers with backspace key and type in new numbers.
Note: Pasting is done by pressing the middle mouse button (often the scroll wheel).

Transcript type (see **Annotation Guidelines**).

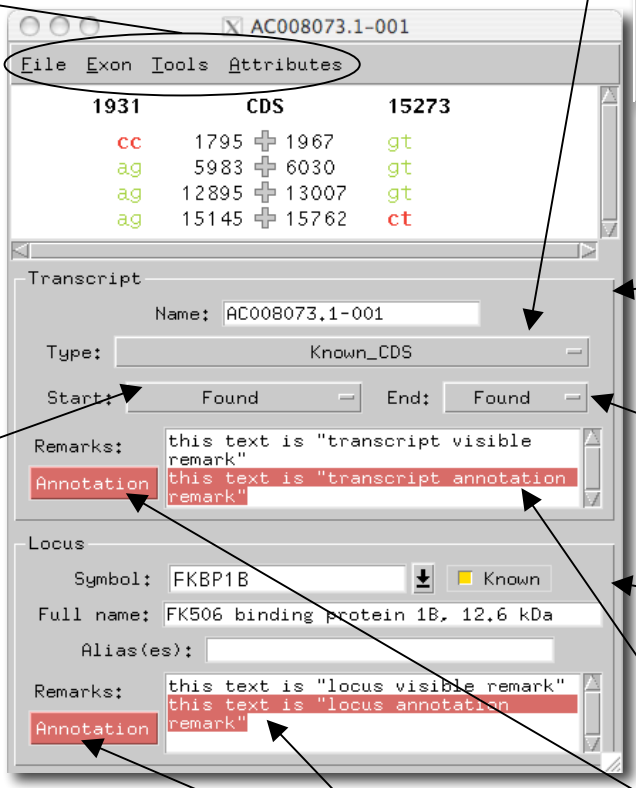
Coding
- Known_CDS
- Novel_CDS
- Putative_CDS
- Nonsense_mediated_decay

Transcript
- Non_coding
- Ambiguous_ORF
- Retained_intron
- Antisense
- Disrupted_domain
- IG_segment
- IG_gene
- Putative

Pseudogene
- Processed_pseudogene
- Unprocessed_pseudogene
- Transcribed_processed_pseudogene
- Transcribed_unprocessed_pseudogene
- Unitary_pseudogene
- Polymorphic_pseudogene
- IG_pseudogene
- Expressed_pseudogene

Transposon
Artifact
TEC
Predicted

The next section describes these menu options.



This section provides information relevant to the **transcript**.

Found
CDS not found - 1
CDS not found - 2
CDS not found - 3
UTR incomplete

Found
Not found

Status of translation stop (CDS only).

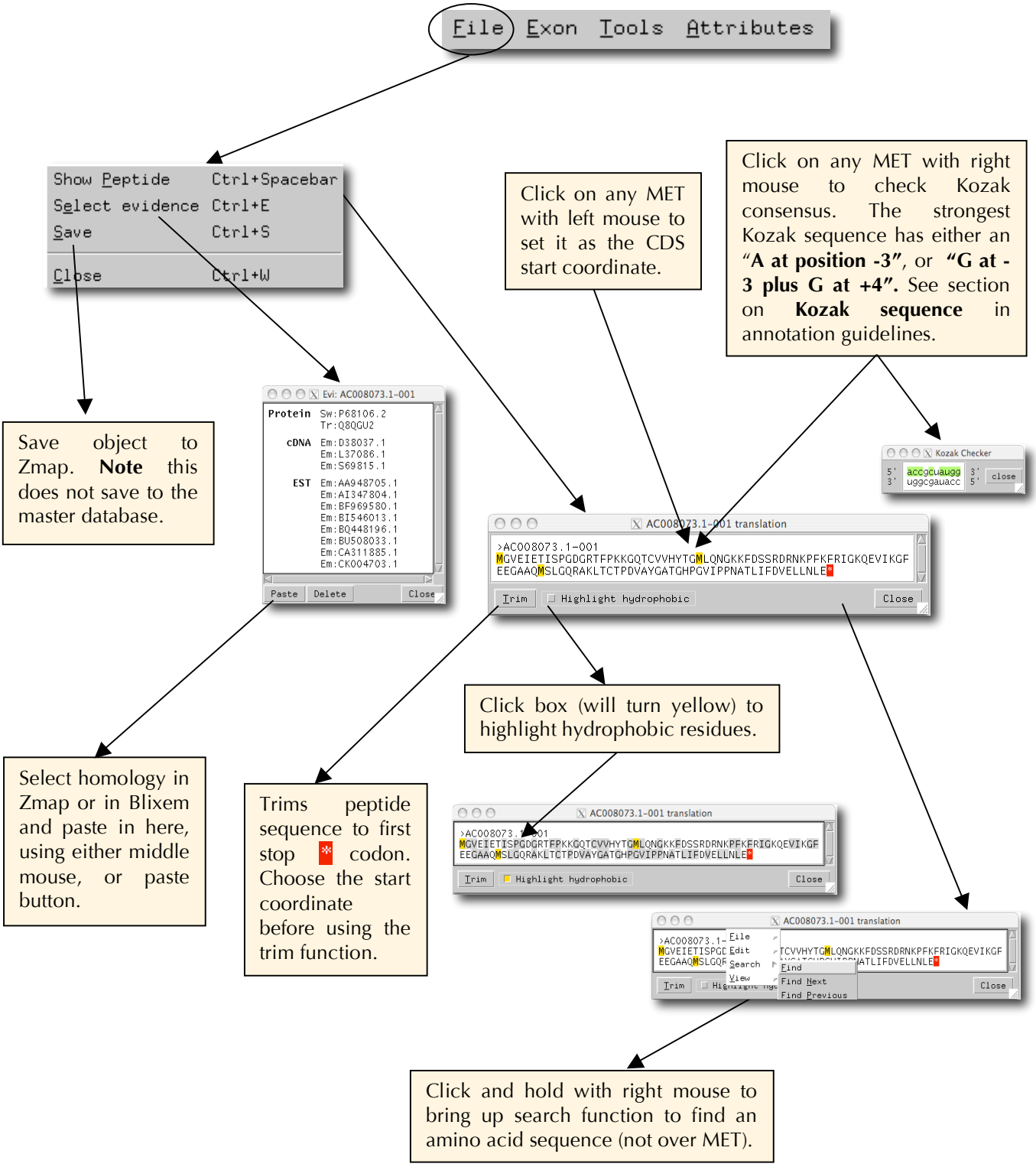
This section provides information relevant to the **locus** (gene).

Status of translation start (CDS only); number indicates translation off-set. The UTR incomplete tag is set if the transcript is cut off within the UTR. For example, if not all of an mRNA used as evidence can be aligned, due to missing genomic sequence etc.

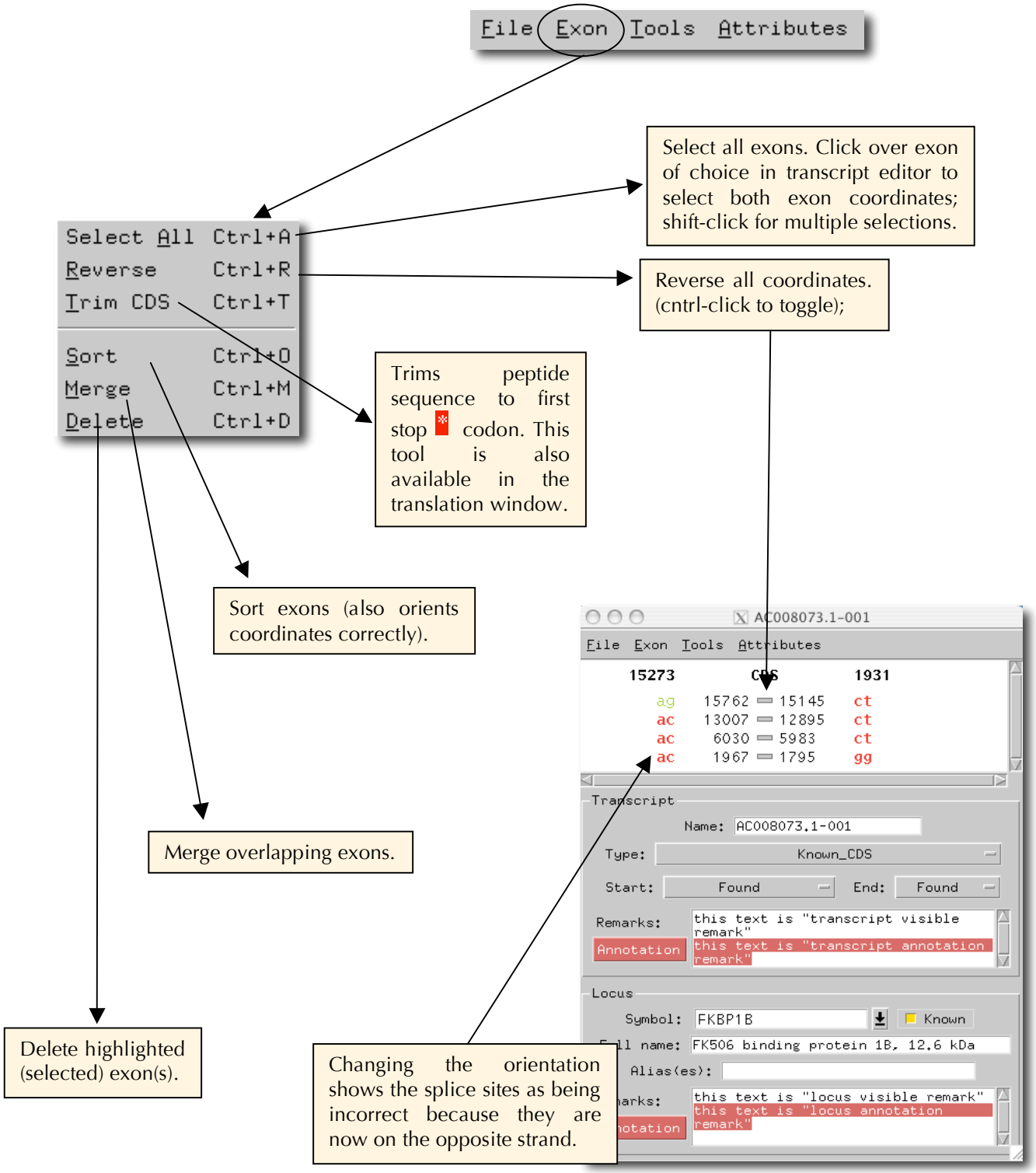
Locus notes. Click on red annotation button to make a comment private, so that it does not appear in the EMBL file.

Transcript notes. Click on red annotation button to make a comment private, so that it does not appear in the EMBL file.

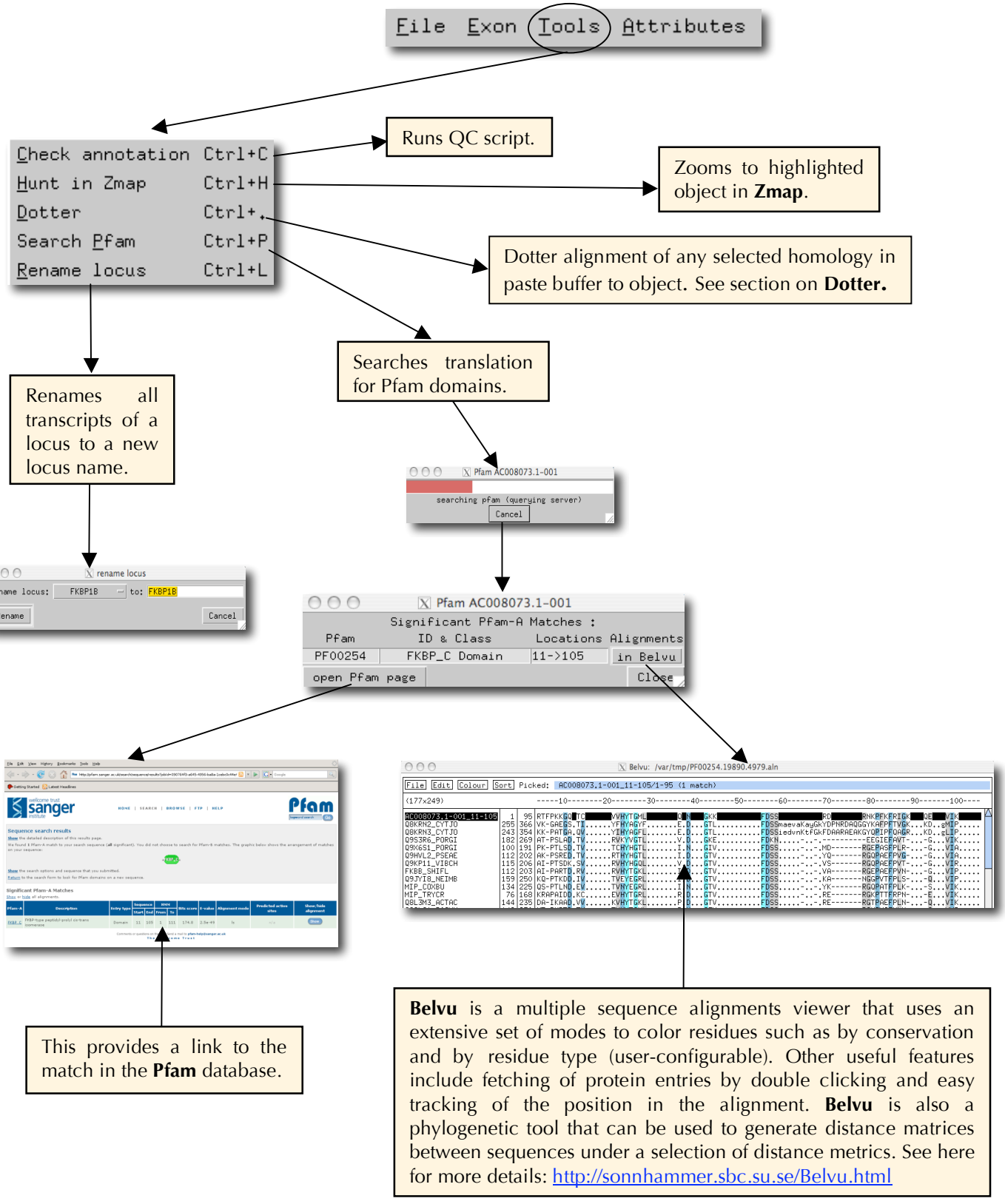
File menu: Saving, closing, plus windows for showing translation and selecting supporting evidence.



