Index

AAtDB 2, 61
Agarose Gel window 46
Arabidopsis thaliana genomic database 2
author information 7

C. elegans genomic database 2
canonical clones 34
chromomsome 21 database 2
chromosome 21 database 2
Clean Up 48
clearing mapping data 22
clone grid display 48
Columns 30, 41
contigs 31, 32

DNA Analysis 44

Genefinder 48
genetic map 12
genetic map cursor 13, 25
genetic map magnification controls 16
genetic mapping data 19, 20
Gmap data 19, 20
grain genes genomic database 2

highlighting 23

likelihood distribution 27
locator 15
locus information 14

Main Keyset window 3
main window 3
mapping data 19, 20
moving around the physical map 36
moving loci 26
multipoint data 20
Mycobacterium genomic database 2

neighbours 9

on-line help 4, 10

paper information 8
physical map 32
pop-up menus 4
preserving windows 9

Query Builder 51, 54
Query by Example 51, 59
Query Commands window 51
query formulation 51
quitting 60

rearrangements 18
restriction digests 44

saving text information 11
search queries 51
sequence display 39

template 6
text dump 9, 11
text search tool 50
text windows 7
two-point data 20, 21

viewing images 61

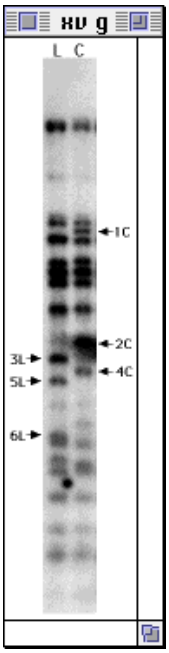
Whole button 16
wildcard characters 6

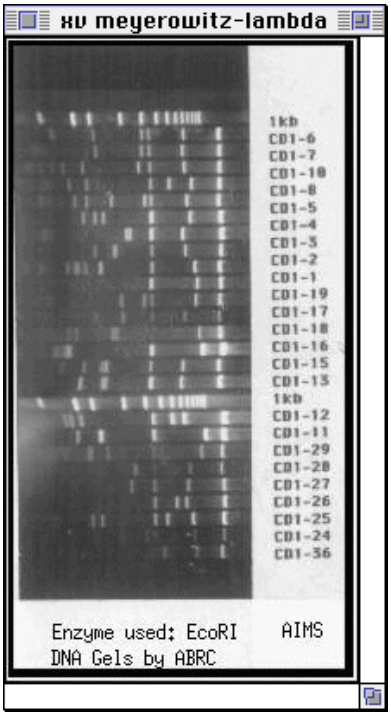
xv 63

Zoom 37
Zoom buttons 16, 43

- When you have finished experimenting with **xv**, click on the **Quit** button.
- Have a look at some of the other pictures available in **AAtDB**.

  These include autoradiograms of Southern blots used to detect RFLPs (image names of the type 'g' or 'm' followed by a number), and photographs of agarose gels (the 'goodman-cosmid' and the 'meyerowitz-lambda' series).

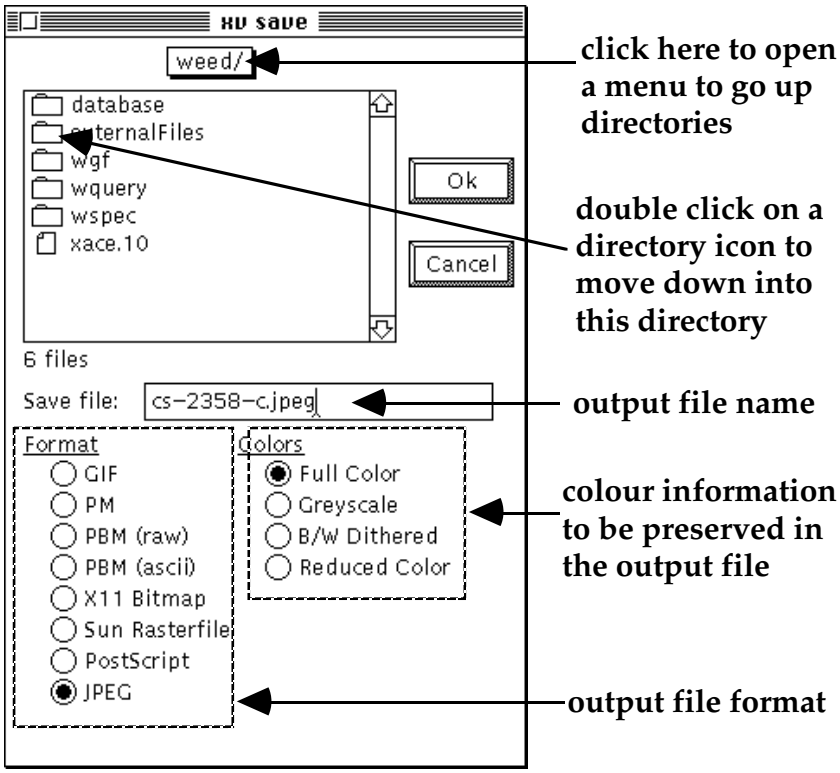




- When done, quit **AAtDB**.

- Experiment with the various features of **xv**, but be careful with the **Save** button.

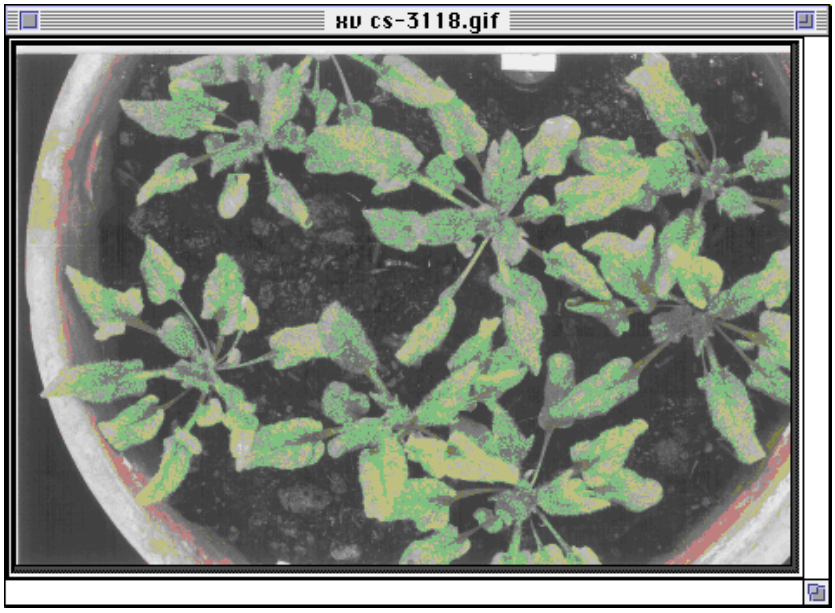
  Note: the program will not allow you to save a modified image under its original name, as this would erase the original picture. It will however allow you to save a modified image under a different name and offers a range of file formats including PostScript. If you really want to save a modified image, make sure to change the destination directory to your own, either by specifying the full Unix path to your directory including the output file name (e.g., /usr46/users/macpherson/picture.ps), or by first finding your way to your home directory using the Macintosh-like **xv** interface:



  In any case, saving a picture is not recommended unless you really know what you are doing.

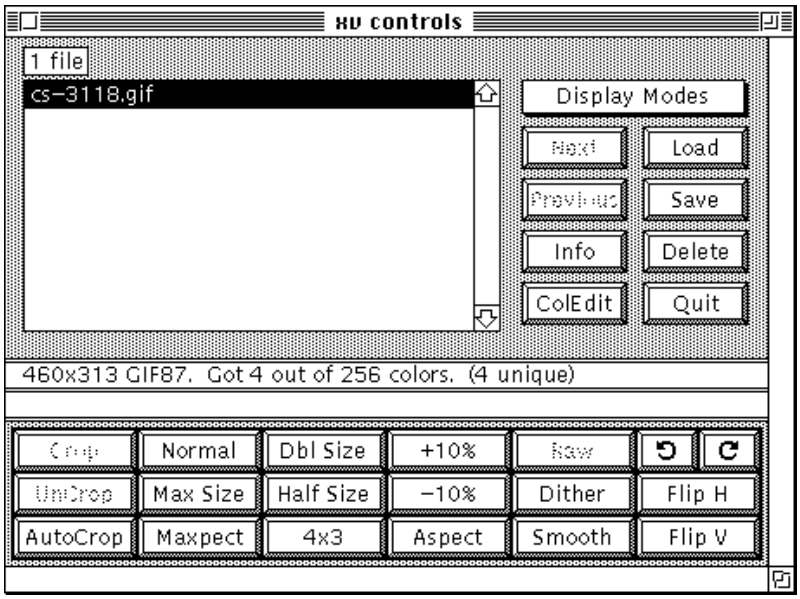
After a variable delay, **xv** displays the picture in a new window.

Note: the picture may not be displayed in full colour on some terminals because of the way these terminals handle colour information.



- Click inside the picture using the rightmost mouse button.

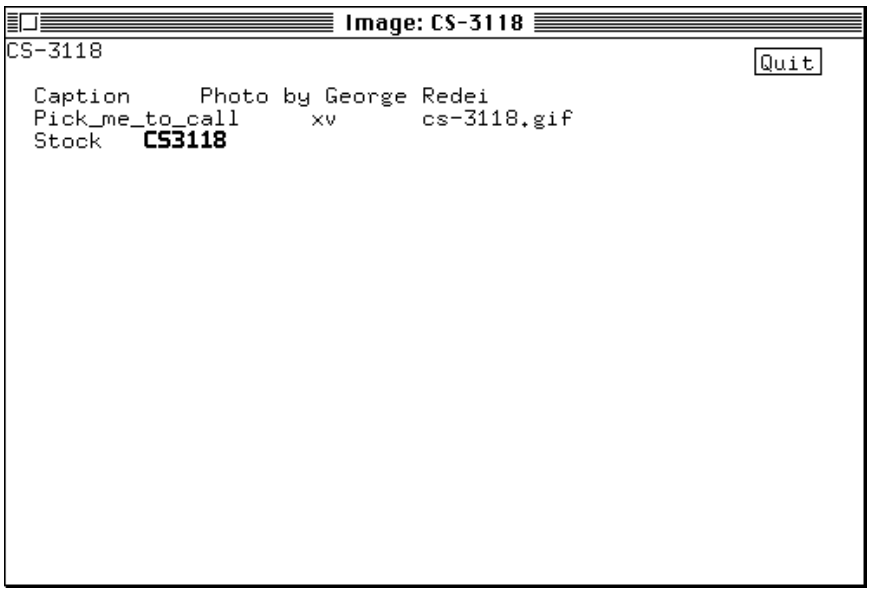
This brings up the **xv** control window.



The buttons in the lower part of the window allow some manipulation of the photo, such as enlargement, reduction and rotation. The **ColEdit** button gives extensive control on the colour content of the picture, which can be useful to see faint details.

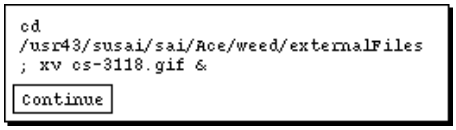
- Double-click on one of the image names starting with ‘CS’.

  This opens a text window:



- Double-click on the words **Pick_me_to_call**.

  This prompts **AAtDB** to start up the image visualization program **xv**, and instructs this program to display the appropriate GIF file. A small acknowledgment window may appear:



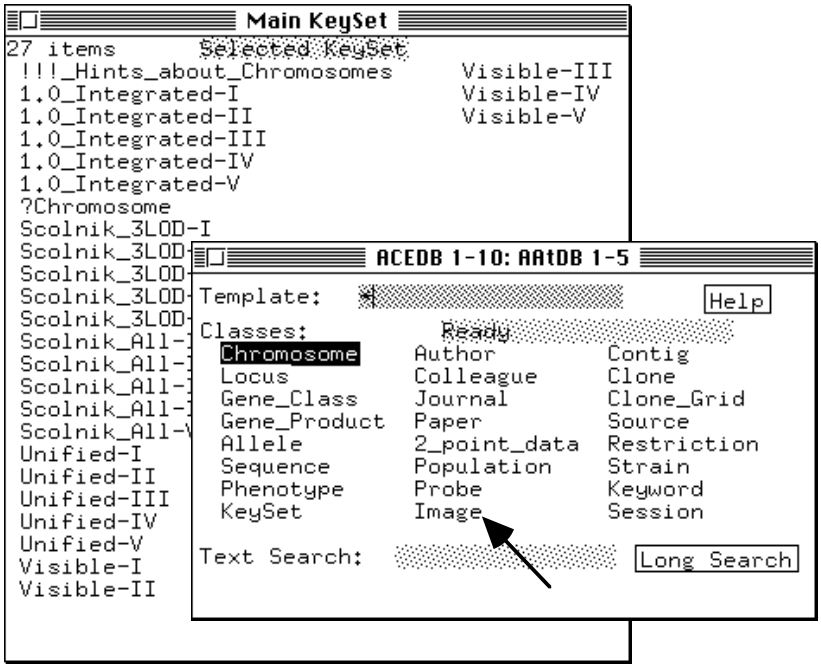
- If the acknowledgement window appears, click on the **Continue** button.

Viewing photographs in AAtDB

AAtDB (the *Arabidopsis thaliana* genomic database) uses the **ACeDB** interface, but offers the extra option of viewing digitized photographs stored in the database (in GIF or JPEG formats). This option is also available in the **Grain Genes** database. The images are not displayed by the **ACeDB** software itself, but by a dedicated image display program called **xv** (written by John Bradley from the University of Pennsylvania).

- Start up **AAtDB**.

  The main **AAtDB** window and the **Main Keyset** window appear on the screen. Note that **AAtDB** does not use exactly the same object classes as **ACeDB**.



- Double-click (left mouse button) on the class name **Image**.

  The available images are listed in the **Main Keyset** window. The first page lists many images with a name starting in 'CS'. These are pictures of various mutant *Arabidopsis* plants.

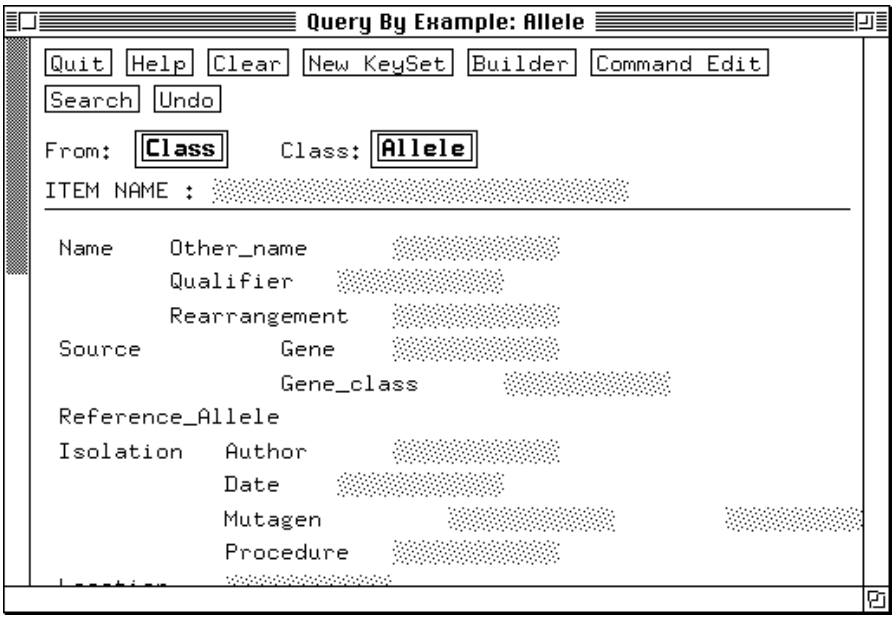
- Click with the *right* mouse button on the **Class:** box (currently containing the word **Paper**).

  A long menu appears listing all the available classes.

  Note: users without large screens may not be able to access the bottom of the menu.

- Select the menu option **Allele**.

  The window now contains a 'blank' description of an allele object.



- Click on the field labelled **Gene** and type **lon-2** <**return**>.
- Bring the **Query Commands** window to the foreground by clicking on the **Command Edit** button.

  The query entry box now contains the search request:

  `Find Allele Gene = "lon-2"`.

- Click on the query entry box and press <**return**> to start the search.
- Have a look at the entries retrieved by the query.
- To close the **Query by Example** window, click on the **Quit** button.

Optional exercise

- Try to design your own queries or series of queries to retrieve the following data:
  - all the authors in the database who contributed a paper to Nature and whose names begin with 'No'.
  - all the sequenced genes mapping to chromosome V.

Quitting ACeDB

- Click in the main **ACeDB** window with the rightmost mouse button and select **Quit** in the pop-up menu.

  A dialog box pops up asking for confirmation.

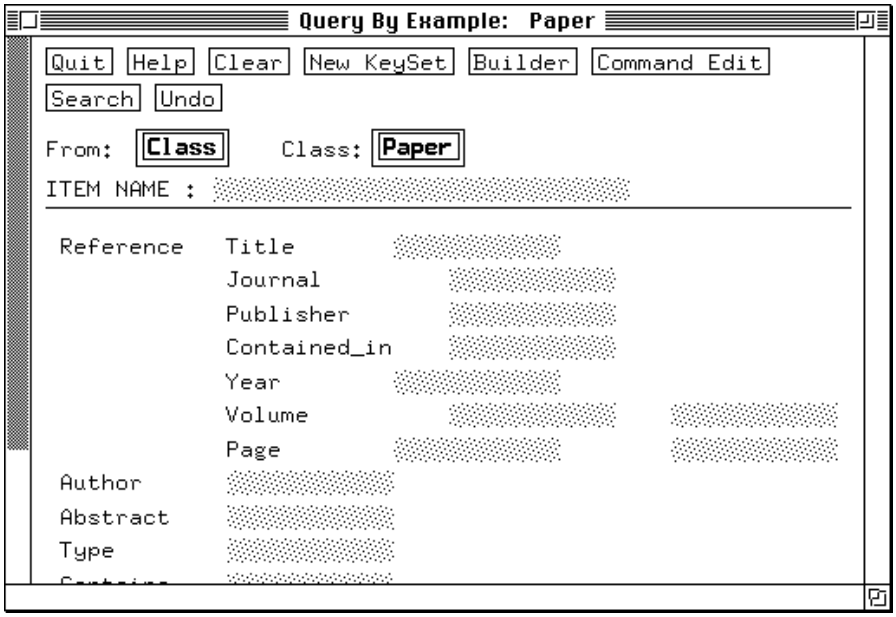
- Click on the **Yes** button.

Query by example

This search utility allows the formulation of queries by 'filling in the blanks' in a 'form'. The following example demonstrates how to formulate a query to retrieve all the catalogued alleles of the gene 'lon-2'.

- Click on the **Clear** button in the **Query Commands** window.
- Click on the **By Example** button.

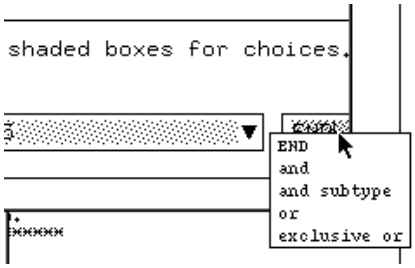
Since the class 'Paper' was the last one searched, the 'form' to retrieve objects of this class appears.



The 'form' contains a series of blank fields corresponding to the various items which may appear in a text object window describing a paper.

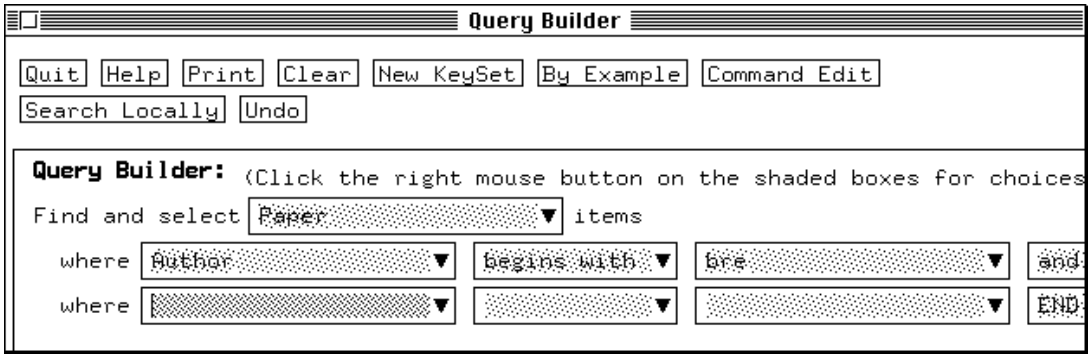
- Click with the *right* mouse button on the next entry box in the **Query Builder** window (currently containing the word **END**).

