DAx allele match sheet tutorial

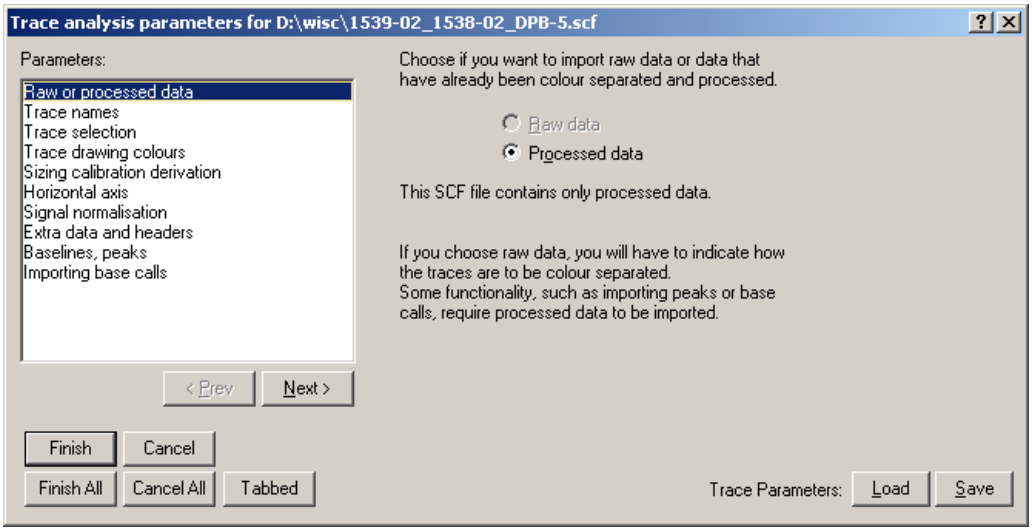
By Shelby O'Connor, Ph.D.

The allele match sheet function in DAx is a feature designed to assist researchers in genotyping their individual subjects for their specific gene of interest. This feature was originally designed to facilitate genotyping of MHC alleles in non-human primates. There are two main functions to this application. First, there is a sequence chromatogram analysis tool. By using this tool, one can open traces from sequencing reactions, identify SNPs, and manually edit sequences. Second, there is the allele match sheet tool. By using the allele match sheet, one can compare sequence traces of interest from a specific individual against a panel of reference alleles. Because one individual will likely have more than one allele at a given locus, the sequence traces will contain mixed bases at specific positions. DAx has the ability to deconvolute these mixed bases and sort out which alleles an individual expresses. This unique function of DAx makes the identification of allelic expression much simpler than using BLAST for genotyping. Additionally, DAx provides even more flexibility because it allows you to compare your FASTA files, sequence chromatograms, or sequence assembly files against your specifically chosen set of reference alleles. Ultimately, these analysis tools should provide researchers with the flexibility and the power they need to genotype individuals for their study.

Sequence chromatogram analysis:

Opening files:

Chromat files can be opened in DAx using the **File | Open** menu option. In the File | Open dialog, under **List files of type**, ABI / AB1 files or SCF files can be chosen. The following settings need to be made in the "Trace analysis parameters" window that appears when files have been selected and the **Open** button clicked.



Trace Analysis Parameters dialog box.

1. Raw or processed data -- If possible, choose processed data (SCF files contain only processed data, but ABI files can contain both raw and processed data)

2. Trace names -- These should be each of A, C, G, T. Note that SCF files always contain the bases in that order, but the order may be different in ABI files. The file will report the correct order, however, and that order should be used.

3. Trace selection -- Select all of A, C, G, and T data

4. Trace drawing colours -- Click the **Use standard colours** button to use green for A, blue for C, black for G and red for T. You may also select your own preferred colours by clicking on the colour buttons next to each of A, C, G and T.

