# ZMap User Manual

Written by and contributions from:
Gemma Barson <gb10@sanger.ac.uk>
Charlie Stewart
Laurens Wilming
Ed Griffiths
James Gilbert
Jennifer Harrow

## Revision History

| Revision | Date | Author |
|---|---|---|
| First revision (ZMap v1.5.0) | 16/03/2015 | Gemma Barson |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Contents

# Introduction

ZMap is a software package that provides a visualisation and editing tool for genomic features. The software is written in C/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 privides facilities for editing and creating features, which can then be saved to a GFF file.

ZMap is part of the Otter annotation suite and contains a protocol for two-way communication with Otter so that it operates seamlessly as an additional display within Otter.

ZMap can call tools from the SeqTools suite of sequence visualisation tools to analyse alignments in more detail.

# Using ZMap

## Running ZMap on a GFF file

To run ZMap standalone on one or more GFF files, simply pass the file name(s) on the command line:

```
zmap file1.gff file2.gff
```

The sequence region will default to the region available in the GFF file. Alternatively, you can specify the sequence/region by passing the `-sequence`, `-start` and `-end` arguments on the command-line. If all of the files contain the same reference sequence, then their contents will be shown in the same view. If the files contain multiple sequences/regions, then they will be shown in multiple views.

## Running ZMap with a config file

If you have several sources, it can be convenient to specify them in a config file rather than typing them on the command-line. You can also then set up many different options and styles to use for these sources. To use a local GFF file as a source, a simple config file would look something like this:

```
[ZMap]
# The list of sources
sources=chr6-18

# Sequence name/start/end can be specified here, on the command-line, or
# omitted if they are to be taken from the file
sequence=chr6-18
start=2696324
end=2864370

# Source stanzas (one per source)
[chr6-18]
url=file:////Users/gb10/work/checkout/zmap/examples/features.gff
```

ZMap can then be started by passing the config file on the command line:

```
zmap --conf_file=/path/to/file
```

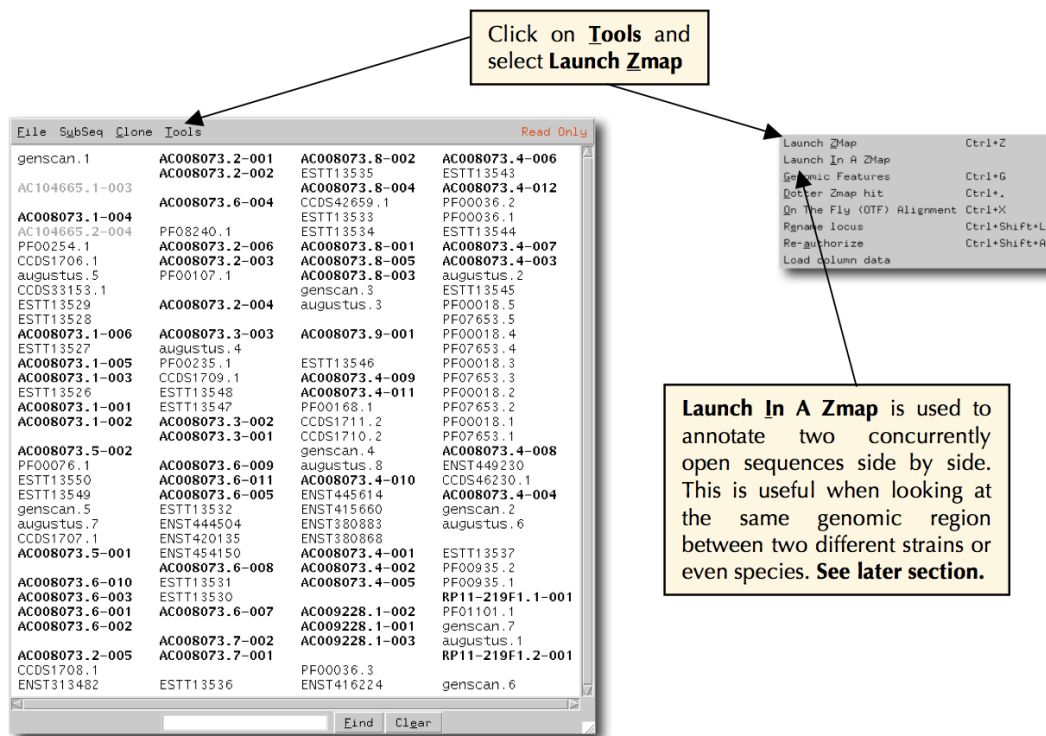Alternatively, the config file can be saved as `/.ZMap/ZMap` and will be picked up by default.

Figure 1: Opening ZMap from Otter

## Operation within Otter

ZMap is opened via the Tools menu bar in Otter (figure 1).

# Main ZMap interface

## 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. See Figure 3.

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). See figure 4.

5

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.