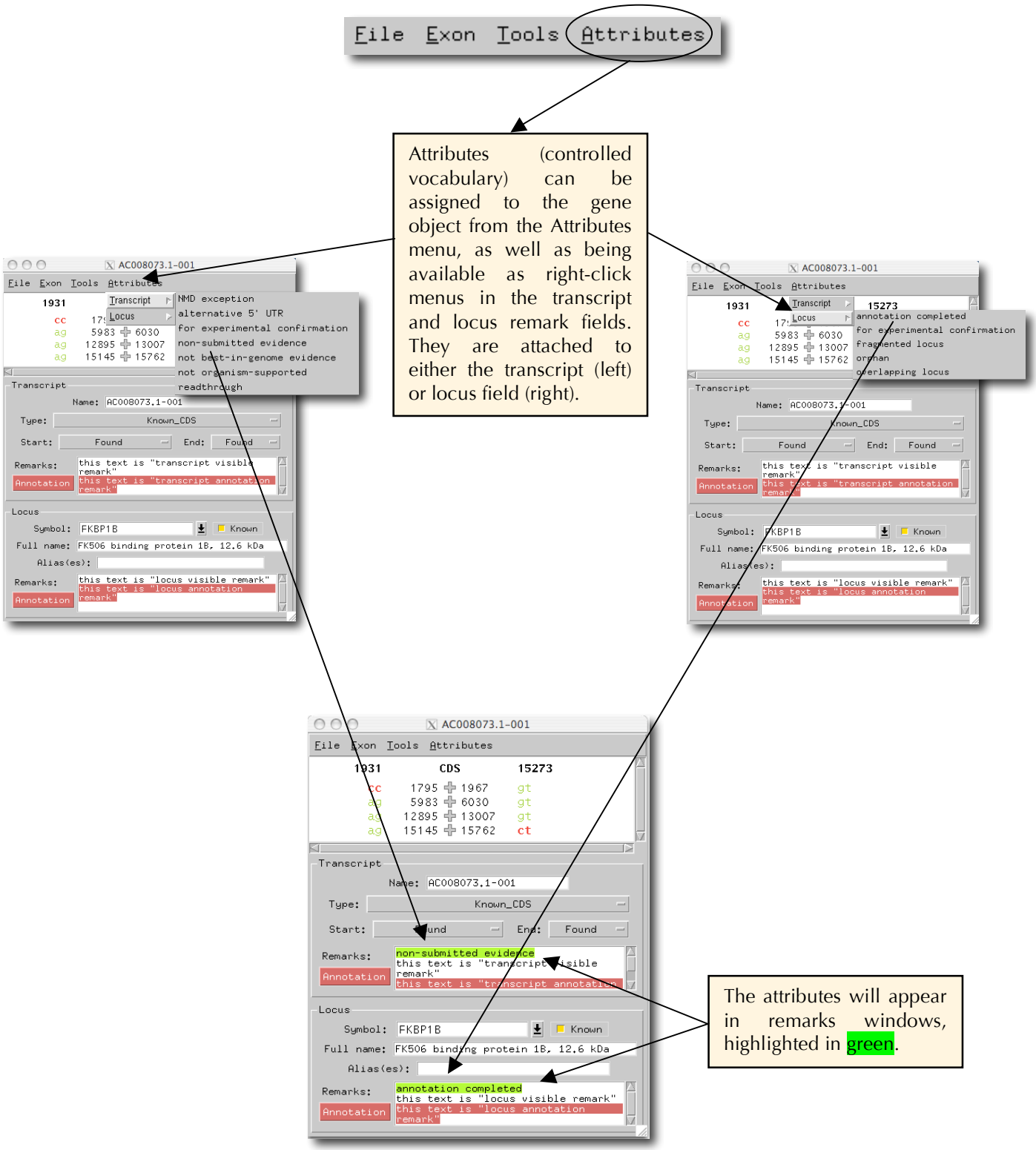
Exon menu: Tools for editing the exons.



Tools menu: Informative operations to run on the transcript.

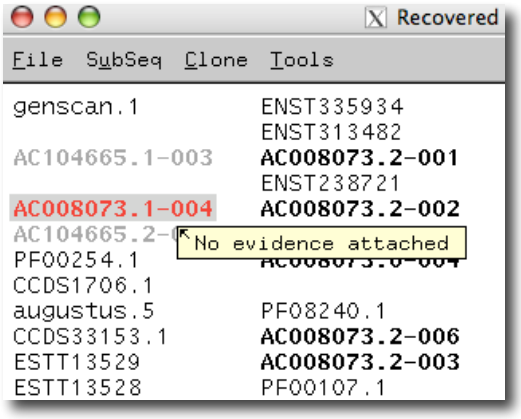


Attributes menu: Controlled annotation vocabulary for the transcript and locus.



Quality control

Otterlace has a built in annotation checking system that checks all manual annotation as it is being created, as well as existing manual annotation, flagging up inconsistent gene objects in red. If you mouse over the offending gene object, you will see a balloon appear explaining any errors that the checking software has found.



This example shows that gene object AC008073.1-004 has no supporting evidence added to it. The gene object will turn black once the checking software finds no inconsistencies.

The complete list of checks carried out is as follows:

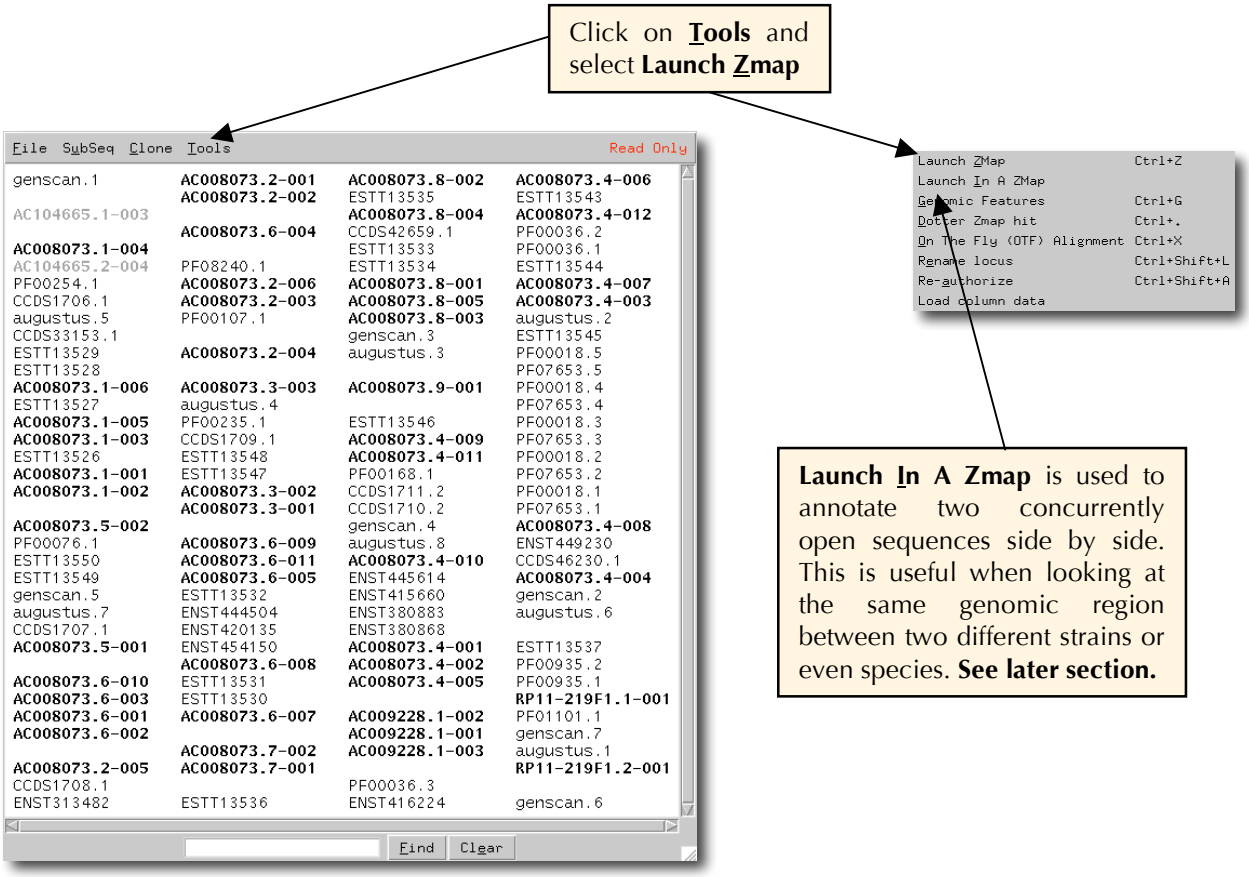
1) No internal stop codons exist in coding object.
2) Transcript has start_not_found set if the translation doesn't begin with Methionine.
3) Transcript has end_not_found set if the translation doesn't end with a stop.
4) The correct selenocysteine remark and coordinates are automatically added if "seleno" appears in an annotation remark for the transcript.
5) Locus has a description (also known as "full name").
6) Transcripts within each locus are all on the same strand.
7) Transcripts do not have a 5' UTR with start_not_found of 1, 2 or 3. (UTR start_not_found has been added as a menu option.)
8) There is evidence attached to each transcript.
9) Nucleotide evidence is only used once in each locus.
10) The same locus name root is not used for transcript names in more than one locus.
11) All the transcript names in the same locus have the same locus name root.
12) Transcript names start with the locus name if the locus name ends with "dot-number" (which means the clonename in such circumstances).
13) Transcript names end "dash-digit-digit-digit".

Zmap

Zmap is a software package that provides a visualisation tool for genomic features. The software is written in C, utilising the gnome toolkit (GTK2) to draw features on a canvas. Zmap accepts input from multiple sources in multiple formats across multiple genomes and is written in a way so that the addition of further formats is made as trivial as possible. Currently the list of formats includes GFF and DAS, which may reside in any one of; a file, an acedb instance, an http server. Multiple genomes and their associated features can be displayed in a single view as aligned blocks providing support for comparative annotation. Zmap does not include any utility for editing the features that it displays. It does however provide a powerful external interface with which to modify the features displayed on the canvas. Using this interface, Otterlace is used to annotate sequences present in the Otter database. This in turn updates to the Vertebrate Genome Annotation (VEGA) website (http://vega.sanger.ac.uk/index.html)

Opening Zmap

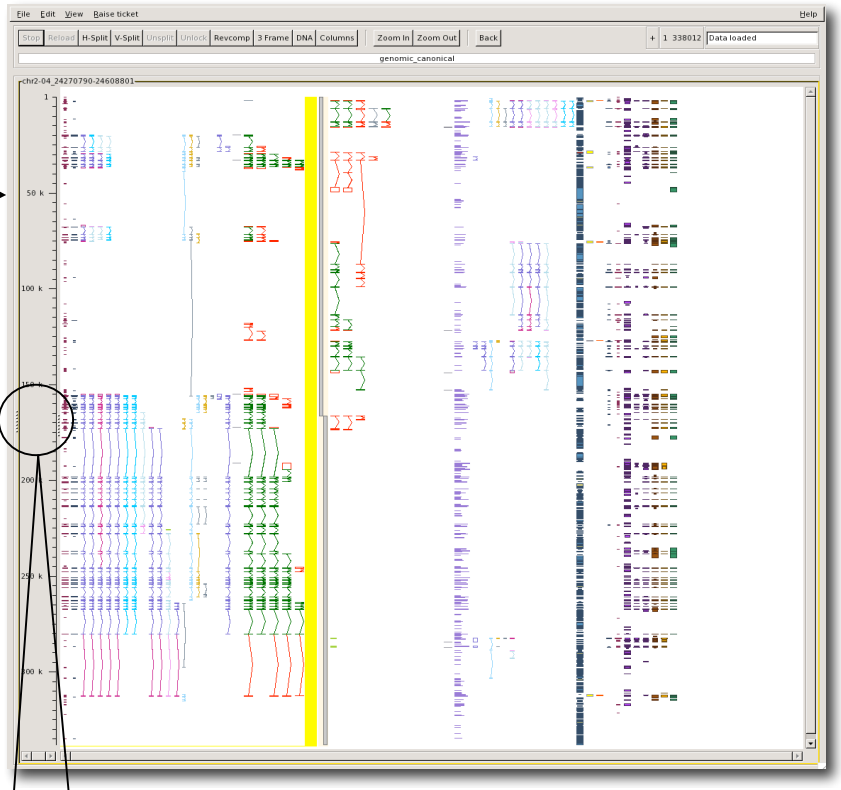
Zmap is opened via the Tools menu bar in Otterlace.



Main Zmap interface

This is the main **Zmap** interface showing an overview of any analysis and annotation that may be present in your region of interest.

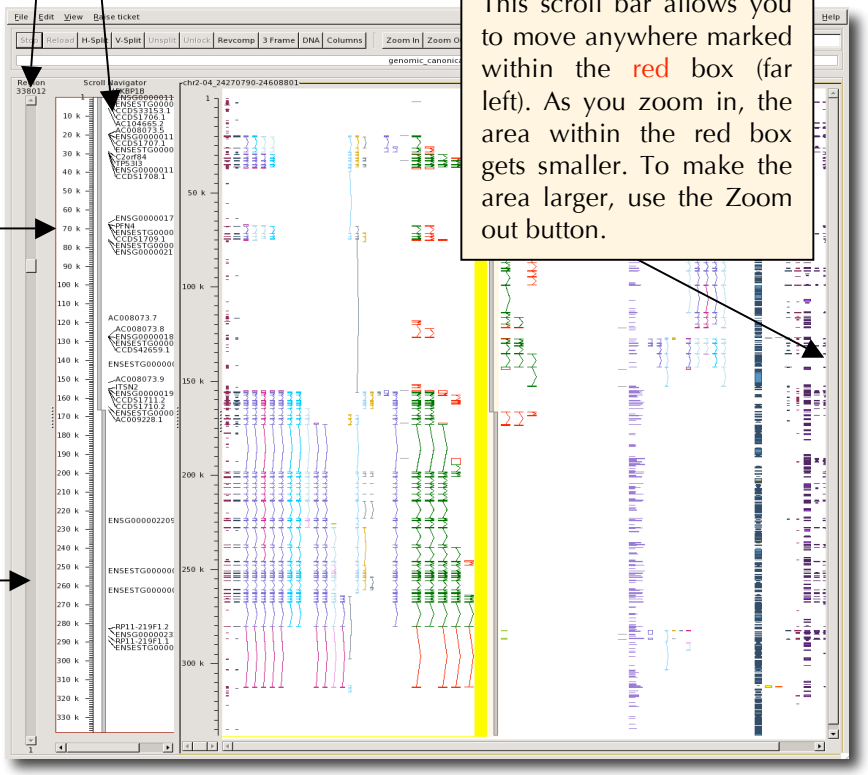
There are various hidden options that you can reveal by dragging the **dotted** regions.



The red box shows the extent of the sequence displayed in the main window showing the analysis, any previously annotated loci or any imported genes that are present in the clone.

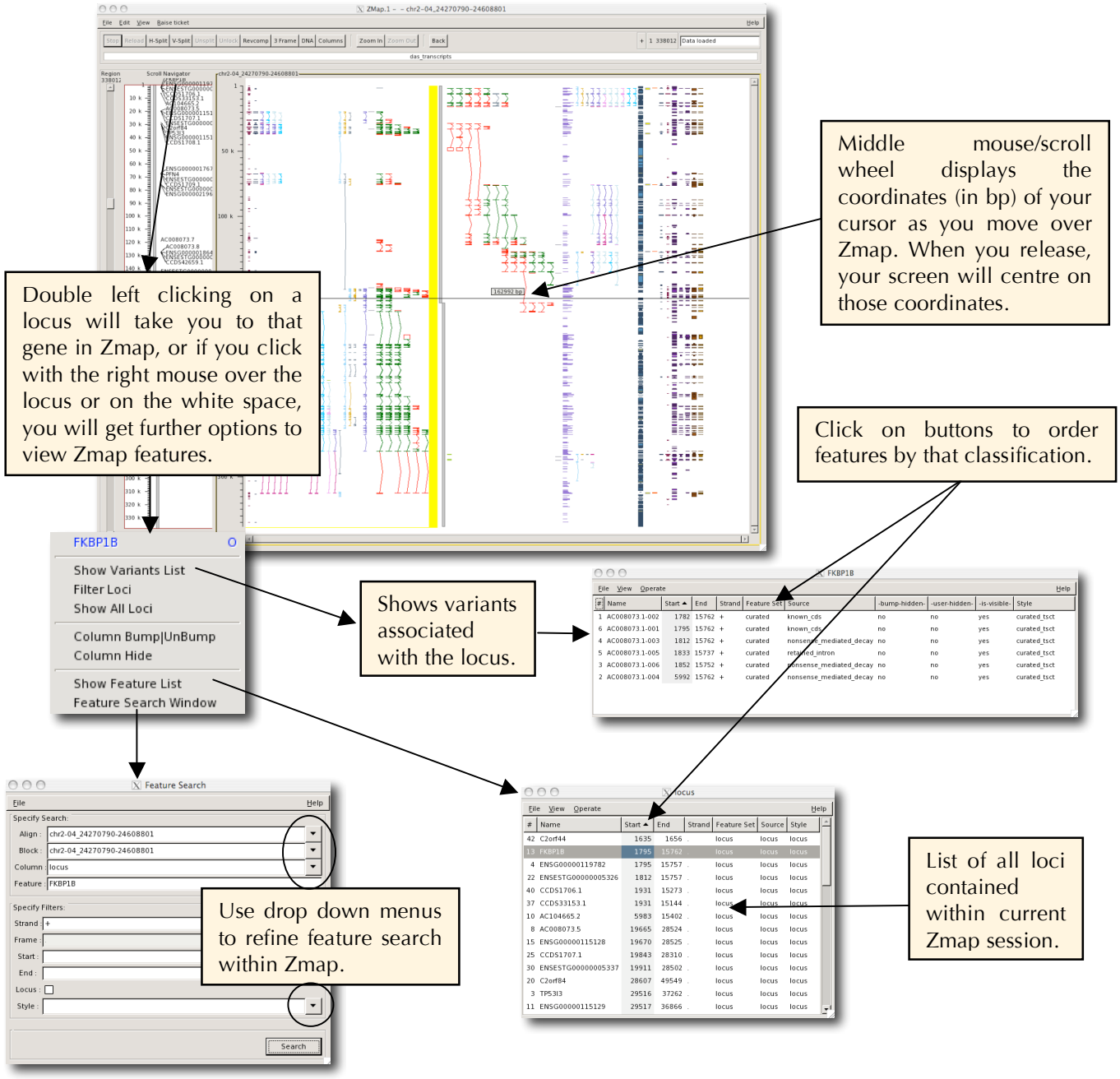
This panel has a scroll bar to show you where you are within the chromosome. It will allow you to jump to different regions. It is generally only useful if you open up very large sections of a chromosome.