5. Sizing calibration derivation -- Choose **Use reported calibration from file**

6. Horizontal axis -- Check **Convert to BP axis**

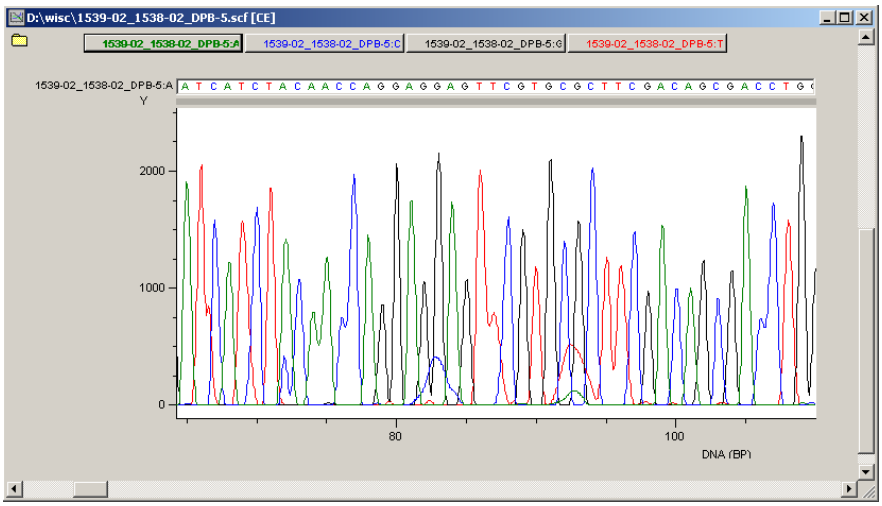
7. Signal normalisation -- Do not choose to normalise data

8. Extra data and headers -- Do not choose to copy file headers to logfile

9. Baselines, peaks -- Do not choose to construct a baseline or to find peaks

10. Importing base calls -- Check **Import base calls**

Once you click **Finish**, the chromat file is opened in a DAx window. You can click **Finish All** if you have multiple files. There are four traces shown, where each trace represents a different nucleotide. By clicking and dragging the left mouse button across the window, you can zoom in. There are up to 10 levels of zoom. By clicking the right mouse button, you can zoom back out.
The bases are shown in a string (or *ribbon*) above the traces. To reverse complement the sequence trace, right-click in the string and select **Reverse complement base calls** in the popup menu that appears.



Graph window with traces and base call ribbon.

Calling mixed bases:

If the menu option **Base calls** is not present in your copy of DAx, you need to use **File | Customise > Extensions** and check **Base call & allele match sheets**.

To call mixed bases, go to **Base calls | Derive SNP letters**. Make sure that all of the traces of interest are checked in the data selection box at left. You can choose to calculate SNP letters using the area or height. This means that you are going to compare either the heights or areas of peaks at a specific position to determine whether that position contains an SNP. By adjusting the **Relative area/height above which to include base in SNP code**, you can adjust the sensitivity of calling mixed bases. Essentially, DAx first looks at an individual position in a string of bases. If there are two peaks at a given position, the sum of the area or height of those two peaks should equal 100%. Suppose the area/height of trace A is 30% of the sum, and the area/height of trace T is 70% of the sum. If you set the limit at 50%, then, this position will be recognized as a T. But, if you set the limit at 20%, then the position is recognized as an A/T mixed base. Thus, by increasing the relative area/height percent limit, you decrease the sensitivity of calling mixed bases. When mixed bases are called, the

standard IUPAC symbols are used.

Bases can also be changed manually. Click in the base call string (or *ribbon*) that is displayed above the graph window. Use the cursor (arrow) keys to navigate left and right. To change a base call, simply type the desired new letter in the correct location.

As always, use the **File | Save as** menu option to save the data after you have modified them.

Allele Match sheet

An allele match sheet is used to identify which alleles are in a population of sequences. In the allele match sheet, an experimental sequence can be paired against a panel of known allele sequences. DAx can determine the match percentage between experimental and reference alleles. DAx can also be used to identify where the mismatches between the experimental and reference alleles lie. Importantly, if the experimental sequence contains mixed bases, DAx can deconvolute these mixed bases into their individual base possibilities (using standard IUPAC nomenclature) and then compare to the reference alleles.

Using the MHC as an example, one could amplify all of the MHC class II DRB alleles in an individual. The sequencing reactions from this experiment will yield a string of nucleotides that match a particular set of DRB alleles. This string of nucleotides can be imported into the allele match sheet. DAx supports ABI or SCF files, DAx files, FASTA sequences, and ACE assemblies. Reference alleles can be added as FASTA or text files. Then, an individual can compare the experimental sequence against a self-selected group of reference alleles to identify which alleles are expressed in their individual.

Adding alleles to the allele match sheet

If the menu option **Base calls** is not present in your copy of DAx, you need to use **File | Customise > Extensions** and check **Base call & allele match sheets**.

Reference alleles first need to be saved as FASTA or text files. They can be added to the allele match sheet from anywhere in the program using **Base calls | Allele match sheet | Add alleles**. If the allele match sheet is the current window alleles can be added using the **Allele match sheet | Add alleles** menu option

Adding experimental sequences to the allele match sheet

If the menu option **Base calls** is not present in your copy of DAx, you need to use **File | Customise > Extensions** and check **Base call & allele match sheets**.