Figure 2: The main ZMap interface

6

Figure 3: Navigating

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.
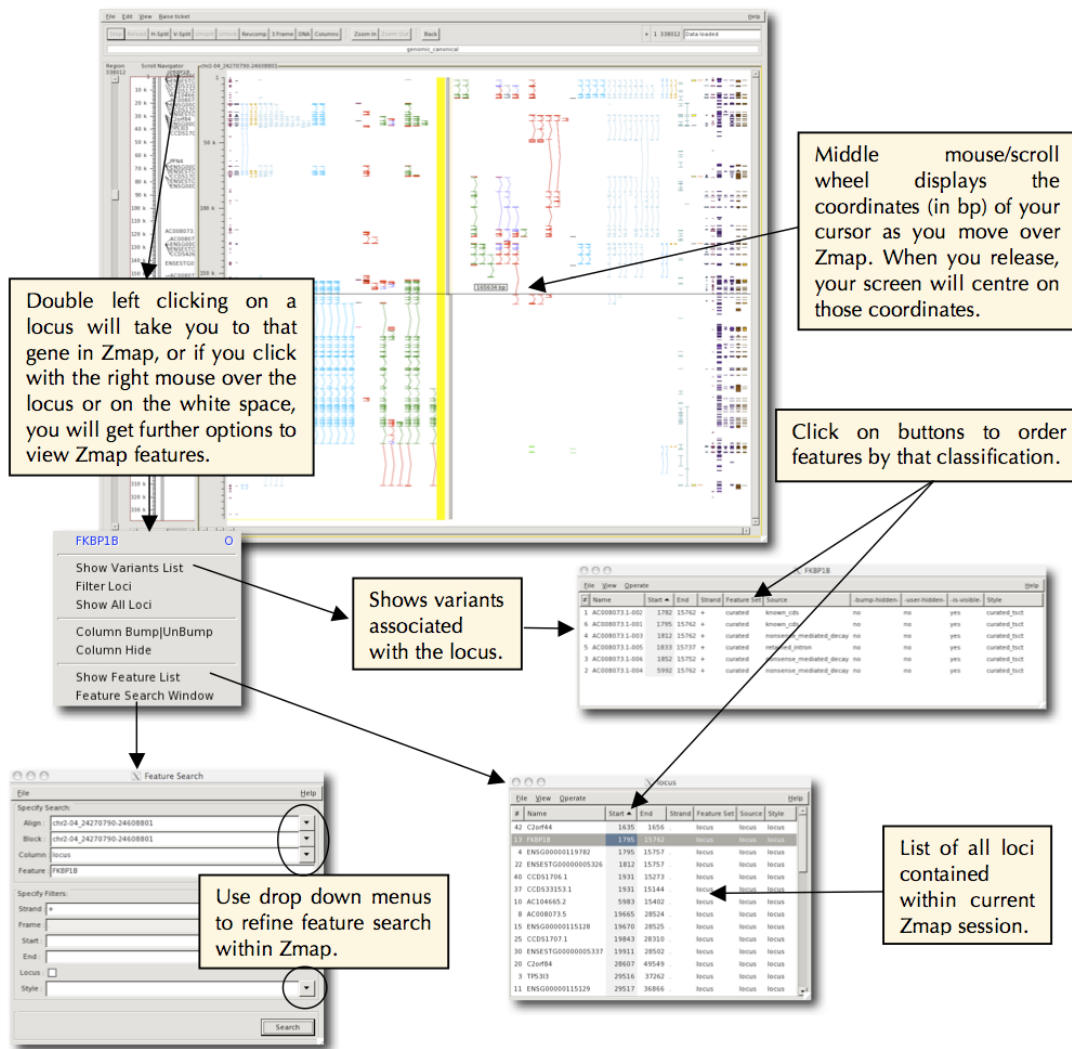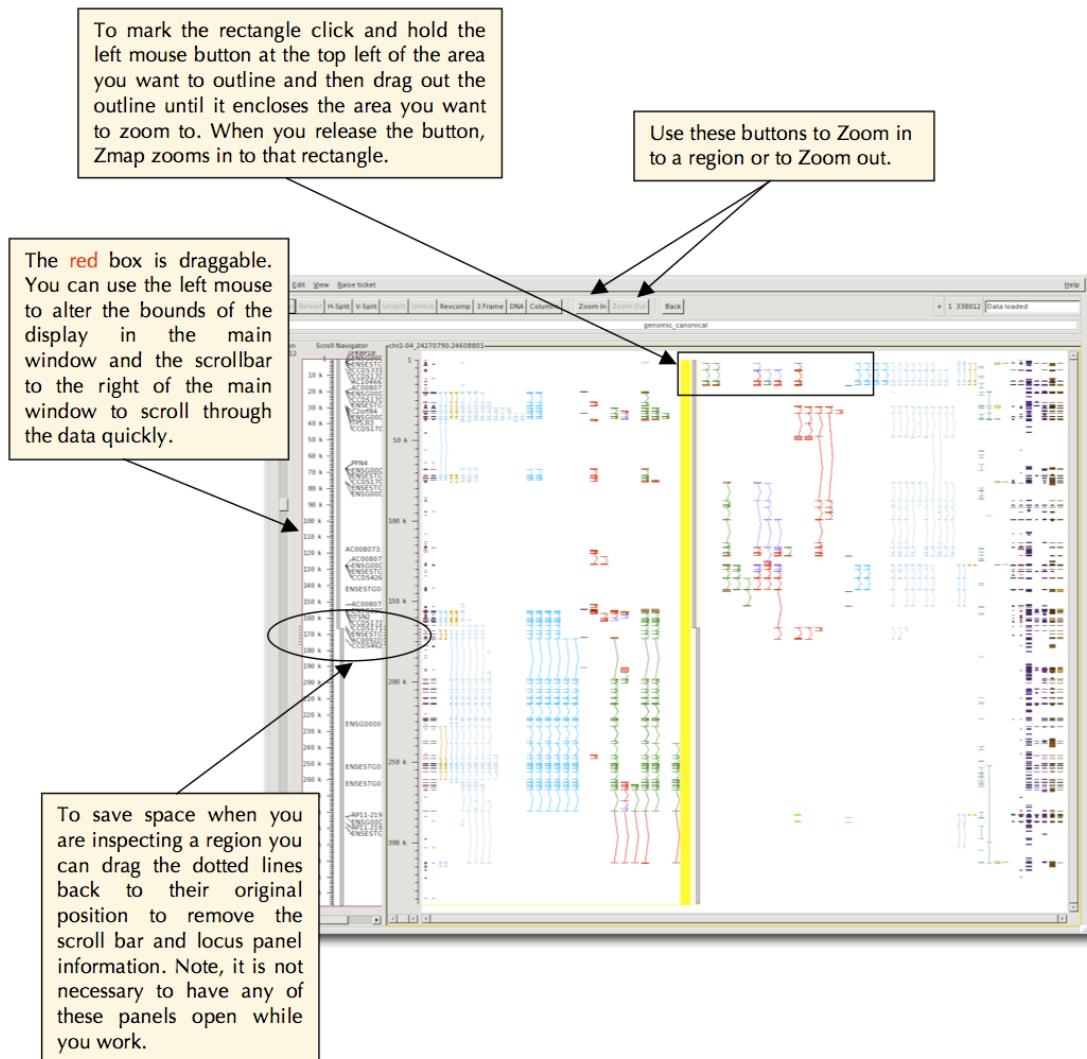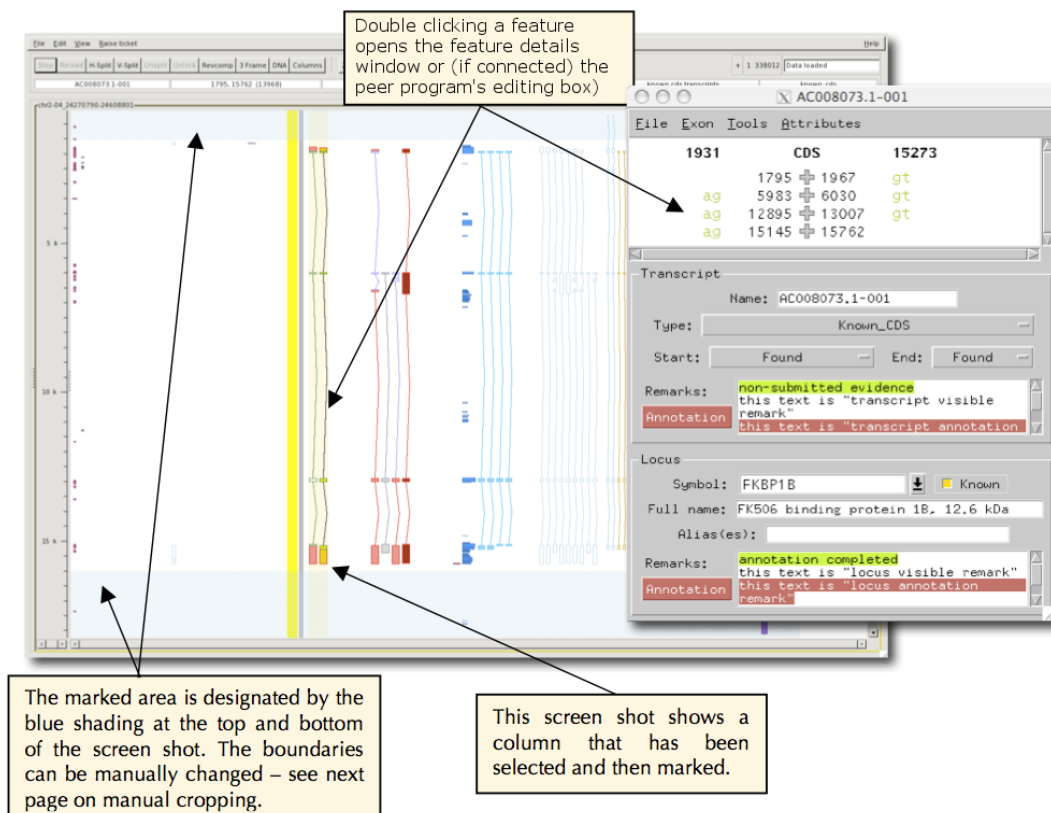
Figure 4: Zooming

8

Figure 5: Focus and Mark

## 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 figure 5):

**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:

9

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 shown in figure 6.
Note - you may see more or fewer features and columns depending on how your preferences are set up. For descriptions of all 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) and purple shows the "coding" region of NMD variants. Grey transcripts (see dotted line) contain exons outside the sequence slice being viewed and should not be confused with Halfwise hits.

