

Contents

Chapter 1 . Introduction	3
Chapter 2 . Creating a Base Calls Sheet	5
Chapter 2 .1 Specifying Files to be included	5
Chapter 2 .2 Specifying Parameters for the Base Calls Sheet	6
Chapter 3 . Using a Base Calls Sheet	9
Chapter 3 .1 Modifying Entries	9
Chapter 3 .2 Customising the base calls sheet	10
Chapter 3 .2.1 Sizing	10
Chapter 3 .2.2 Column selection	11
Chapter 3 .2.3 SNP selection	11
Chapter 3 .2.4 Sorting	12
Chapter 3 .3 Displaying the data file	12
Chapter 3 .4 Base Calls Sheet menu options	13
Chapter 3 .4.1 Base calls sheet Create ..	13
Chapter 3 .4.2 Base calls sheet Add files ..	13
Chapter 3 .4.3 Base calls sheet Configure ..	13
Chapter 3 .4.4 Base calls sheet Load ..	13
Chapter 3 .4.5 Base calls sheet Save ..	13
Chapter 3 .4.6 Base calls sheet Show	14
Chapter 3 .4.7 Base calls sheet Print ..	14
Chapter 3 .4.8 Base calls sheet Close	14
Chapter 4 . Base Call Colour-Views	15
Chapter 4 .1 The Colour-View Window	15
Chapter 4 .2 Changing alignment using the mouse	17
Chapter 4 .3 Colour-View Window Menu Options	17
Chapter 4 .3.1 Edit Copy bitmap, Edit Create bitmap file	17
Chapter 4 .3.2 View Distinct Colours, View Mixed Colours	17
Chapter 4 .3.3 View Halftones	17
Chapter 4 .3.4 Colour-view Print	17
Chapter 4 .3.5 Colour-View Sort by sequence	18
Chapter 5 . Finding Base Sequences	19
Chapter 5 .1 Step 1: Loading the data files	19
Chapter 5 .2 Step 2: Looking for sequences	21

Chapter 1 . Introduction

DAX can read sequence files such as the files that are created by the ABI Genescan® system, as well as Staden *.SCF files. SNP (single nucleotide polymorphism) locations in these files can subsequently be analysed.

This manual presents a step-by-step guide to set up a so-called Base Calls Sheet, which visually depicts the SNP locations, with trace plot, area / height values, and base ratios. Also shown will be the base calls colour-view, which allows quick visual inspection of the alignment of many sequences.

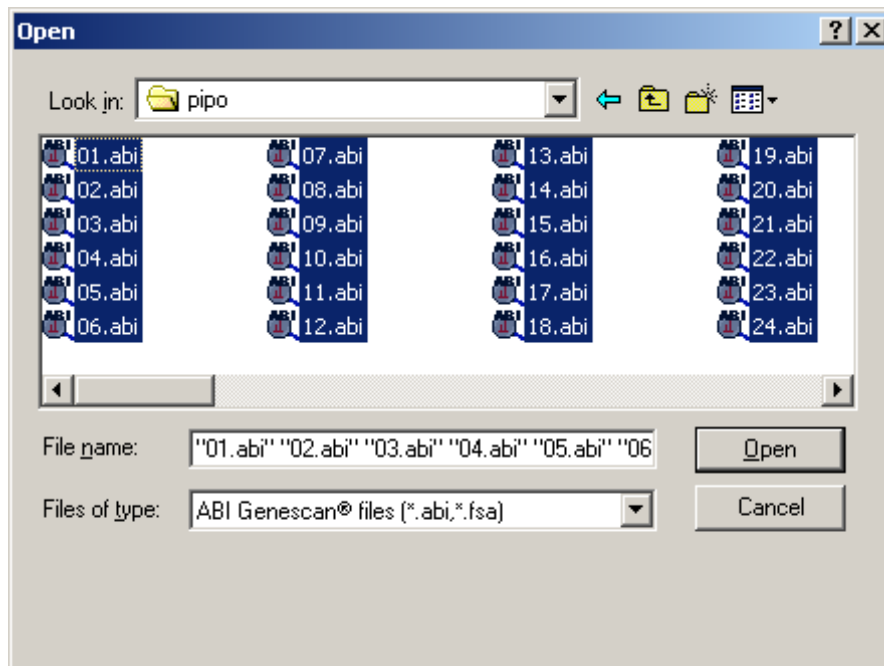
NB To be able to use the Base Call Sheet menu options, make sure that base call sheets are active, using the File | Customise > Extensions menu option.

Chapter 2 . Creating a Base Calls Sheet

This chapter describes the steps needed to create the Base Call Sheet.