Experimental files first need to be saved either as ABI or SCF files, DAx files, FASTA files, or ACE assembly files. They can be opened from the main window with **Base calls | Allele match sheet | Add files**. There are some differences when opening the different file types:

- ABI, SCF, or DAx files -- These chromat files are identical to the files that can be opened directly into a DAx window using **File | Open**. To open them in a graph window from within the allele match sheet, double click on the file name or right click, and choose the **Load file & show trace** menu option. SNP letters can now be derived using the **Base calls | Derive SNP letters** menu option. Bases can be

manually adjusted by clicking in the base calls ribbon and entering base calls. Importantly, a bar is displayed above the traces with triangular arrow markers denoting the start and end of the sequence to be compared to the reference alleles. These arrows can be clicked and dragged to change the start and end of the sequence being considered. The relevant parts of the allele match sheet will be recalculated after a slight delay. When opening an ABI or SCF file, the start and end of the sequence are initially set to exclude **N** calls at the start and end of the sequence in the file. When opening a DAx file, the start and end locations are retrieved as the last entered values.

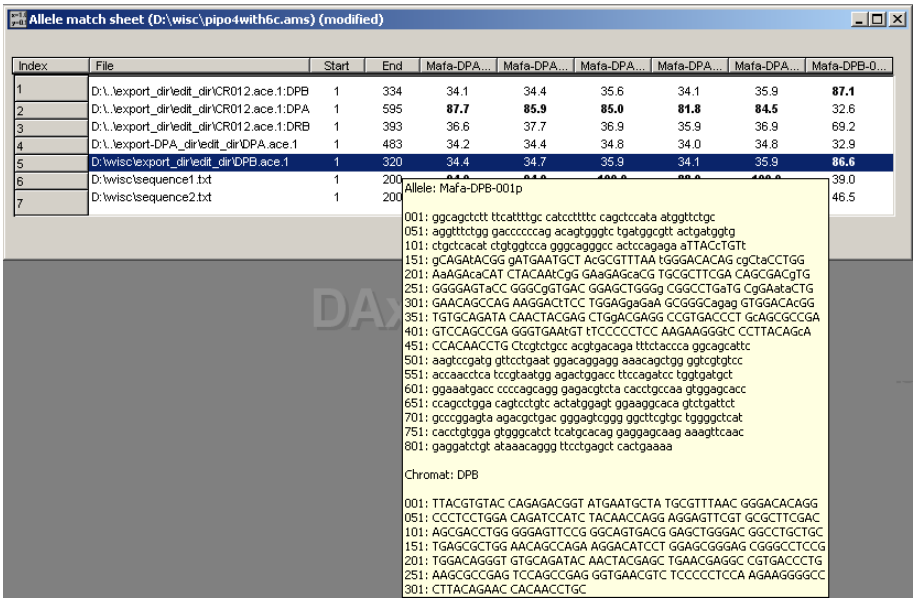
- ACE assembly files -- These are sequence assembly files that have been exported from a sequence alignment program. These files contain references to multiple chromat files. When saving an assembly in the sequence alignment program, each contig of multiple chromat files is essentially saved as an individual alignment. When the ACE assembly is opened in the allele match sheet, each contig becomes its own experimental allele to compare against the panel of reference alleles. When an ACE assembly is opened from the allele match sheet (by double clicking or using the right click popup menu), the base calls are loaded as read-only, and cannot be changed.

- FASTA files -- These files are simply text files. DAx can compare the nucleotides in the FASTA file to the reference alleles directly. FASTA files can be generated from sequence alignments in other specialised sequence alignment programs. However, the FASTA files cannot be viewed or edited in DAx.

Viewing the nucleotide matches and mismatches in an allele match sheet

Nucleotide matches and mismatches can be viewed by either placing the mouse over a match percentage in the allele match sheet, or by double clicking on the match percentage in the allele match sheet.