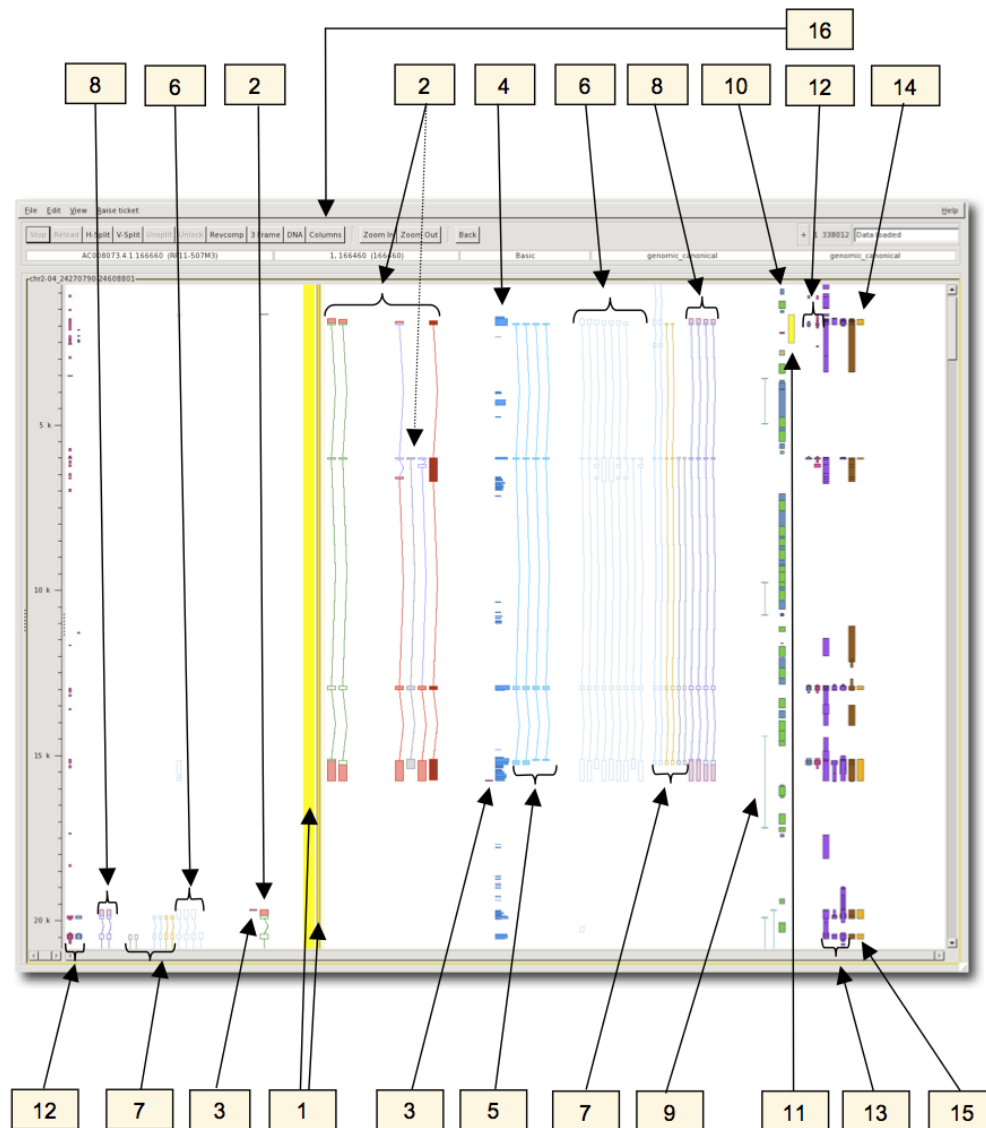3. Curated features, such as PolyA features are seen as horizontal black lines.
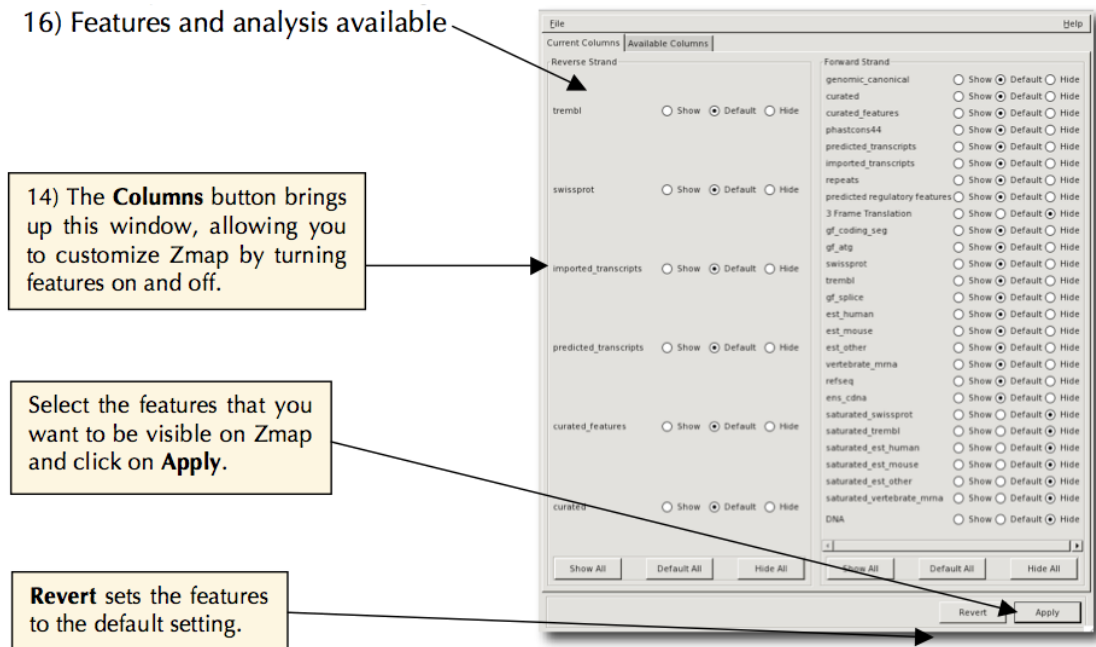
Figure 6: Features

Figure 7: Columns dialog

4. Phastcons44 - conserved regions detected using multiple sequence alignments of 44 organisms.

5. Imported annotation from CCDS (human and mouse only).

6. Imported transcripts via DAS source. Here PASA_ESTs are shown.

7. Predicted transcripts such as Genscan (pale blue), Augustus (gold) and Halfwise predictions of Pfam (grey).

8. Imported annotation from Ensembl.

9. gis_pet_ditags and chip_pet_ditags are indicators of transcript boundaries.

10. Repeats ( blue=Line , light green=Sine , gold=other ), tandem repeats are red.

11. CpG islands appear as yellow boxes.

12. Protein matches are strand specific - SwissProt are light blue and Trembl pink.

13. 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.

14. mRNA matches contains all species and are displayed as brown blocks,

15. RefSeq matches are the orange blocks.

16. Features and analysis available (see figure 7).

## Toolbar

Many functions are available from the toolbar section in ZMap. See figures 8 and 9.

This window sets the range for **Blixem**. The default setting is 200,000 bp. However, you can set it to a more appropriate range for the clones you are annotating. The range must be reset when you start a new Zmap window.

Access to help menu

Contact Helpdesk

When any of the features are clicked on, information about them will be displayed in the panels along the top of the screen e.g. the feature name or accession number, coordinates, length of match, % identity, exon length, etc.

AC008073.1 is a curated transcript with type known_cds.

Place the mouse over the buttons to get further information about its function, such as to **reverse complement** your sequence.

Use the **Back** button to undo the last marking or zooming action.

Some buttons have further options when you right click over them.

Figure 8: Toolbar feature details

13

The **DNA** button will show the nucleotide sequence. If you click on an exon, the sequence is highlighted in red. 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.