This scroll bar allows you to move anywhere marked within the red box (far left). As you zoom in, the area within the red box gets smaller. To make the area larger, use the Zoom out button.



Navigating in Zmap and zooming options

1) Navigate by using the scroll bars or the middle mouse button. By clicking the middle mouse anywhere in Zmap you will see a horizontal line. You can move this up and down and the relative position in bp will be displayed along the line. When the button is released, the window will refresh, centering on the position of the line. You can also click in the window to make it active and use the scroll wheel to navigate up and down or achieve the same result using the scroll bar on the right hand side of the window. If you release the mouse outside the Zmap window, you can then check the sequence position displayed, without re-centering.

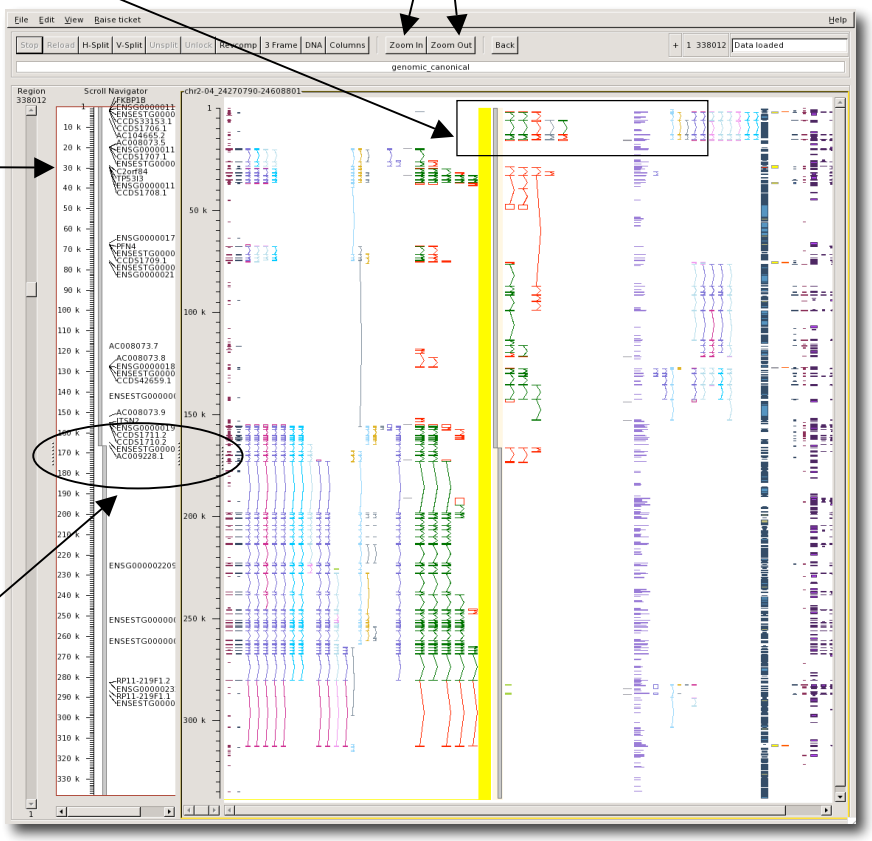


2) Zoom in by using the Zoom in/Zoom out buttons at the top, or by drawing a rectangle around the area of interest with the left mouse button. Use the "**z**" key on the keyboard to zoom to whatever feature is highlighted. Use the "**Z**" key to zoom to a whole transcript if you have an exon (s) highlighted or all HSPs if you have one HSP highlighted (HSPs are the "blocks" that you see in the homology columns, such as ESTs and protein hits).

To mark the rectangle click and hold the left mouse button at the top left of the area you want to outline and then drag out the outline until it encloses the area you want to zoom to. When you release the button, Zmap zooms in to that rectangle.

Use these buttons to Zoom in to a region or to Zoom out.

The red box is draggable. You can use the left mouse to alter the bounds of the display in the main window and the scrollbar to the right of the main window to scroll through the data quickly.

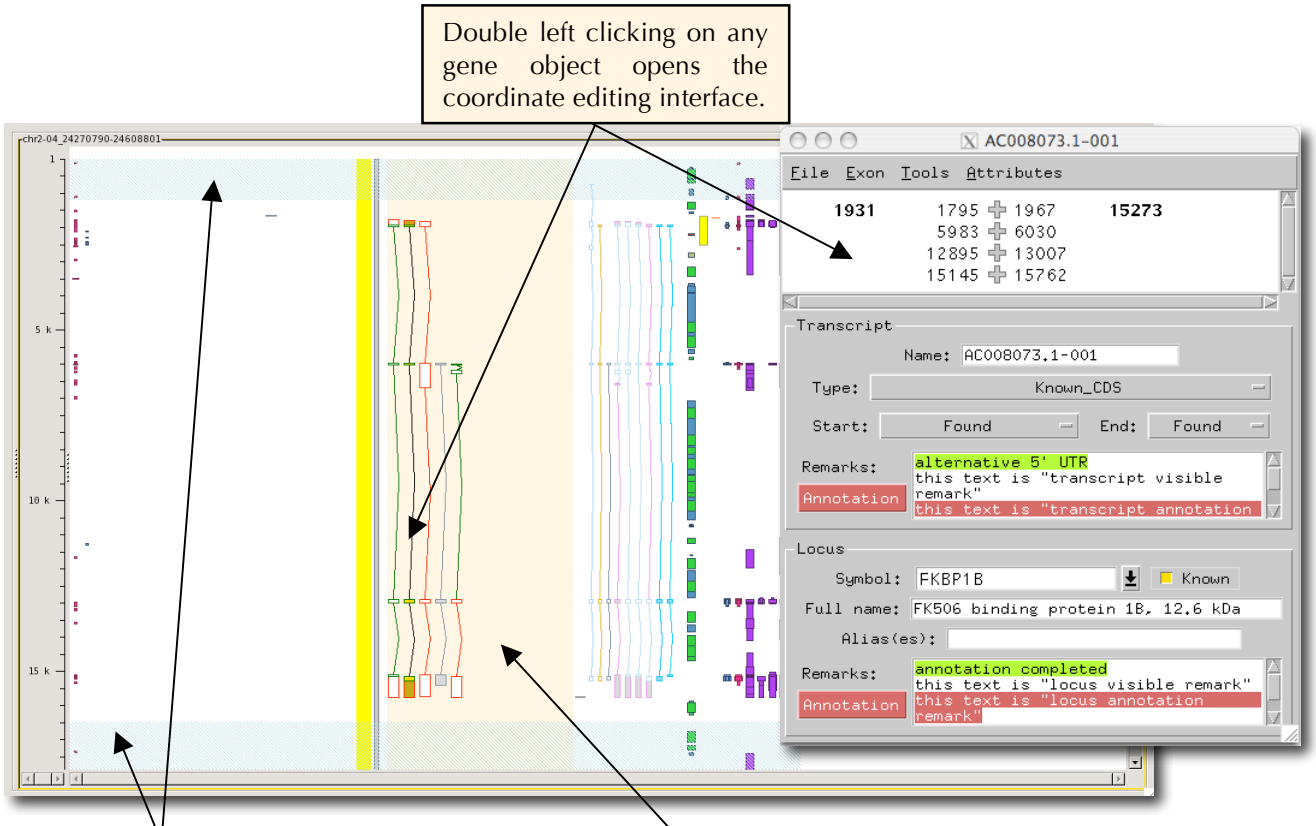


To save space when you are inspecting a region you can drag the dotted lines back to their original position to remove the scroll bar and locus panel information. Note, it is not necessary to have any of these panels open while you work.

The Focus Feature vs the Marked Feature

If you click on a column background then that column becomes the "focus" column and you can do various short cut operations on it such as pressing "b" to bump it. If you click on a feature then that feature becomes the "focus" feature and similarly you can do various short cut operations on it such as zooming in to it. (Note when you select a feature then its column automatically becomes the focus column.)

While the focus facility is useful, the focus changes every time you click on a new feature. Sometimes you want to select a "working" feature or area more permanently. To do this you can "mark" the feature or area and it will stay "marked" until you unmark it. 'Marking' an area within Zmap to work on is essential, allowing you to work much faster. The "marked" area is left clear while the unmarked area above and below is marked with a blue overlay (see screen shot below):



Mark a feature

1) Select a feature to make it the focus feature.
2) Press "m" to mark the feature, the feature will be highlighted with a blue overlay.

Feature marking behaves differently according to the type of feature you highlighted prior to marking and according to whether you press "m" or "M" to do the marking:

1) If you press “m”, the mark is made around all features you have highlighted, e.g. a whole transcript, a single exon, several HSPs.
2) If you press "M" to do the marking around transcripts the whole transcript becomes the marked feature and the marked area extends from the start to the end of the transcript.
3) If you press "M" to do the marking around alignments all the HSPs for that alignment become the marked feature and the marked area extends from the start to the end of all the HSPs.
4) If you press "M" to do the marking around all other features: the feature becomes the marked feature and the marked area extends from the start to the end of the feature.
5) If no feature is selected but an area was selected using the left button rubberband then that area is marked.
6) If no feature or area is selected then the visible screen area minus a small top/bottom margin is marked.

Mark an area

1) Select an area by holding down the left mouse button and dragging out a box to focus on that area.
2) Press "m" to mark the area.

Manual cropping of the marked borders

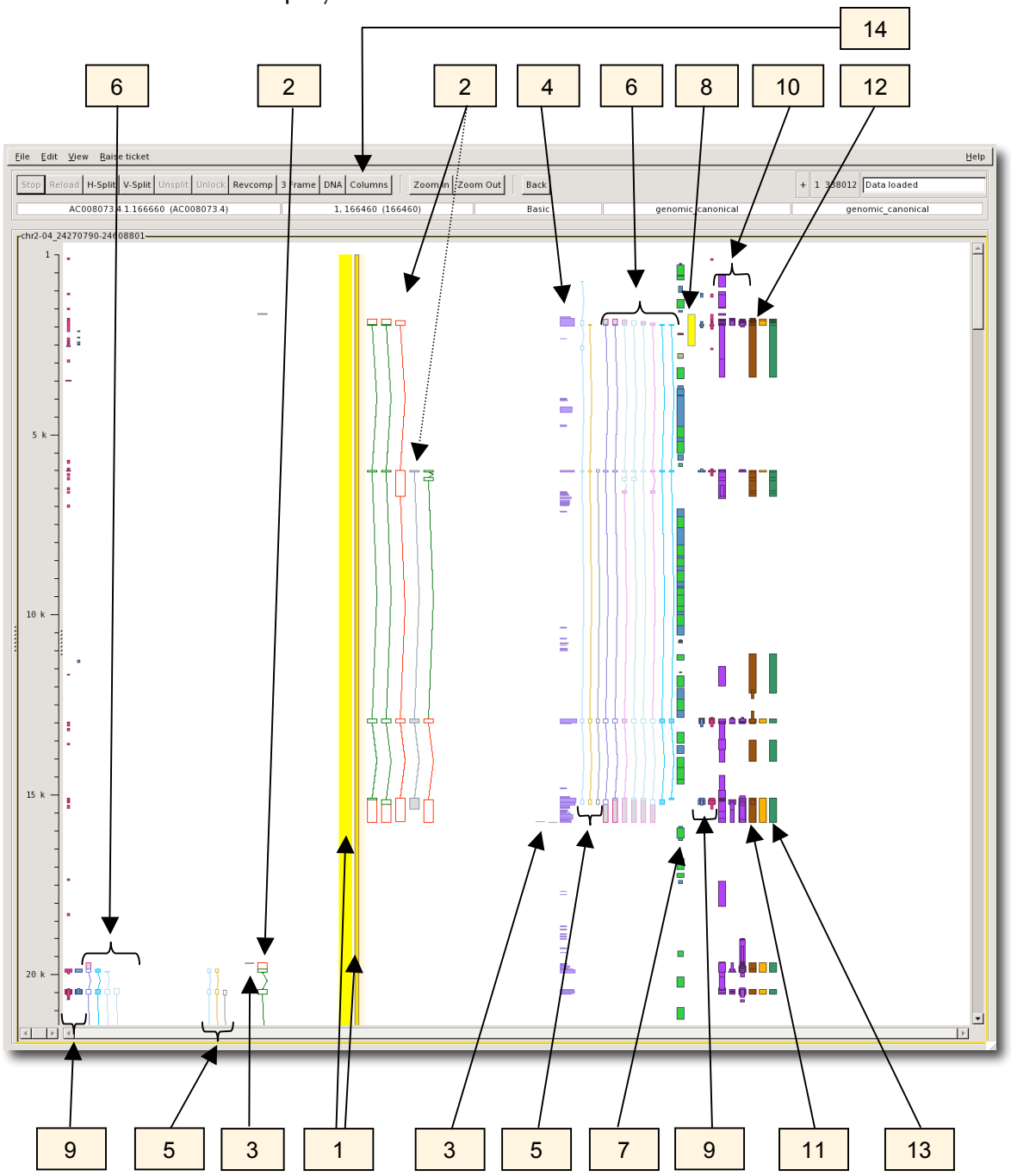
You can manually change the borders of the marked area by putting your cursor over this area and using the cropping tool by clicking and holding with the left mouse button and dragging to make the area bigger or smaller.

Unmark a feature

Press "m" or "M" again, i.e. the mark key toggles marking on and off.