  This opens a small menu listing a few Boolean operators:



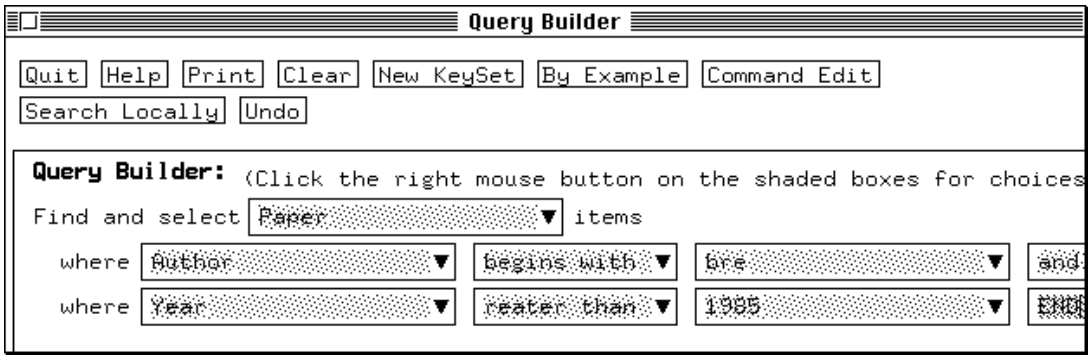
- Select **and** in the menu.

  A new set of entry boxes appears to enter further conditions.



- Click on the small triangle in the first new entry box (labelled 'where') and select **Year** from the list.
- Open the selector window for the next entry box (the 'condition' box) and select **is greater than**.
- Click on the next entry box and type **1985** <**return**>.

  The Query Builder window now contains the complete query.



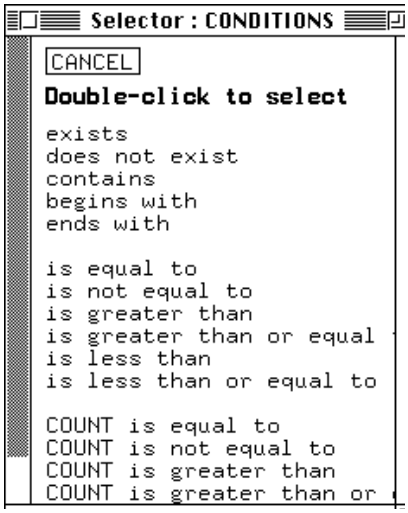
  The complete query should also appear in the **Query Commands** window as **`Find Paper Author = bre* & Year > 1985`**.

- Bring the **Query Commands** window to the foreground, click on the query entry box and press <**return**>.

  After some delay, the objects satisfying the query are listed in a keyset window.

- Bring the **Query Builder** window to the foreground and click on the **Quit** button.

The new selector window offers a choice of 'conditions' which are used to formulate the search expression.



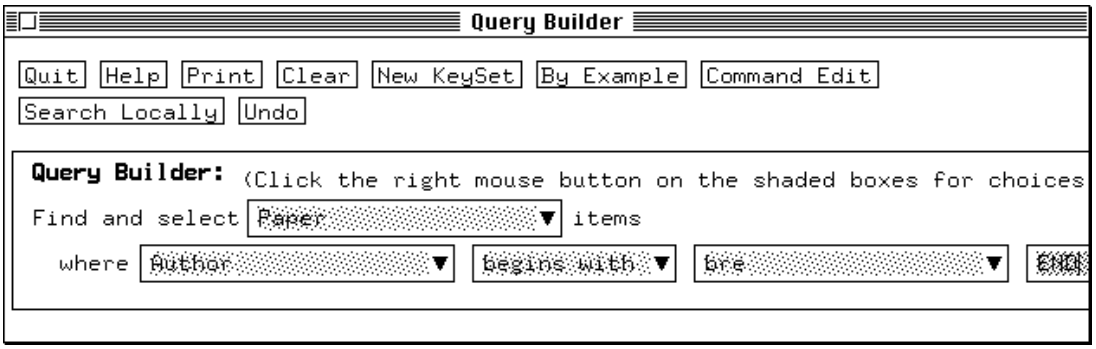
In this case, we want the author name to start with 'Bre'.

- Have a look at the conditions available then double click on **begins with**.

  The selector window disappears and 'begins with' appears in the entry box.

- Click on the entry box to the right of the one containing the words 'begins with' and type in **bre** <**return**>.

  The query so far is formulated in 'plain English' in the **Query Builder** window (it should be: 'Find and select Paper items where Author begins with bre END'). It also appears in query syntax in the entry box of the **Query Commands** window (**`Find Paper Author = bre*`**).

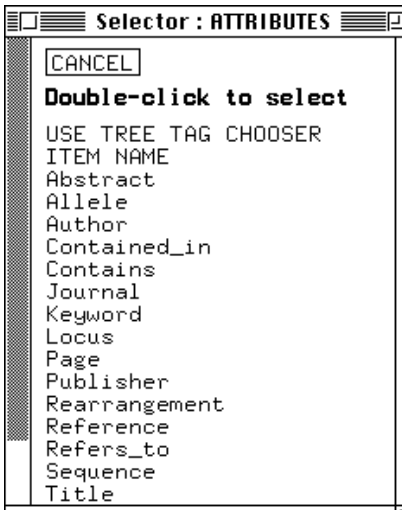


The second line of entry boxes is used to formulate the search expression. This expression is made up of an attribute of the object to be retrieved (e.g., its name, or the value of one of its tags), and of a condition which this attribute must satisfy in order for the object to be retrieved (e.g., a tag of the type 'sequence' must exist in the object, or the object name must start with 'a').

The first entry box on the second line (labelled **where**) is used to select the object attribute to be subject to the condition.

- Click on the small triangle in the **where** entry box.

The attributes available for objects in the class 'Paper' are listed in a new **Selector** window. They include all the tags which can be found in objects of the class 'Paper', as well as the name of the object itself (**ITEM NAME**).



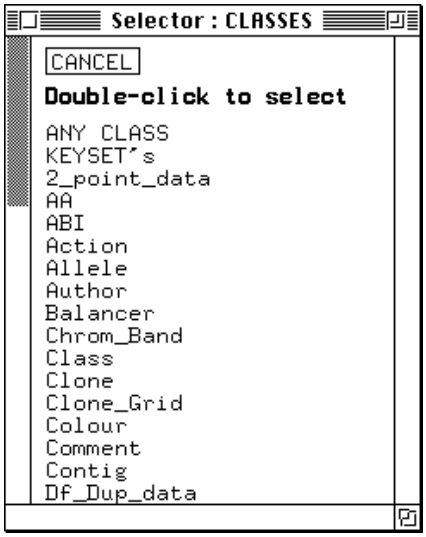
Note: the **USE TREE TAG CHOOSER** option provides an alternative method of choosing an attribute by double clicking on a tag in a 'blank' text window of the appropriate class.

- Double click on **Author**.

Author now appears in the entry box in the **Query Builder** window, and also in the **Query Commands** window.

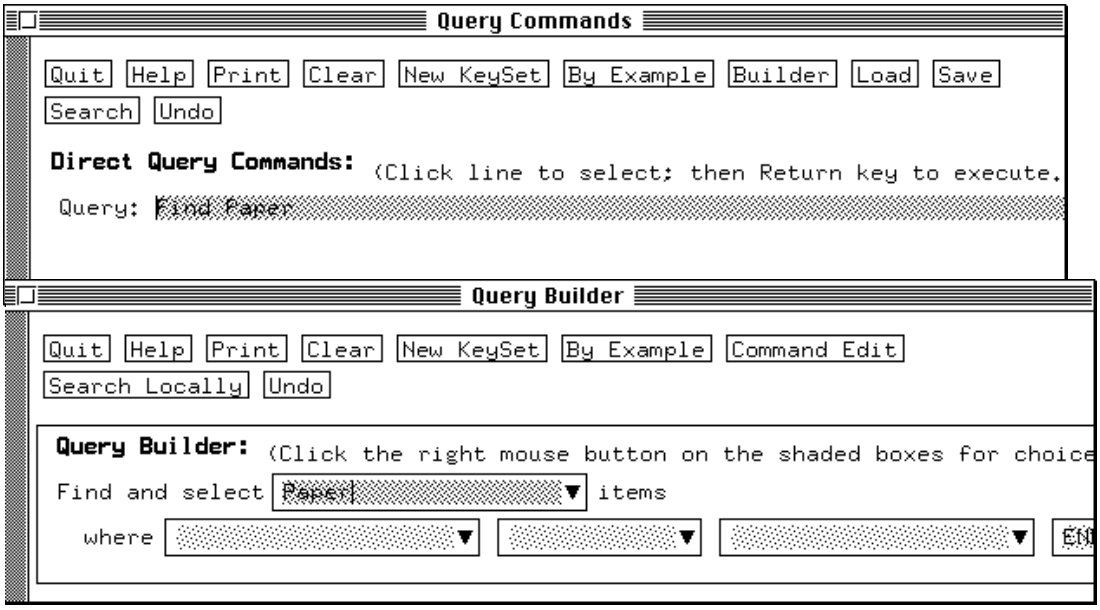
- Click on the triangle in the next entry box, on the right of the one now containing the word **Author**.

A new scrolling window appears. This window lists all the available search classes.



- Scroll down the list and have a look at the available search classes.
- When done, double-click on the class name **Paper** with the left mouse button.

The **Selector: CLASSES** window disappears and the word **Paper** appears in the entry box in the **Query Builder** window. The words **Find Paper** appear in the entry box in the **Query Commands** window.



Note: clicking on an entry box with the *right* mouse button opens a menu listing all the possible entries for this box, as an alternative to clicking on the small triangle to bring up a selector window. This alternative method is faster and more convenient for users with large monitors, but may be cumbersome for other users due to the size of the menu.

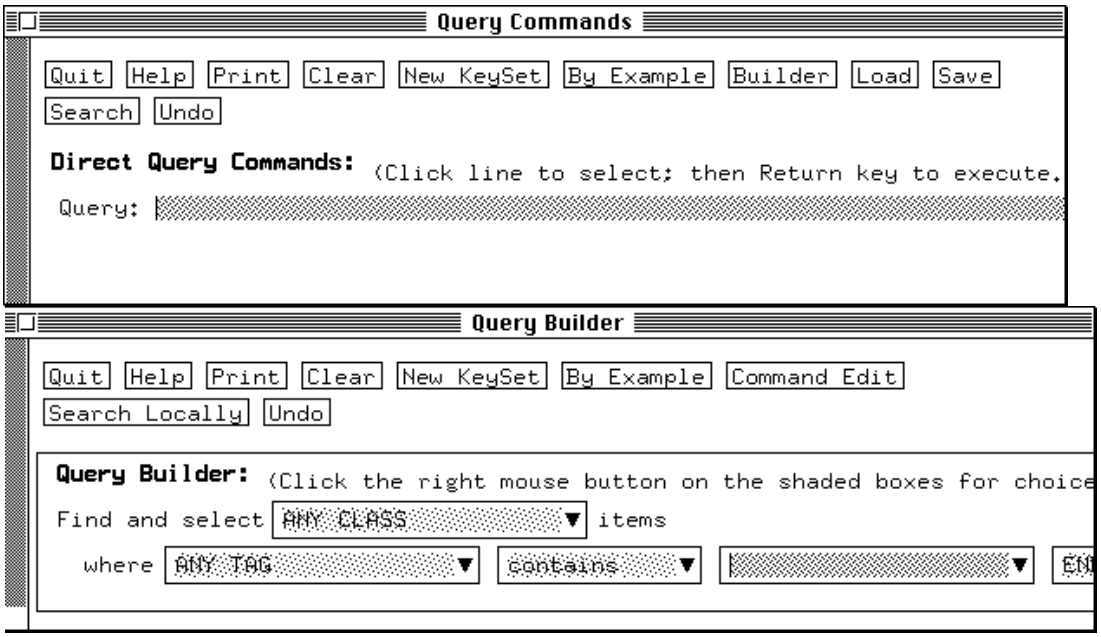
The Query Builder

The **Query Builder** utility greatly simplifies query formulation, especially for users unfamiliar with the organization of the database. With this utility, search commands can be chosen from menus rather than typed in. The following example demonstrates how to build a query to retrieve all the papers in the database published after 1985 with at least one author with a surname starting with 'bre'.

- Click on the **Clear** button in the **Query Commands** window to clear the previous queries.

  Note: this does not change the active keyset.

- Click on the **Builder** button to open the **Query Builder** window.
- If necessary, move the new window so both it and the **Query Commands** window are visible.

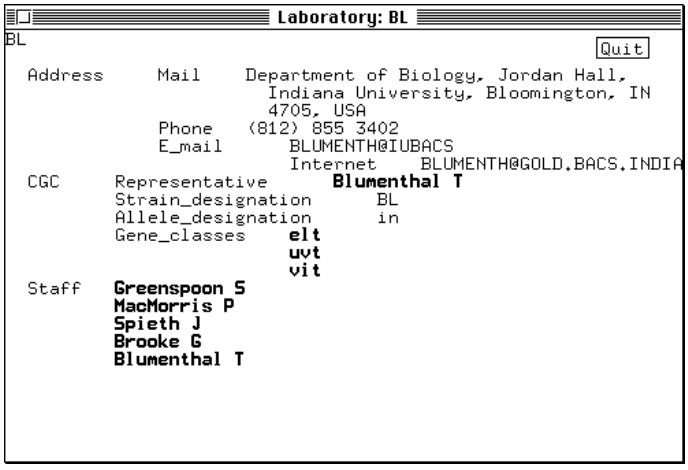


- Click with the left mouse button on the small triangle on the right of the **Find and select** box (currently containing the words **ANY CLASS**).

The following example illustrates the difference between tags and classes:

- Click on **Clear** to reset the **Query Commands** window.
- Enter the query **>?author br*** to list all author names starting with 'br'.
- Enter the query **>laboratory** in the next **Query** box.

  The program lists (by abbreviation) all the laboratories associated with the authors in the active keyset.
- Double-click on one of the laboratory abbreviations to display its text window.



Note that the author names are listed under the heading 'Staff'.

- Click on the third query box in the **Query Commands** window and enter the query **>author**.

  Since there is no tag by that name in the **Laboratory** window, nothing is retrieved.
- Click on the first **Query** box in the **Query Commands** window (which should contain the query '>?author br*) and press <**return**> to repeat the first query.
- Do likewise for the second query.
- Enter **>staff** as the third query.

  This time, the authors associated with the laboratories in the active keyset are correctly listed. Note that these objects are members of the 'Author' class (double-click on one of these names and check the title of the window which appears).

Note: a 'tag' is a heading in a window, not a class name (although the two can be the same). For example, 'Author' is both a class (which can be searched with the command **>?author**) and a heading in the 'Paper' text windows (which can be searched using the command **>author** provided there are objects of the class 'Paper' in the active keyset). Author names are also listed in the 'Laboratory' text windows, but under the tag 'Staff' rather then 'Author'.

If no keyset is specified, the expression is applied to all the objects in the current keyset.

- Click on the **Query** box and type **>?map** <**return**>.

  This query retrieves all the objects in the class 'Map' and lists them in the **Main Keyset** window. In this case, maps of the *C. elegans* chromosomes are listed. A second query box appears underneath the first one to enter further queries.

- Click on this second **Query** box and type **locus = cat-1** <**return**>.

  The program searches the active keyset (the chromosomes) and reports those objects containing the cat-1 locus (only the X chromosome qualifies). A third query box appears.

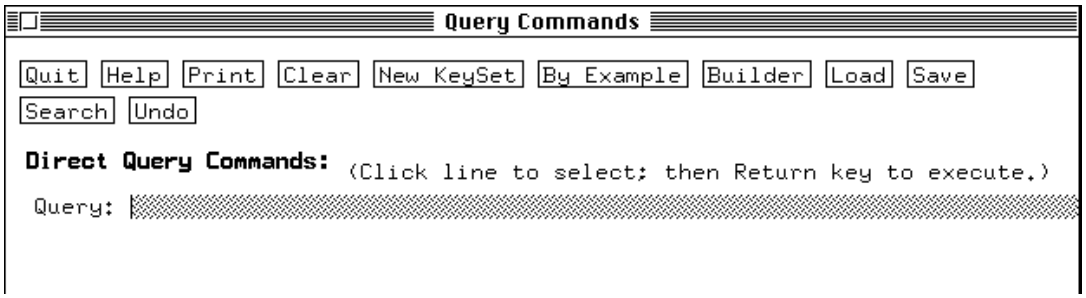
- Click on the third **Query** box and type **>locus allele = a*** <**return**>.

  This time the program searches for loci in the active keyset (chromosome X) with an allele whose name starts with 'a'. Because the query started with **>**, the loci (and not the map) are reported.

Query tool

- Select **Query** from the pop-up menu of the main window.

  This brings up the **Query Commands** window.



ACeDB queries can be entered directly into the **Query** box, using a powerful but moderately complex syntax. **ACeDB** also provides two utilities, **Query Builder** and **Query by Example**, which greatly simplify query formulation.

Direct query formulation

The query syntax is described in the help page **Query_syntax** (to read it, select **Help** from any pop-up menu, click on the **Menu** button then click on **Query_syntax**). In a nutshell, a query is a logical expression which is either true or false. The query tool searches a designated keyset (a set of objects) and reports all those for whom the expression is true. The program can understand a number of operators including AND, OR, NOT, **=**, **>** (greater than) and **<** (less than). The wildcard characters *****, **#** and **?** can also be used. The keyset to be searched is specified using the following commands:

>?*ClassName expression*: (or **Find** *ClassName expression*)

searches all the objects in the class *ClassName* and retrieves those for whom the logical expression *expression* is true (e.g., **>?locus a*** retrieves all the objects in the class 'Locus' whose name starts with 'a'. **>?locus allele = a*** retrieves all the loci with an allele whose name starts with 'a'). The **Query builder** (demonstrated below) provides the simplest way to obtain a list of available classes.

>*Tag expression*: (or **Follow** *Tag expression*)

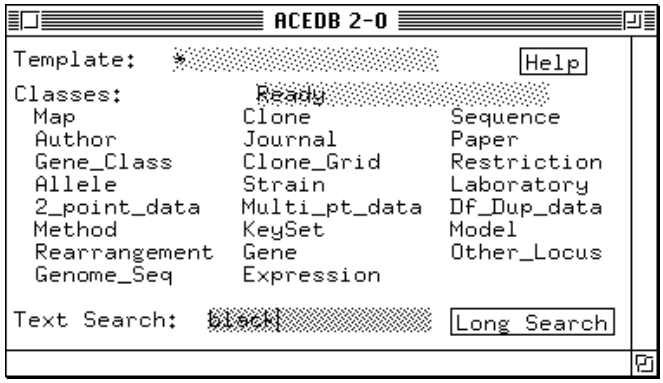
searches the objects under the heading *Tag*, inside the objects of the active keyset, i.e., those currently displayed in the **Main Keyset** window, or a new user-defined keyset window, and retrieves those for whom the expression *expression* is true (e.g., **>gene a*** will search the objects in the current keyset and retrieve the genes with a name starting with 'a' inside these objects). The simplest way to list the available tags is to use the **Query Builder**.

Search tools

ACeDB provides a simple search tool to locate objects containing a particular text string, as well as an advanced tool which allows the formulation of complex queries.

Text search tool

- Click on the **Text Search** box in the main **ACeDB** window and type in **black** <**return**>.