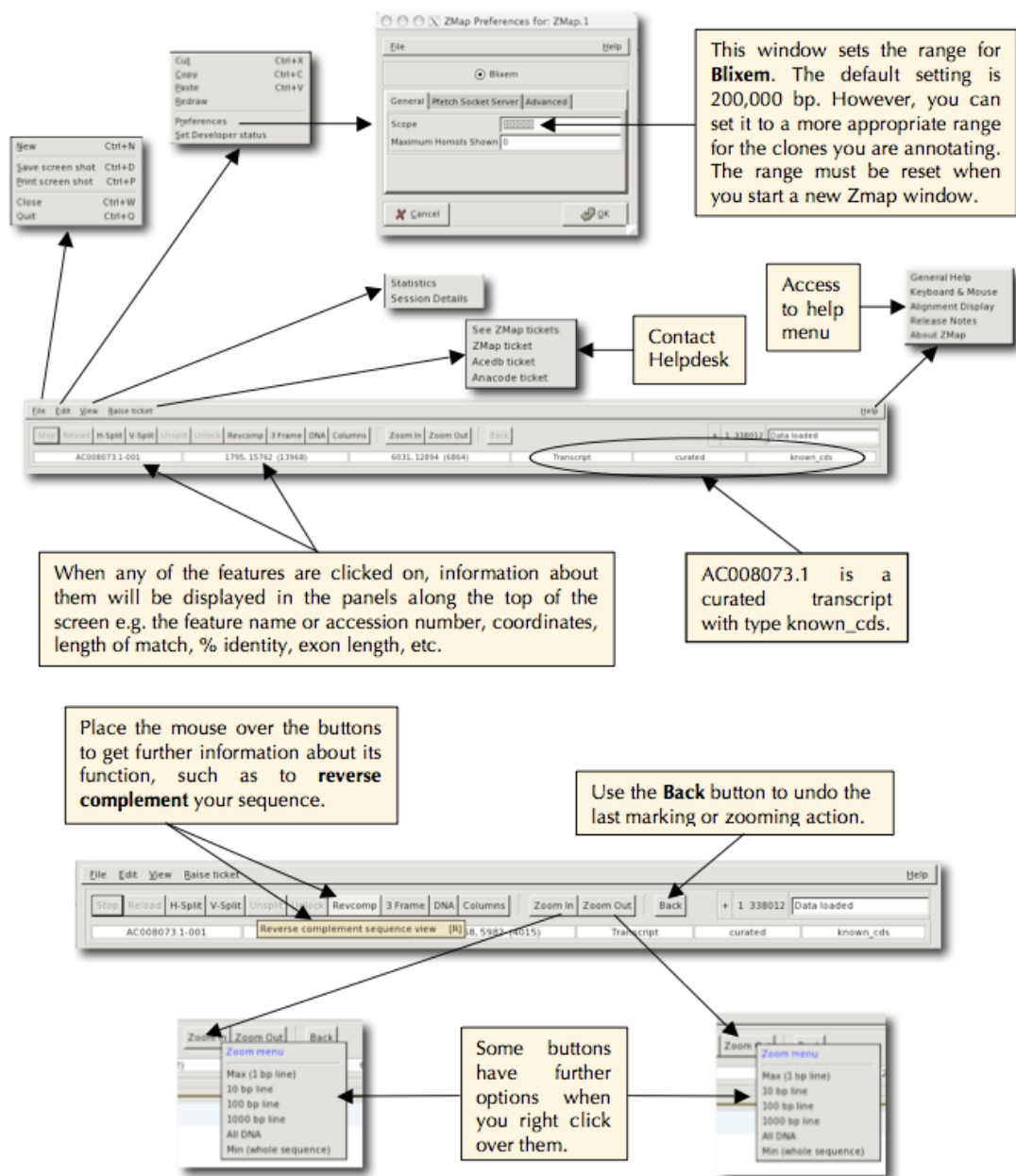The **3 Frame** button will show the amino acid sequence in each of the three reading frames. If you click on an exon, the sequence is highlighted in red.
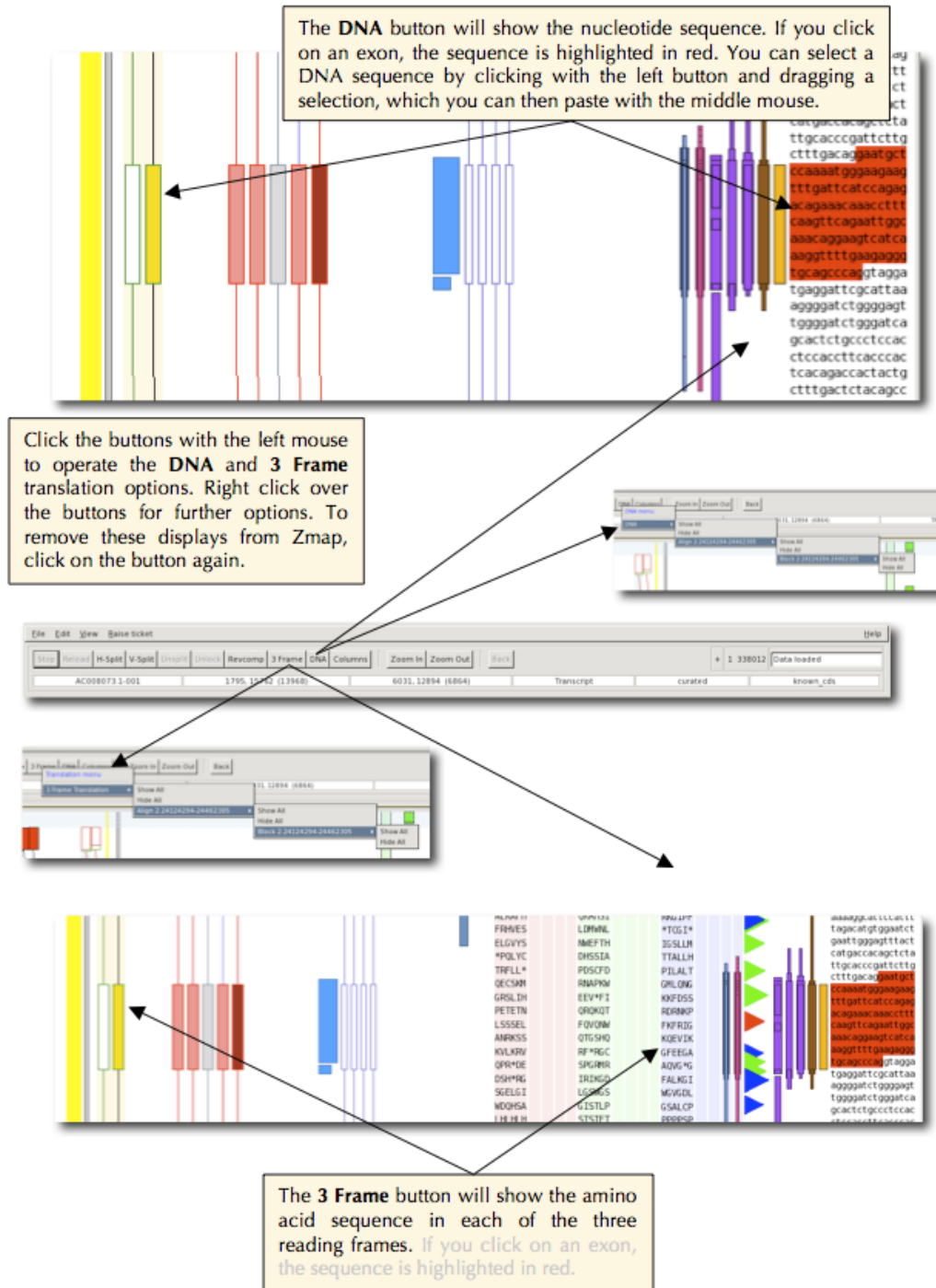
Figure 9: Feature display details

### 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 (see figure 10).

### 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). See figure 11.

### 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 figure 12).

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.