General Zmap display features

Different features are displayed in distinct columns as follows:



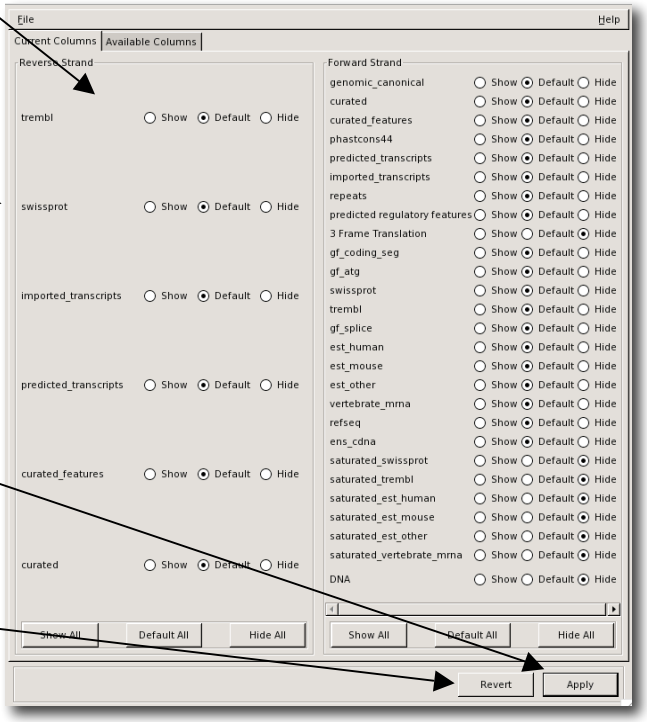
Note - you may see more or fewer features depending on how your preferences are set up. For descriptions of other column types such as DAS sources, visit this URL - http://scratchy.internal.sanger.ac.uk/wiki/index.php/Otterlace_filter_descriptions

1) The thick yellow line represents the genomic sequence; everything to the left represents the negative strand and everything to the right the positive strand. DNA matches (i.e. ESTs, mRNAs and RefSeq) and repeats are all displayed to the right of the center although they may align to either strand. The thin bar to the right is the clone that the genomic sequence is made up from. Double click on this to access the DE editing window.
2) Annotated transcripts; green is coding (CDS), red is non-coding (UTR and transcript variants). Grey transcripts (see dotted line) contain exons outside the sequence slice being viewed and should not be confused with Halfwise hits.
3) Curated features, such as PolyA features are seen as horizontal black lines.
4) Phastcons44 – conserved regions detected using multiple sequence alignments of 44 organisms.
5) Predicted transcripts such as Genscan (pale blue), Augustus (gold) and Halfwise predictions of Pfam (grey).
6) Any imported annotation is located here such as Ensembl hits and CCDS.
7) Repeats (blue=Line , light green=Sine , gold=other), tandem repeats are red.
8) CpG islands appear as yellow boxes.
9) Protein matches are strand specific - SwissProt are light blue and Trembl pink.
10) EST matches are displayed as purple blocks and are broken down into human ESTs, mouse ESTs, and other ESTs from other organisms. 5' reads are on the left and 3' on the right.
11) mRNA matches contains all species and are displayed as brown blocks,
12) RefSeq matches are the orange blocks.
13) Ensembl aligned mRNAs.
14) Features and analysis available

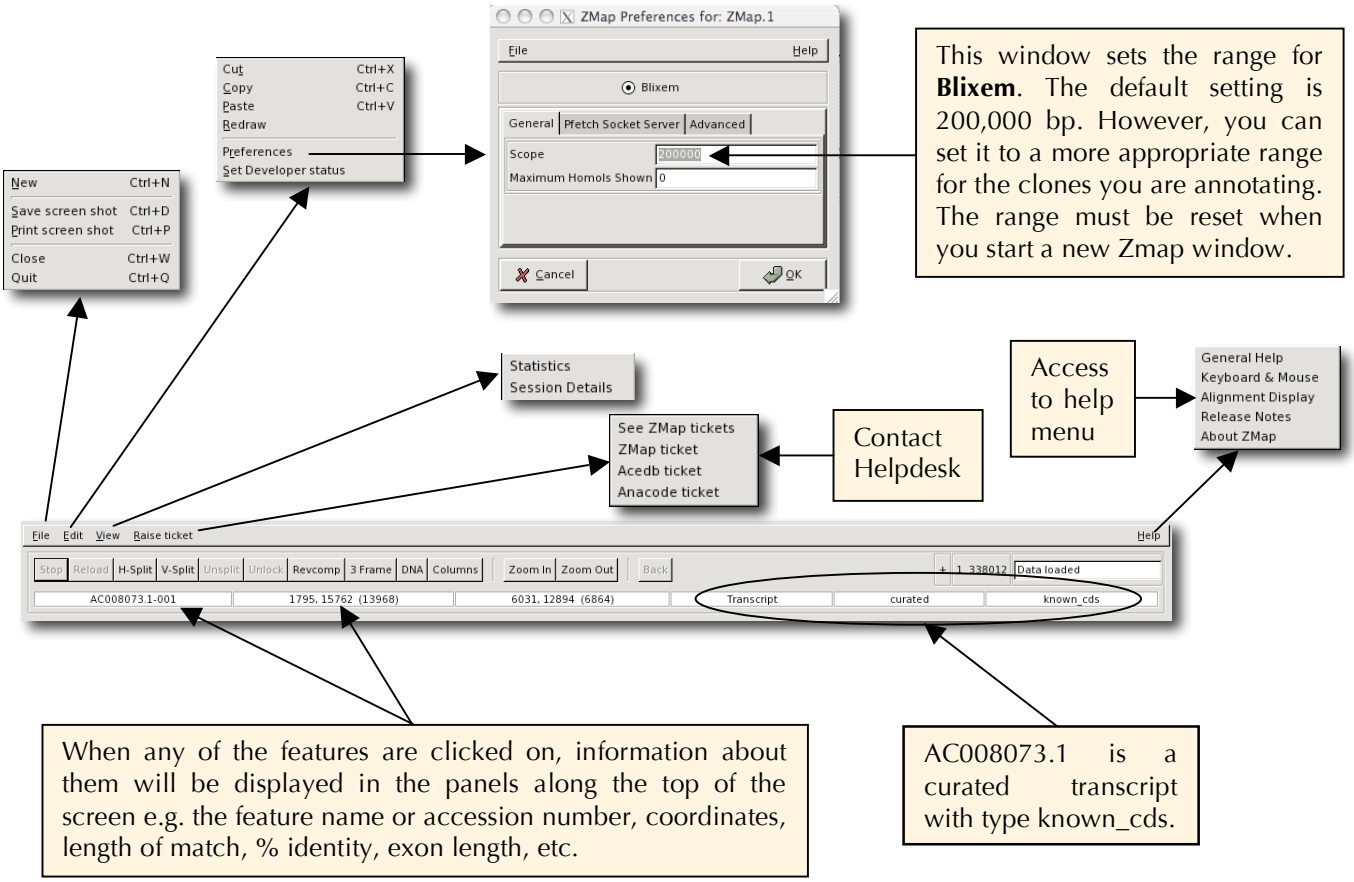
14) The **Columns** button brings up this window, allowing you to customize Zmap by turning features on and off.

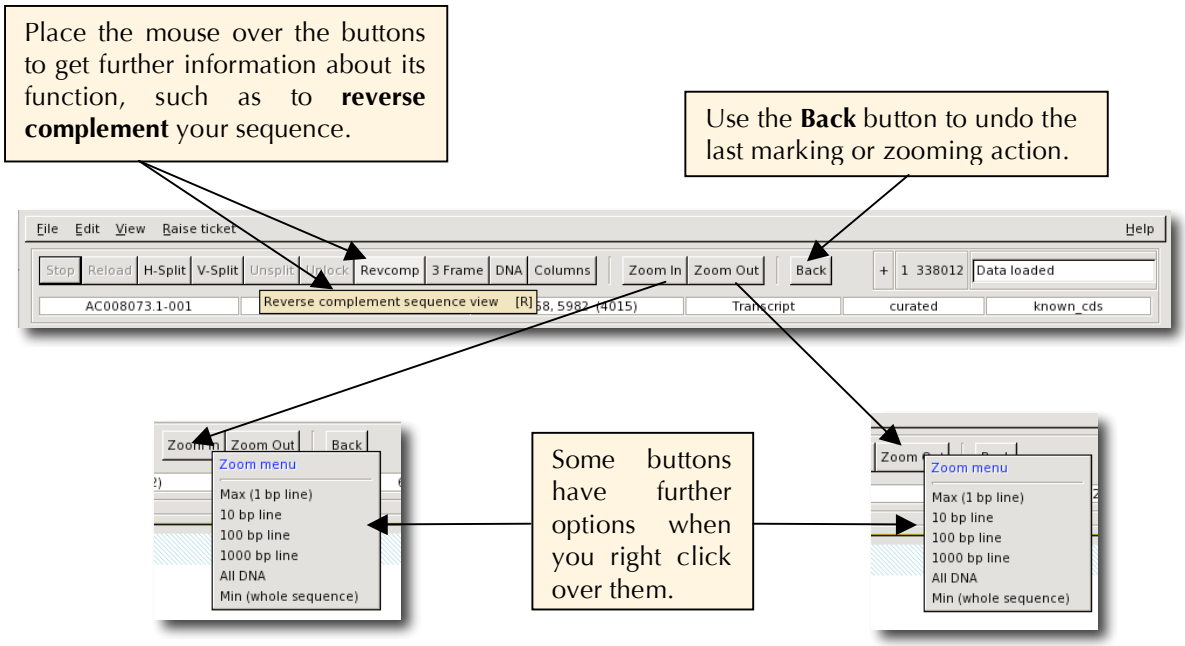
Select the features that you want to be visible on Zmap and click on **Apply**.

Revert sets the features to the default setting.

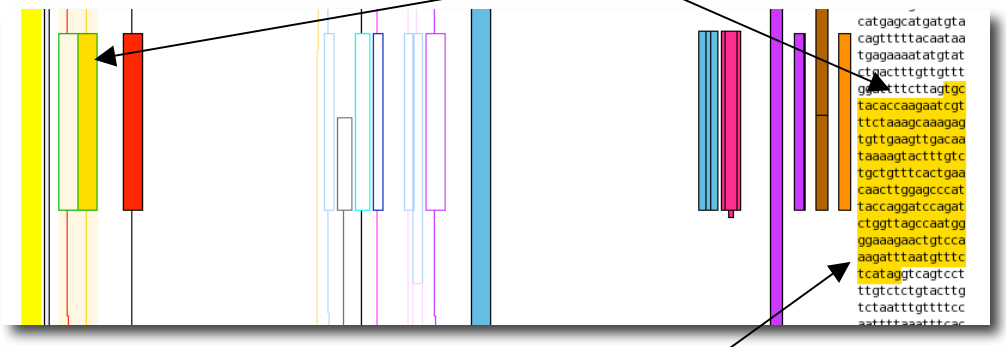


Functionality of the features at the top of the Zmap display.

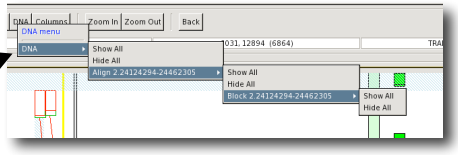


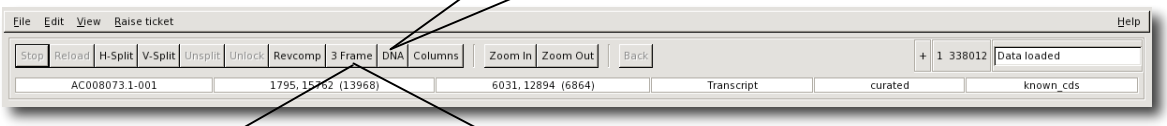


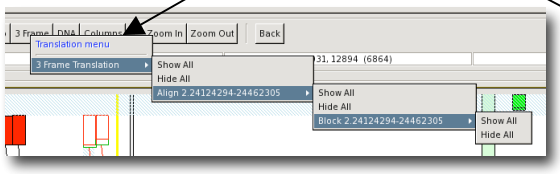
The **DNA** button will show the nucleotide sequence. If you click on an exon, the sequence is highlighted in orange. You can select a DNA sequence by clicking with the left button and dragging a selection, which you can then paste with the middle mouse.

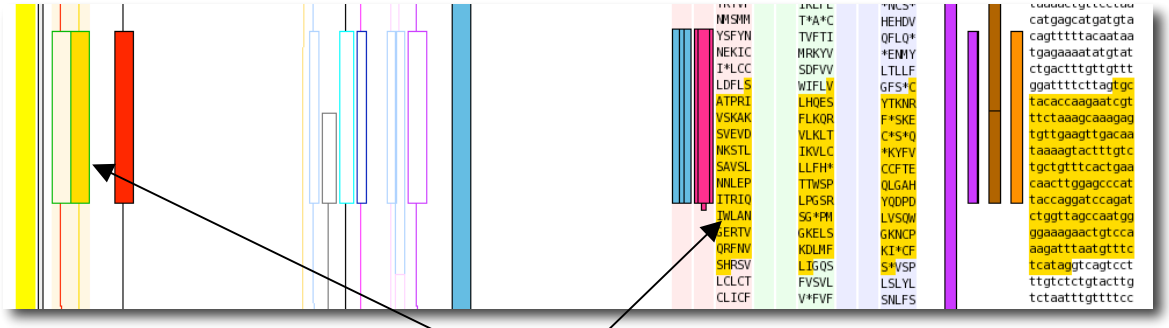


Click the buttons with the left mouse to operate the **DNA** and **3 Frame** translation options. Right click over the buttons for further options. To remove these displays from Zmap, click on the button again.





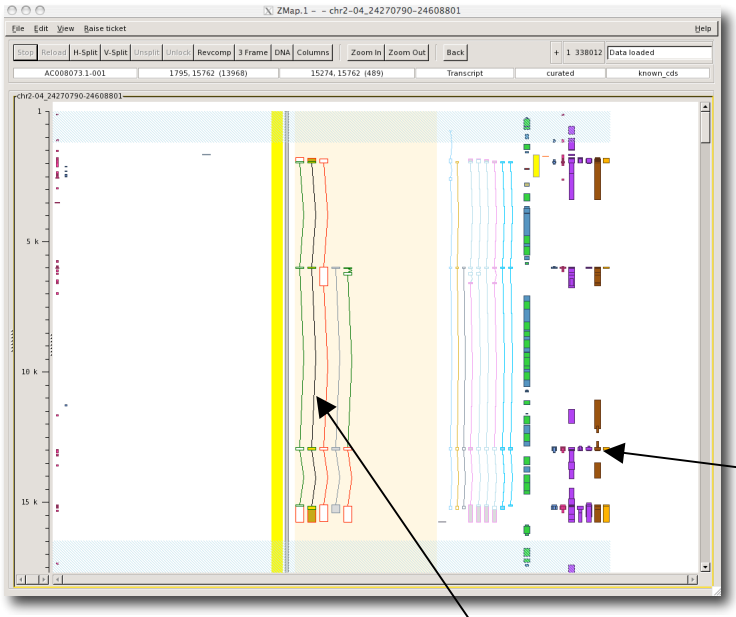


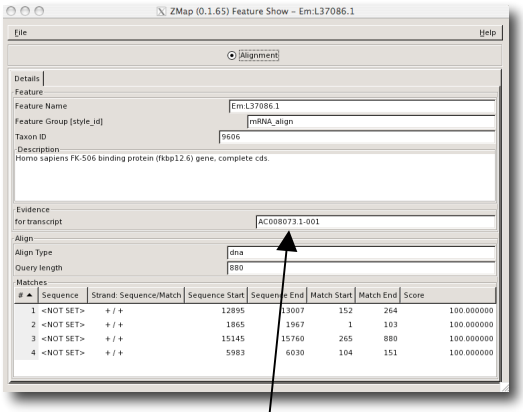


The **3 Frame** button will show the amino acid sequence in each of the three reading frames. If you click on an exon, the sequence is highlighted in orange.

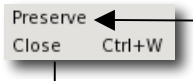
Show feature details

Right click on a gene object or 'o' key when highlighted to see information on otter IDs and Ensembl IDs. For BLAST hits, double click on the HSP to get the feature interface where you will find details on alignment and on what HAVANA object the HSP has been assigned to, if any:

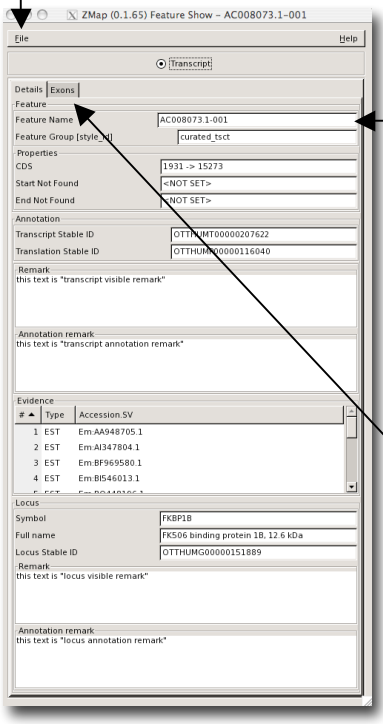




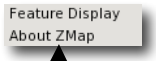
Feature Details for an HSP will show alignment information as well as any gene object it has been assigned to as evidence.



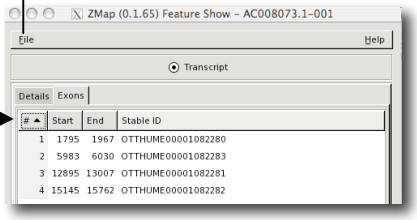
Prevents window from being reloaded.



Left click once on a gene object and hit return to reveal the **Feature Details** interface, where you can see the stable IDs (also available by right clicking and selecting **Show Feature Details** from the popup menu).