Chapter 2 .1 Specifying Files to be included

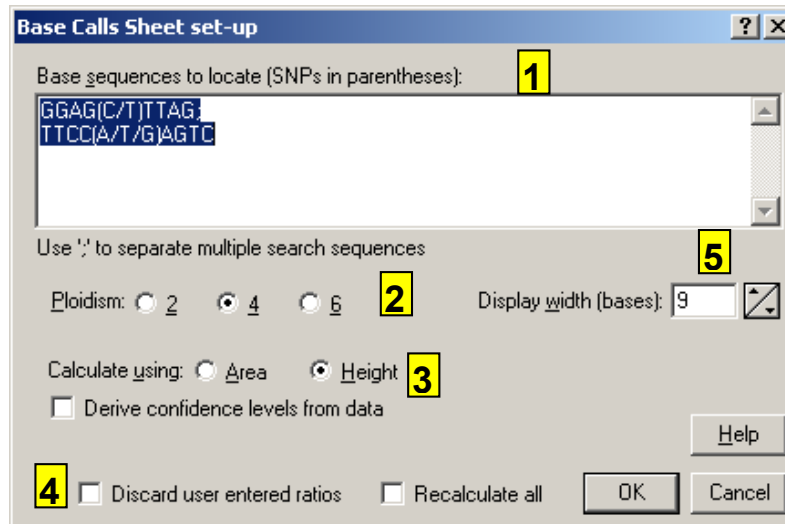
To create a base calls sheet, invoke the **Base calls sheet | Create base calls sheet** menu option. A **File Open** dialog box will be displayed. In this dialog box, select the files that are to be included in the sheet, then click the **Open** button. It is also possible to select (highlight) a folder name and click the **Open** button. In that case, all files in the selected folder(s) will be added to the sheet.



As soon as the **Open** button is clicked, a sheet will be created, containing the selected files. DAX will then display the Base Calls Sheet editing dialog box, in which parameters will need to be entered.

Chapter 2.2 Specifying Parameters for the Base Calls Sheet

The Base Calls Sheet edit dialog box looks like this.



1. In the **Base sequence to locate** box, enter a string defining the base sequence, including SNP locations. SNP locations should be specified by listing the bases for the location in between parentheses, separated by forward slashes. For example:

GGAG(C/T)TTAG; TTCC(A/T/G)AGTC

This defines two SNP locations, the first with **C** and **T** bases, the second with **A**, **T** and **G** bases.

The sequence may contain any characters, but only the following characters have meaning:

- A, C, G, T define bases to look for
- N will match any base
- (/) are used in defining an SNP location
- the semi-colon ; is used to separate multiple sections of bases. Each section is located independently of all other sections. Sections can contain one or more SNP locations
- spaces and new line characters are ignored, so you can use new line characters to nicely shape the sequence. A new line will be started automatically after approximately 60 bases

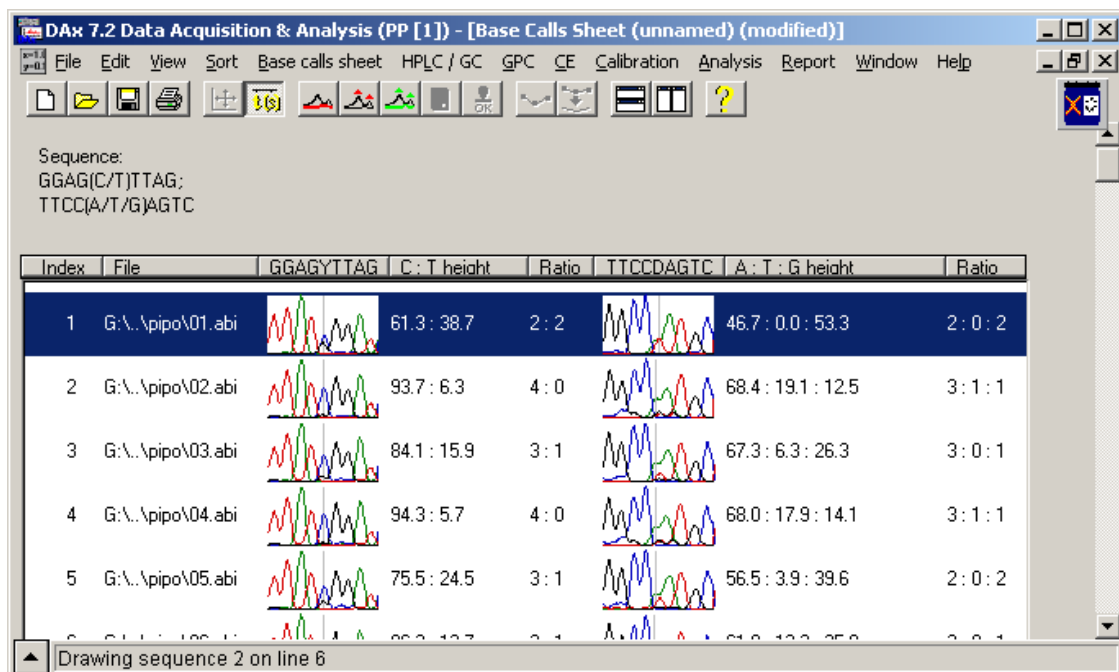
2. Specify the **ploidism**: diploid, tetraploid, or hexaploid.
3. SNP locations are inspected for the ratio of the signal strengths

of the bases. Specify if this is to be done using the **area** or the **height** of the base traces.

4. DAx will call the ratios for the SNP locations. Subsequently, these ratios can be changed by the user. When the base call sheet is edited (specifically, by changing between using area and height), user called (changed) ratios will normally be retained. By checking **Discard user entered ratios** those ratios will be overwritten.
5. A base call sheet contains small plots of bases around each SNP location. Specify the number of bases to display as the **Display Width**.