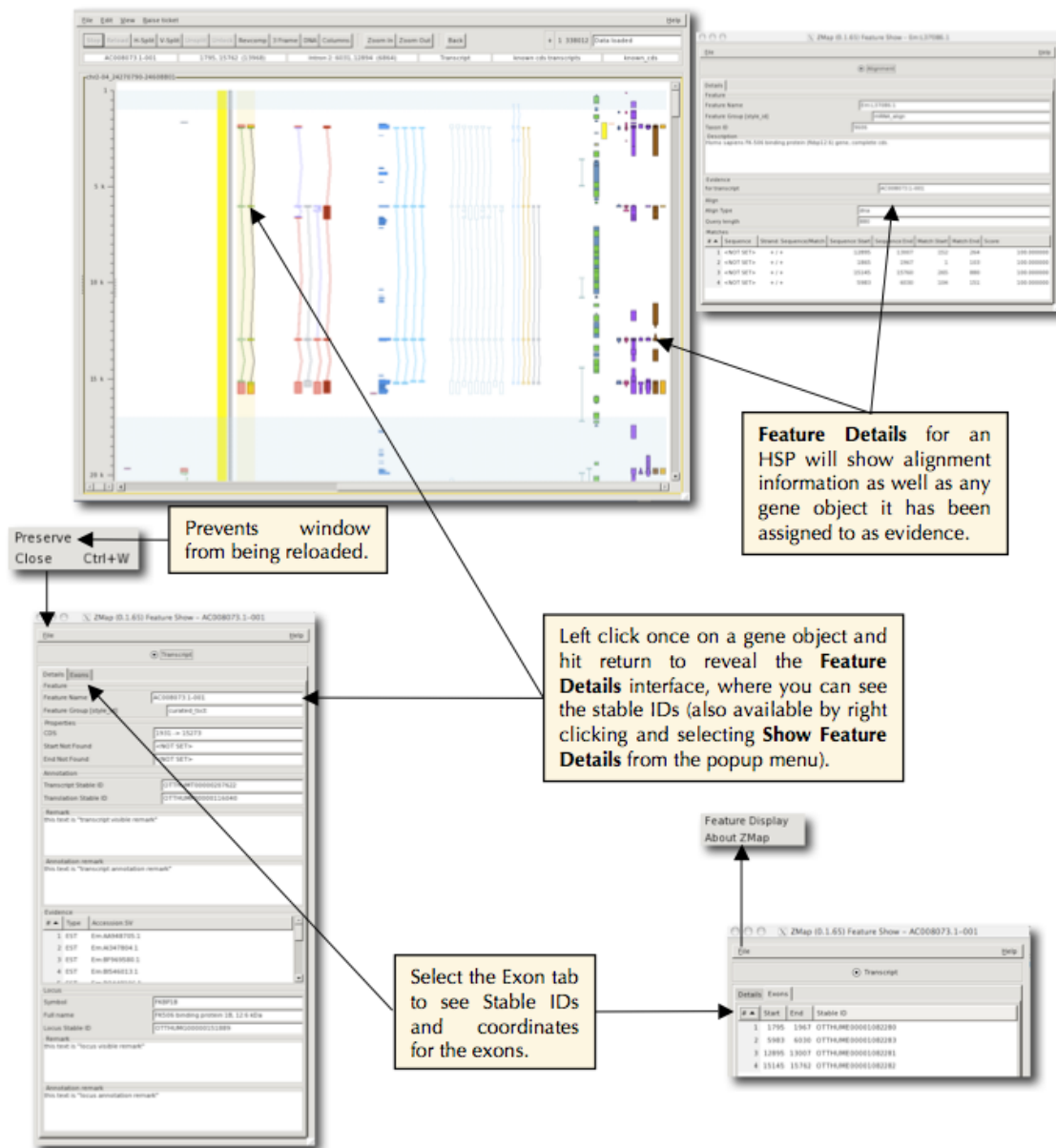3. **Red**: the matches are not colinear, e.g. 100, 230 —> 141, 423

**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.

Figure 10: Feature details dialog
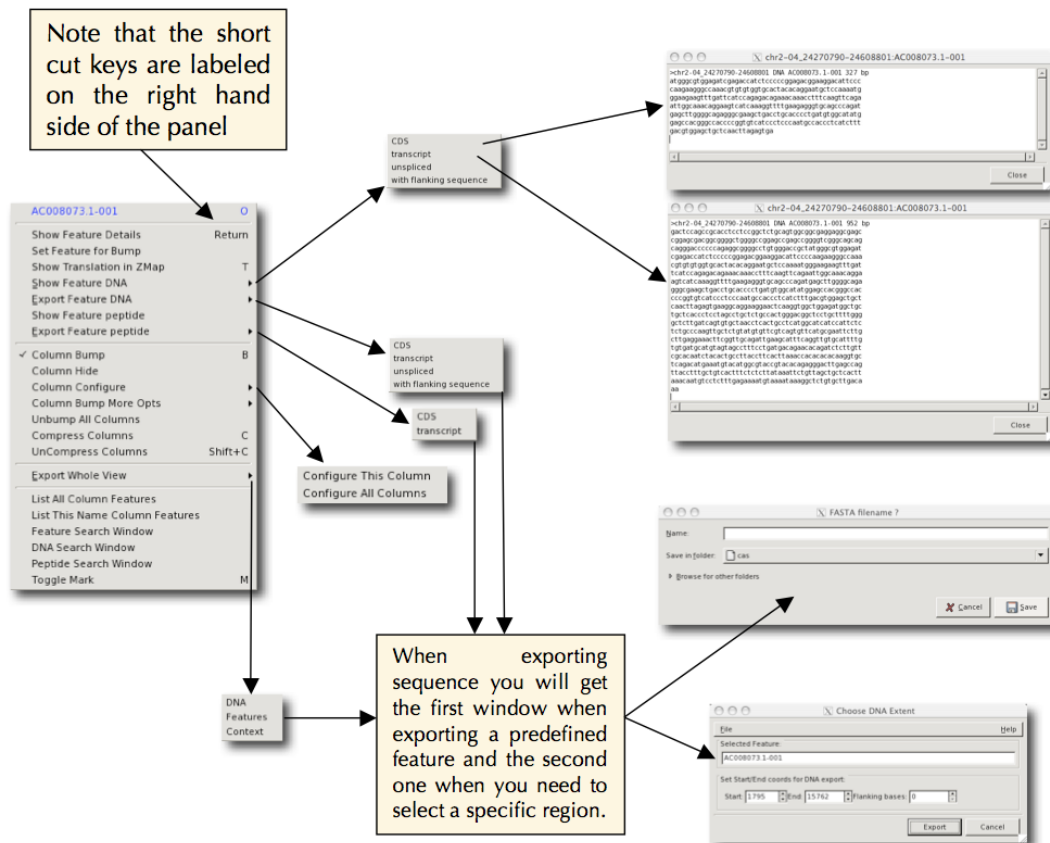
Figure 11: Exporting features

17

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 Blast match of interest (in this case an EST) for more menu features.

Pfetch returns the EMBL flatfile for that sequence.