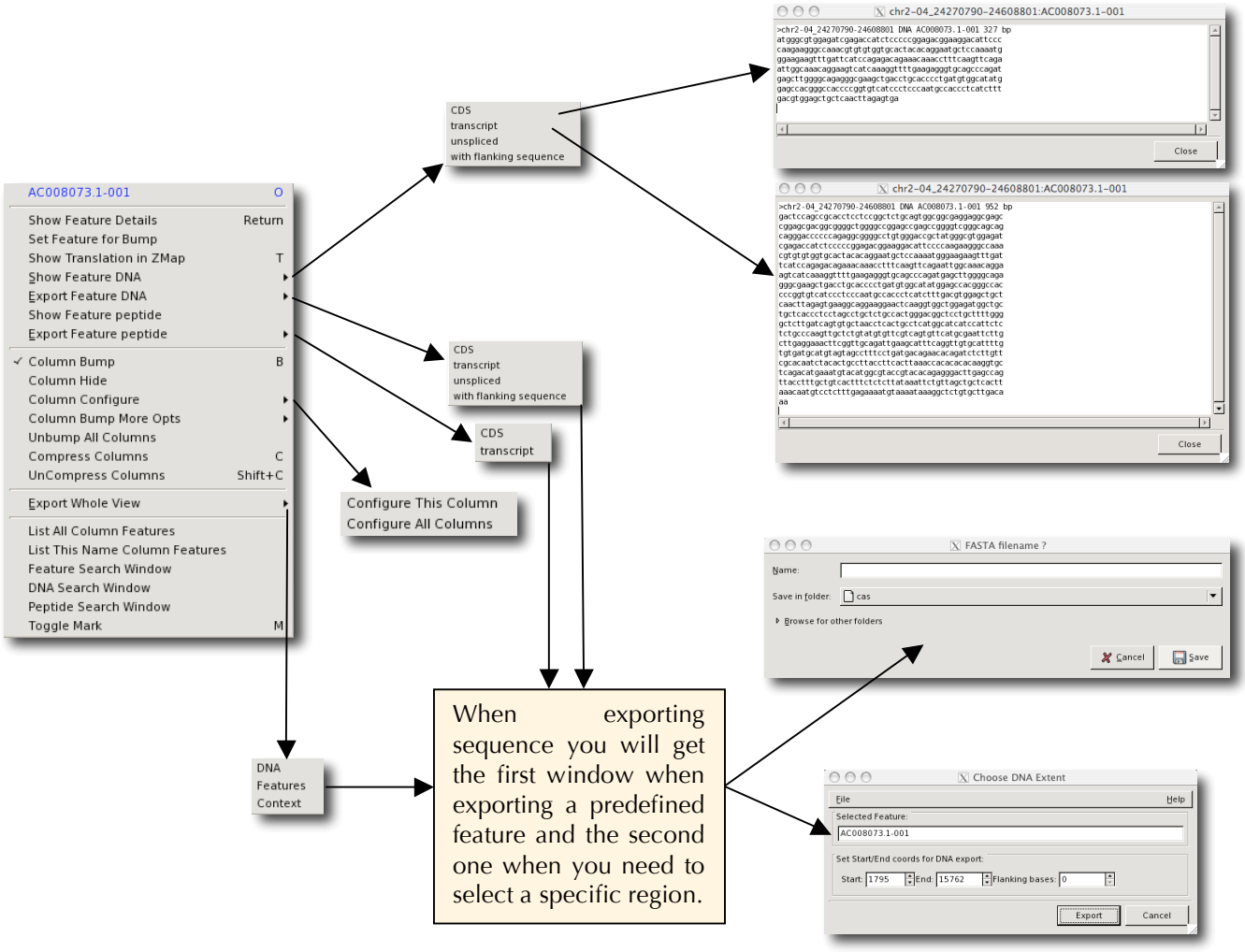


Select the Exon tab to see Stable IDs and coordinates for the exons.



Exporting features for gene objects

As described on the previous page, if you right click over any feature (or type “o” when a feature is highlighted) you get further information. These screen shots show how you can view and export an annotated sequence to your home directory in various different ways, such as dumping features directly. In the main Zmap window, right click on an annotated gene object. From the drop down menu select **Export Feature DNA** and choose sequence required from CDS, transcript, unspliced and with flanking sequence. Alternatively select **Export Feature peptide** and choose either CDS or transcript. Here you can see how to **Show Feature DNA** for annotated gene object AC008073.1-001 in FASTA format; firstly, the section of the transcript that corresponds to the CDS and secondly the whole transcript, including the untranslated region (UTR).



Bumping features

This section describes how to select a feature, mark it and then zoom in to it and examine evidence that overlaps that feature. The default setting for Zmap is to show HSPs drawn on top of each other. This saves space on the canvas making it easier to see the general features of the region of interest. The bump option allows you to see the HSPs as multiple alignments.

1. Click on the feature you are interested in (perhaps a transcript)
2. Mark it by pressing "m"
3. Zoom in to the feature by pressing either "**z**" or "**Z**" (as described previously).

Now when you bump an evidence column to look at matches that overlap the feature you will find that bumping is much faster because only those matches that overlap the feature get bumped and you also have fewer matches to look at. The quickest way to bump a column is:

1. Click on the column to select it.
2. Bump it by pressing "b" (if you press "b" again the column will be unbumped). If you have marked a feature then bumping is restricted to matches that overlap that feature, otherwise bumping is for the whole column.

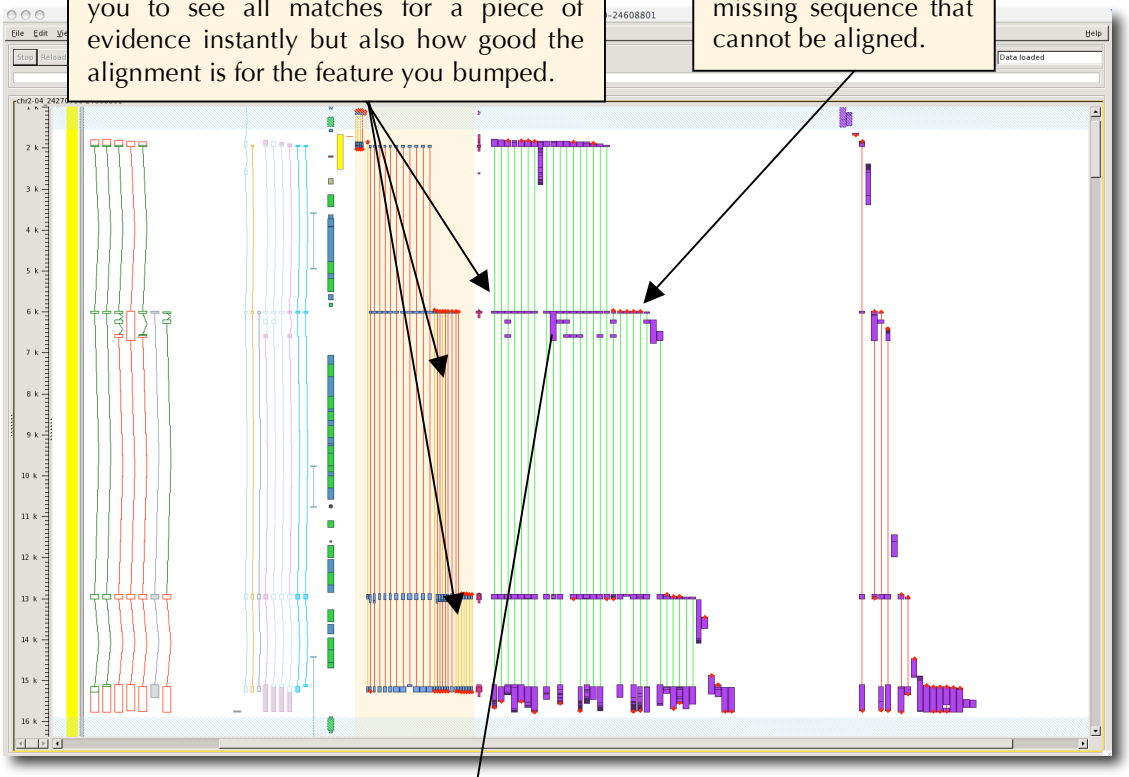
If you use the default bumping mode (i.e. you pressed "b") then you will find all matches from the same piece of evidence are joined by coloured bars, the colours indicate the level of colinearity between the matches (see next screen shot).

1. **Green**: the matches at either end are perfectly contiguous, e.g. 100, 230 ---> 231, 351
2. **Orange**: the matches at either end are colinear but not perfect, e.g. 100, 230 ---> 297, 351. Matches may also be this color when there are extra bases in the alignment, e.g. around clone boundaries.
3. **Red**: the matches are not colinear, e.g. 100, 230 ---> 141, 423

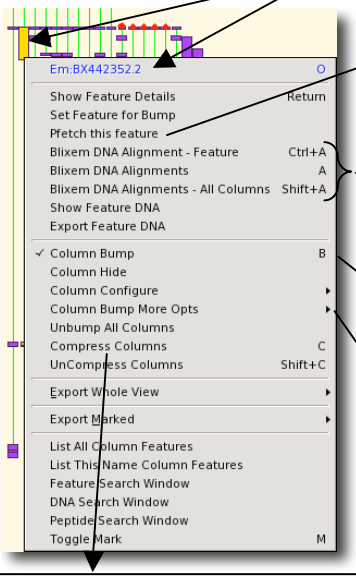
Alignment quality of the HSPs is depicted by the width of every alignment displayed since the width is a measure of that HSP's score. Therefore, the wider it is the closer the score is to 100%. The precise score is displayed in the Zmap details bar by clicking on the alignment. If HSPs are missing either the first or last Blast alignments in the set, they are marked with a red diamond at their start/end respectively. This indicates if they do not start at the first base/amino acid and/or do not end with the last base/amino acid of the alignment sequence. The screen shot below shows what options you get when you right click over a homology – note that you can also select an HSP and type "o". You also get further options such as retrieving the EMBL file for that homology using pfetch and starting **Blixem**, see later section (note, HSPs do not need to be bumped to use **Blixem**).

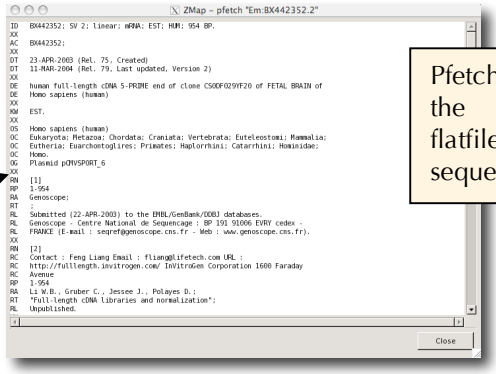
Note the different coloured lines for bumped homologies. The colouring allows you to see all matches for a piece of evidence instantly but also how good the alignment is for the feature you bumped.

Note the red diamonds warning of missing sequence that cannot be aligned.



Right click on the EST of interest for more menu features.



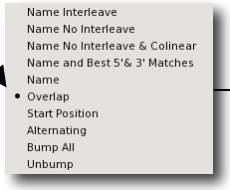


Pfetch returns the EMBL flatfile for that sequence.

Allows you to inspect the sequence of just the chosen feature, all of the column features or all columns, aligned horizontally down to either the nucleotide or amino acid level against the genome. See later section on **Blixem**.

The ✓ shows that the column is bumped. Select it again to unbump it.

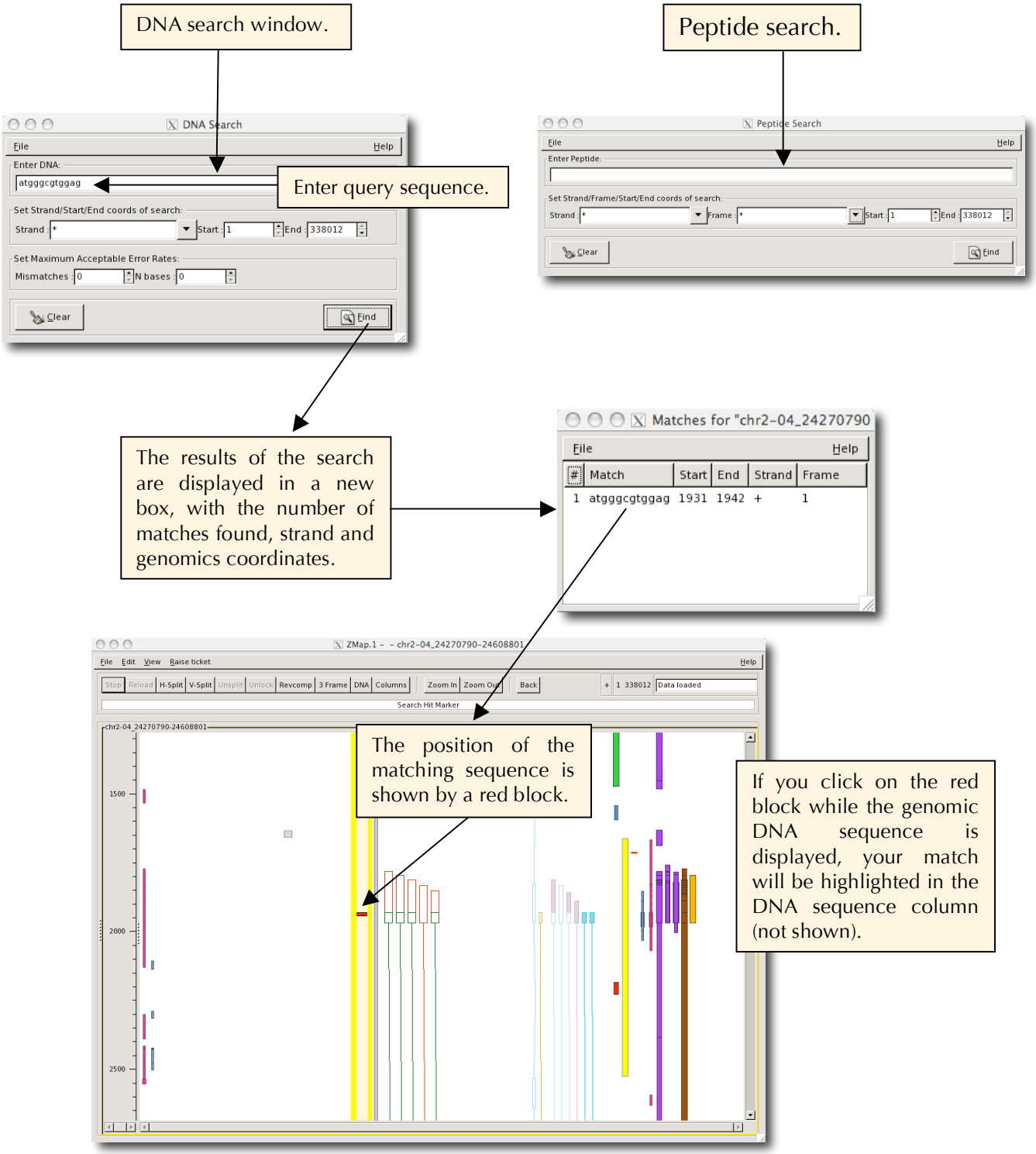
The Compress function removes excess white space by hiding columns that have no features in them, apart from those that have been set to "Show" in the "Columns" menu.



This menu allows you to change the way that bumping is displayed. There are multiple bump options, but the default is the most useful.