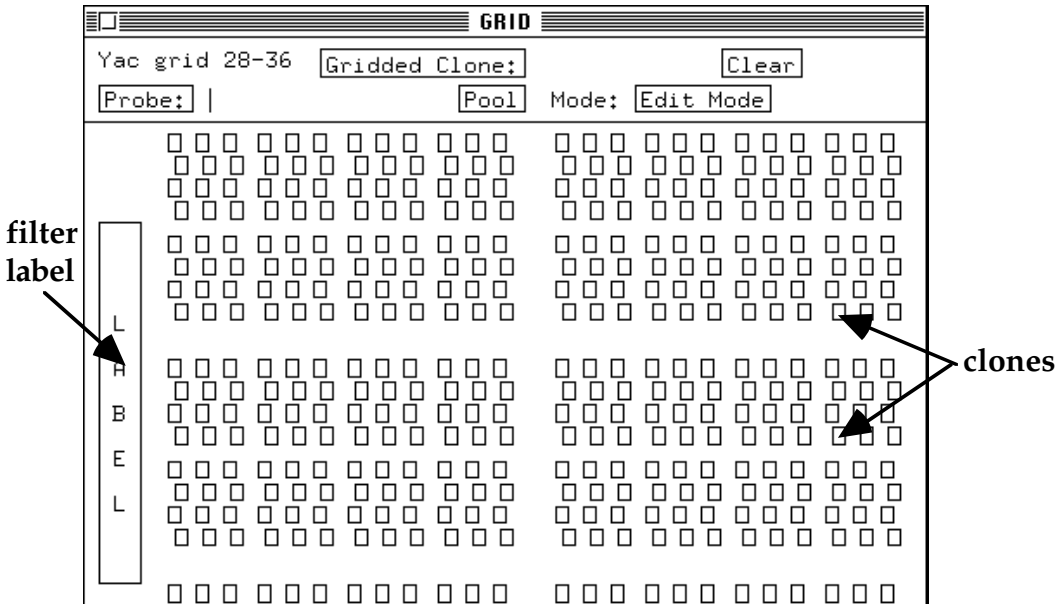
A new list of objects is displayed in the **Main Keyset** window.

- Browse through these objects by double-clicking on their name.

The text string 'black' appears in all these objects, either as the word 'black' or as part of 'Blackburn'.

A graphical representation of the filter appears in a new window.

Note: some of the buttons in the upper part of the window are visible only on colour and greyscale screens.



Individual clones bound to the filter are represented by boxes layed out in the same pattern as the clones on the filter. The 'LABEL' box shows where the filter was labelled to help orienting the filter.

When **Map mode** is highlighted, double-clicking on a clone box either brings up the physical map window centered on the corresponding YAC clone, or a text window with the name of the YAC clone if no clone was gridded at this position.

- Use **Quit** in the pop-up menu to close the clone grid window.

The remaining buttons of the sequence window are the **Reverse Complement** button, which changes the sequence to its reverse complement, and the **Genefinder** button. This button is in fact a pop-up menu and must be opened using the rightmost mouse button. It allows access to an automatic gene finder utility to detect coding sequences by codon usage, splice site and open reading frame analysis. However the functionality of **Genefinder** is limited when 'browsing' through **ACeDB** because some of its commands modify the database and are restricted to the database editors.

- Use the **Quit** option in the pop-up menu to close the sequence window.
- Click in the main **ACeDB** window with the rightmost mouse button and select **Clean up** in the pop-up menu.

This closes all **ACeDB** windows except the main window.

Clone grid window

The last type of window in **ACeDB** is the clone grid window which provides a graphical representation of the gridded filters used to locate clones by hybridization.

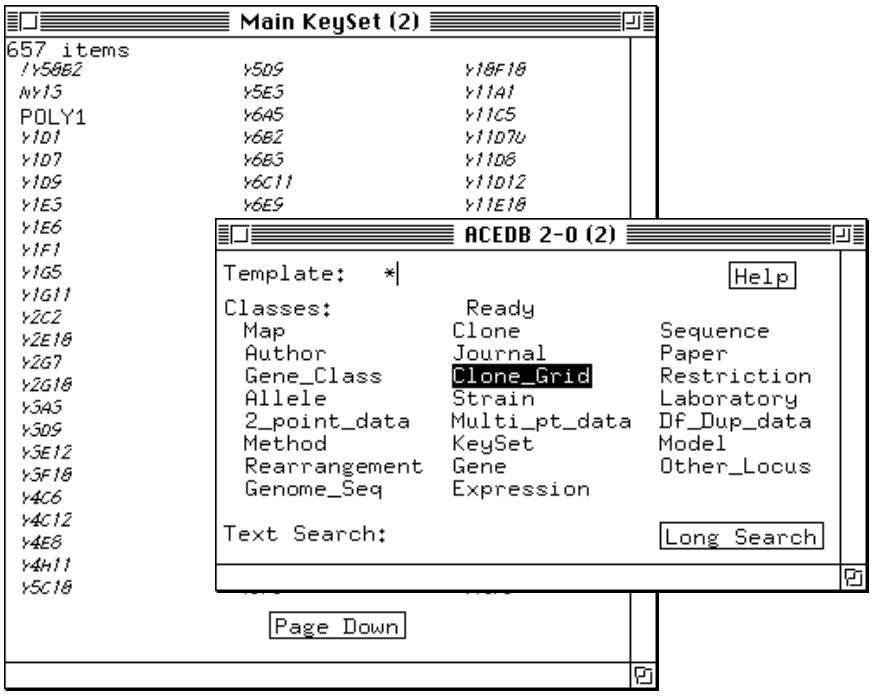
- Double-click on the **Clone_Grid** class in the main **ACeDB** window.

The filters are listed in the **Main Keyset** window. At this writing, most of the filters have not been entered into the database and are listed in italic.

- Use the template search facility in the main **ACeDB** window to locate the filter **Y28-36**.

This filter has been entered in the database and is not listed in italic type.

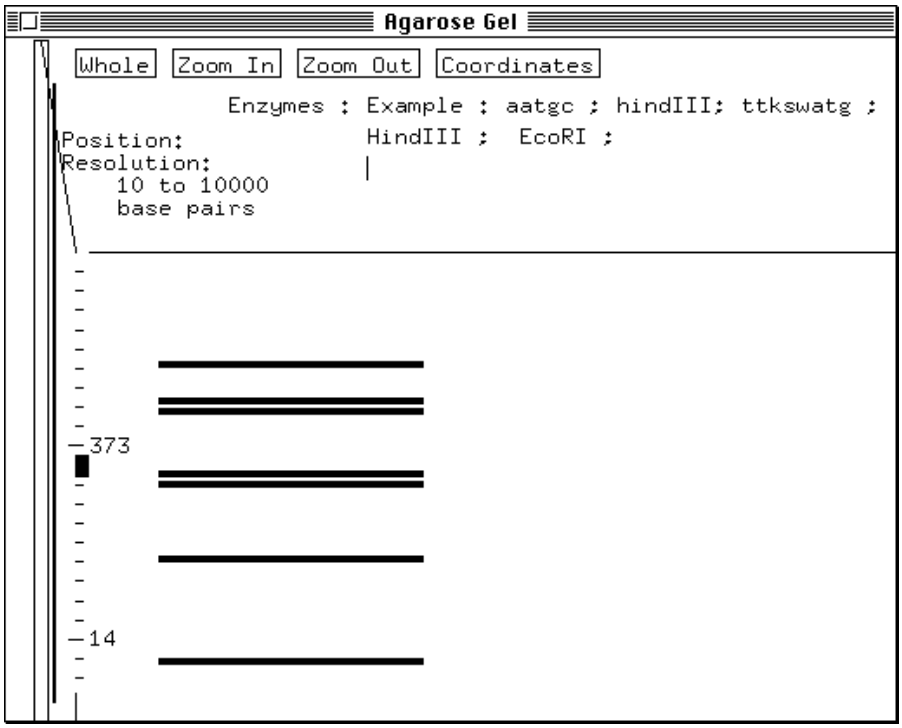
- Double-click on **Y28-36** in the **Main Keyset** window.



- Click on the text entry box and type in **hindiii; ecori** then press <**return**>.

  A series of bands appear in the window. They simulate an agarose gel on which the specified digest has been electrophoresed. A second restriction enzyme entry box appears underneath the first one to allow the specification of a second digest to be displayed in another lane of the gel.

  Note: the specified restriction sites are also shown in the sequence window.



- Click on the **Coordinates** button.

  Each band is now labelled with the length of the corresponding restriction fragment in base-pairs.

- Double-click with the leftmost mouse button on the band corresponding to the largest fragment (1541 bases-long).

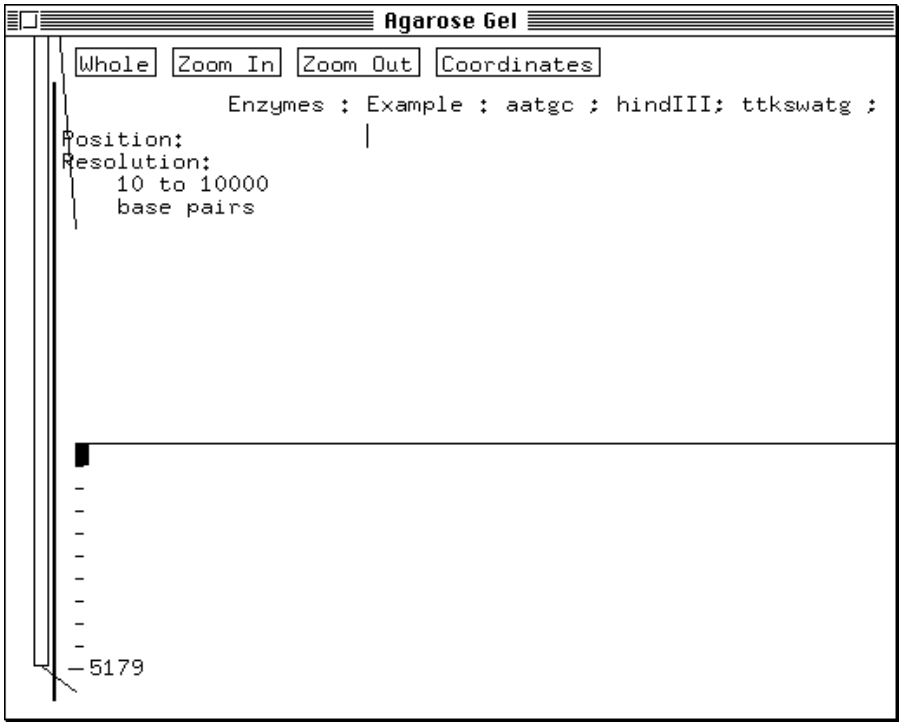
- Bring to the sequence window to the foreground.

  The part of the sequence corresponding to the selected band is now highlighted in dark blue, both in the summary bar and in the sequence listing.

- Click on the **Clear** button.

- Click on the **Show gels** button.

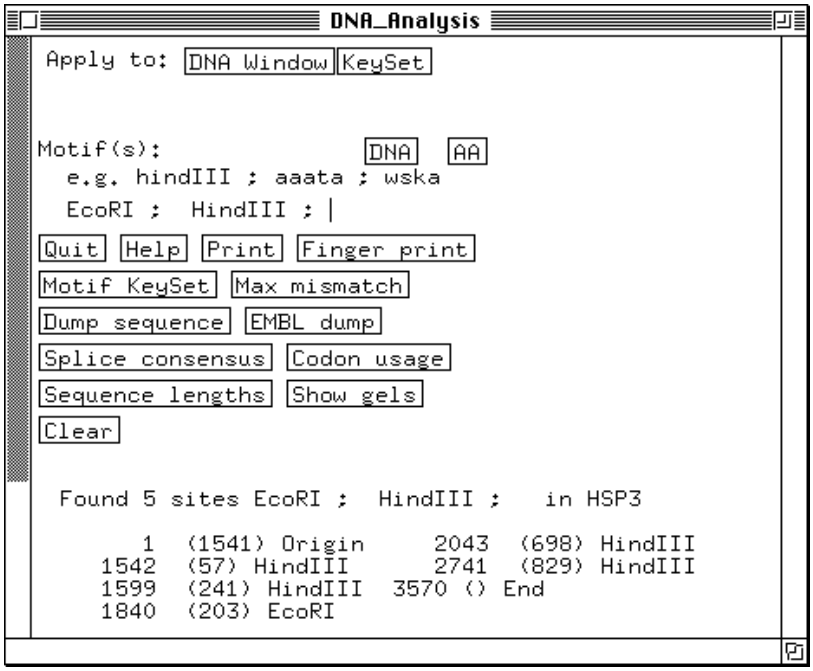
  A new window titled **Agarose Gel** appears.



The controls in this window are similar to those of the genetic map window (coarse and fine control sliders, **Zoom In**, **Zoom Out** and **Whole** buttons). For the time being, no data are displayed in the window. Restriction enzyme names can be typed in the yellow text entry box near the top right corner of the window.

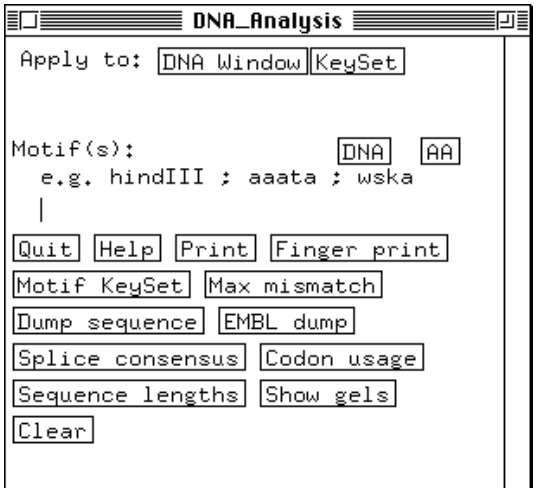
- Bring the **DNA Analysis** window back to the foreground.

  The results of the restriction digest (cutting sites and fragment sizes) are shown in the lower part of this window (you may have to change the size of the window to see it all).



- Open the pop-up menu and select the option **Analysis window**.

  This brings up a new window offering a number of options for sequence analysis. The following example demonstrates how to simulate a restriction digest of the selected gene with HindIII and EcoRI.



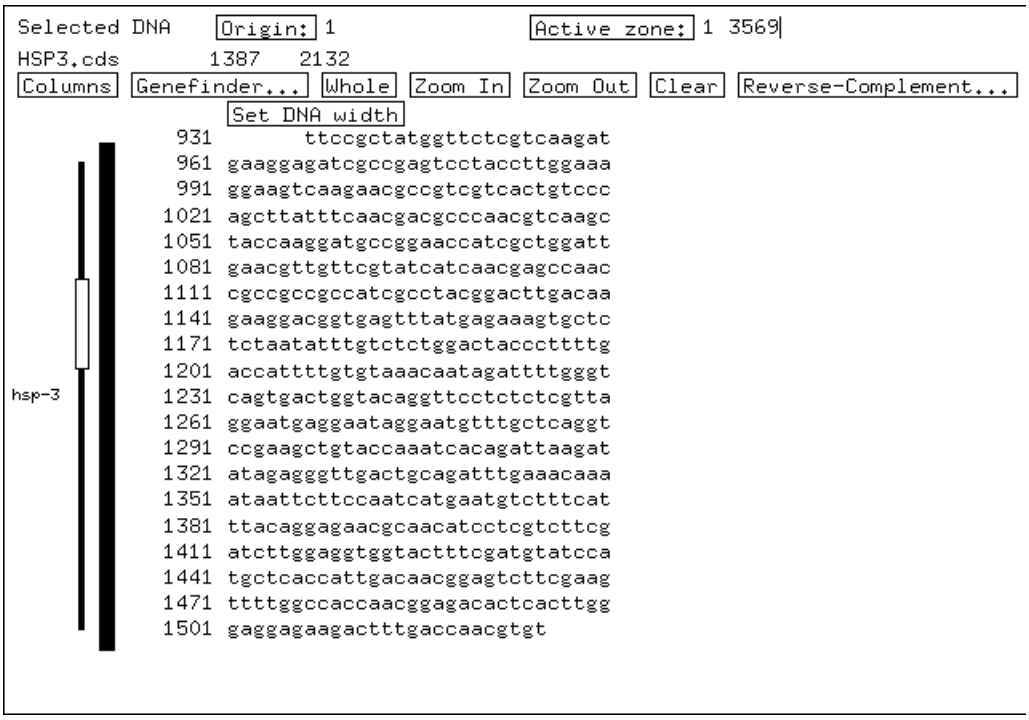
- Click on the yellow box and type in the enzyme names separated by a semicolon: **hindiii ; ecori**.
- Press <**return**>.

  The sequence window is brought back to the foreground. The specified restriction sites are highlighted in the sequence (when they are in a currently displayed part of the sequence) and are indicated by slashes on the summary bar.

- Click on the **Clear** button to erase the restriction digest results from the sequence window.

- Click twice on the **Zoom In** button.

  This magnifies the screen to a level where the whole sequence can be displayed and the ellipses disappear.



- Click on the **Clear** button to remove any highlighting.
- Click on the **Zoom Out** button twice to return to the original magnification.
- Click on the **Set DNA Width** button and change the displayed width to **50**.

The scale bar and the various features are no longer displayed. They are replaced by a DNA sequence and a new button enabling the user to set the number of sequence characters to display per line. Because the current window magnification is too low, the whole sequence cannot be displayed. The ellipses (**...**) at the end of each line indicate that additional sequence exists but is not shown.

- Click once on one of the exons in the exon-intron graph.

  The corresponding sequence is highlighted in the sequence listing.

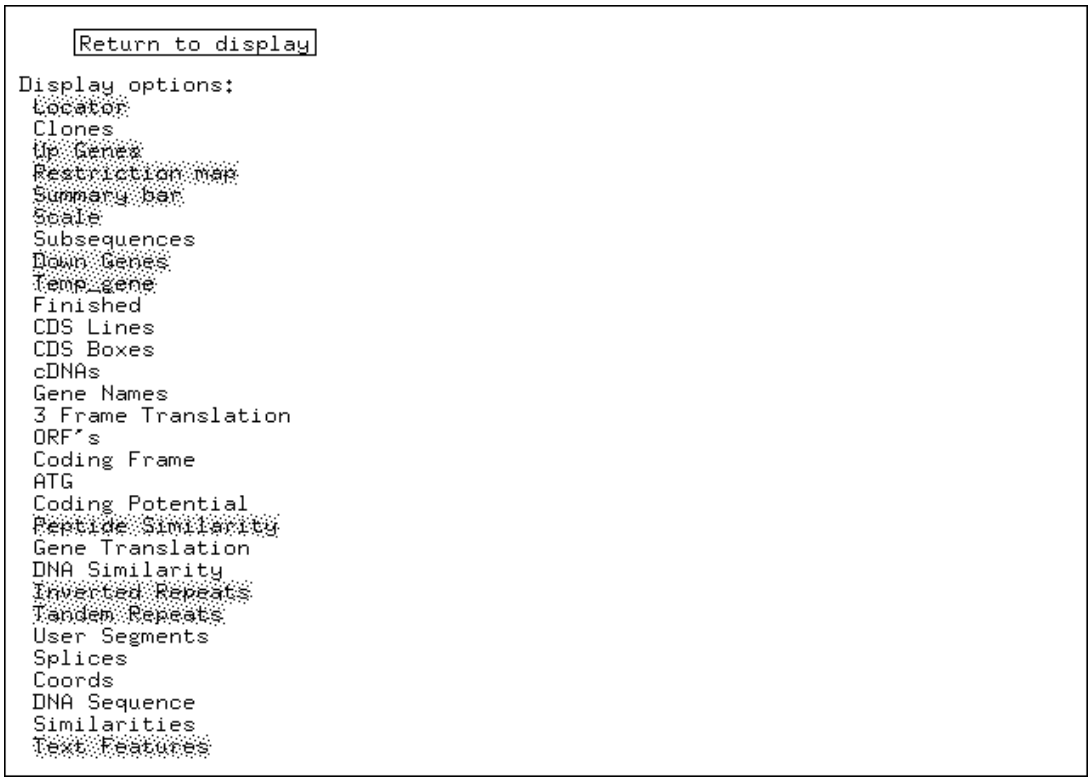
- Experiment with the **Columns** button to see the data types that can be displayed in the sequence window.