Allows you to inspect the sequence of just the chosen feature or all of the 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.
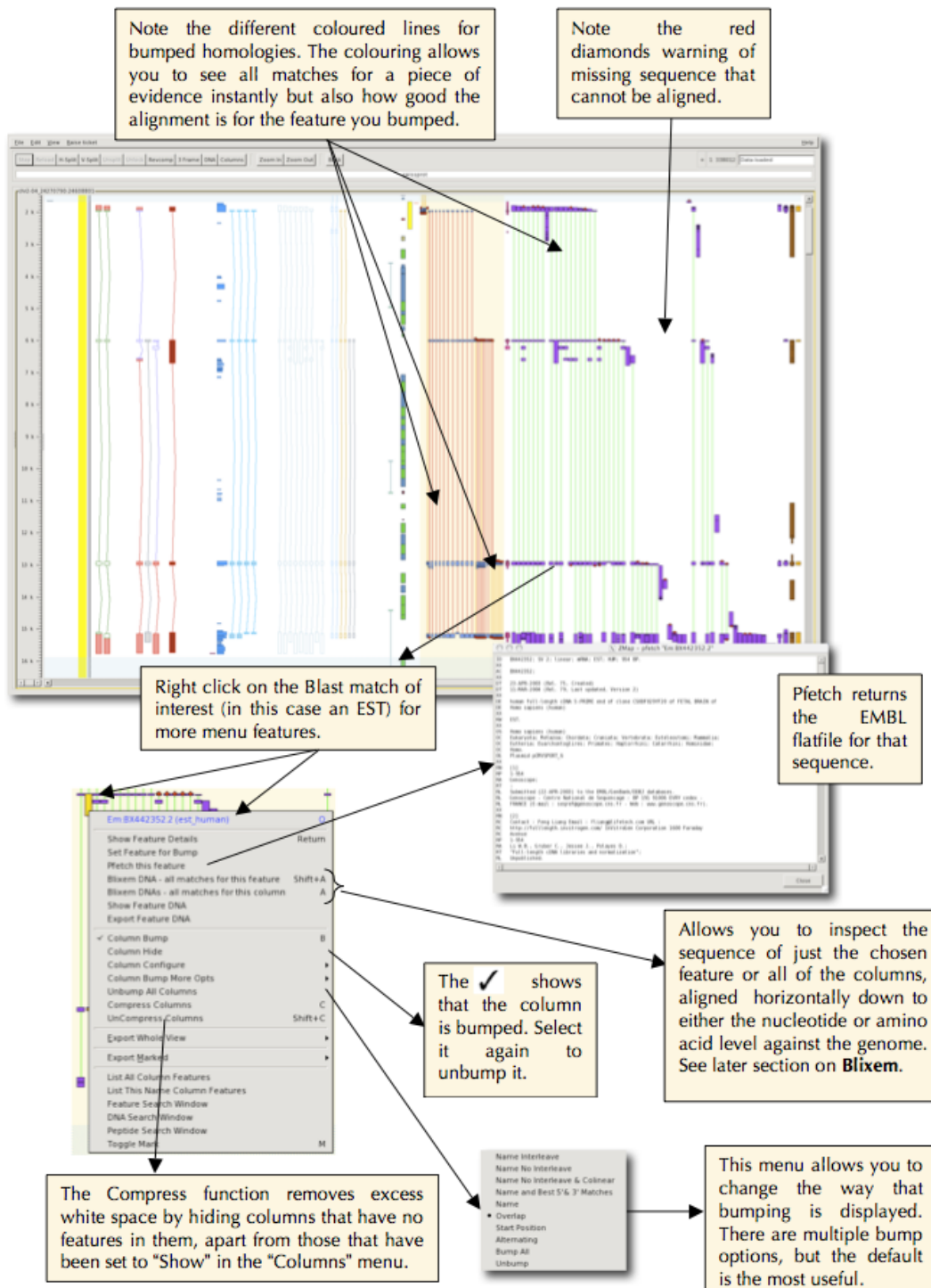
Figure 12: Bumping features

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).

### 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 in figure 13.

### 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. See figures 14 amd 15.

### 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. See figure 16.

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. See figures 17 and 18.

### 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
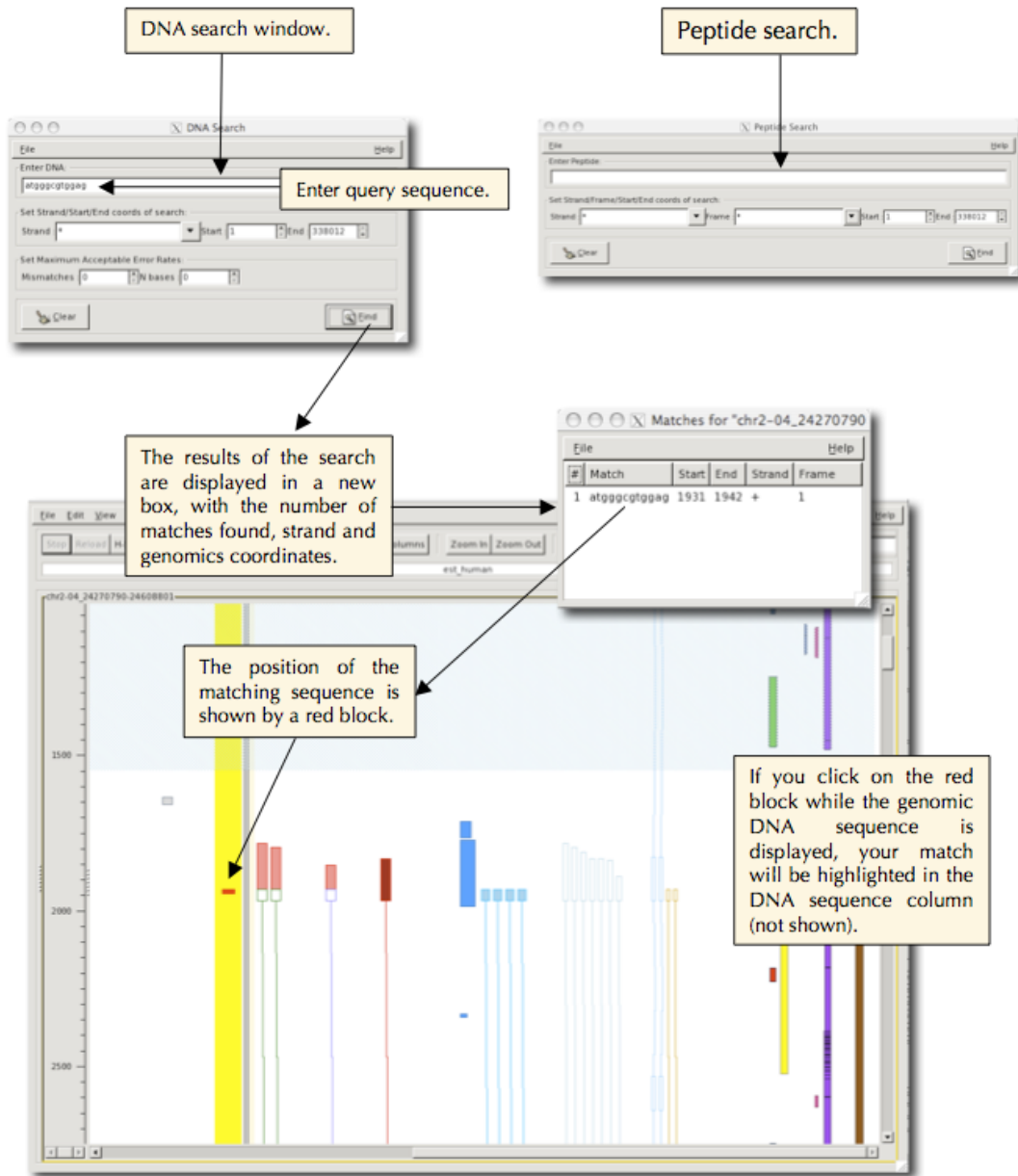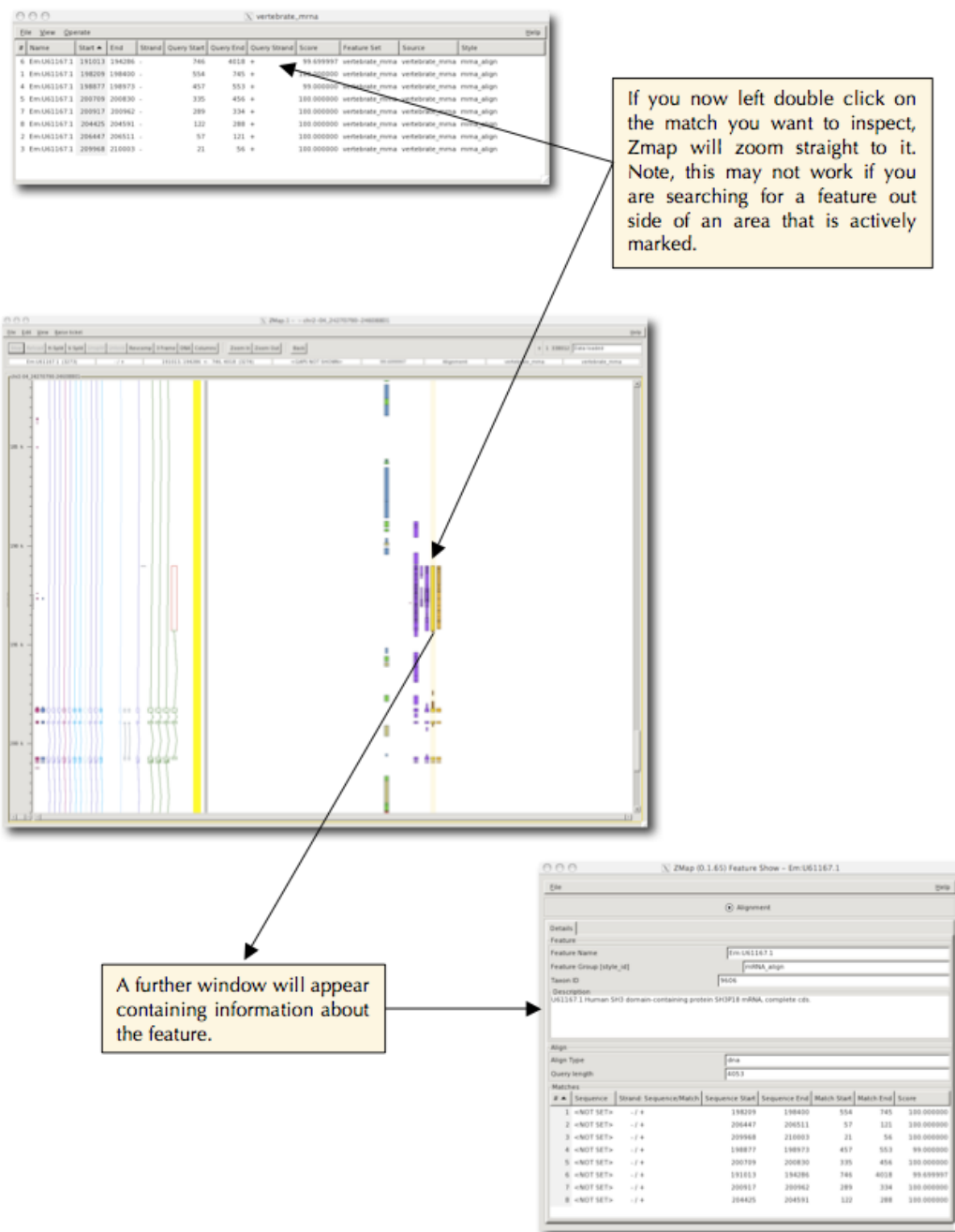
19