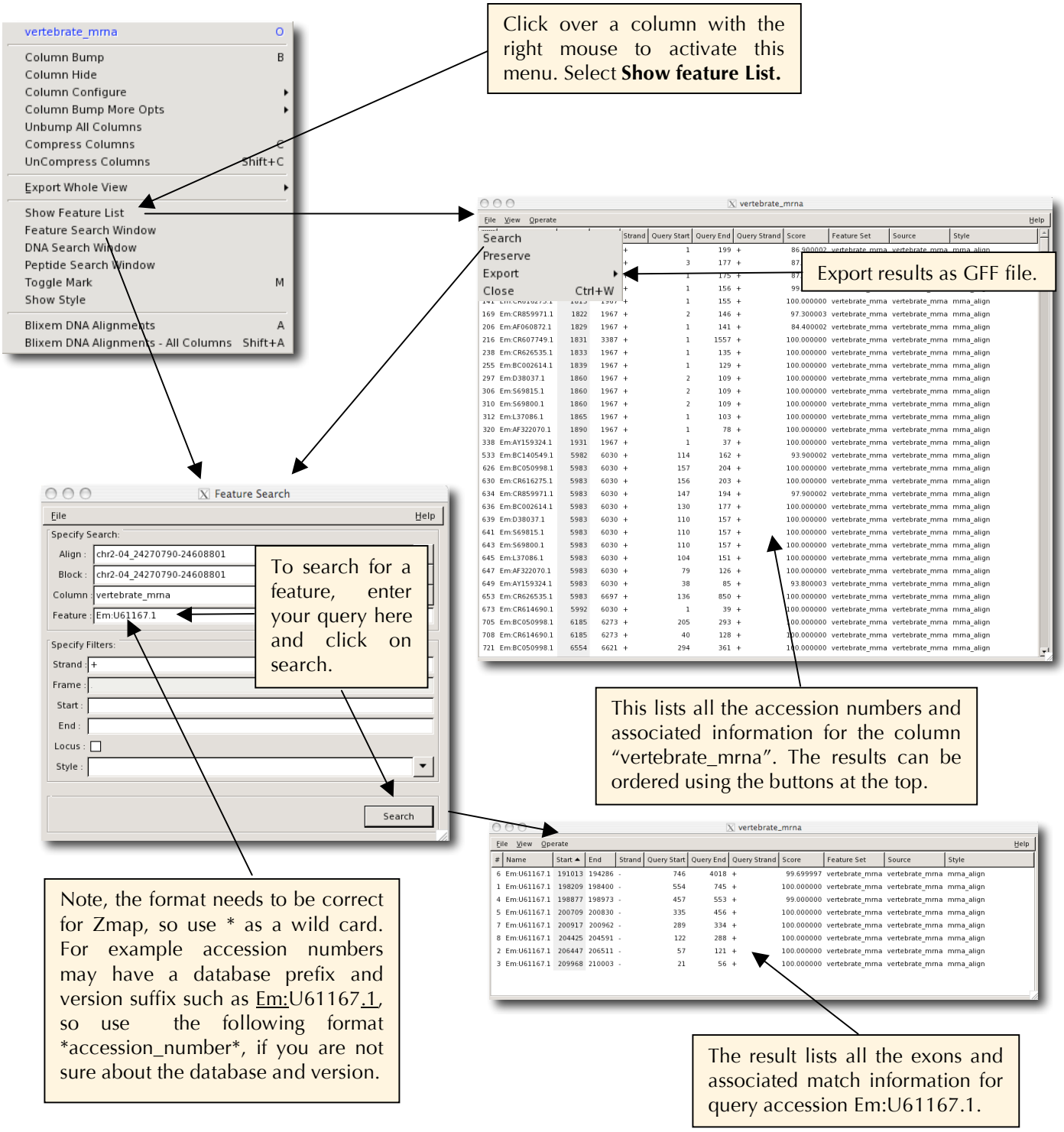
Searching for a sequence in Zmap

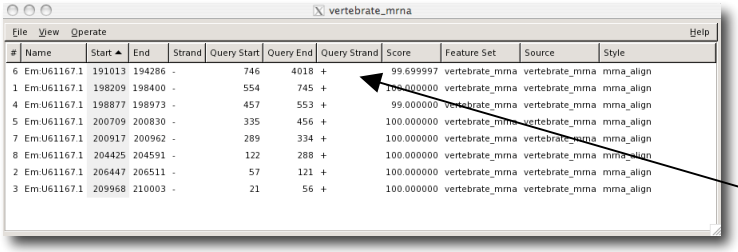
DNA and peptide search windows are provided from within Zmap and can be accessed by right clicking on Zmap space and selecting the option at the bottom of the menu. Both search windows are shown below:



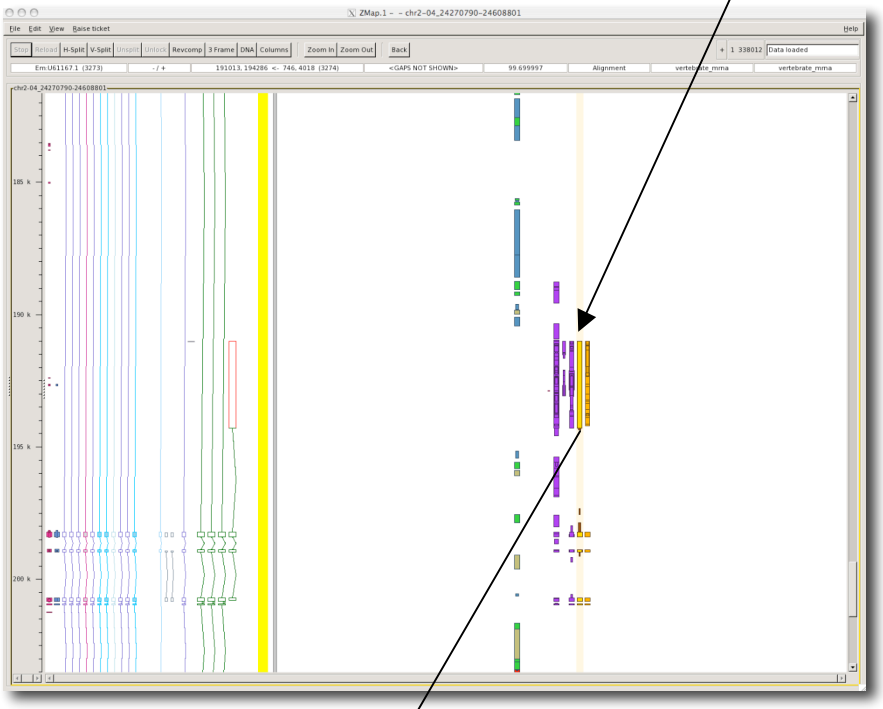
Searching for a feature in Zmap

This option allows you to list all the features contained in a column in one window. There are further options for you to search within these results to find a specific feature. The list of column features can be exported as a GFF file via the File menu.

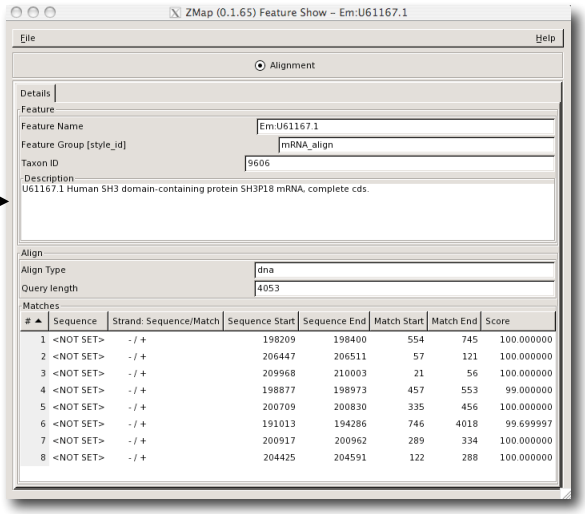




If you now left double click on the match you want to inspect, Zmap will zoom straight to it. Note, this may not work if you are searching for a feature out side of an area that is actively marked.



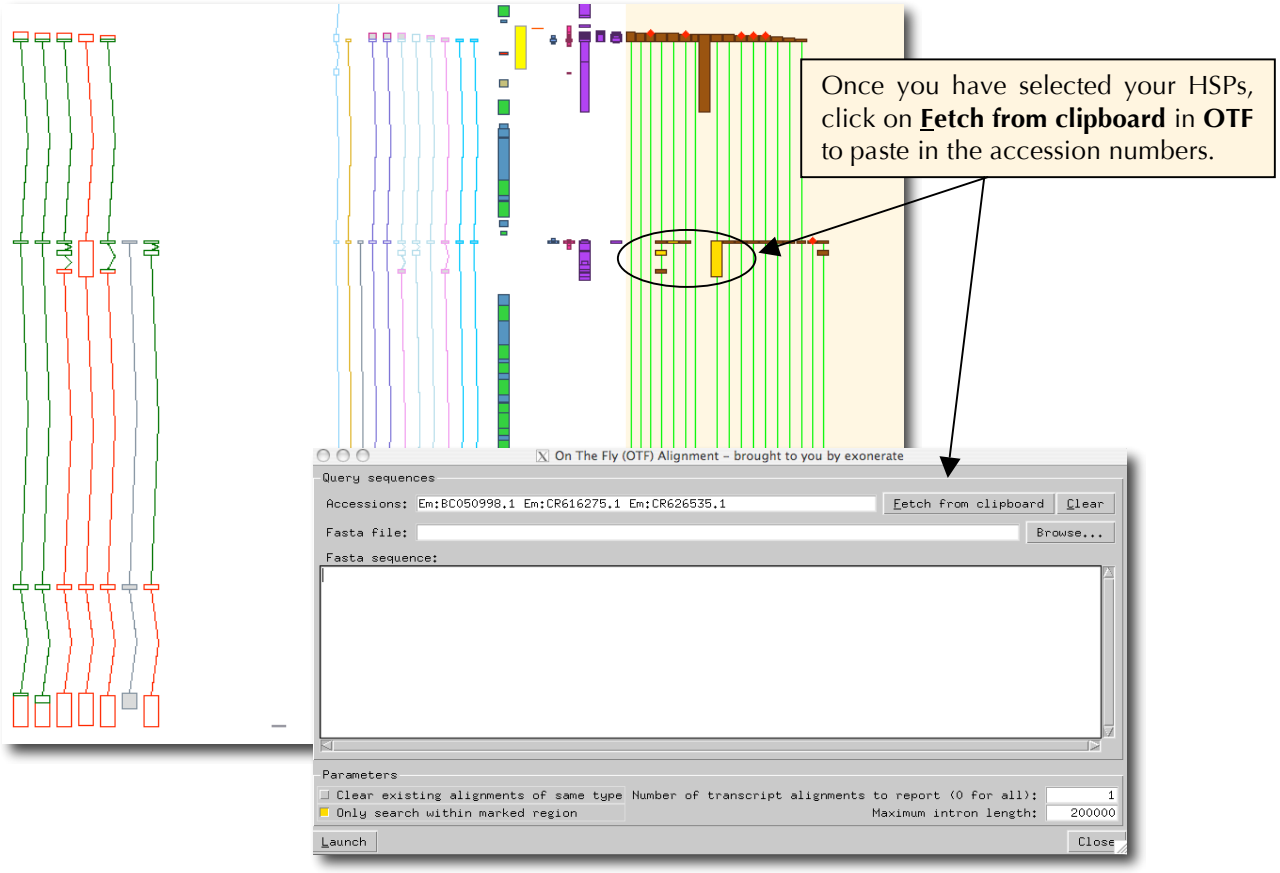
A further window will appear containing information about the feature.



Selecting single or multiple features and hiding/showing them

1) If you left click once on a feature in Zmap, you will highlight all of its exons, the coordinates of which are now stored in the paste buffer and can be copied elsewhere, such as into the transcript editing window in Otterlace.

2) You can select multiple features by holding the Shift key down and left clicking with mouse (same as for multi select on the Mac, Windows etc). This option will highlight a single exon at a time for each feature, but the accession numbers of each feature and the individual exon coordinates are held in the paste buffer. This is a particularly useful way of selecting Zmap hits to use in the OTF alignment tool, as all selected homologies will be held in the paste buffer and automatically pasted into the OTF accession window. Each of the exon coordinates can also be pasted into the transcript editing window in Otterlace.

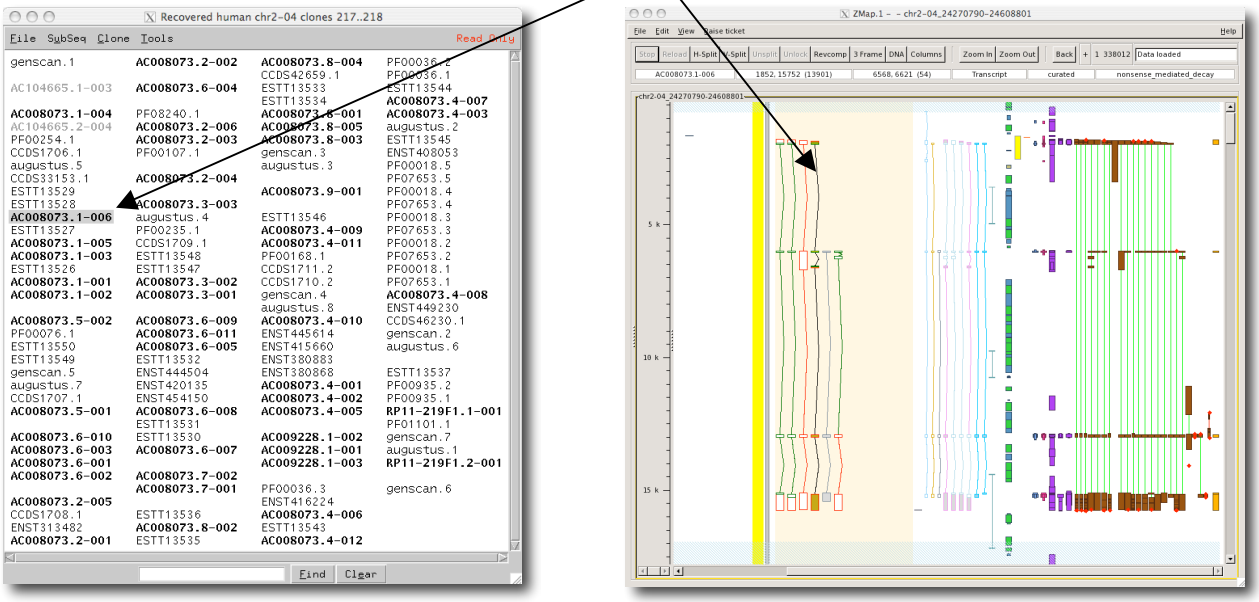


3) You can remove selected features in Zmap by pressing **Delete** on the keyboard and restore them by pressing **Shift-Delete** (note on the Mac you need to press **Fn-Delete** and **Shift-Fn-Delete**). This is a particularly useful way of removing evidence that you have already assigned to a transcript object.

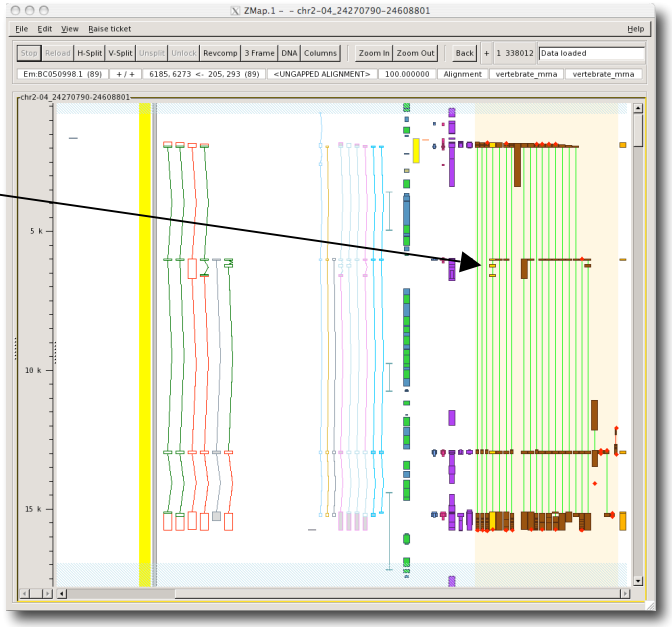
Rapid variant construction

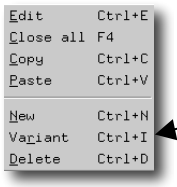
Otterlace and Zmap can be used together to generate variant objects quickly. Existing transcript objects can be used as a template for a new object while a Zmap HSP can be used to provide the coordinates for the new variant. The new object will take its transcript type from the parent.

1) Select the object that will form the foundation to the new variant, either by highlighting the object in Otterlace or clicking on the object in Zmap.