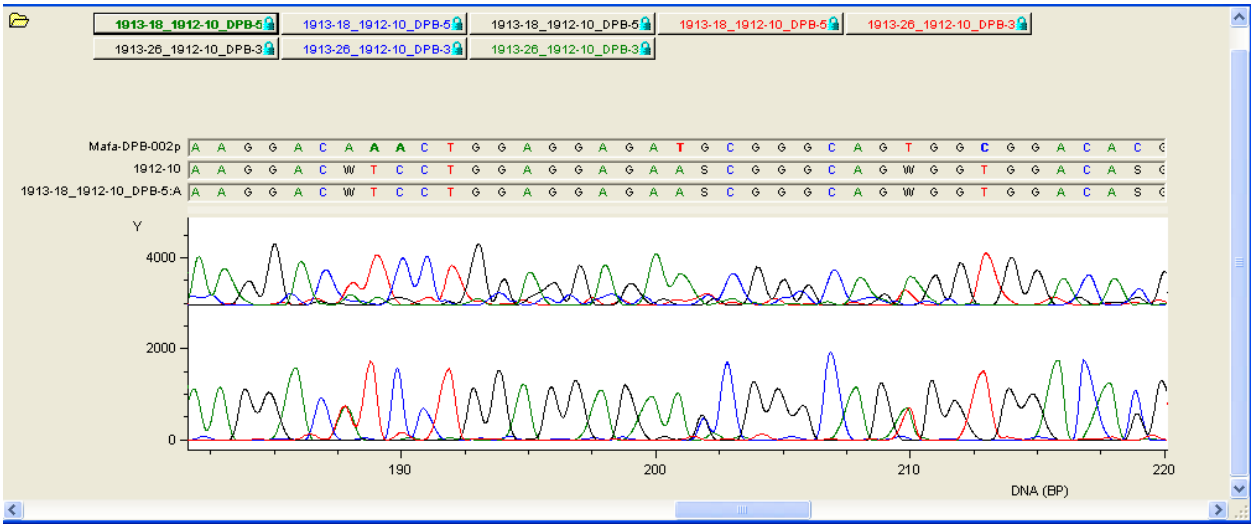
When you place the mouse over a match percentage, a yellow pop-up window appears. In this window, matching bases are displayed in capital letters and the mismatched bases in lower case letters. The contents of this window are called an **allele match report**; a copy of this report can be placed on the clipboard by clicking the right mouse button and choosing the **Copy allele match report** menu option in the popup menu that appears. To stop displaying the pop-up window, use the **View | Pop-up sequences** menu option.



Example of an allele match sheet with a popup window showing specific of a match between a chromat and an allele.

If the experimental file contains chromat files (either ABI, SCF, or ACE), double clicking on the match percentage will open a window that will contain the sequence traces. Up to three nucleotide strings will be displayed at the top of the window. The top string is the string of nucleotides for the reference allele. The middle string is the consensus sequence from the multiple traces (this is displayed only for ACE files). The bottom string contains the base calls contained in the data set. Mismatches between the reference allele and the consensus sequence are shown in bold in the reference sequence. Mismatches between the consensus and the experimental data set are shown in bold in the consensus sequence. These apparent bold mismatches take into account the IUPAC mixed base nomenclature.

To reverse complement the sequence trace, right-click on the match percentage and select **Reverse complement base calls**. This action will reverse-complement the sequence trace and the corresponding adjustment in the match percentage will be displayed in the Allele Match Sheet.



Example of a graph window with three nucleotide ribbons. The top one shows the allele, the middle one the consensus, the bottom one the currently active set of traces. Note the mixed base letters, and the bold letters where the allele does not match the consensus string.

Display of the nucleotide strings in graph windows can be turned on/off using the **View | Data set base calls**, **View | Consensus base calls**, and **View | Allele base calls** menu options.

Other viewing options in the allele match sheet

- Changing the axes for the experimental and reference alleles -- Reference alleles can be viewed in columns using the **View | Alleles in columns** menu option. Reference alleles can be viewed in rows using the **View | Alleles in rows** menu option.

- The allele match sheet can be sorted in both horizontal and vertical directions. To sort by match percentage, click in the header item above a column or to the left of a row. Clicking a second time reverses the sort order. If alleles are in columns, the sheet can be sorted by file name by clicking in the header item of the file name column, and the sheet can be sorted in the order in which files were added by clicking in the header item of the index column. Further sorting options are available by clicking the right mouse button in the sheet or on the headers above and to the left of the sheet.

- Alleles (but not files) can be arranged manually. Choose the **Manually arrange alleles** menu item in the header (when alleles are in columns) or sheet (when alleles are in rows) popup to start manually arranging alleles. Re-invoke same popup menu and use the **Move to top/left**, **Move up/left**, **Move down/right**, and **Move to bottom/right** menu options to change the order of the alleles.

- Highlighting the percent matches -- To easily identify the highest match percentages,

these numbers can be made bold. Using the **View | Bold matches menu options**, a limit can be put on which matches will be shown in bold. These cut-offs are above 95%, 90%, 80%, 75%, 50%, or never. *(In the example above, matches over 80% are displayed as bold.)*

- Viewing an allele map -- Using the **View | View as map** menu option, one can use colours to identify the match percentages. These colours can be configured using the **View | Attributes** menu option.

- Wrap lines -- To view the allele match sheet in a smaller window, line wrapping can be turned on using the **View | wrap lines** menu option.

Saving an allele match sheet

To save the allele match sheet, use the **Allele match sheet | Save** menu option.

Alternatively, the data can be exported using the **File | Export** menu option. In the File | Export dialog, if you choose "Tabbed files (*.txt)" under **List files of type**, the exported file can be opened in Microsoft Excel.