  The program can display many features and views of the gene, including start and stop codons, open reading frames, protein translation etc.

- When finished, select the following data types to be displayed: Locator, Restriction Map, Summary Bar, Coords and DNA sequence.

A lot more information can be displayed in this window by changing its settings using the **Columns** button.

- Click on the **Columns** button.



The various types of information which can be displayed (when available for this particular sequence) are listed here. The currently selected data types are highlighted in blue. These include the locator, the scale bar, the summary bar and the text features (the features listed in the feature table of the sequence record). Some data types are highlighted but do not appear in the sequence window because no such data is available for this particular sequence (for example, inverted and tandem repeats).

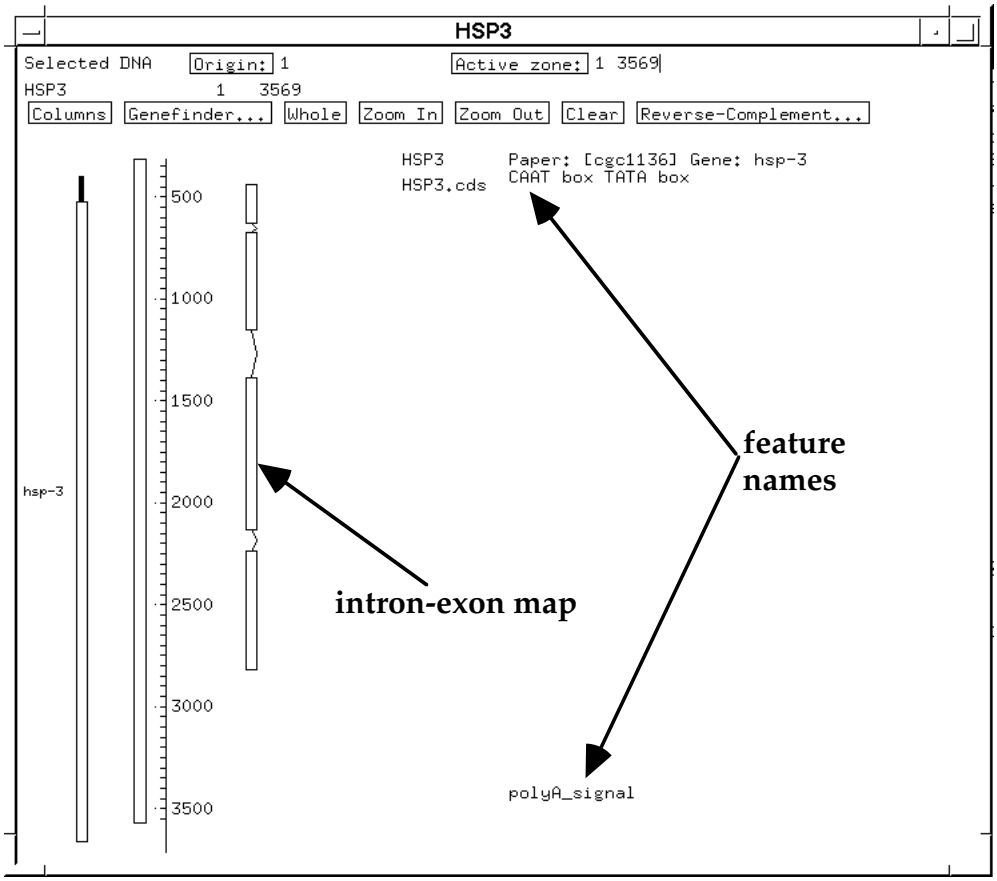
- Click on **Text Features** and on **Scale**.

  These items are no longer highlighted and will not be displayed when the **Return to display** button is clicked.

- Click on **DNA Sequence**.

  This item is now highlighted.

- Click on the **Return to display** button.

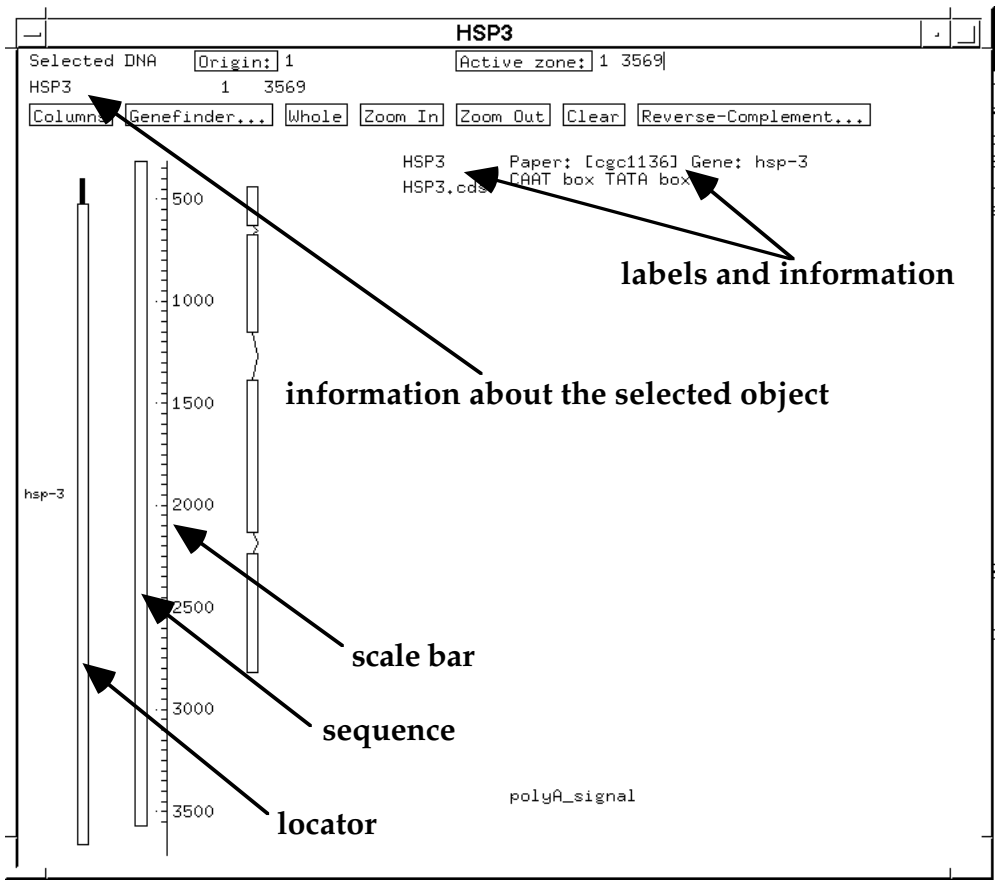


The type of information displayed in the window depends on the data available for the sequence and the settings of the sequence window. In this case, the program displays graphical representations of the features listed in the GenBank database entry for this sequence. These include the intron-exon arrangement of the gene, and various DNA signals (e.g., TATA box, polyA signal) indicated by small violet boxes (these are not visible in the figure above, but should be on a colour screen).

These features are labelled on the right side of the window. The labels and the graphical representations are linked: clicking once either on the graphical representation or on the label highlights both, and displays some additional information at the top left of the window (under 'Selected DNA').

The sequence window

Many of the controls in the sequence window are similar to those in the genetic map window: the **Zoom In**, **Zoom Out** and **Whole** buttons have the same function. The locator on the left can be used to select the region of the sequence to be displayed and its magnification by clicking on the left of the slider with the middle mouse button. Clicking on the right of the locator with the middle mouse button allows precise movement along the sequence.

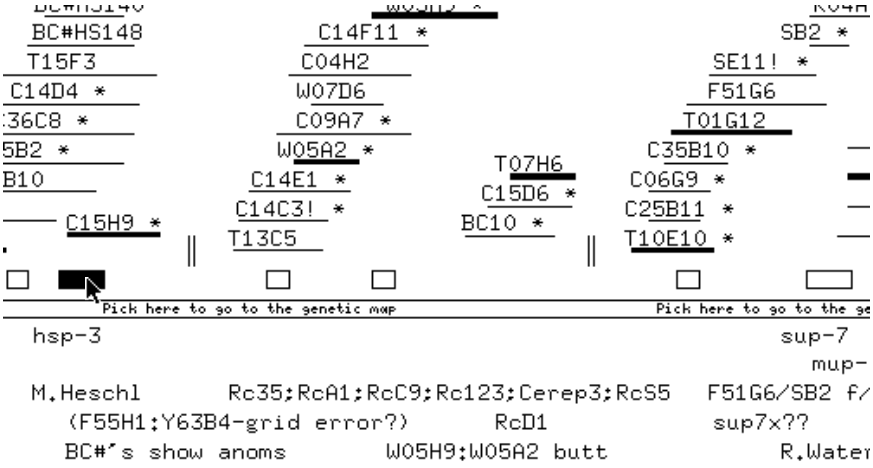


The yellow bar next to the locator is called the summary bar: it represents the DNA sequence and, and with the appropriate window settings, can be used to display various features such as restriction sites. Its scale (in base pairs) is given by the bar on its right.

Some information is given on the right side of the window, including the name of the gene and its literature reference and labels indicating various sequence features (TATA box, poly-A signal). Additional information can be obtained by double-clicking on these data items to bring up relevant text windows.

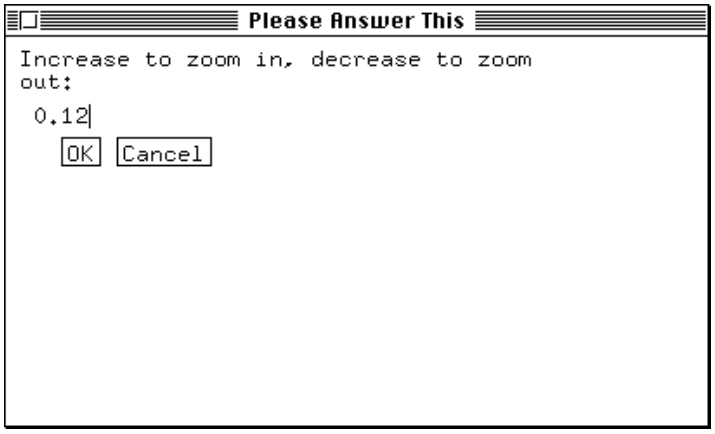
One line of information about the currently selected item is given at the top of the window (highlighted in blue on colour screens).

- Double-click on the yellow box above the name **hsp-3** (on the left side of the window) to bring up the sequence window associated with this sequence.



- Click on the physical map window with the rightmost mouse button and select **Set Zoom** in the pop-up menu.

  A new window prompts you for a magnification value.



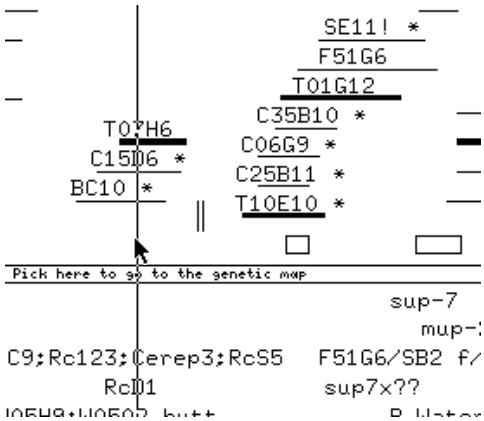
- Enter **0.4** as the new magnification value and click **OK**.

  The physical map is redrawn at higher magnification.

- Position the mouse pointer in the upper portion of the screen then press and hold down the middle mouse button.

  A single line appears on the screen at the location of the mouse pointer. This line is used for fine recentering of the map.

- Move the mouse to position the line in the middle of the clone **T07H6**.

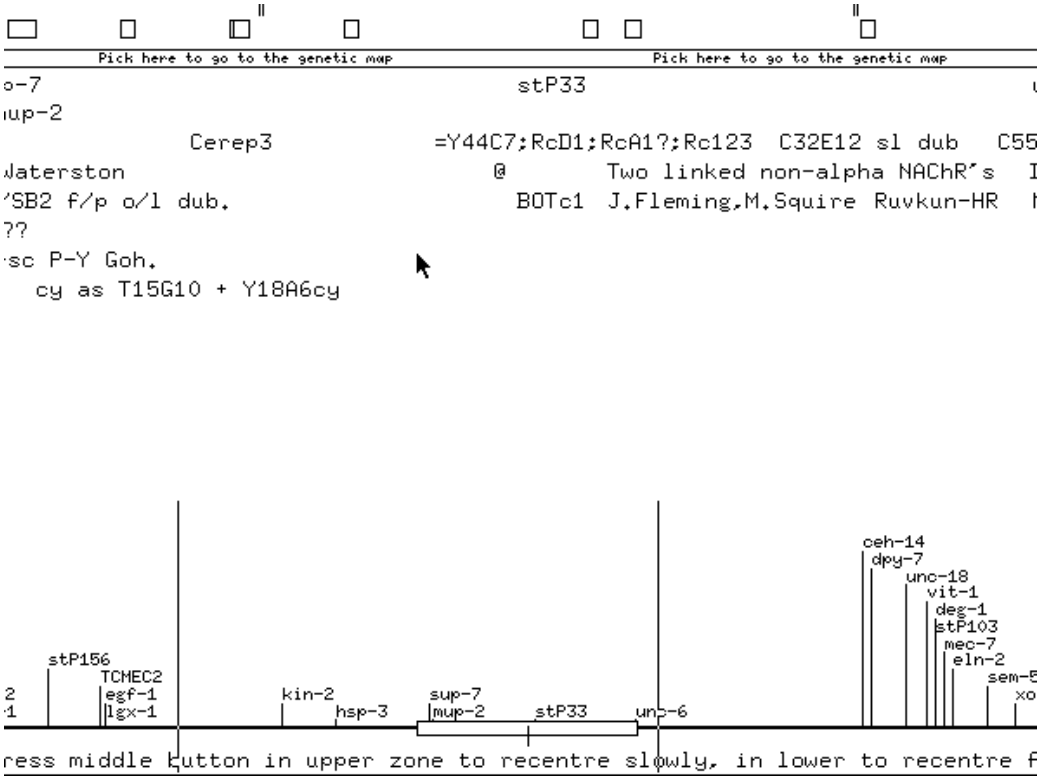


- Release the mouse button.

  The map is now centered on **T07H6**.

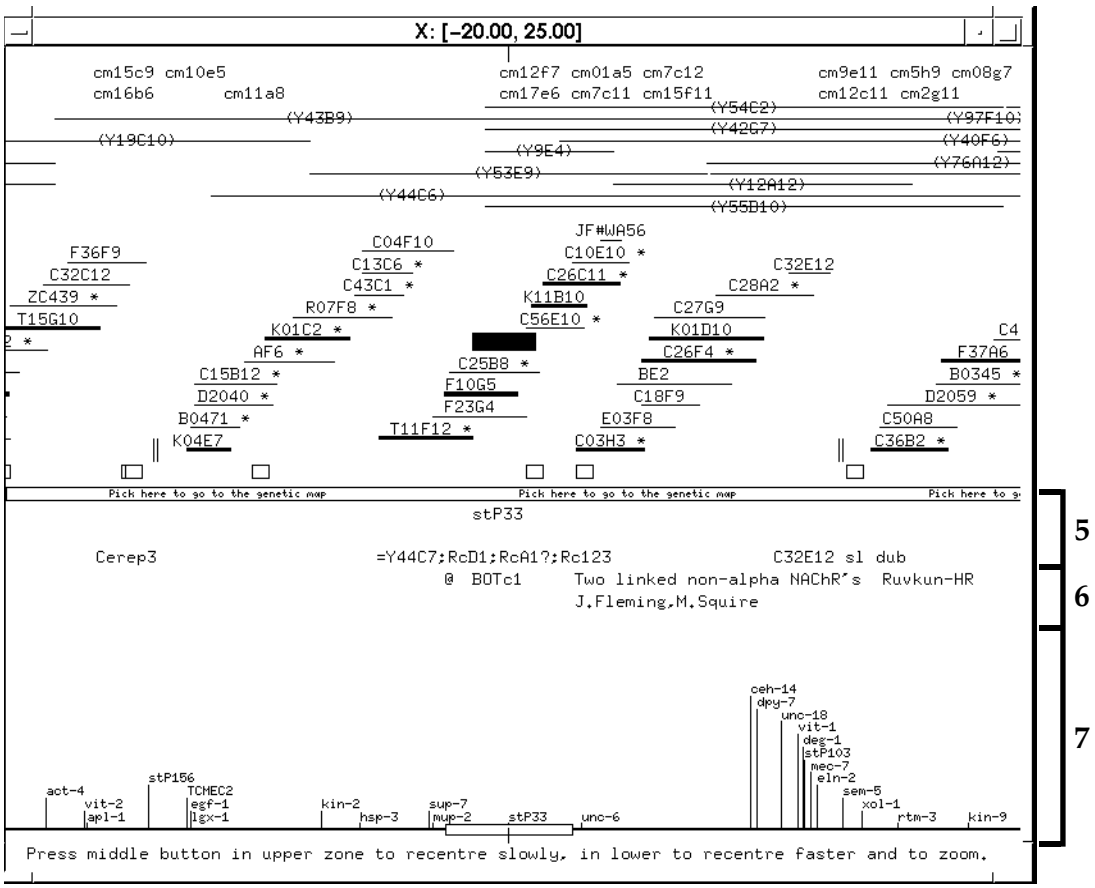
- Position the mouse cursor on or near the green slider.
- Press and hold the middle mouse button.

  Two parallel vertical lines appear on the screen around the mouse pointer. These lines define a 'window' of the physical map to be displayed: the portion of the miniature contig view lying between these lines represent the portion of the physical map which will be seen in the upper part of the window when the mouse button is released. While the middle mouse button is depressed, horizontal movements of the mouse pointer translate into movements of the window along the contig, and vertical movements widen or narrow the window, therefore reducing or magnifying the selected portion of the physical map.
- Move the mouse pointer to the left and up so that the window defined by the two lines is centered on **sup-7** and is about twice the size of the green slider.



- Release the mouse button.

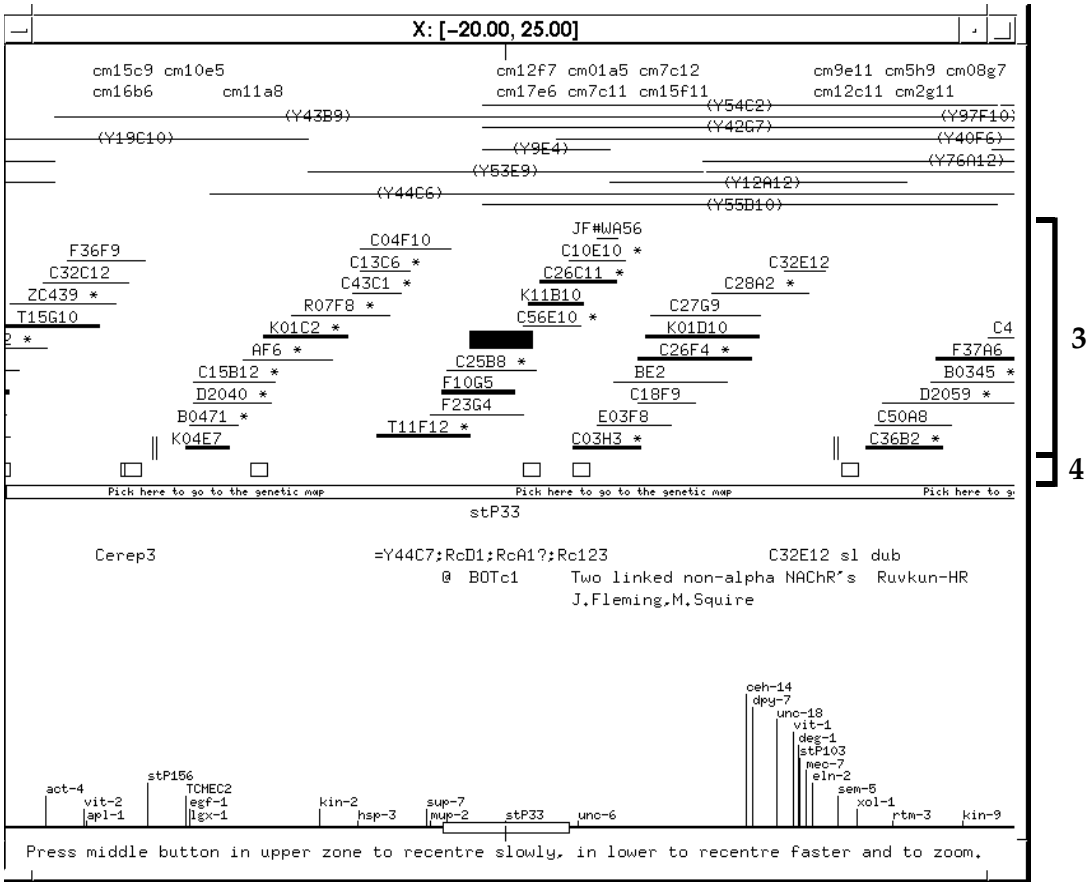
  The window is redrawn at about half its original magnification and with **sup-7** at the centre.