When all parameters have been specified, click the **OK** button to start analysing the files. This analysis can be a lengthy process. While it is running, a dialog box with a cancel button is displayed, so that the analysis can be cancelled.

When the analysis is done, a base calls sheet is displayed on screen, similar to the one shown below. Notice the small trace plots for each SNP location in each file.



NB It is not *required* to enter a base sequence in the configuration dialog box. In this case, the base calls sheet is created without SNP locations. A colour-view (cf. Chapter 4) might next be used to *derive* the sequence and SNP locations.

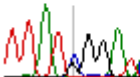

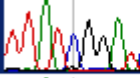



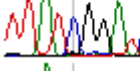

Chapter 3 . Using a Base Calls Sheet

Base Calls Sheets are displayed as tables with the following columns.

- The **Index** column displays a running index.
- The **File** column displays the name of the file.
- For each specified SNP location, the **trace plot** column displays a small graph of the SNP location and surrounding bases.
- For each specified SNP location, the **area / height** column displays the relative areas or heights of the bases that are included in the SNP.
- For each specified SNP location, the **ratio** column displays the ratio between the bases in the SNP. DAX will attempt an initial call for the ratio, but the user can always change it by selecting another ratio.

Chapter 3 .1 Modifying Entries

The values displayed in the ratio column can be modified by the end user. Click on the down arrow, and select a new ratio value.

File	GGAGYTTAG	C: T height	Ratio	T
G:\..\pipo\01.abi		61.3 : 38.7	2 : 2	
G:\..\pipo\02.abi		93.7 : 6.3	4 : 0	
G:\..\pipo\03.abi		84.1 : 15.9	3 : 1	
G:\..\pipo\04.abi		94.3 : 5.7	0 : 4	

To use the keyboard to change the ratio, first hit the **F2** key, then use the cursor keys to change the ratio. Hit the **F2** key again to be able to use the cursor keys to navigate between lines. It is also possible to select a new ratio by hitting the number key of the desired ratio – hitting the number **3** will select a **3 : 1** ratio.

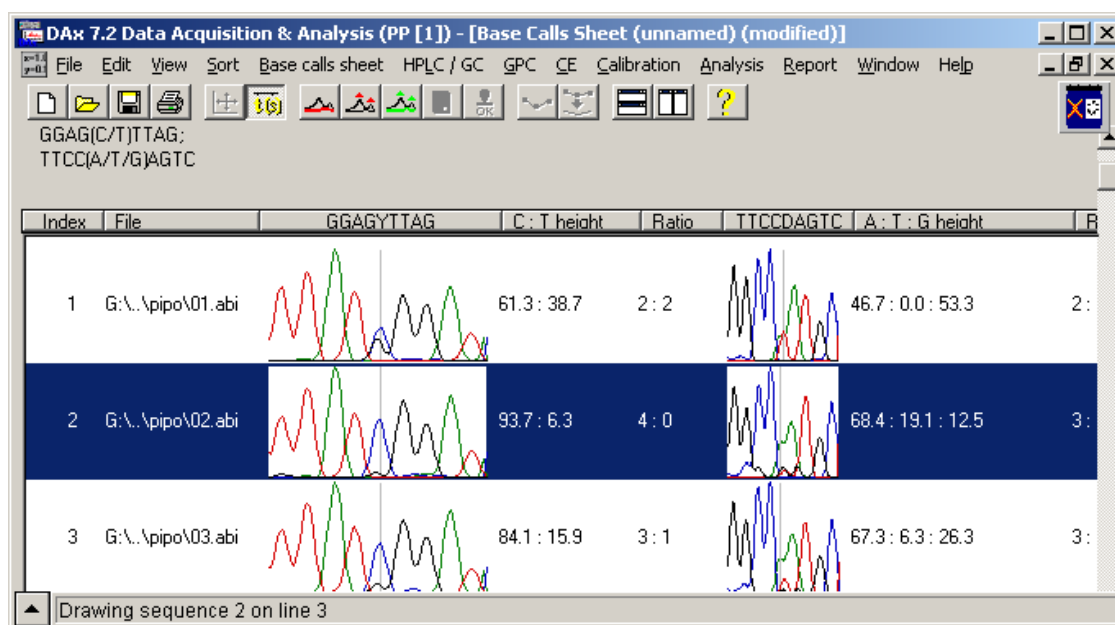
Chapter 3.2 Customising the base calls sheet

The way the base calls sheet is displayed can be customised in a number of ways.

Chapter 3.2.1 Sizing

The size of the columns can be modified by clicking and dragging on column divider lines in the header at the top of the sheet. Column widths can also be changed using the **View | Fit columns** and **View | Shrink & fit columns** menu options. The former option expands column widths as needed to accommodate all data, the latter will also shrink columns if possible.

The size of the rows can be modified by clicking and dragging the mouse between rows. The mouse cursor will change shape to a North-South arrow to indicate the ability to change the line height.



Example of Base Calls Sheet with significantly increased line height. The width of the first plot column has also been increased.