DNA search window.

Peptide search.

Enter query sequence.

The results of the search are displayed in a new box, with the number of matches found, strand and genomics coordinates.

The position of the matching sequence is shown by a red block.

If you click on the red block while the genomic DNA sequence is displayed, your match will be highlighted in the DNA sequence column (not shown).

Figure 13: Searching for a sequence

Figure 14: Searching for a feature

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 outside of an area that is actively marked.

A further window will appear containing information about the feature.
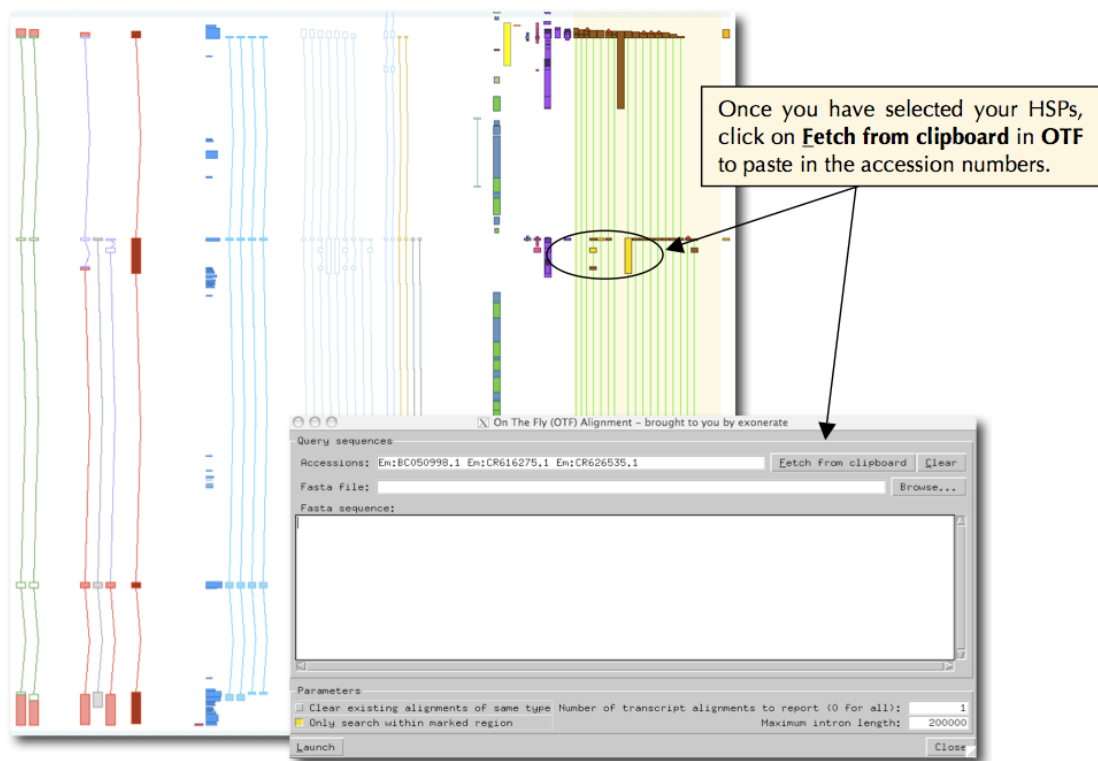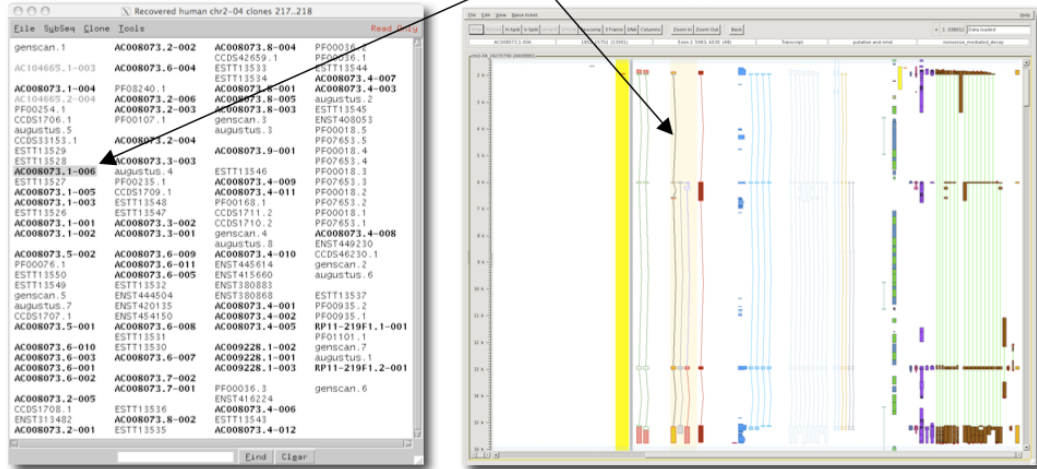
Figure 15: Feature search results