5: The genetic markers located at this portion of the chromosome are listed in this section, just under the green bar. The gene name is positioned so that its first letter lines up with the first letter of the name of its clone of origin (e.g., **stP33** and **RW#L33**, which is not visible in the above figure but should be highlighted in red on a colour screen).

6: This section contains selected remarks and comments about particular regions of the physical map. The remarks are aligned with the name of the clone they apply to (and any relevant gene name). A number of remarks which were considered of little interest to most users are not shown in the window, but can be displayed by selecting **Show All Remarks** in the pop-up menu.

7: The bottom part of the window contains a miniature view of the contig, with the genetic markers serving as points of reference. The green slider represents the region currently displayed in the physical map window. The slider is analogous to the locator in the genetic map, and can be used to select a new region of the physical map to display.

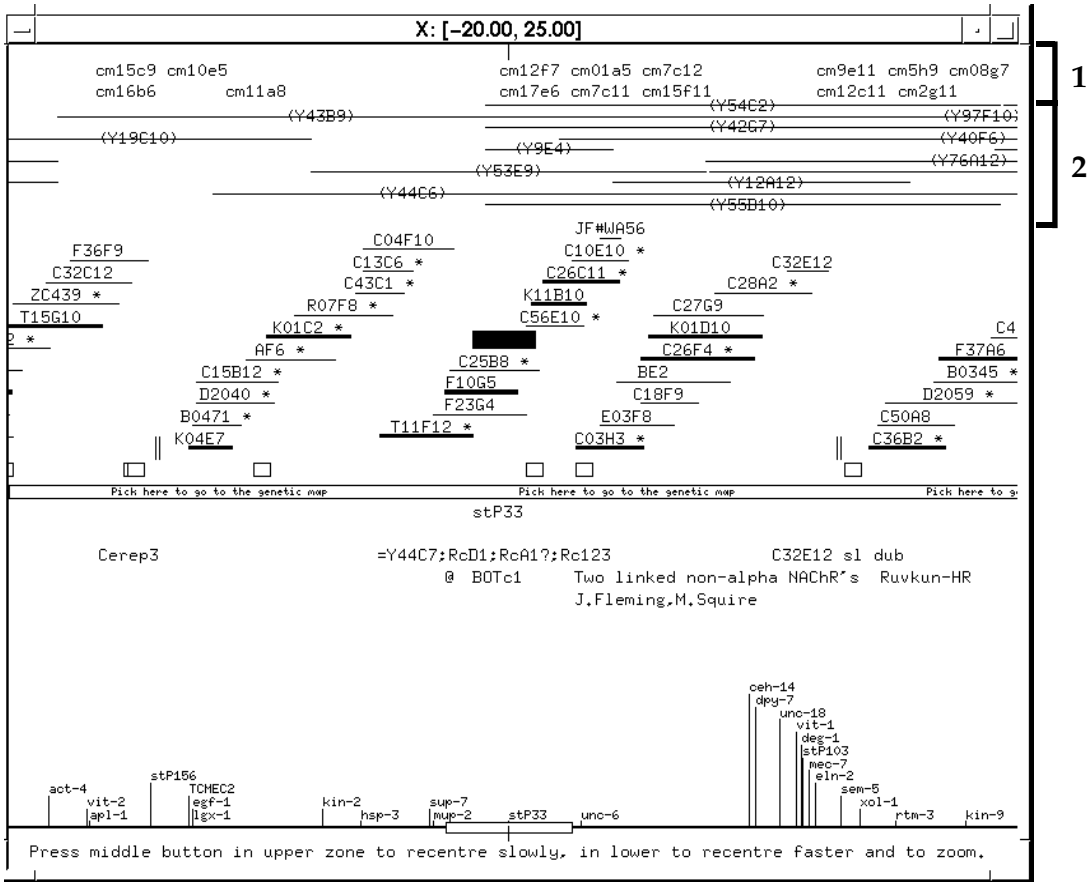


3: The short horizontal lines in this section represent cosmid clones. Some are marked with an asterisk: they are *canonical* clones, which represent a set of identical or near identical clones. Double-clicking on a 'canonical' clone lists the hidden (or 'buried') clones it represents together with other information. Cosmids represented by a thicker line (e.g. **F10G5**) are (or are canonical for) clones which have been used to determine the positions of the YAC clones by hybridization.

Note: the hidden clones can also be displayed using the **Show All Buried Clones** option in the pop-up menu.

4: The small boxes in this section (just above the green bar) represent sequences associated with some of the clones in the contig.

The layout of the physical map window is divided into sections (indicated by numbers at the right of the following figures):



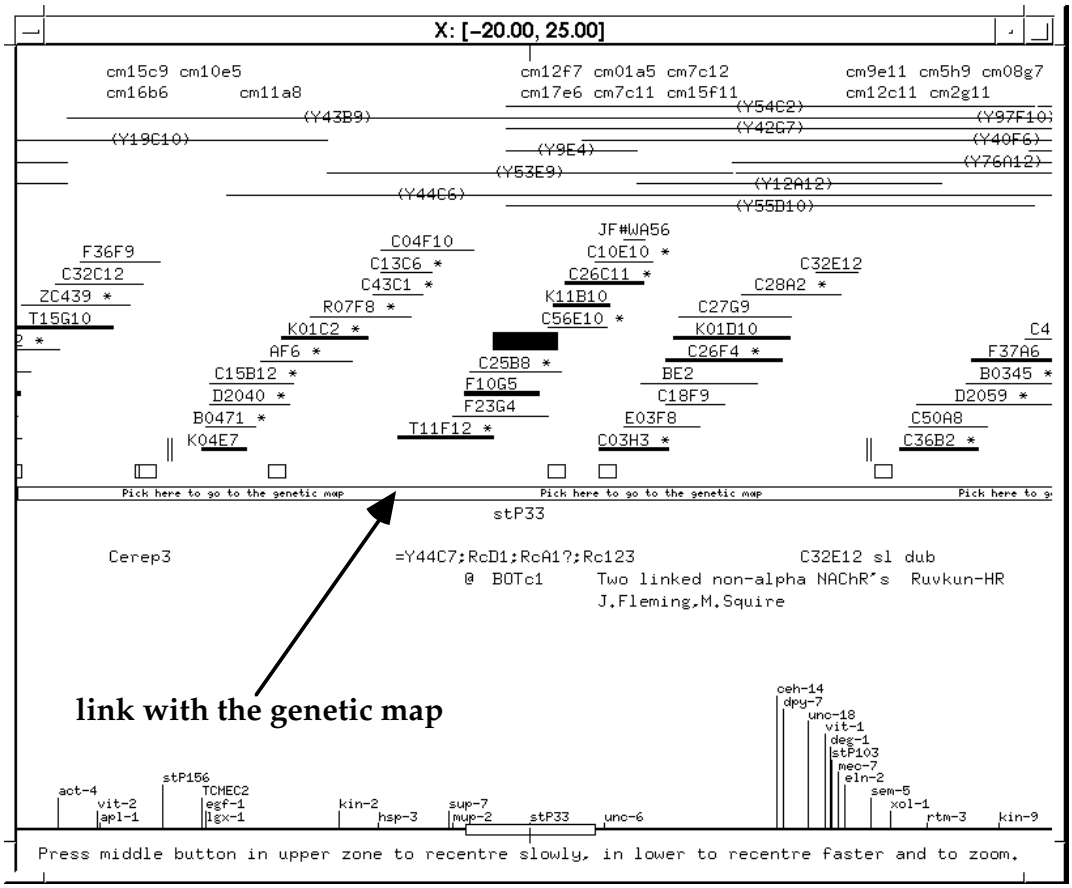
1: This section, just below the title bar, lists the clones whose location was determined by hybridization to filter-bound YAC clones.

Note: the location of these clones may not be definitive as in some cases a given clone will hybridize to several places in the genome, presumably because of sequence similarities. A list of the YACs a given clone hybridizes to can be seen by double-clicking on the clone name.

2: This section contains a series of horizontal lines identified by a label in parentheses at the centre of the line. These lines represent the YAC clones spanning the region currently displayed.

The physical map

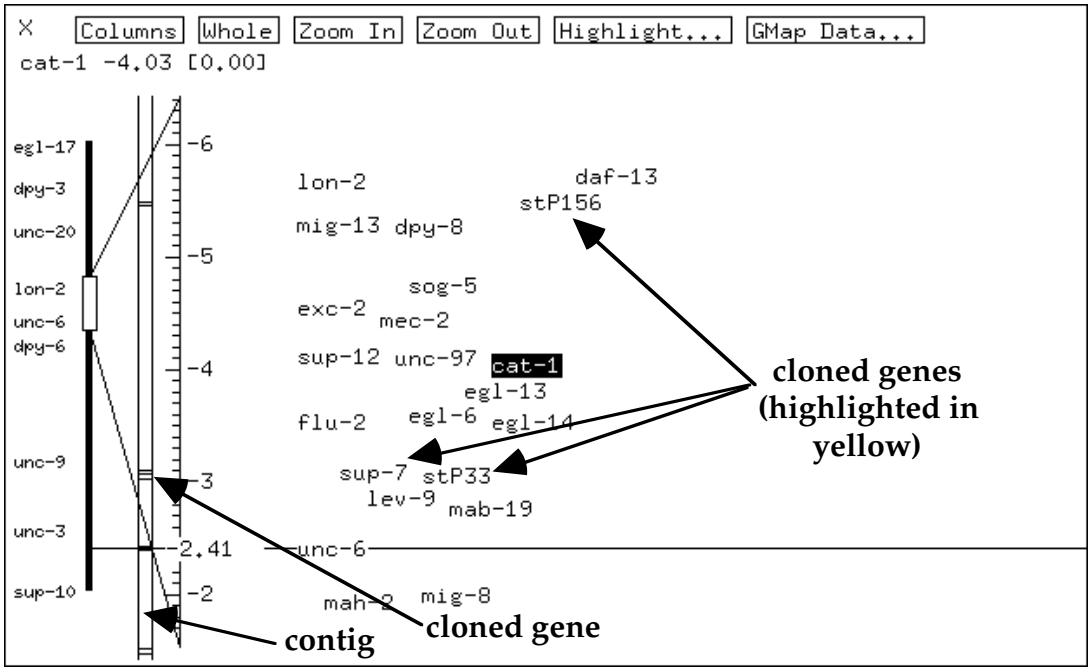
Physical mapping involves the ordering of a large number of cloned inserts into overlapping assemblies called contigs. Overlaps are determined by matching restriction digest patterns and by hybridization. The inserts can be cloned in cosmid vectors or into yeast artificial chromosomes (YACs) which can contain much larger inserts.



When the physical map appears, it is centered on the marker which was selected in the genetic map (in this case, **stP33**). On colour systems, the name of the clone this marker was located to is highlighted in red, and other items of information connected to it are shown in pink (or purple, depending on the colour display). The title bar of the window lists the region of the chromosome currently displayed on the screen.

Note: because of the window size, it may be compressed on some systems with limited screen space, and some objects may overlap.

The screen is divided by a green bar. Double-clicking on this bar will open the genetic map window. Clicking once on an item in the window will select this item: the item will be highlighted in red and related items will be highlighted in pink. Double-clicking on an item will bring up a text window with information on this item.



The yellow bar with small bands immediatly to the left of the scale bar represents the connection with the physical map. In this case the bar is continuous and covers the whole chromosome because the physical map for this chromosome is complete, and all the clones have been joined into one continuous contig (a contig is an assembly of overlapping clones).

On colour systems, some marker names are highlighted in yellow (e.g. **sup-7**, **stP33**). They represent genes which have been cloned and have been assigned a position in the physical map as well as in the genetic map. These markers act as points of reference to anchor the physical map and the genetic map together. The bands on the contig bar correspond to these markers.

- Double-click on the contig bar on the band corresponding to the marker **stP33** (indicated by the 'cloned gene' arrow in the above figure).

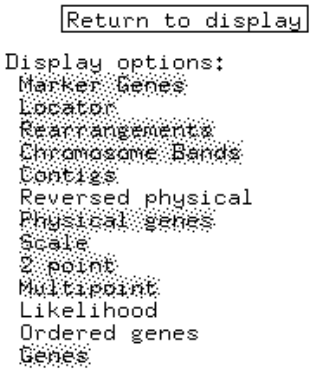
  Note: as the band is clicked, the code **RW#L33** appear at the top of the window, highlighted in blue. This is the name of the clone in the physical map containing the marker **stP33**.

  This opens a new window displaying the physical map for this region of the chromosome.

The genetic map display offers a way to 'filter' the type of information shown on the screen. This feature can be useful when the genetic map display gets very crowded.

- Click on the **Columns** button at the top of the window.

  A list of the different types of information which can be displayed in the genetic map window appears.



  The highlighted items (in blue on colour screens) are those which are to be displayed on the screen, either by default (e.g. the locator, the rearrangement bars, the scale bar and the genes), or if the appropriate command is issued (e.g. 2 point and multipoint data). The display of the various types of objects can be toggled on and off by clicking on the list.

  Note: the highlighting may not be visible on monochrome screens.

- Click on **Rearrangements** to 'unselect' it.
- Click on **Return to display**.

  The rearrangement bars no longer appear in the physical map display.

- Click on **Gmap data...** with the rightmost mouse button and select the menu option **Clear Map Data**.
- Click on a blank spot in the genetic map with the rightmost mouse button.

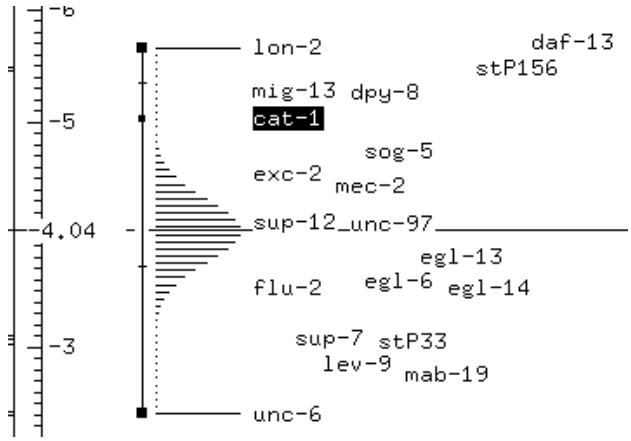
  A new submenu appears, containing some standard option such as **Quit** and **Help**.

- Select the menu option **Recalculate**.

  This recalculates the genetic map from all the available mapping data. The markers **unc-6** and **cat-1** are moved back to their original locations.

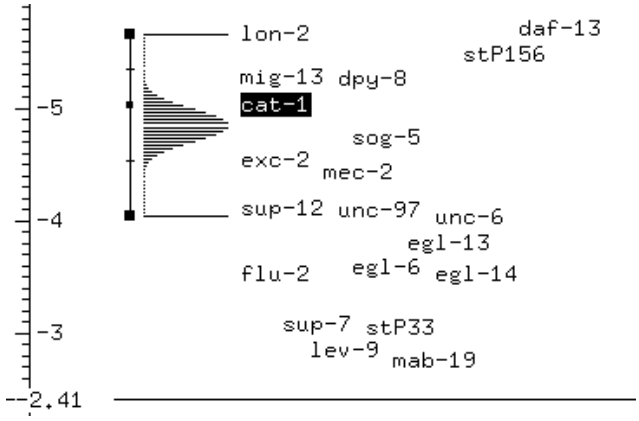
- Click on **cat-1** with the right mouse button and select the menu option **Show data**.

  The mapping data used to determine the location of **cat-1** is displayed. In this case, the position of **cat-1** was determined using a three-point mapping experiment which linked **cat-1** to the markers **lon-2** and **unc-6**.



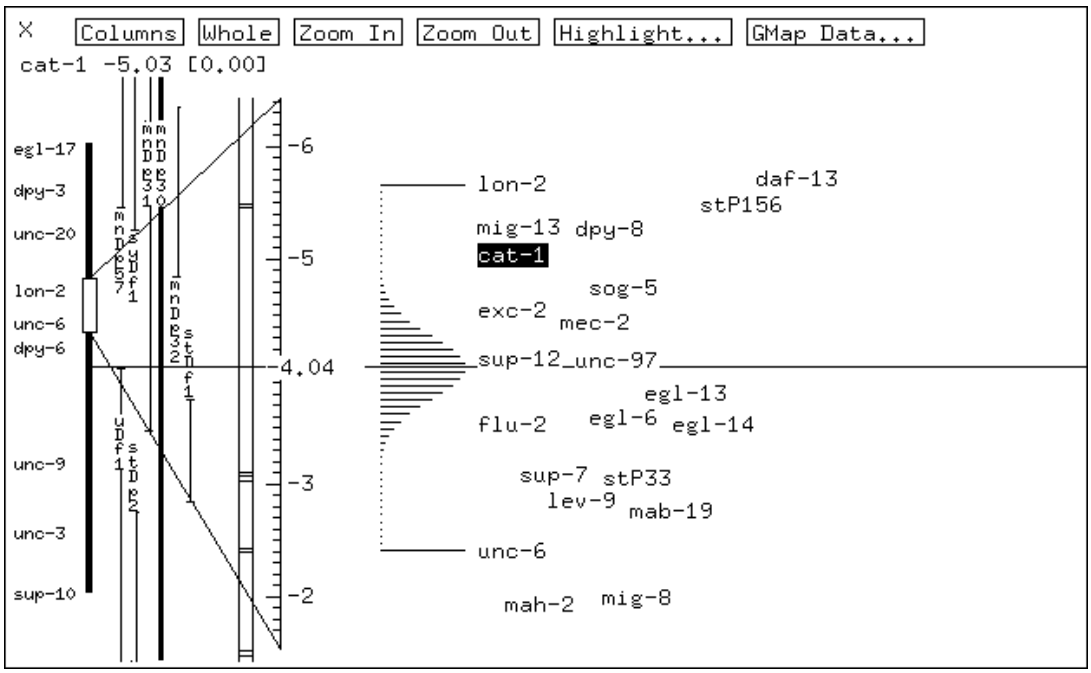
- Click on **unc-6** with the rightmost mouse button and select the menu option **Move to cursor**.

  The cursor and **unc-6** exchange positions. Since the position of **cat-1** was calculated in relation to the position of **unc-6**, the likelihood distribution curve for **cat-1** is altered.



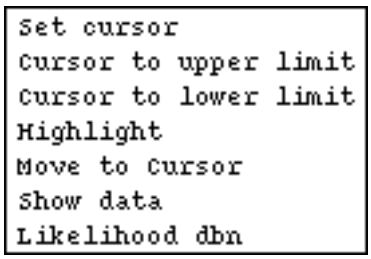
- Click on **cat-1** again with the right mouse button and select **Likelihood dbn**.

  A graph of the likelihood distribution for the location of **cat-1** appears on the genetic map. The graph indicates that the marker's previous location was the most likely given the available linkage data.



- Click on the marker **cat-1** with the *right* mouse button.