Chapter 3 .2.2 Column selection

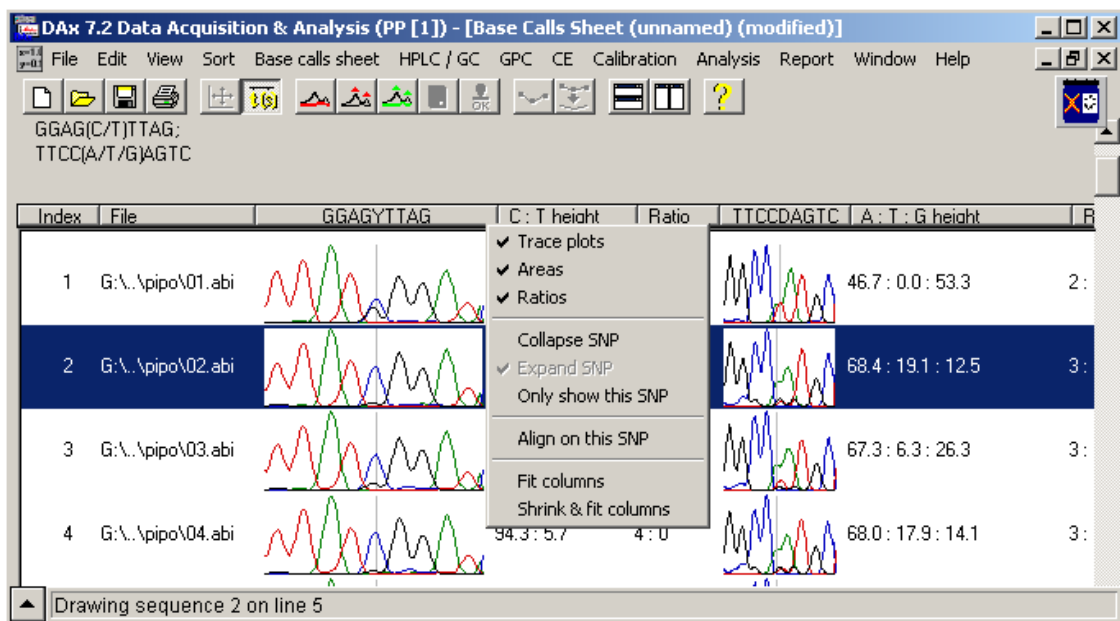
Use the **View | Trace plots** menu option to show or hide the **trace plot** column in the base call sheet. Use **View | Only active traces** to determine if all bases should be included in the plots, or only the bases that are present in the SNP.

Use the **View | Areas / heights** menu option to show or hide the **area / height** column. Use the **View | All areas / heights** option to determine if the areas / heights for all bases should be displayed (in ACGT order), or just the area / height of the bases that are present in the SNP.

Use the **View | Ratios** menu option to show or hide the **ratio** column.

Chapter 3 .2.3 SNP selection

By right clicking the mouse in the header at the top of the base calls sheet, a popup menu can be displayed.



In this popup menu, use the **Collapse SNP** option to collapse the trace plot, areas / heights and ratio columns for the pertinent SNP into a place holder column (marked >>). Use the **Expand SNP** menu option, or click on the >> place holder, to show the columns again.

Use the **Only show this SNP** option to collapse all SNP columns except the ones for the pertinent SNP.

Chapter 3 .2.4 Sorting

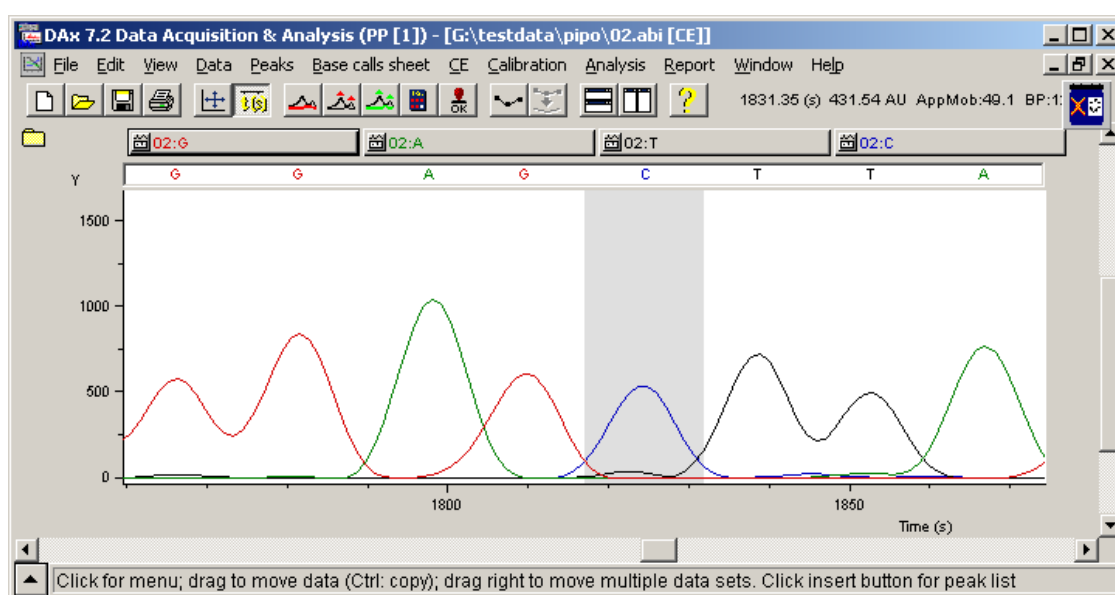
The base calls sheet can be sorted by clicking on column headers. Clicking a second time reverses the sort order. Clicking in the **File** or **trace plot** column causes the lines to be sorted by file name. Clicking in the **areas / heights** or **ratio** column causes the lines to be sorted by the area / height or ratio values of the pertinent SNP location, respectively.