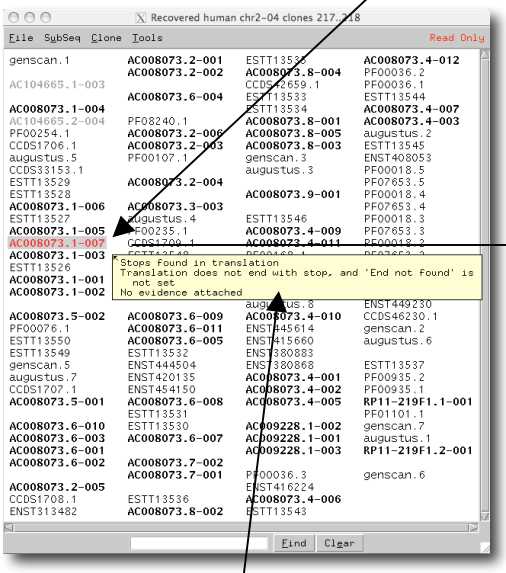


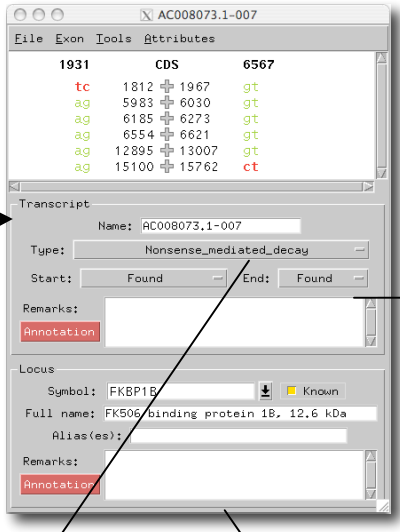
2) Click on the HSP that will give its coordinates to the new variant object.



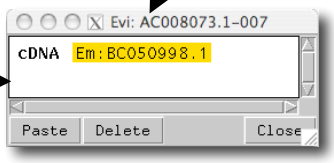


3) Now either use the key-stroke short cut or click on Variant. You will see a new object appear in your main window.

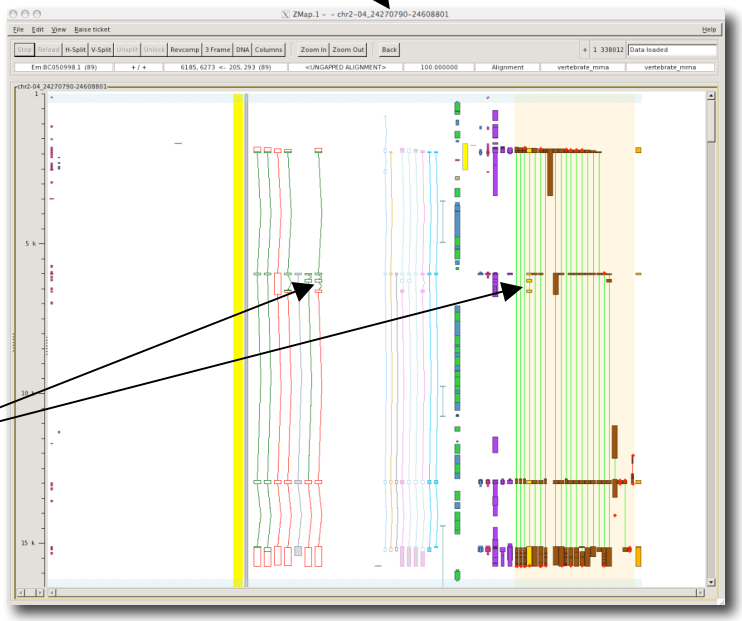




4) The evidence is attached automatically to the new gene object.



4) The new object will inherit its structure from the HSP. **However, you must always check the splice sites of your object in Blixem in case the alignment is incorrect.** Start/end coordinates (if a coding object) and transcript type are inherited from the parent, so these may not be relevant and may need to be changed. Note, that the new object is coloured red due to a number of errors. The checking software will not recognise evidence until the object is saved.



5) Once the errors have been removed, save the object to see it appear on Zmap (the evidence used has been highlighted).

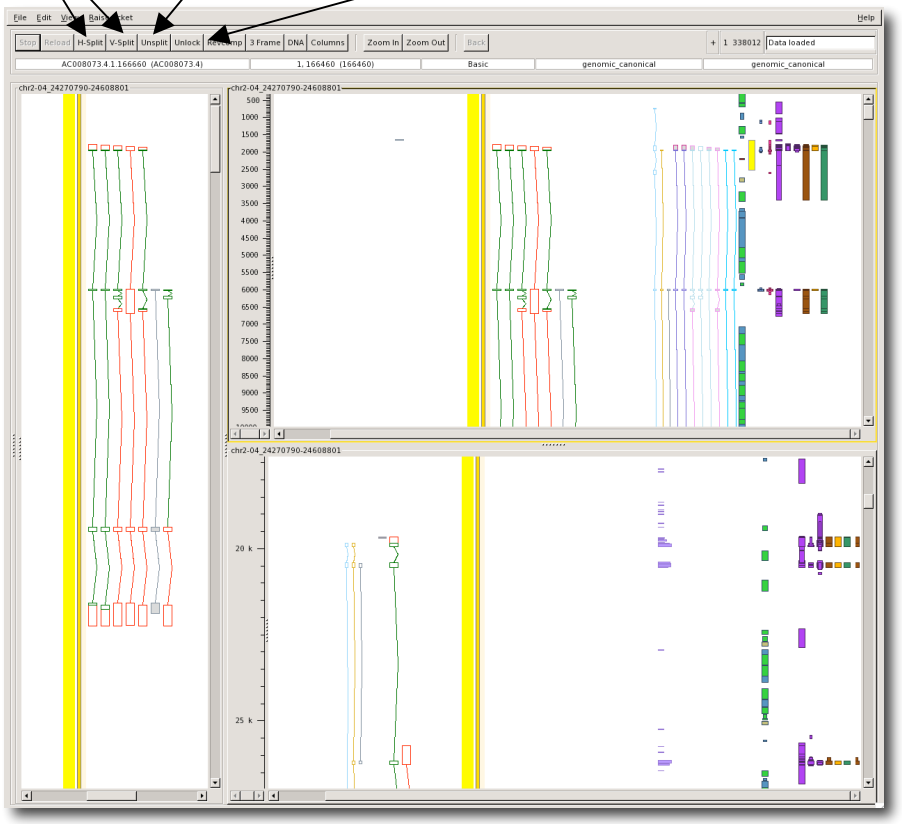
Splitting windows in Zmap

Use the **split** window function to effectively reduce the size of the window when looking at homologies. This is of particular use when you have to deal with very large introns because you can essentially reduce the introns to whatever size you wish, or when there are very many HSPs, because you can keep your gene object in view and static, but still scroll across the evidence.

The screen can be split horizontally or vertically (as shown) multiple times. An active window must be selected for **splitting**.

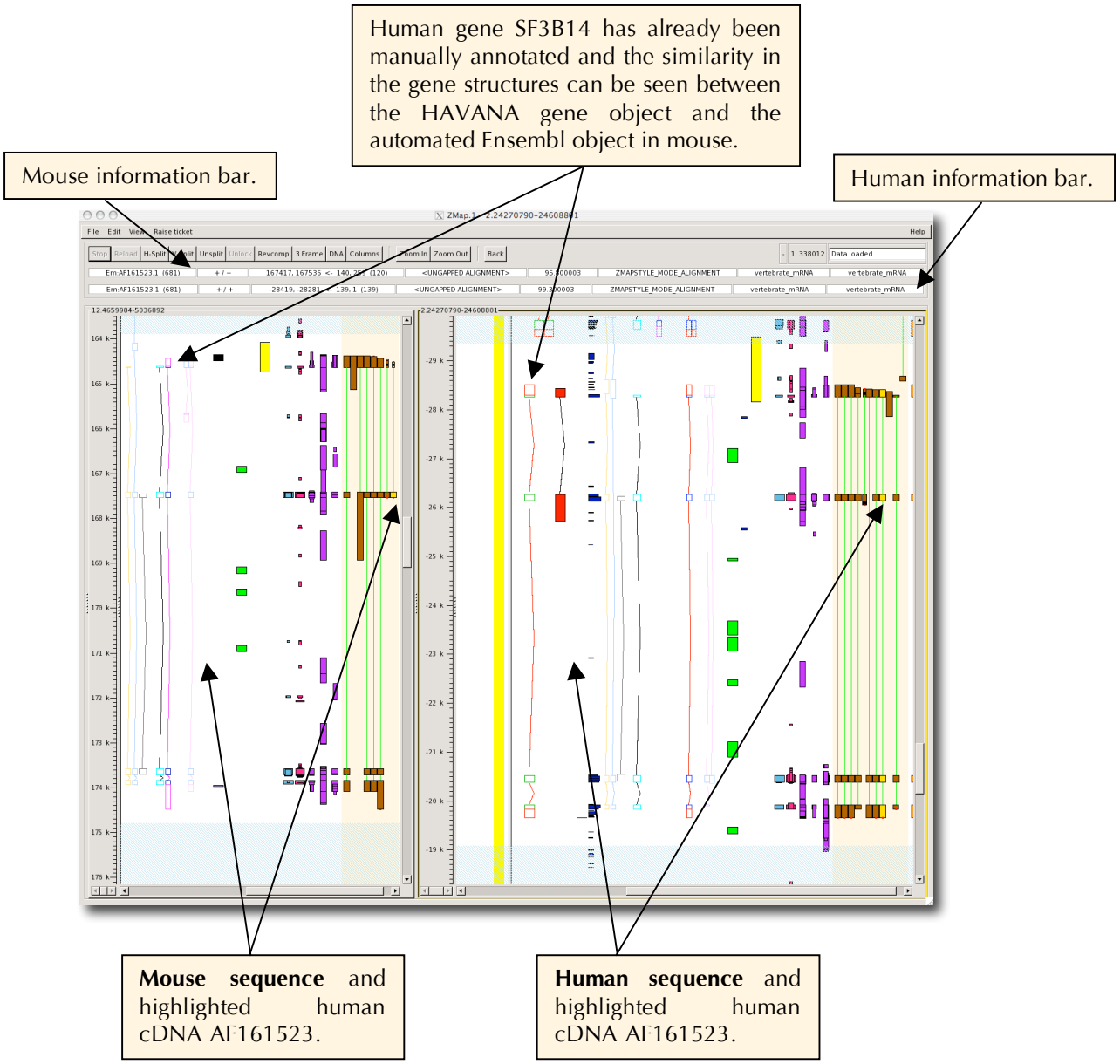
Unsplit will remove the last split window.

The windows will be locked together when you first open them. To scroll independently within each window, use the **Unlock** button.



Launching in a Zmap

This function allows you to open two or more sequences alongside each other (such as a human region and the syntenic region in mouse, or two haplotypes), so that simultaneous investigation can be carried out. To do this you will need to open both sets of clones in the same Otterlace session. To open both Zmap windows in one window as shown below, you need to select "Open in a Zmap" option in one clone set. These clones will open to the left of the already open Otterlace session. This screen shot shows human gene SF3B14 and the syntenic region in mouse. The gene copy and paste function (referred to in the Otterlace section) is of much use here, saving time when building gene objects.



Zmap keyboard and mouse shortcuts.

In general Zmap will be faster for zooming, bumping etc if you make good use of the built in short cuts. These can often avoid the need for Zmap to redraw large amounts of data that you may not even be interested in. For example, click once (highlight) on a feature and a carriage return will bring up evidence. Another example is to press T for translation.

All windows

Short Cut	**Action**
Cntl-W	close this window
Cntl-Q	quit ZMap

Zmap Window

Short Cut	**Action**
Control keys	
+ (or =), -	zoom in/out by 10%
Cntl + (or =), Cntl -	zoom in/out by 50%
up-arrow, down-arrow	scroll up/down slowly bit
Cntl up-arrow, Cntl down-arrow	scroll up/down more quickly
left-arrow, right-arrow	scroll left/right slowly
Cntl left-arrow, Cntl right-arrow	scroll left/right more quickly
page-up, page-down (Mac users should use fn and up/down arrow)	up/down by half a "page"
Cntl page-up, Cntl page-down	up/down by a whole "page"
Home, End (Mac users should use fn and left/rights arrows)	Go to far left or right
Cntl Home, Cntl End (Mac users will have to configure their keyboards for this)	Go to top or bottom
Delete, Shift Delete	Hide/Show selected features.
Enter	Show feature details for highlighted feature.
Shift up-arrow, Shift down-arrow	Jump from feature to feature within a column.
Shift left-arrow, Shift right-arrow	Jump from column to column.

Alpha-numeric keys

b	bump/unbump current column within limits of mark if set, otherwise bump the whole column.
B	Bump/unBump current column within limits of the visible feature range.

c	compress/uncompress columns: hides columns that have no features in them either within the marked region or if there is no marked region within the range displayed on screen. Note that columns set to "Show" will not be hidden.
C	Compress/unCompress columns: hides all columns that have no features in them within the range displayed on screen regardless of any column, zoom, mark etc. settings.
h	Toggles highlighting (good for screen shots).
m	mark/unmark a range which spans whichever features or subparts of features are currently selected for zooming/smart bumping
M	Mark/unMark the whole feature corresponding to the currently selected subpart (e.g. the whole transcript of an exon or all HSPs of the same sequence as the highlighted one) for zooming/smart bumping
o or O	show menu Options for highlighted feature or column, use cursor keys to move through menu, press ESC to cancel menu.
r	reverse complement current view, complement is done for all windows of current view.
t or T	translate highlighted item, T hides Translation.
w or W	zoom out to show whole sequence
z	zoom to the extent of any selected features (e.g. exon/introns, HSPs etc) or any rubberbanded area if there was one.
Z	Zoom to whole transcript or all HSPs of a selected feature.

Zmap Mouse Usage

Left	**Middle**	**Right**
Single mouse button click		
highlight a feature or column Plus drag: draw a rectangle around an object for zoom	horizontal ruler with sequence position displayed, on button release centre on mouse position. Release mouse outside Zmap window to prevent re-centering.	show feature or column menu – for options such as pfetch, show feature DNA, show peptide, export peptide
Double mouse button click		
display details of selected feature. Double click on object to get edit window	same as single click	same as single click
Shift + mouse button click		
highlight a subpart of a feature (e.g. a single exon or alignment match) OR multiple highlight	same as single click	same as single click

Tips for a speedier Zmap

1. Specifically: zoom and mark within Zmap early on after launching. Either select a gene object and press 'z' to zoom OR select a rectangle to zoom in by dragging the left mouse button around it. Reverse complement now if necessary, then press 'm' to mark the region.

2. The quickest way to zoom out of Zmap again is to right mouse click on the 'zoom out' buttons at the top of zmap and choose one of the options (this is definitely much quicker that doing individual 'zoom outs' with the left mouse button). Likewise for 'zooming in' again (or use keyboard equivalents).

3. Bump within a marked region only. Bumping without marking is slow and removes the lines connecting Blast matches.

4. When you have finished working within a marked region, unbump the evidence you have been working on (e.g. ESTs) and unmark that region before you go on to select the next region to mark and bump – or you could miss visualising the evidence in the new region.

5. If you want to get rid of some white space try the compress 'c' function or alternatively toggle off some of the columns. Warning – this may hide features as well. If a column (e.g ESTs) is bumped and you want to lose it temporarily, it is quicker to turn the column off (when you turn it on again it will still be bumped when it re-appears) than unbump then rebump again later.