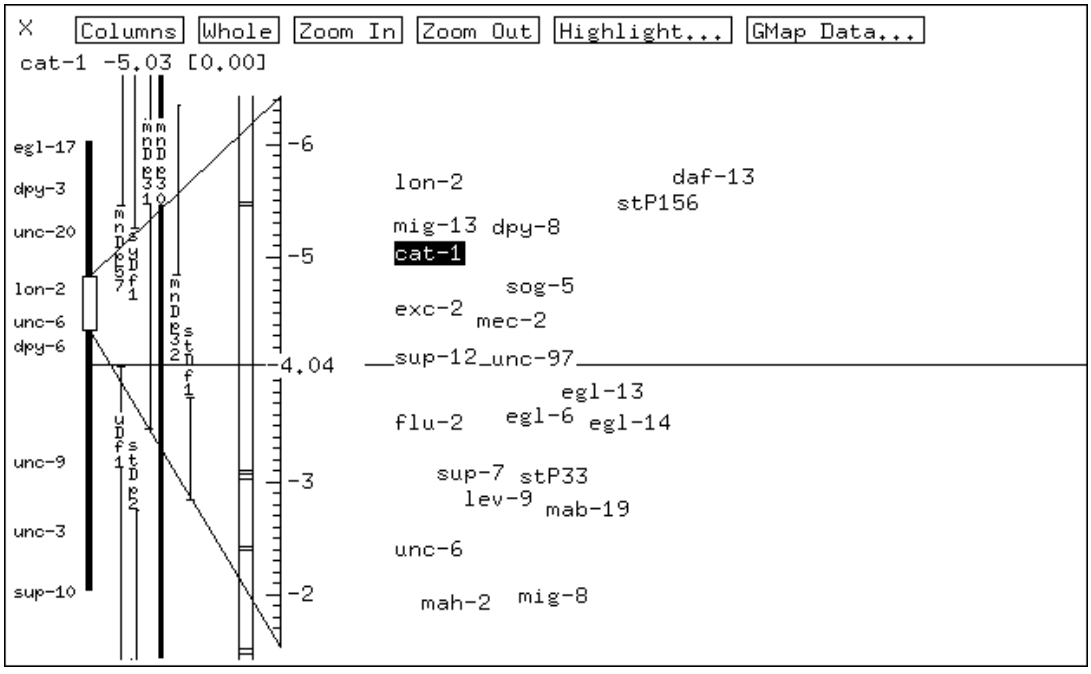
  A submenu appears:



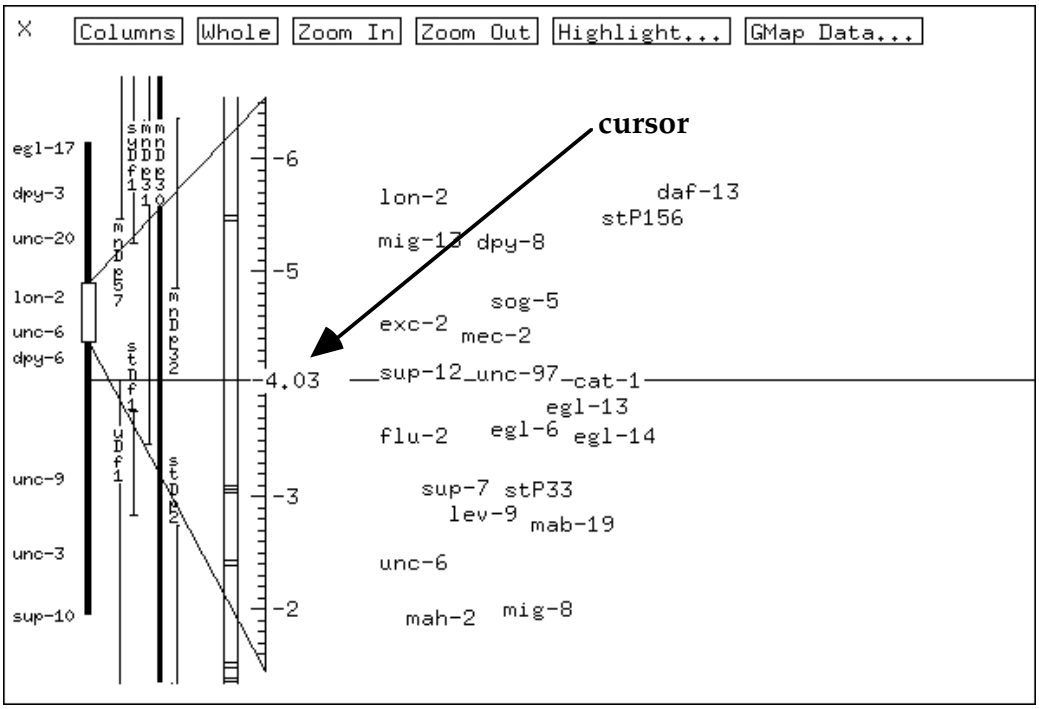
  This menu can be used to move the cursor, either to the location of the marker (**Set cursor**) or to the upper or lower limits of the marker's range of likely locations, for those markers whose position is not precisely known (**Cursor to upper limit** and **Cursor to lower limit**). The submenu can also be used to highlight the marker (**Highlight**), move it to the current cursor position (**Move to cursor**) or show its genetic map data (**Show data**, which is equivalent to clicking on the **Gmap data...** button with the left mouse button). The last menu item, **Likelihood dbn**, plots a graph showing the likelihood distribution of where the marker may be located on the chromosome, based on all the map data.

- Select **Move to cursor**.

  The marker **cat-1** and the cursor exchange locations (this provides an easy way to return the displaced marker to its original location if desired, by selecting **Move to cursor** again).



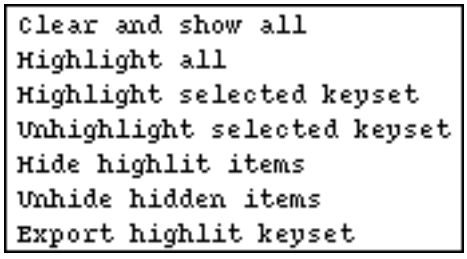
The genetic maps presented by **ACeDB** show the arrangement of markers along the chromosomes which best fit the overall data. Users without database editor privileges cannot alter these genetic maps permanently. However, it is possible to 'move' markers experimentally. Before moving a marker, the new position must be specified using the cursor.



- Click on the cursor with the left mouse button and drag it up the scale bar to position **-5** approximately.

- Bring the genetic map window back to the foreground and click on the **Highlight** button with the *right* mouse button.

  A new menu opens:



- Select **Highlight selected keyset**.

  The marker name **cat-1** (the only one in the current view in the list in the **Main Keyset** window) is highlighted in magenta.

- Select **Hide highlit items** in the **Highlight...** menu.

  The marker **cat-1** is now invisible.

- Select **Unhide hidden items** in the **Highlight...** menu.

  Cat-1 reappears.

- Open the **Highlight...** menu and select **Highlight all**.

- Select **Unhighlight selected keyset**.

  This time, all items in the genetic map are highlighted except for **cat-1**.

- Select **Hide highlit items** in the **Highlight...** menu.

  All the highlighted features disappear, leaving only **cat-1** in the genetic map.

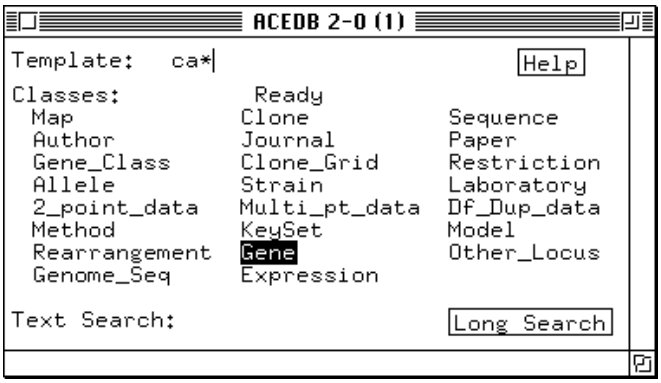
- Open the **Highlight...** menu and select **Clear and show all**.

  This restores the default view showing all markers.

  Note: the **Highlight...** menu is especially useful to highlight the markers retrieved using the **Query** facility of **ACeDB**, which will be demonstrated later.

The button labelled **Highlight...** is used to highlight or to hide specific features of the genetic map window. It is best used in conjunction with the main **ACeDB** window.

- Bring the main **ACeDB** window to the foreground (do not quit the genetic map).
- Click on the **Template** box and type **ca*** (do not press <**return**>).

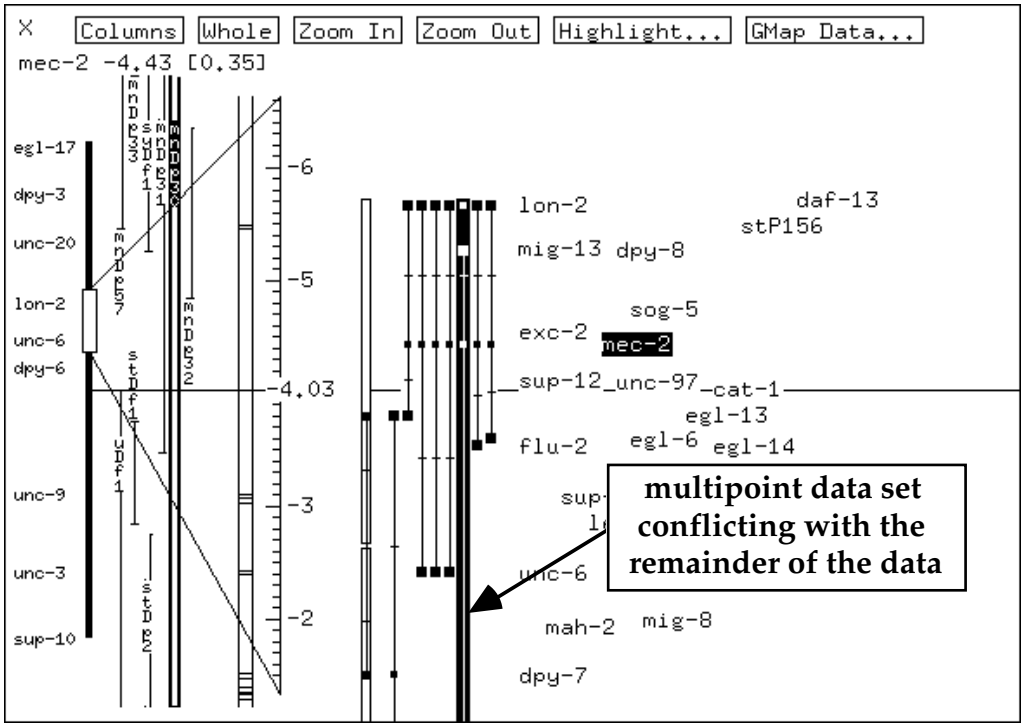


- Double-click on the class name **Gene**.

The **Main Keyset** window comes to the foreground and lists those genes with a name starting with 'ca'. The window indicates that this list is the current 'selected keyset'.

- Click on the marker **mec-2** with the leftmost mouse button to select it, then click on the **Gmap data...** button.

  The mapping data used to position **mec-2** on the genetic map is shown, highlighted in green on colour screens. The previously displayed mapping data (for **egl-13**) remain on the screen, but is not highlighted.



  One of the multipoint data bars for **mec-2** is shown in dark (dark green on colour screens). This bar corresponds to a multipoint mapping experiment which predicted a marker order contradicting the remainder of the data.

  Note: on colour screens, some rearrangement bars may become highlighted when the genetic mapping data for a marker are displayed. This indicates rearrangements which are known to cover or not to cover the selected gene: the rearrangements covering the gene are highlighted in green and the rearrangements not covering it are highlighted in blue.

- Click on the **Gmap data...** button with the *right* mouse button.

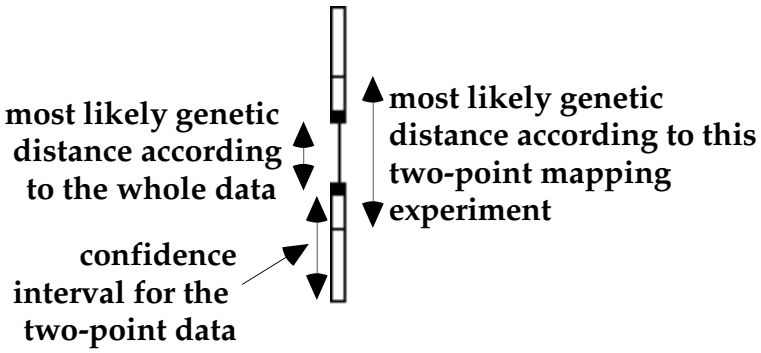
  This opens a pop-up menu offering access to some of the advanced features of the genetic map display.

- Select the top menu item (**Clear Map Data**).

  The two-point and multipoint data bars are erased.

The two-point and multipoint data bars show how closely the experimental data match the overall genetic map. The markers themselves are represented by small black squares whose position indicates the most likely marker location, as calculated using all the available data.

For each two-point data bar, the genetic distance between these markers which was calculated using only the data from this particular experimental cross is indicated using short horizontal lines. This distance may not necessarily agree with the overall data. Confidence intervals for the two-point data are indicated by rectangles.



For example, in the above figure, the 2-point data linking the two markers predicts a distance between these markers which is higher than that calculated by taking the whole genetic map into account (the distance between the horizontal lines is larger than the distance between the two black squares).

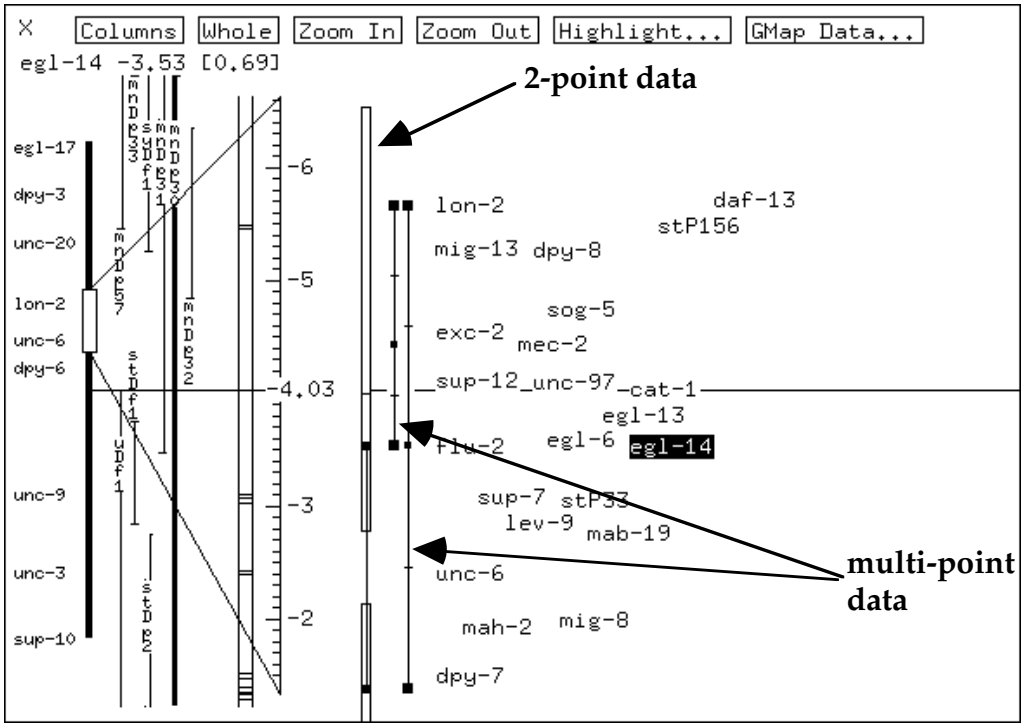
For multipoint mapping data, only the marker order predicted by the experimental data is shown.

- Click once on the marker name **egl-14** to select it.

  The selected gene is highlighted in red on colour systems.

- Click on the **GMap Data...** button.

  A graphical representation of the mapping data available for this marker appears. Both 2-point data (calculated from the observed recombination frequency between a pair of markers) and multipoint data (obtained by analysing the inheritance of three or more markers) are shown. In this case, the position of **egl-14** was deduced from one two-point mapping and two multipoint mapping experiments.



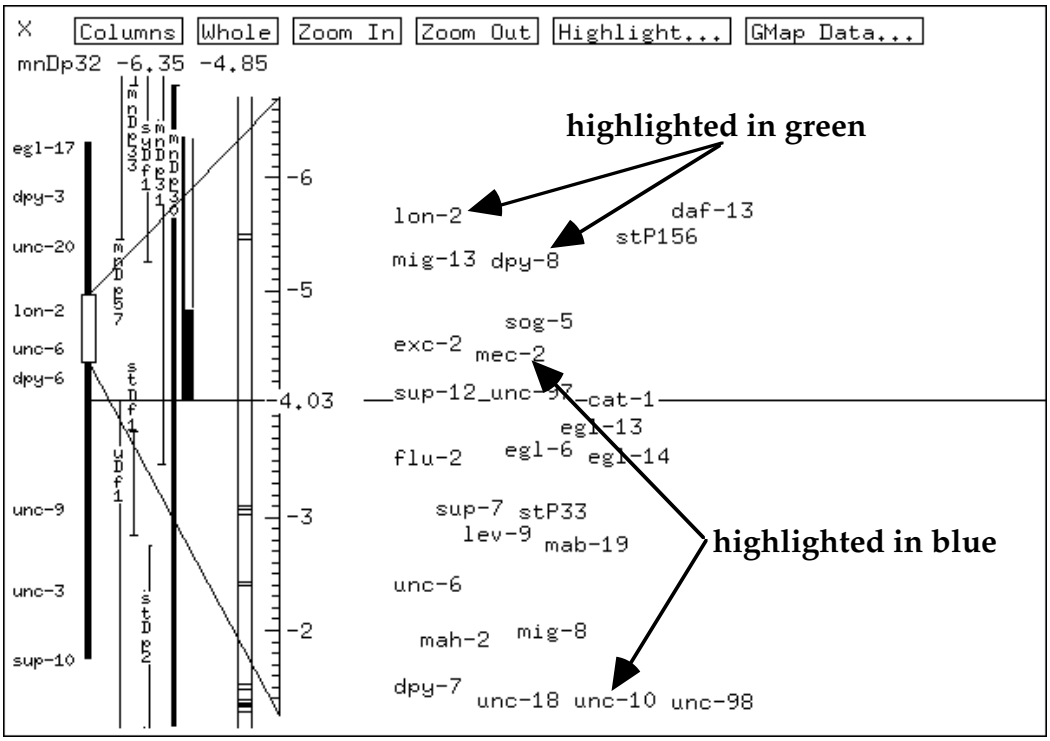
- Click once on the bar representing the results of the two-point mapping experiment.

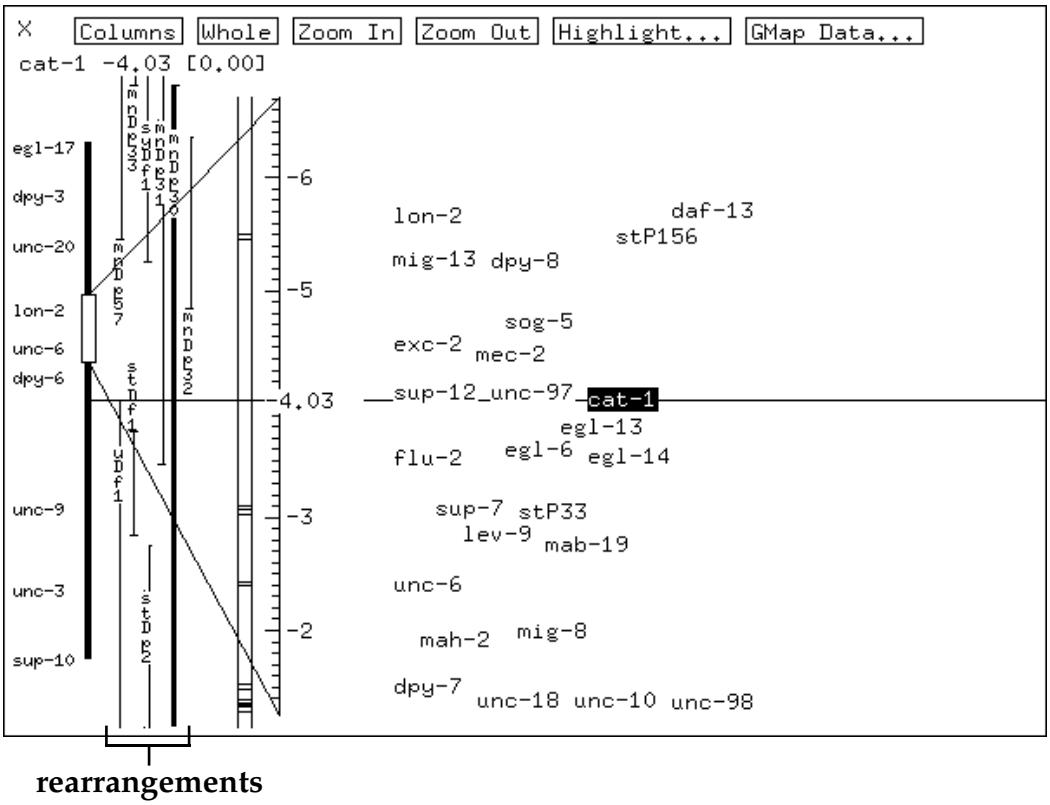
  A one-line summary of the data is presented at the top of the window. Double-clicking on a mapping data bar brings up a text window describing the mapping data.