Clicking on the **Index** column has special functionality: it causes the lines to be sorted by the **sum** of the areas or heights of **all** SNP locations.

When sorting by area / height or ratio, or when sorting by clicking in the trace plot column, lines that do not have data about the pertinent SNP location will be put at the bottom of the sheet.

Chapter 3 .3 Displaying the data file

By clicking the right mouse button in the base calls sheet, then selecting the **Load file & show trace** menu option, the data file is loaded and zoomed to the location of a specified SNP.



Example of data file zoomed to the location of a specified SNP.

Chapter 3 .4 Base Calls Sheet menu options

Base Calls Sheets have the normal text window menu options, such as **File | Export** to write text files to disk, and **Edit | Select All** and **Edit | Copy using tabs** to put a copy of the table on the clipboard. Refer to the DAX Reference Manual for details on these options.

Chapter 3 .4.1 Base calls sheet | Create ..

Use this menu option to create a new base calls sheet. The steps as detailed in Creating a Base Calls Sheet will be followed.

Chapter 3 .4.2 Base calls sheet | Add files ..

Use this menu option to add more files to an existing base calls sheet.

Chapter 3 .4.3 Base calls sheet | Configure ..

This option is used to modify the base calls sheet by invoking the Base Calls Sheet Edit dialog box.

Chapter 3 .4.4 Base calls sheet | Load ..

This menu option is used to load a previously saved base call sheet from disk.

Chapter 3 .4.5 Base calls sheet | Save ..

This menu option is used to save the current base calls sheet to disk.

If the base calls sheet has been changed but has not been saved to disk yet when the software is about to terminate you will be asked to confirm the loss of the sheet. Saving the sheet to a floppy disk does not prevent this - you must save to hard disk.

Chapter 3 .4.6 Base calls sheet | Show

This option shows the base calls sheet.

Chapter 3 .4.7 Base calls sheet | Print ..

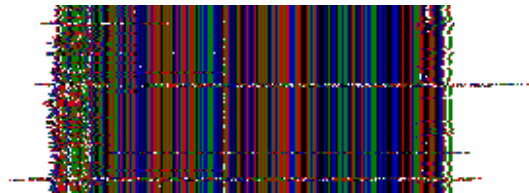
This option is used to print the contents of the base calls sheet. The print dialog will contain the option **Page tiling**. This option is useful when base calls sheets with high numbers of SNP locations are printed. Without tiling, not all information would fit on the printer page. Tiling uses side-by-side pages to ensure that all information is printed.

Chapter 3 .4.8 Base calls sheet | Close

This option is used to remove the base calls sheet from memory.

Chapter 4 . Base Call Colour-Views

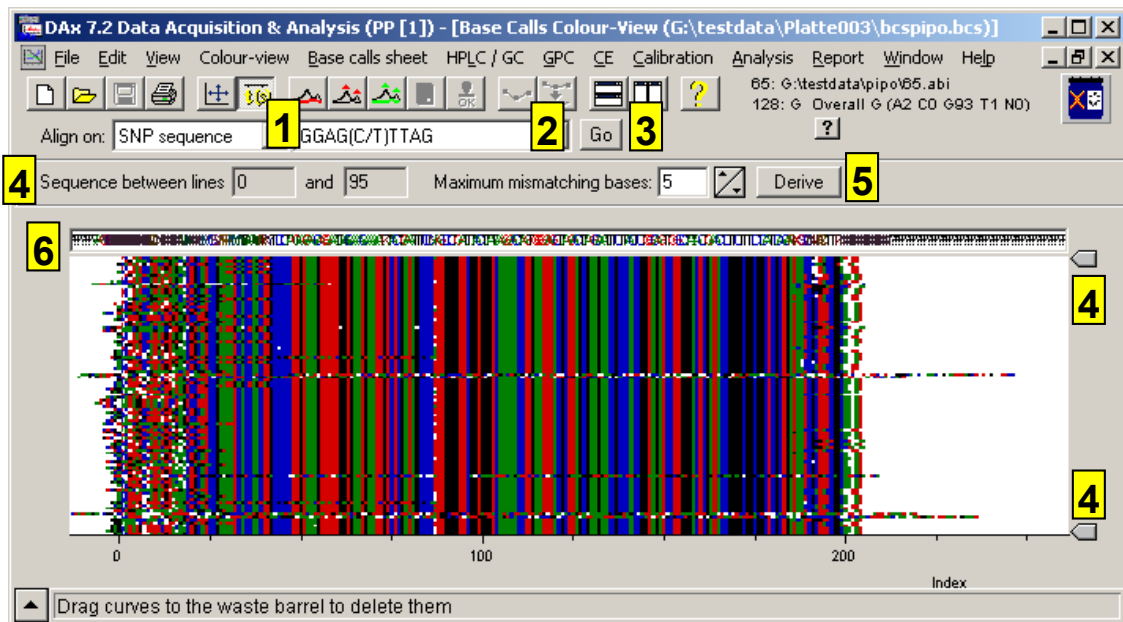
DAX has the ability to create *colour-views* of base call sheets. By assigning a colour for each of the bases A, C, G and T, an image can be created that readily shows the alignment of the sequences in the base calls sheet. DAX attempts to optimally align the sequences, but the user can change the alignment in several ways.