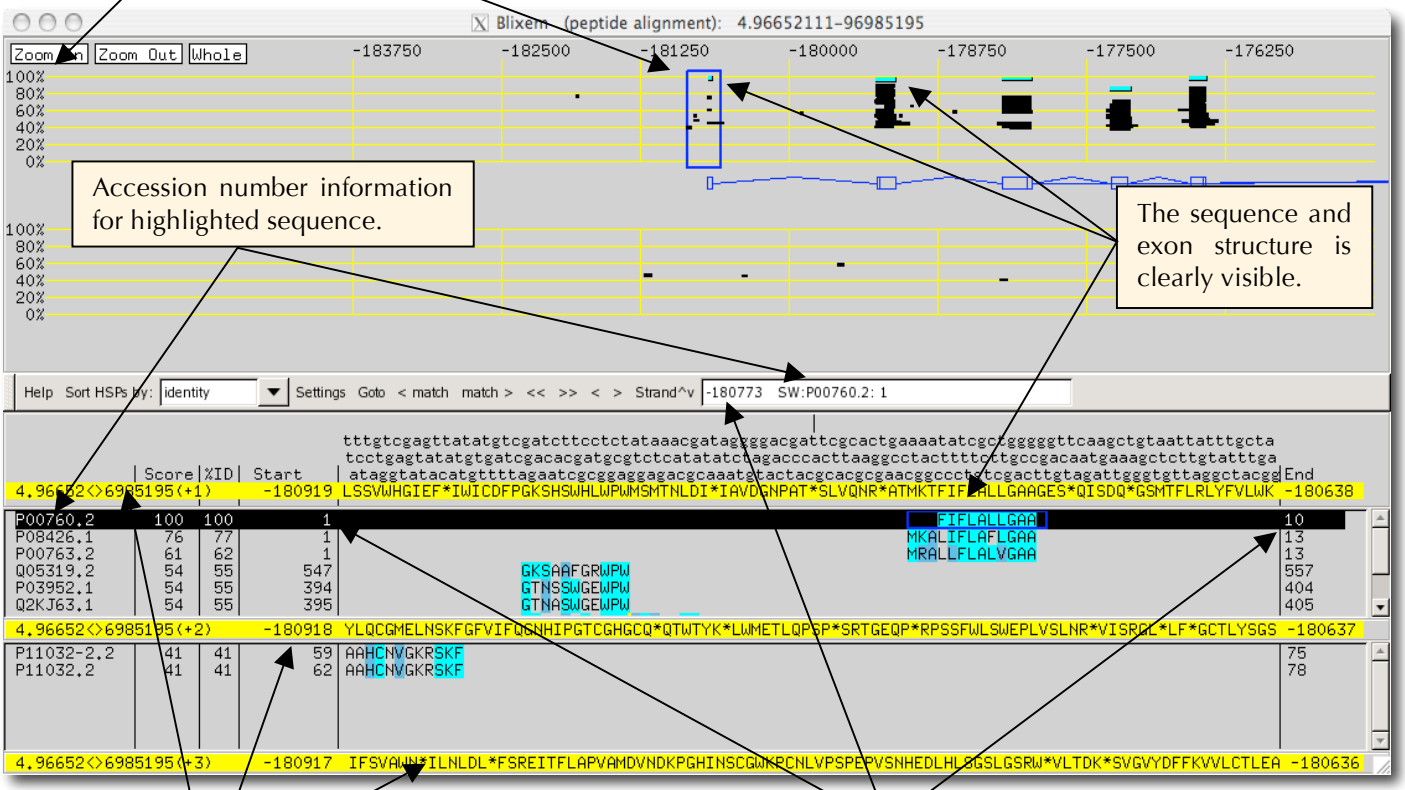
6. Jumping to genes/objects: If you expand the left hand 'scroll navigator' overview' you can jump directly to genes and objects by double-clicking on them.

Blixem

Blixem, which stands for "BLast matches In an X-windows Embedded Multiple alignment", is an interactive browser of pairwise Blast matches displayed as multiple alignments. It is not strictly a multiple alignment tool, rather a 'one-to-many' alignment. It is used to check the Blast alignments of nucleotide and amino acid sequences against a reference sequence. The Blixem range is set using the preferences option in Zmap – default is 200kb, so it is generally necessary to expand the range.

Homologies can be ranked from 0% to 100% (see scale). As such, Blixem is useful for inspecting gene families.

Navigation in Blixem is similar to that in Zmap. Click with the middle mouse button and a hairline appears; release the button and the display refreshes, centered on where the cursor was. Alternatively use the < match << < etc. buttons or drag the blue box in the overview panel with the middle mouse button.

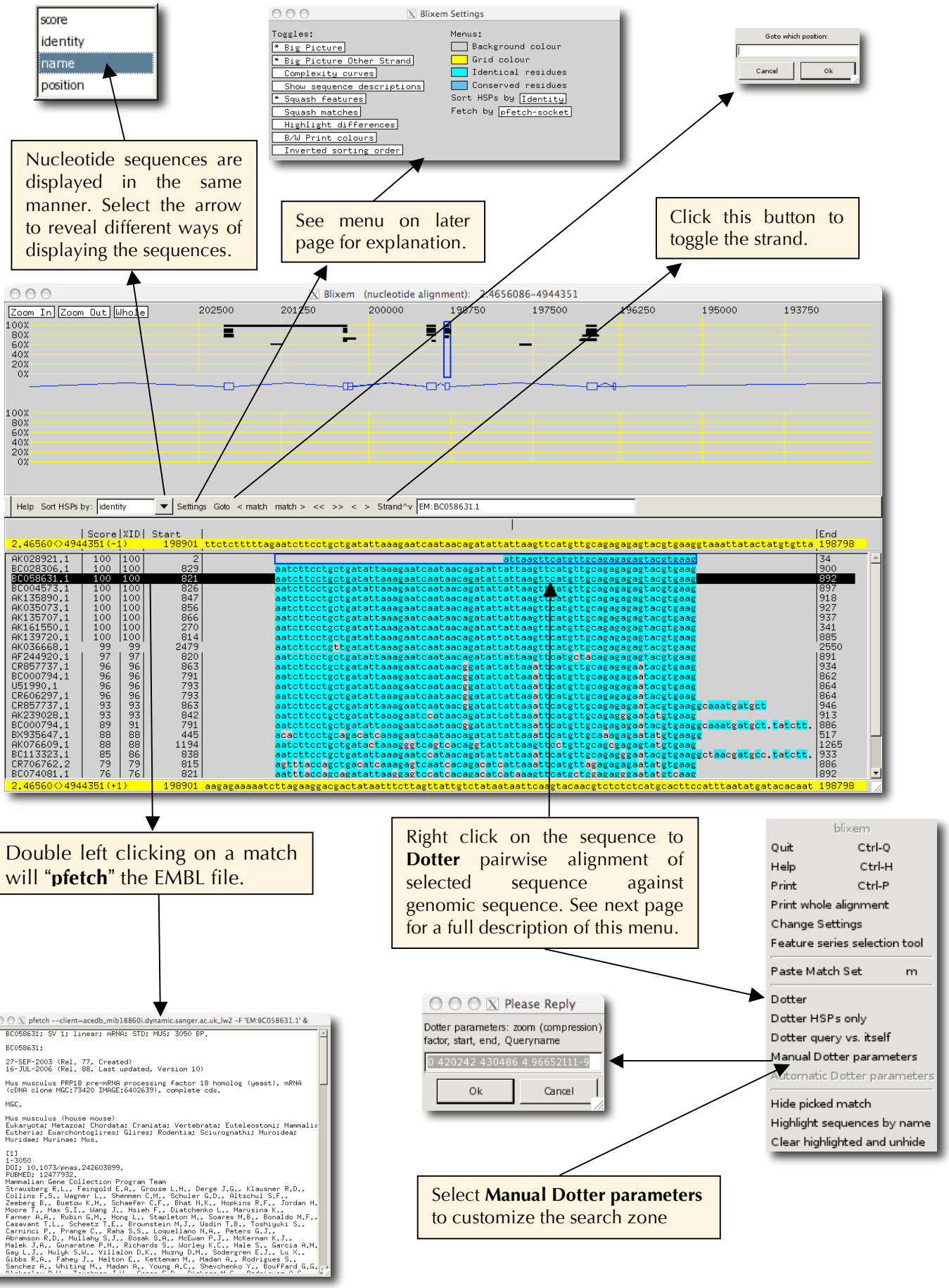


Accession number information for highlighted sequence.

The sequence and exon structure is clearly visible.

3 frame translation (protein Blixem only) with coordinates for that window.

Coordinate information for sequence and clone set.


score
identity
name
position
Nucleotide sequences are displayed in the same manner. Select the arrow to reveal different ways of displaying the sequences.
Blixem Settings
Toggles:
* Big Picture
* Big Picture Other Strand
Complexity curves
Show sequence descriptions
* Squash features
Squash matches
Highlight differences
B/W Print colours
Inverted sorting order
Menus:
Background colour
Grid colour
Identical residues
Conserved residues
Sort HSPs by Identity
Fetch by pfetch-socket
See menu on later page for explanation.
Goto which position:
Cancel
Ok
Click this button to toggle the strand.
Blixem (nucleotide alignment): 2.4656086-4944351
Zoom In
Zoom Out
Whole
Help
Sort HSPs by: identity
Settings
Goto
< match
match >
<<
>>
<
>
Strand^v
EM:BC058631.1
Double left clicking on a match will "pfetch" the EMBL file.
Right click on the sequence to Dotter pairwise alignment of selected sequence against genomic sequence. See next page for a full description of this menu.
blixem
Quit Ctrl-Q
Help Ctrl-H
Print Ctrl-P
Print whole alignment
Change Settings
Feature series selection tool
Paste Match Set m
Dotter
Dotter HSPs only
Dotter query vs. itself
Manual Dotter parameters
Automatic Dotter parameters
Hide picked match
Highlight sequences by name
Clear highlighted and unhide
Please Reply
Dotter parameters: zoom (compression) factor, start, end, Queryname
0 420242 430486 4.96652111-9
Ok
Cancel
Select Manual Dotter parameters to customize the search zone

Blixem main menu, press right mouse button anywhere. **Note** that options in grey have no current working function and as such there are no further details on them.

Quit	Exit program.
Help	Get brief help.
Print	Print the currently displayed window.
Print whole alignment	Print all matches. May produce many pages.
Change Settings	Start the Settings tool.
Feature series selection tool	Under development.
Dotter	Do a dotter dotplot with last picked matching sequence.
Dotter HSPs only	Display blast matches in a Dotter dotplot.
Dotter query vs. itself	Call Dotter for a region of the query vs. itself. This is useful to analyse internal repeats etc.
Manual Dotter parameters	Use if Dotter estimates the start and end coordinates wrongly.
Automatic Dotter parameters	
Hide picked match	Removes highlighted match in Blixem.
Highlight sequences by name	Define a template, e.g. *human, to highlight all human proteins. This works on the name field in the leftmost column. Note that the names start with a database prefix that is hidden in blixem, so it's a good idea to always start your template with the wildcard *.
Clear highlighted and unhide	Reset all picked and highlighted sequences.

Blixem settings menu / settings tool (press right mouse button pull-down on "Settings", or click once on it with left mouse button)

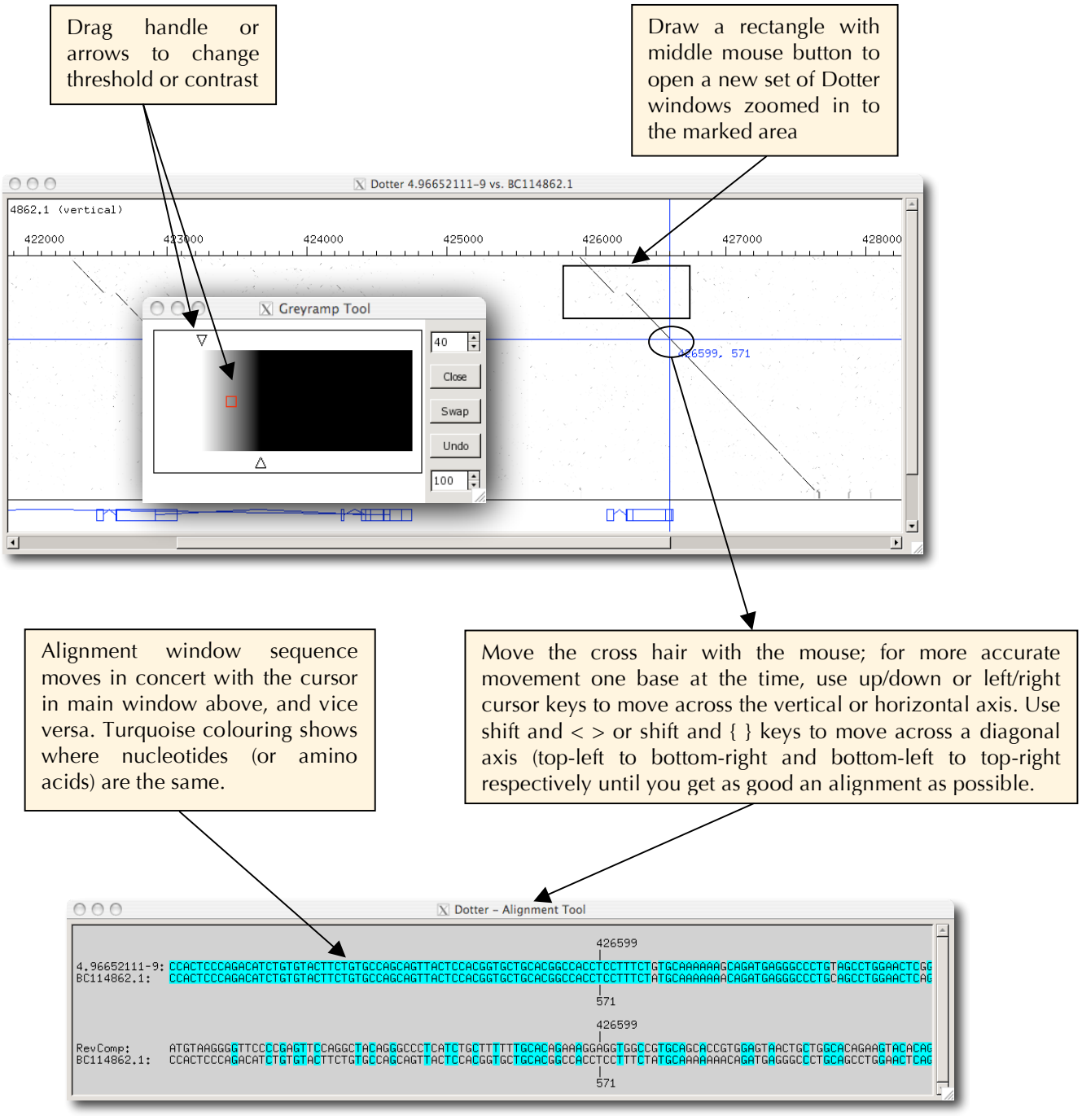
Click on Sort HSPs by identity:

Sort by score	Sort all proteins with the highest-scoring first.
Sort by identity	Sort all proteins with the most identical first.
Sort by name	Sort all proteins alphabetically.
Sort by position	Sort all proteins with the most N-terminal first.
Big Picture	Toggle Big Picture (top display) on/off.
Big Picture Other strand	Toggle between single and double strand display in the Big Picture.
Complexity curves	Draws plots of low complexity at 3 different window sizes.
Show sequence descriptions	
Squash features	
Squash matches	
Highlight differences	Shows identical residues as a dot (.) and draws mismatching residues in bright blue.
B/W Print colours	Good for printing on black/white printers.
Inverting sorting order	Inverts order of sequences
Menus:	Change the colours of the Blixem display by right clicking on the box and selecting your choice of colours.
Fetch by	Changes method of fetching sequences.

For further information on Blixem, visit this website:
http://sonnhammer.sbc.su.se/Blixem.html

Dotter

Dotter is a graphical dotplot program for detailed comparison of two sequences. Here, every residue in one sequence is compared to every residue in the other sequence. The first sequence runs along the x-axis and the second sequence along the y-axis. In regions where the two sequences are similar to each other, a row of high scores will run diagonally across the dot matrix. For further information on Dotter, visit this website: http://sonnhammer.sbc.su.se/Dotter.html



Annotation resources

AspicDB – useful analysis of splice junctions http://t.caspur.it/ASPicDB/

CCDS http://www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi

Ensembl genome browser http://www.ensembl.org/index.html

Entrez Gene for nucleotide and protein sequence, cloning, gene information etc http://www.ncbi.nlm.nih.gov/sites/gquery

HORDE database for olfactory receptors http://genome.weizmann.ac.il/horde/

Swiss Institute of Bioinformatics has many tools for analysing nucleotide and protein sequences http://www.expasy.ch/

UCSC genome browser http://genome.ucsc.edu/cgi-bin/hgGateway

UniProt has protein sequence information http://www.uniprot.org/

Vertebrate Genome Annotation Browser for manual annotation http://vega.sanger.ac.uk/index.html