It is possible to review some of the genetic data underlying the map using the **GMap Data...** button.

- Try clicking (once) on the various rearrangements until the one called **mnDp32** is highlighted.
- Click on the **GMap Data...** button with the leftmost mouse button.

On colour systems, this highlights a few markers. Those highlighted in green (e.g., **dpy-8** in this case) are those genes known to map inside the rearrangement boundaries. The genes known to map outside the rearrangement boundaries are highlighted in blue (**mec-2** for example in this case).





rearrangements

The vertical bars between the coarse control slider and the yellow bar represent known chromosomal rearrangements. On colour systems, it is possible to distinguish between deletion rearrangements which appear as a single line and duplications which are shown as a thicker bar or as a thin bar with a light grey 'shadow' (the 'shadow' may not be visible on monochrome screens).

- Click once on one of the rearrangement bars with the leftmost mouse button.

  The selected rearrangement is highlighted in red and relevant information (name and location of its ends) appears at the top of the screen in blue highlight.

- Double-click on a rearrangement bar.

  A new text window appears with information about this rearrangement and links (in bold type) to other windows containing further information.

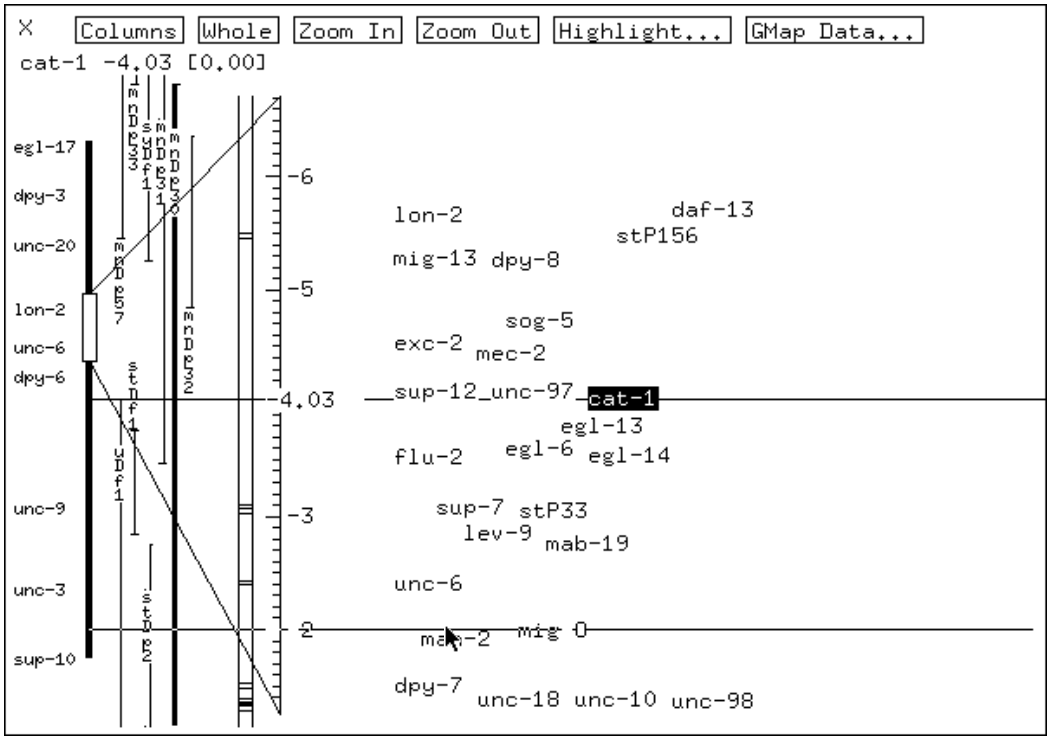
- Click on the **Quit** button to return to the genetic map.

Precise movement along the chromosome is possible using the middle mouse button.

- Click anywhere to the right of the locator with the middle mouse button and keep this button depressed.

A horizontal bar appears under the mouse pointer. This line indicates where the map will be recentered when the slider is released.

- Drag the line down to the marker **mig-8** and release the mouse button.



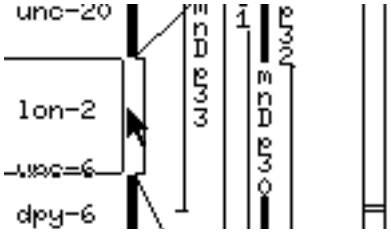
The map is redrawn, vertically centered on **mig-8**.

- Use the middle mouse button to center the map back on **cat-1**.

The locator can also be used to control the magnification of the display.

- Click on the left side of the locator slider with the *middle* mouse button and keep the button depressed.

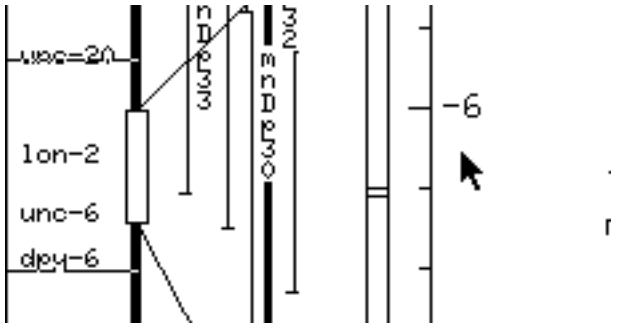
Two parallel horizontal lines appear on the screen at the ends of the slider. These lines define a 'window' of the genetic map to be displayed: the portion of the chromosome lying between these lines represent the portion of the genetic map which will be seen in the right part of the window when the mouse button is released. While the middle mouse button is depressed, vertical movements of the mouse pointer translate into movements of the window along the chromosome, and horizontal movements widen or narrow the window, therefore reducing or magnifying the selected portion of the genetic map.



- Move the mouse pointer right and left without releasing the middle mouse button to see the effect of the movements on the horizontal lines.

Moving the pointer to the left of the slider brings the lines closer together, therefore 'zooming into' the display. Moving the pointer to the right of the slider 'zooms out' of the display.

- Position the mouse pointer to the right of the slider so the horizontal lines are separated by approximately twice the length of the slider, and release the mouse button.



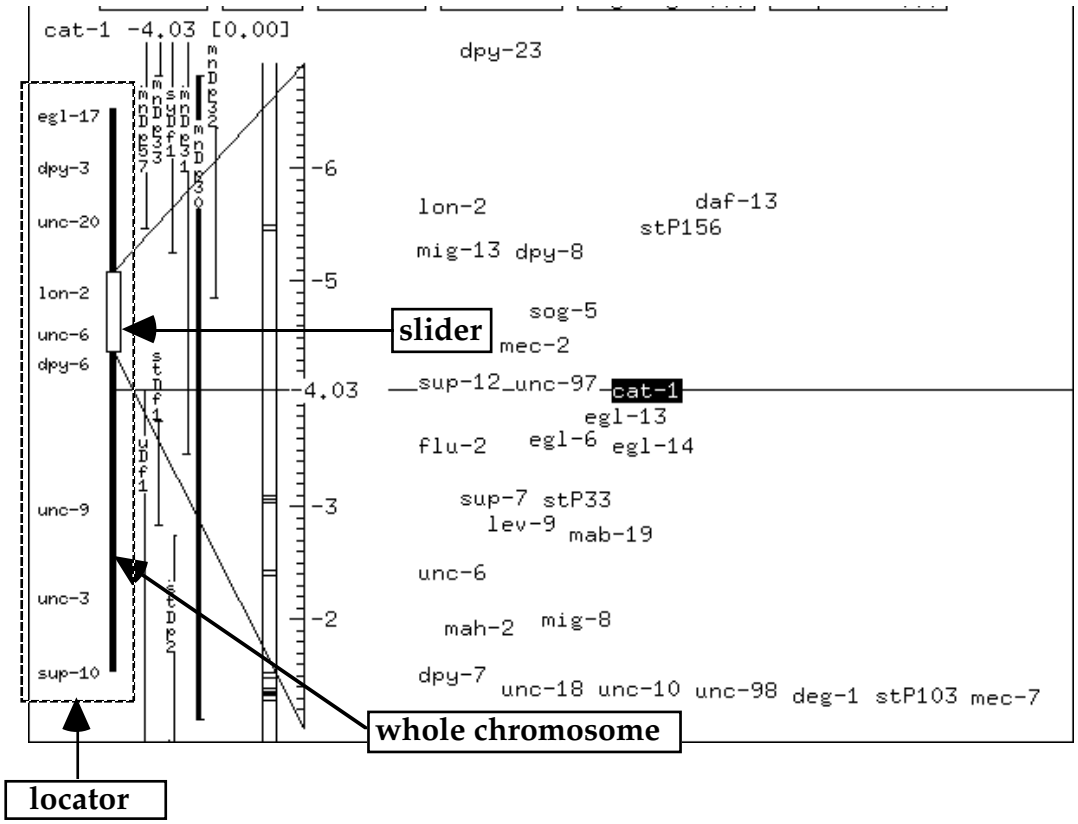
The genetic map display is reduced by a factor of about 2. The size of the slider changes to reflect the change in size of the portion of the chromosome now displayed on the screen.

The map magnification is also controlled by the buttons **Zoom In**, **Zoom Out** and **Whole** at the top of the window.

- Click on the **Zoom In** button.

This brings the magnification back to around its original value.

Note: clicking on the **Whole** button zooms out to show the whole chromosome and all its loci.



Rapid movement along the chromosome is achieved using the locator on the left side of the genetic map. The locator consists of a black bar representing the whole chromosome and of a slider (green on colour systems) which corresponds to the portion of the chromosome currently displayed. A few genetic loci are marked along the locator to give a rough indication of the current slider position.

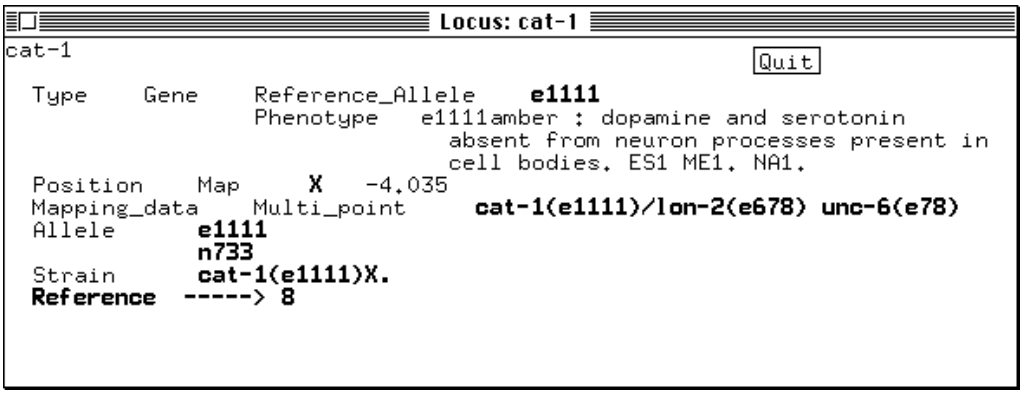
The slider can be dragged along the black bar using the lefmost mouse button to move along the chromosome.

- Click on the slider with the left mouse button and drag it down (without releasing the mouse button) so it is centered on the locus **unc-9** (written on the left side of the locator).
- Release the mouse button.

  The map now shows the portion of the chromosome around locus unc-9.
- Drag the slider back to its original position (centered approximately between **lon-2** and **unc-6**).

- Double-click on the marker name **cat-1**.

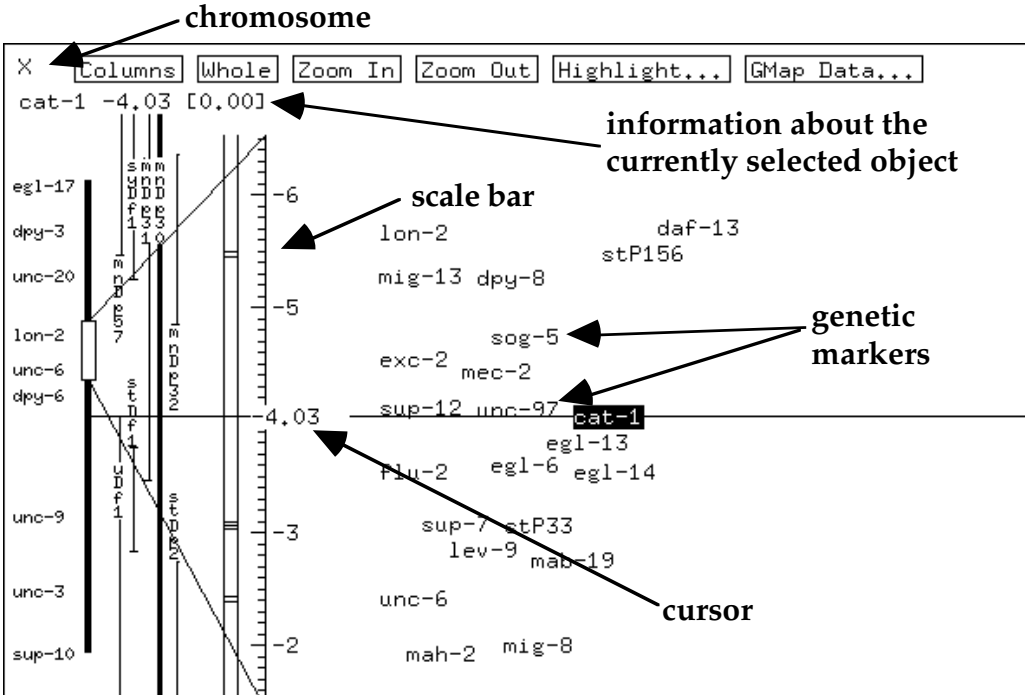
  This brings up a new text window containing the known information about this marker.



  All the information linked to this marker can be accessed from this window by double-clicking on the appropriate entry (in bold type).

- Click on the **Quit** button to return to the genetic map.

A new window opens, displaying graphically the location of this gene in the genetic map of its chromosome of origin (in this case, chromosome X).



The right hand side of the map lists the genetic markers in this region of the chromosome. The position of these markers on the chromosome as determined using linkage analysis is indicated by their vertical positioning (relative to the scale bar). Their horizontal position gives an indication of the confidence with which the map position is known. The further to the left a marker is in this window, the more certain its position on the genetic map. For example, in this case, the position of **flu-2** is known with more confidence than that of **egl-14**.

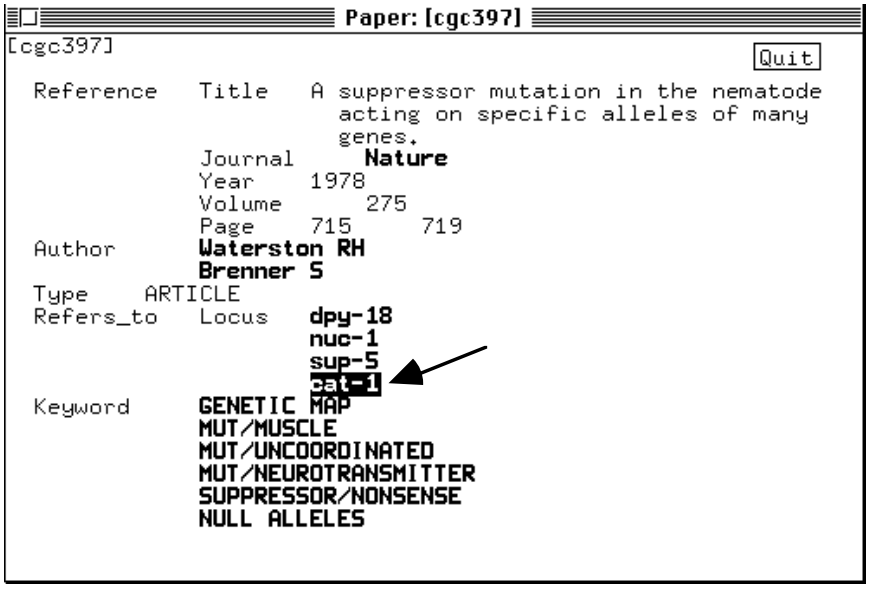
The currently selected gene (**cat-1**) is highlighted in red, and markers assigned to a clone in the physical map are shown with a yellow background (e.g., **sup-7** and **stP33**). Information on the currently selected object is displayed at the top of the window (highlighted in blue on colour systems). In this case, since a locus is selected, the position of the locus on the chromosome is given.

A cursor is present on the scale bar. On colour screens, it appears as a green box labelled with a number indicating its current position (on monochrome screens, only the number may be visible). A horizontal line running across the screen marks the current cursor position. The function of the cursor will be demonstrated later on in the tutorial.

Note: marker positioning may not be very accurate: **ACeDB** does not overlap marker names, and, when the map is crowded, moves the names slightly horizontally or vertically to make them fit better. This effect is reduced or eliminated when the view is magnified using the **Zoom In** button. It is therefore a good idea to examine markers of interest in close view.

Genetic maps

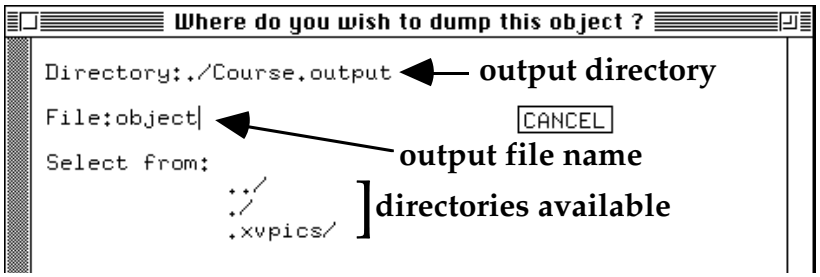
- Double-click on the word **cat-1** (one of the loci referred to in the paper).



Saving the contents of a text window

- Select **Text dump** in the pop-up menu of the 'Paper' window.

  A new window appears:



  The **Directory** box lists the current directory in which the text file is to be saved. You can change this directory by clicking on its name, deleting it and typing in a new name, or just by double-clicking on the desired directory name in the list of available directories.

  The **File** box contains the name of the output file (in this case, 'object' by default).