Example of colour-view of base calls sheet. The lines are sorted by file name. The colours have been changed to hide the underlying data.

Chapter 4 .1 The Colour-View Window

Colour-views are displayed in the colour-view window.



Above the colour-view, the *alignment bar* can be seen. This toolbar contains options that are used to change the alignment of the sequences.

1. Sequences can be aligned in several ways.

- **current line order** does not change the line order in the base calls sheet, but tries to obtain optimum alignment by moving each line left or right (by a maximum of 128 bases).
 - **optimum line order** determines which line in the base calls sheet has the highest total alignment with all other lines. This line is made the first line; and so on. This process can be quite lengthy, especially if the base calls sheet contains more than 100 lines. After the lines have been re-ordered, optimum alignment is again reached by shifting sequences left or right.
 - **SNP sequence** aligns on one of the SNP search sequences from the base calls sheet (the SNP sequence is selected using the combo box **2**).
 - **specified sequence** aligns on a base sequence specified by the user.
 - **using the mouse**, see chapter 4 .2
3. The **Go** button starts the alignment process. Remember that the lines in the base calls sheet and the lines in the colour-view have the same order; when the colour-view lines get realigned, so will the lines in the base calls sheet.
 4. When the alignment is finished, or when the **Derive** button is clicked, DAX will attempt to construct an overall sequence of bases. The sliders **4** to the right of the colour-view are used to limit this process to certain lines of the colour-view.
 5. The value entered as **Maximum mismatching bases** determines how many bases of a certain type must be present for the overall sequence to take that base type into account. For instance, if 20 lines contain A, and 5 lines contain C, the overall call will be M if maximum mismatching bases is below 5, and A otherwise.
 6. This box contains the overall sequence that DAX determined. This sequence could be used as input for the base calls configuration dialog box (cf. Chapter 2 .2). In this way, the colour-view could be used to *derive* SNP locations from aligned sequence data.

Chapter 4 .2 Changing alignment using the mouse

By pressing the **Shift** key, then clicking and dragging the mouse on the colour view, you can change both the alignment and order of the sequences.

Dragging up or down changes the order of the base call sheet's lines.

Dragging left or right changes the alignment of the sequence on which the mouse was clicked.

Chapter 4 .3 Colour-View Window Menu Options

Chapter 4 .3.1 Edit | Copy bitmap, Edit | Create bitmap file

These menu options place a copy of the colour-view on the clipboard as a device independent bitmap, or create a device independent bitmap file. The bitmap can subsequently easily be inserted into documents and presentations (that is how the bitmap in this document was inserted).

Chapter 4 .3.2 View | Distinct Colours, View | Mixed Colours

When SCF files are used, the base calls have confidence levels attached to them. These confidence levels are used to mix the pure base colours if the **Mixed Colours** option is active.

Chapter 4 .3.3 View | Halftones

Halftone output gives better looking colour-views, especially if many sequence lines are present. However, drawing becomes slower.

Chapter 4 .3.4 Colour-view | Print

Prints a copy of the colour-view.

Chapter 4 .3.5 Colour-View | Sort by sequence

This menu option sorts the lines in the base calls sheet, and by extension the lines in the colour-view, alphabetically on sequence.