Figure 16: Show/hide features

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.
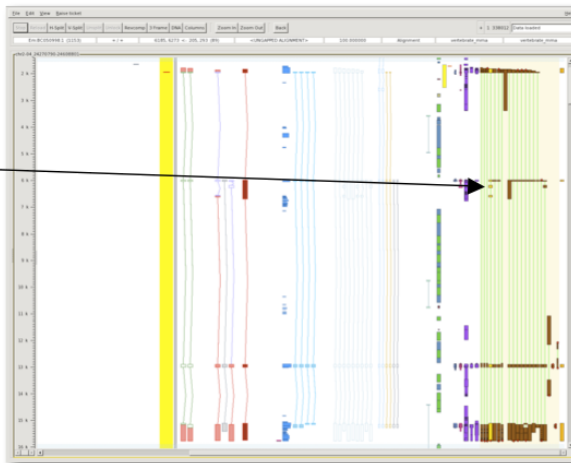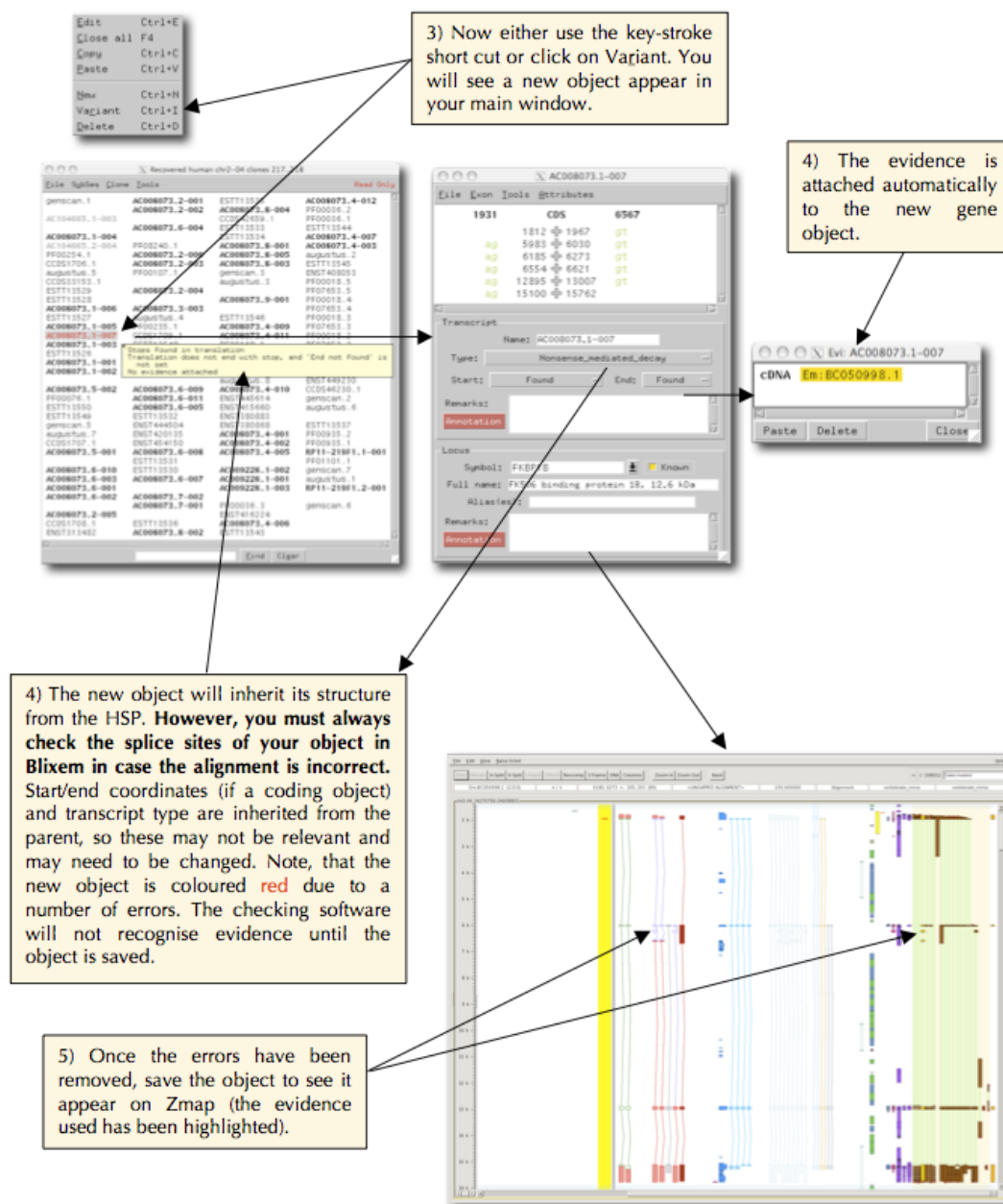
Figure 17: Variant construction

24

Figure 18: Variant construction

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. See figure 19.

### 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 "Launch 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. See figure 20.

## 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 |

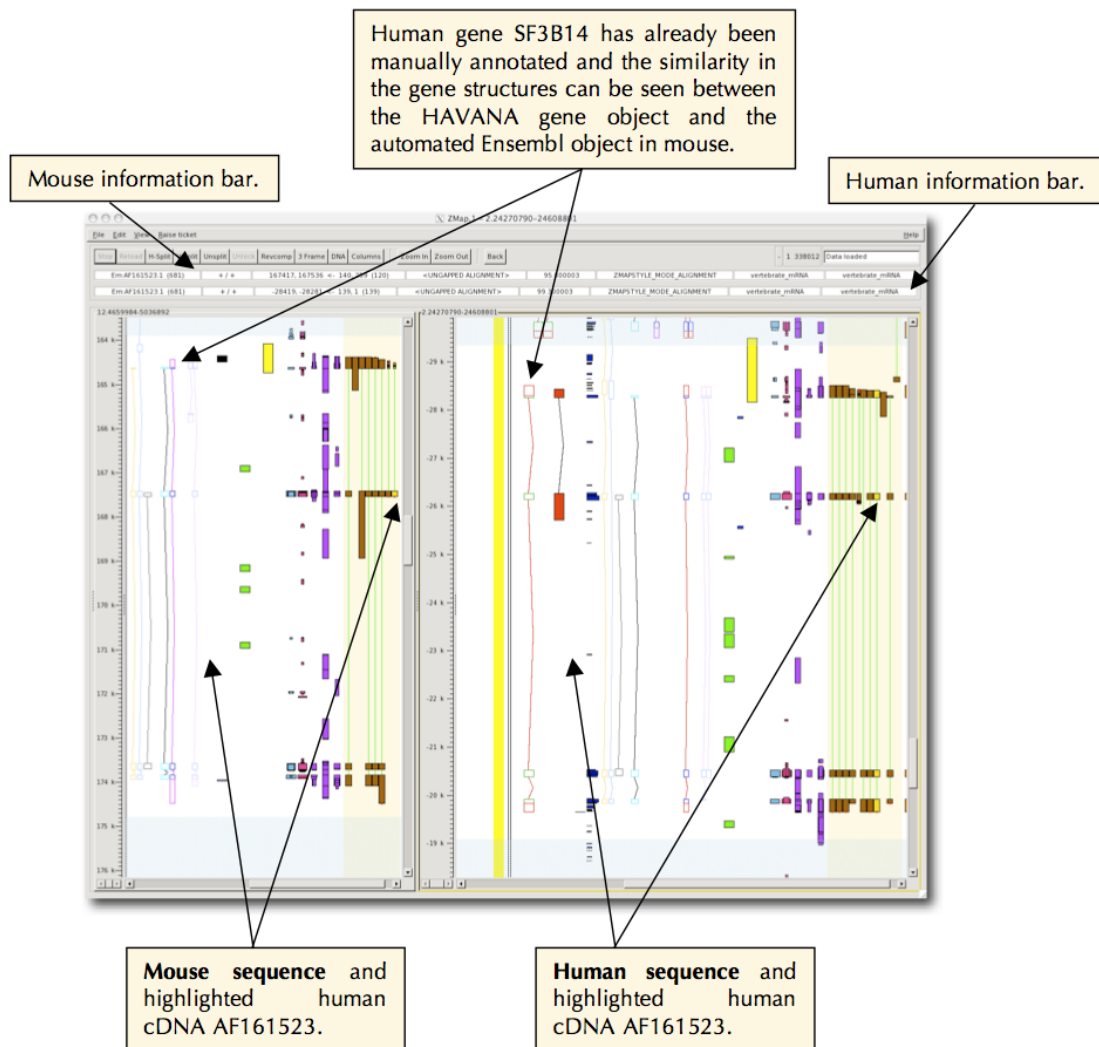The screen can be split horizontally or vertically (as shown) multiple times. An active window must be selected for **splitting**.

**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.

Figure 19: Splitting windows

Figure 20: Launch in an existing ZMap

| | |
|---|---|
| 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

| Short Cut | Action |
|---|---|
| a | Blixem all sequences in column |
| A | Blixem only highlighted sequence in column |
| 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.