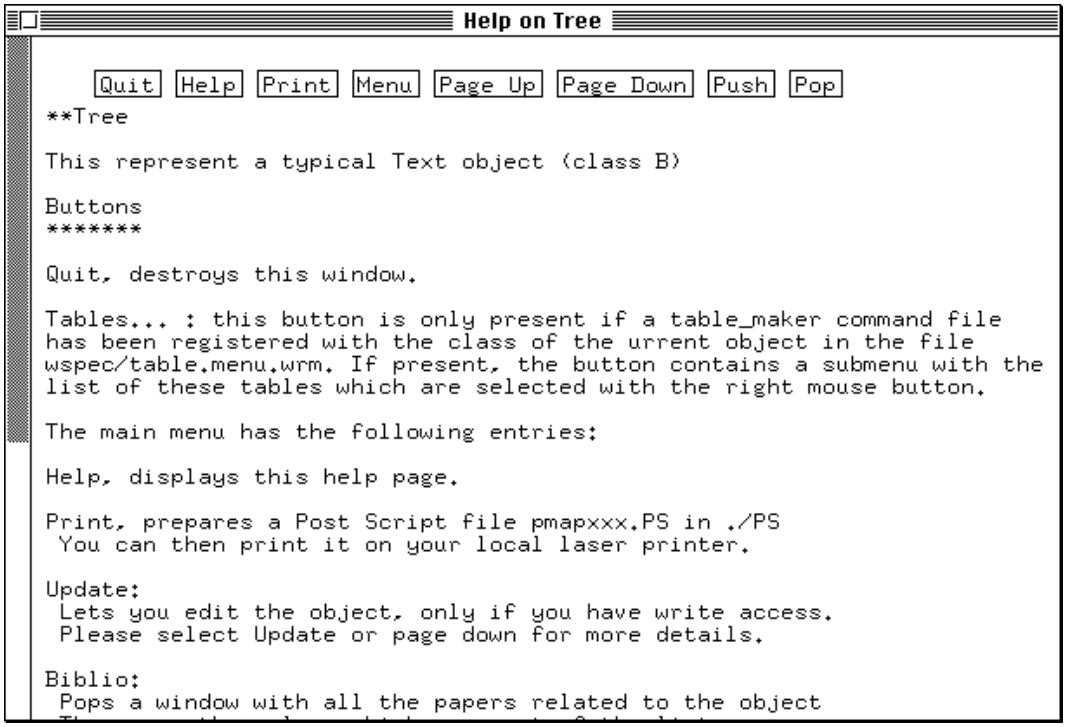
- Click on the **File** box and delete **object**. Type **paper** instead as the output file name.
- Press <**return**> to save the information (you may need to press it twice).

  ACeDB notifies you that the contents of the window were saved as a text file in your area. In this case, the name of the file is 'paper.txt' (**ACeDB** adds the '.txt' automatically).

On-line help

- Select **Help** in the pop-up menu.

The program opens a help window describing the buttons and menu options available in a text window.



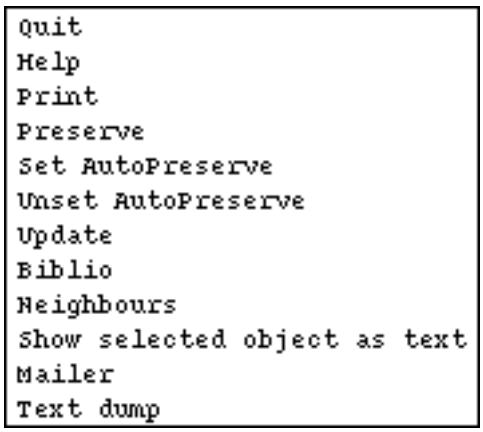
The following buttons are available at the top of the window:

Quit: leave help
Help: displays the first page of the on-line help.
Print: prints the contents of the window (unavailable on some systems).
Menu: lists the available help topics (a topic can then be selected by clicking on its name).
Page Up: displays the previous topic in the on-line help.
Page Down: displays the next page in the on-line help.
Push and **Pop:** pushes the help window to the background or brings it to the foreground (may not work on some systems).

These functions are also available in the pop-up menu.

- Scroll through the help page using the scroll bar on the left.
- Select **Quit** in the pop-up menu to leave help.

- Have a look at the pop-up menu for this window by clicking inside it with the rightmost mouse button.



This menu is typical of most text windows in **ACeDB**. It offers the following options:

Quit: closes the window.

Help: brings up a relevant on-line help window.

Print: sends the contents of the current window to your local printer. This feature may be disabled on some systems.

Preserve, **Set AutoPreserve**, **Unset AutoPreserve**: these options determine whether the window remains when another window of the same type is called. For example, in this case, if you wanted to display the information about another journal article, this information would be displayed in the same window and would replace the current article, unless **Preserve** or **Set AutoPreserve** had been previously selected.

Update: this is used to change the contents of **ACeDB**'s database, and may be disabled on some systems.

Biblio: displays some bibliographic information about the window when available.

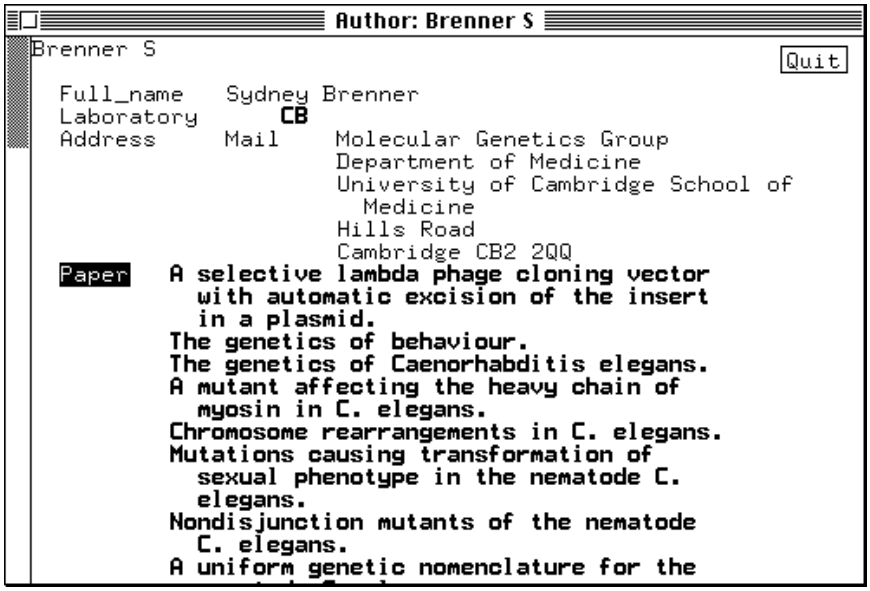
Neighbours: lists the links available from the current window (that is, the items in bold type).

Show selected object as text: displays the information linked to a highlighted item (same as double-clicking on the item).

Mailer: sends the text of the window as an email message to a specified address (inoperative on some systems).

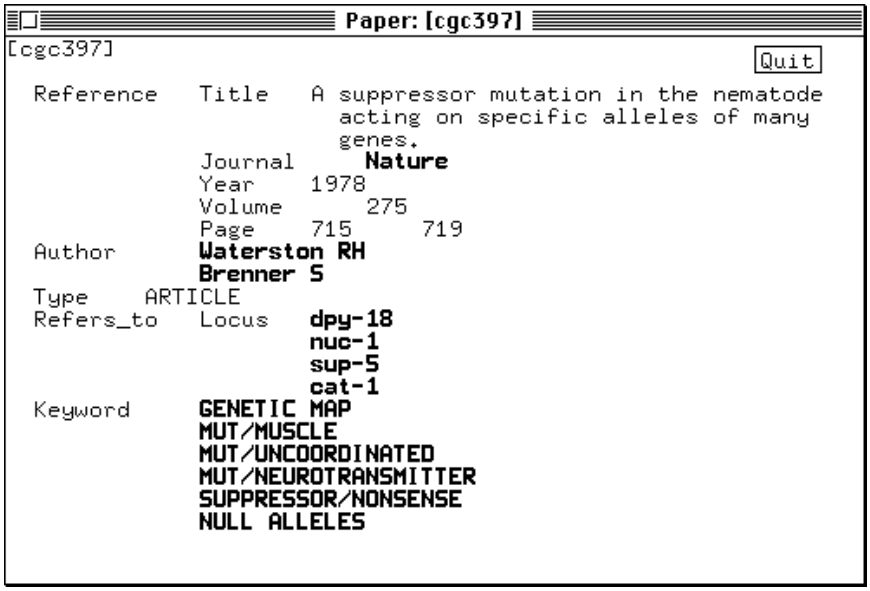
Text dump: saves the information in the window as a text file.

The titles of the papers are now listed in the window. A scroll bar appears on the left of the window to allow the perusal of these titles (the left mouse button is used to scroll down, the right mouse button to scroll up and the middle mouse button to drag the slider).



- Scroll down the list until you can see the title **A suppressor mutation in the nematode acting on specific alleles of many genes**.
- Double-click on this title.

This brings up a new window describing the selected paper.

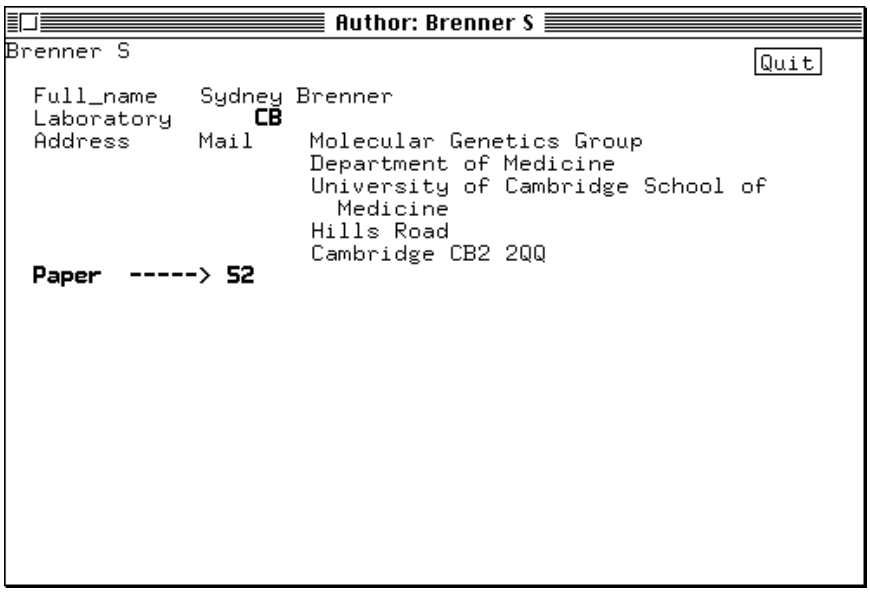


Again, double-clicking on the items in bold type displays the information linked to these items. For example, double-clicking on the journal name (**Nature**) brings up a list of all the articles in the database published in **Nature**. In the same way, double-clicking on one of the keywords displays the titles of all the other articles in the database which list this keyword. In turn, the information about any one of these articles can be retrieved by double-clicking on its title in the list.

Text windows

- Double-click (with the leftmost mouse button) on the words **Brenner S** in the **Main Keyset** window.

  A new window appears listing the information on this author:



  Some information is given in bold type. This indicates that these items of information are linked to additional information, either in this window or in other windows. Text linked in such a manner is known as 'hypertext'. For example, **Paper -----> 52** indicates that this author is associated with 52 papers in the database.

- Double-click on **Paper -----> 52**.

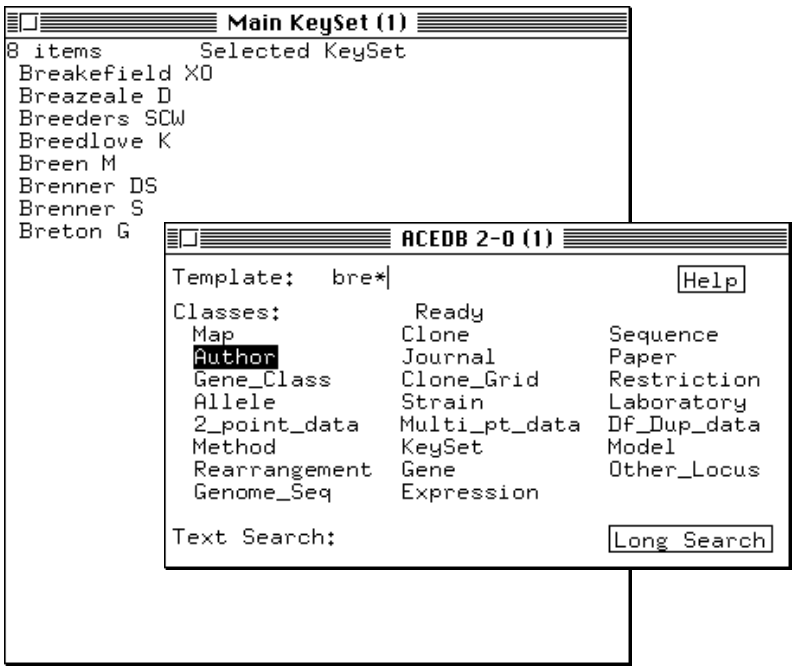
It is possible to limit the quantity of information displayed by specifying a template in the main window. For example, to list only the authors whose name start with 'Bre':

- Click in the main window, on the word **Template**.

  This activates the cursor in the template entry box (on a colour system, the box turns yellow). An asterisk (*) is currently in this box. This is a wildcard character which can replace any group of characters. As a result, all available entries for the selected class are listed in the **Main Keyset** window.

- Delete the * character, and type **bre*** instead (do not forget the *), then press <**return**>.

  Only those authors with a name fitting the template (that is, 'bre' followed by anything) are now listed in the **Main Keyset** window.

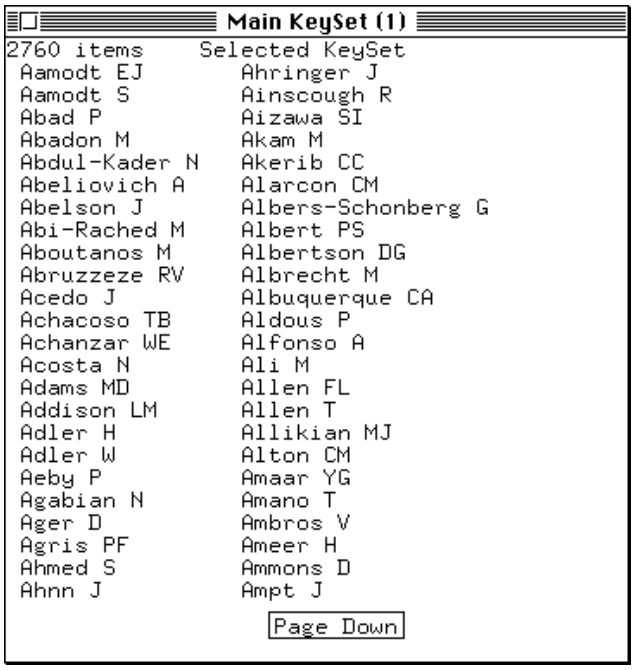


Note: the 'wildcard' characters available to specify a template are: *, which can replace any group of characters, **?** which stands for any single character and **#** which can be used to specify any numeral (any group of digits). If no wildcard character is used, only those entries matching the template exactly are listed.

Selecting classes

- Double-click (with the leftmost mouse button) on the word **Author** in the main window.

  The **Main Keyset** window changes to display a list of contributing authors.



  Because the author list is so large, only the first page can be displayed in the window.

- Click on the **Page Down** button at the bottom of the **Main Keyset** window.

  The next page of the author list is shown. In addition to the **Page Down** button, a **Page Up** button is now available at the top of the window to allow moving back one page.

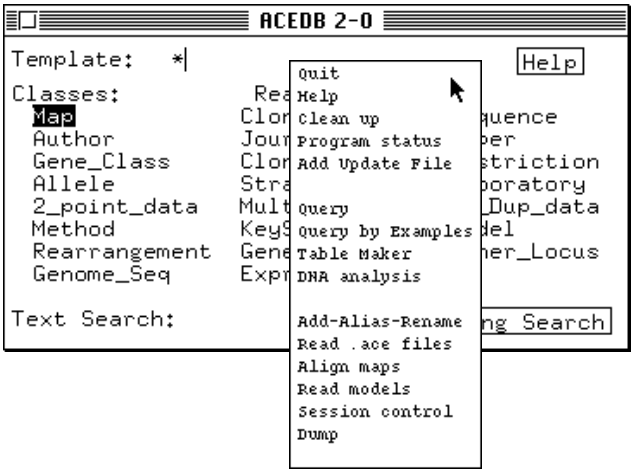
  Note: as you move through the author list, you may notice some of the author names are listed in italics. This indicates that no additional information on these authors is available in the database.

Pop-up menus

Each **ACeDB** window is controlled by a pop-up menu specific to this window. This menu is accessed by clicking inside a window with the rightmost mouse button.

- Click anywhere in the main **ACeDB** window with the *rightmost* mouse button.

The pop-up menu for the main window appears:



The menu options are different for each type of window, but two items are available for all windows. The **Help** option brings up an on-line help window relevant to the window the mouse pointer is currently in. Choosing the **Quit** option from the pop-up menu of the main window will quit **ACeDB**. Choosing it from any other window will close that window only.

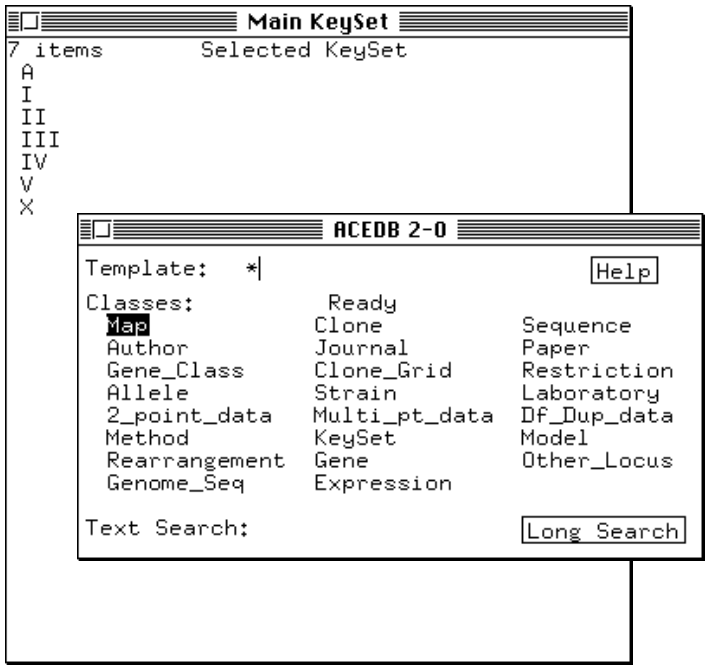
- Close the pop-up menu without selecting any option by moving the mouse pointer out of the menu before releasing the rightmost mouse button.

Starting up

The exact procedure to start up **ACeDB** depends on the way the program was set up on your system. If you are not sure, ask your local system administrator.

- Start up **ACeDB**.

ACeDB initially opens two windows: a main selection window labelled **ACEDB** (followed by a version number) and a **Main Keyset** window:



The main selection window lists the various *classes* of data available, that is, the main types of information stored in the database. The members of the currently selected class are listed in the **Main Keyset** window. By default, the **Map** class is automatically selected when **ACeDB** is first started: the word **Map** is highlighted in the main window. Since in the *C. elegans* database, the genetic maps are organized on a chromosome basis, the *C. elegans* chromosomes are listed in the **Main Keyset** window.

The main window also contains a 'message box' where the program displays various messages. At the moment the word 'Ready' appears in this box.

INTRODUCTION

ACeDB is a database of genome mapping information for the nematode worm *Caenorhabditis elegans*. It is accessed through an object-oriented graphical user interface which is also used by the *Arabidopsis thaliana* (**AAtDB**), the human chromosome 21, the grain genes and the Mycobacterium databases.

ACeDB is a large program and the following exercises only demonstrate its basic features. However, the graphical nature of the program makes it easy to explore and experiment with. In addition, extensive help is available on-line at every step.

A complete mapping project brings together genetic mapping (the process of locating the genes responsible for given phenotypes, for example using recombination data) and the physical mapping of restriction enzyme sites. **ACeDB** allows the retrieval of these data at various levels, from whole chromosomes down to individual genes. This includes information about alleles, genetic loci, clones, contigs (assemblies of overlapping clones), sequences and much more. Each particular item of information is classified according to its type (e.g. chromosome, allele, author, journal, clone, gene etc.). Related items of information are linked together, so it is possible, for example, starting from a given gene, to retrieve the information available about its physical mapping, its genetic map location, its various alleles, the DNA clone on which it was mapped, the journal article describing it, the authors of this article, the laboratory they worked in and so on.

Note: the basic **ACeDB** software is easily adapted to store all kinds of information pertaining to genome mapping. For example, **AAtDB** (the *Arabidopsis thaliana* database) contains digitized photographs of various *Arabidopsis* mutants and of autoradiograms, which can be viewed on the screen using a graphical display program called **xv**.

TUTORIAL
INTRODUCTION TO ACEDB

Bruno Gaëta, PhD

Australian National Genomic Information Service
JO3, University of Sydney, NSW 2006, Australia
Email: gaeta@angis.su.oz.au
FAX: +61 692 2948

December 1993, ACeDB version 2.0