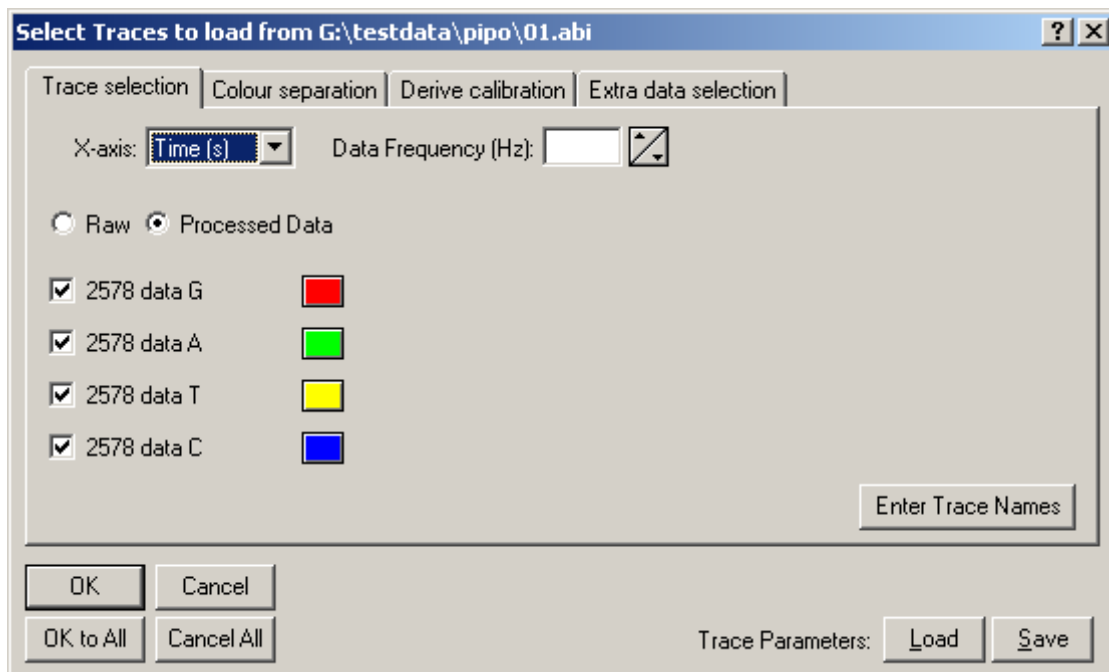
Chapter 5 . Finding Base Sequences

The previous chapters dealt with setting up a Base Calls Sheet, where many files are studied, using specified SNP locations. DAX also has the ability to load one or more data files, and look for the presence of certain base sequences in them.

Chapter 5 .1 Step 1: Loading the data files

To open data files, invoke the **File | Open** menu option. A dialog box appears. At the bottom of the dialog box, in the **Files of Type** box, select **ABI Genescan ® files** or **SCF Files**. Select the files (one or more) you want to load. If you select more than one file, it makes sense to check the **Multiple Windows** box. Do not check the **Add to Window** or **AutoAnalyse** boxes. Now click the **Open** button.

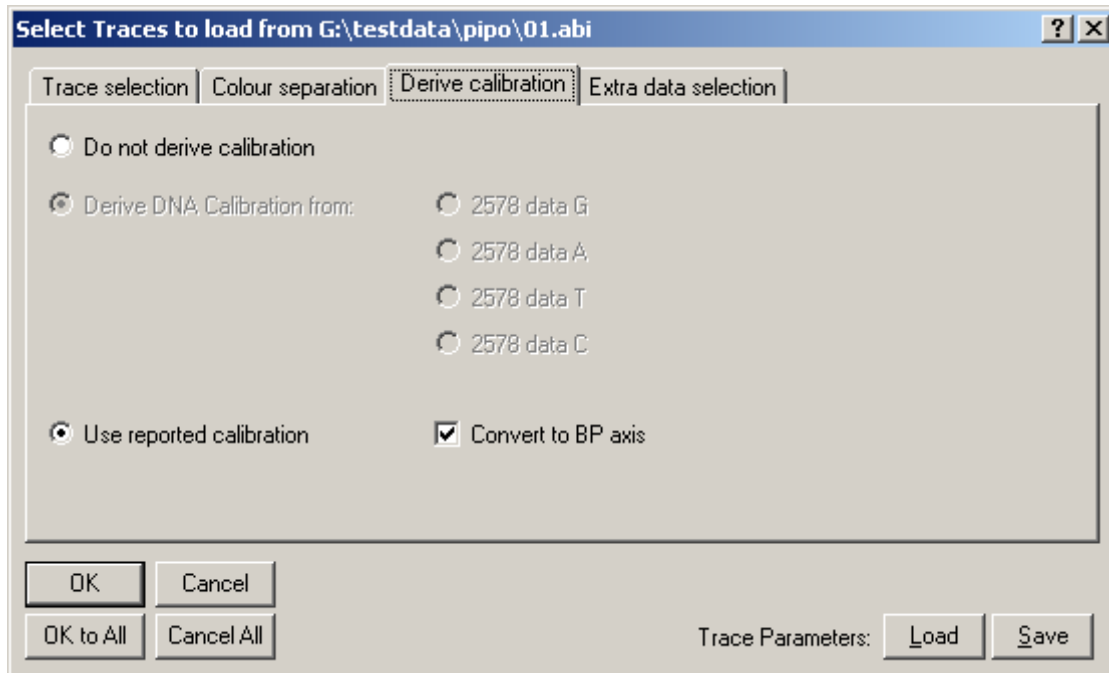
A **Trace Analysis Parameters** dialog box will appear.



Use the **Trace selection** tab to indicate which colours, present in the file, should be loaded. For purposes of Base Sequence analysis, you should load all four colours, loading **Processed data**.

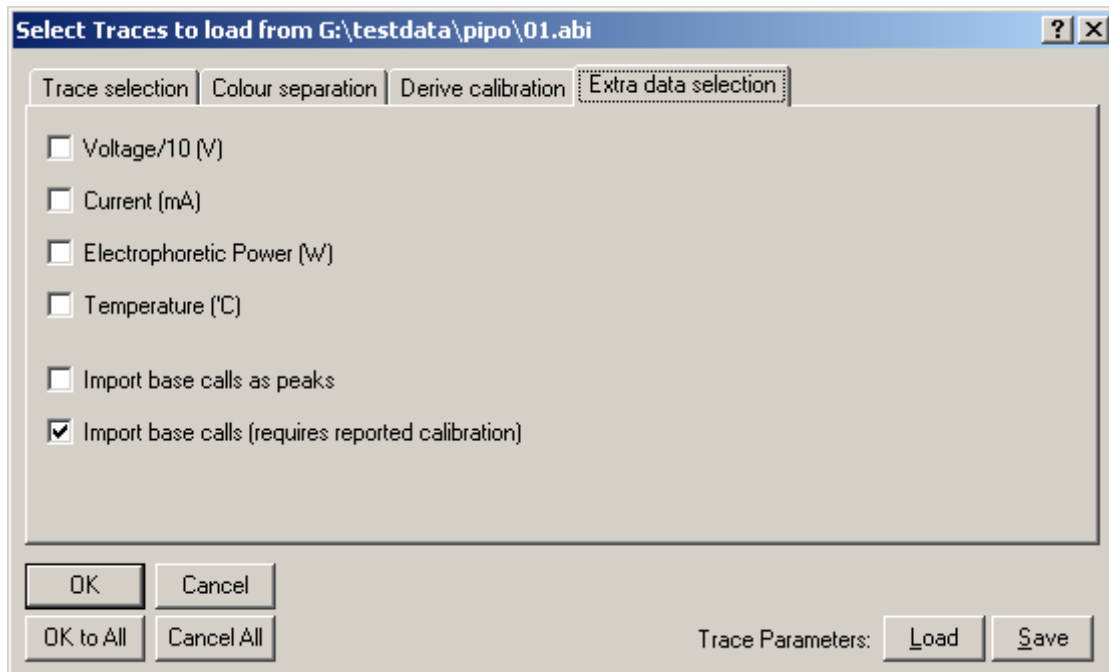
When **Processed data** are loaded, the **Colour separation** tab is not used.

The **Derive calibration** tab is important. Base call files contain a calibration, relating data points to base pair index. It is important to load this calibration, by checking **Use reported calibration**.



Normally, the data will be displayed with a time or index horizontal axis. Checking **Convert to BP axis** you can cause the data to be displayed with a base pair horizontal axis.

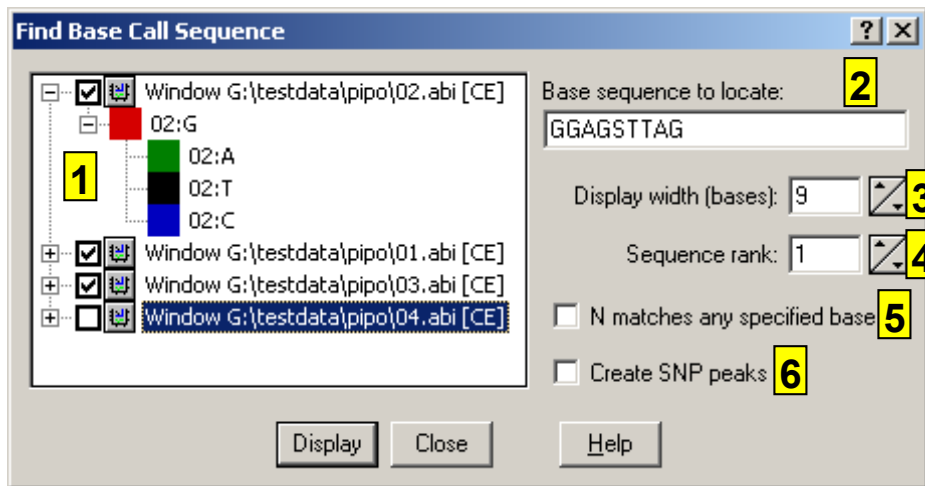
The **Extra data selection** tab is also important. On it, you must check **Import base calls (requires reported calibration)**. This causes the base sequence in the file to be loaded into DAX. It is in this sequence that you can search for specified base sequences.



Finally, click the **OK** button to load the data file. If you have selected multiple data files to be loaded, you can click **OK to All** to load all data files without having to fill in the trace analysis parameters dialog box.

Chapter 5 .2 Step 2: Looking for sequences

Use the **Base calls sheet | Locate base sequence** menu option to look for specified base sequences in the files you have loaded. This invokes the following dialog box.



1. Select the graphics windows in which you want to look for the base sequence here. Upon entry, all windows will be selected.
2. Enter the sequence of bases you are trying to locate in the selected graphic windows here. Use the standard A, C, G and T characters for bases. Specifying ? matches any base. Specifying S denotes an SNP location. For those locations, DAX can insert peak markers into the data (if **Create SNP Peaks** is checked).
3. When a specified sequence of bases is found in a graph window, the corresponding part of the window will be displayed. You can specify the window width, in bases, here. If the base call sequence that is being located is longer than the specified width, its length will be used.
4. If a relatively short sequence of bases is specified, it may occur more than once. Specify a rank to display the second, third, etc occurrence.
5. Checking **N matches any specified base** causes any base called as N to match any base in the specified sequence. For instance, if you specify to look for AAA, a called sequence of NAN will match.
6. Check the **Create SNP peaks** box to create peaks for base locations denoted by an S in the specified base sequence. These peaks can then be used to perform calculations, such as peak height and area calculations.

Click the **Display** button to locate the specified sequence in all selected graphics windows. The windows will be zoomed to the sequence, if it is found.