Flapjack Documentation

Release 1.16.10.31

Information and Computational Sciences

May 24, 2018

Getting started

1 Quickstart **3**
- 1.1 Importing data . . . 3
- 1.2 Browsing the data . . . 3
- 1.3 Interacting with Flapjack . . . 4

2 Projects & Data Formats **5**
- 2.1 Data sets, maps, and genotypes . . . 5
- 2.2 Flapjack views . . . 6
- 2.3 Phenotypes/Traits . . . 6
- 2.4 QTLs . . . 6
- 2.5 Graphs . . . 7

3 Pedigree Information **9**

4 Import Data **11**
- 4.1 Importing data . . . 11
- 4.2 Advanced data import options . . . 13

5 Duplicate Markers **15**

6 Export Data **17**

7 Quick Export Data **21**

8 Export Image **23**

9 Genotype Visualization **27**
- 9.1 Genotype display . . . 27
- 9.2 Chromosome overviews . . . 28
- 9.3 Line information . . . 28
- 9.4 Chromosome display . . . 29
- 9.5 Line and marker overviews . . . 29

10 Modes and Views **31**
- 10.1 Navigation mode . . . 31
- 10.2 Marker mode . . . 31
- 10.3 Line mode . . . 32

11 Selecting Lines and Markers **33**
11.1 Marking selections . . . 33
11.2 Selecting monomorphic markers . . . 33

12 Show/Hide Lines **35**
12.1 Hiding lines . . . 36
12.2 Restoring lines . . . 36

13 Show/Hide Markers **37**
13.1 Hiding markers . . . 38
13.2 Restoring markers . . . 38

14 Filtering Markers **39**
14.1 Filter missing markers . . . 40
14.2 Filter missing markers by line . . . 40
14.3 Filter heterozygous markers by line . . . 41
14.4 Filter monomorphic markers . . . 41

15 Chromosome Map Scaling **43**
15.1 Global map scaling . . . 43
15.2 Local map scaling . . . 43
15.3 Classic map scaling . . . 43

16 Colour Schemes **45**
16.1 Nucleotide model . . . 45
16.2 Simple 2 colour model . . . 45
16.3 By similarity to line (2 colour) . . . 45
16.4 By similarity to marker (2 colour) . . . 46
16.5 By allele frequency . . . 46
16.6 Random colour schemes . . . 46

17 Customize Colours **47**

18 Allele Frequency Threshold **49**

19 Bookmarks **51**
19.1 Creating a bookmark . . . 51
19.2 Deleting bookmarks . . . 52

20 Toggle Visible Displays **53**

21 Create New View **55**

22 Database Link Settings **57**
22.1 Genotype File Headers . . . 58

23 Chromosome Visualization **59**

24 Phenotype Visualization **61**
24.1 Importing phenotypic data . . . 61
24.2 Phenotype summary information . . . 61
24.3 Visualizing alongside genotype data . . . 62
24.4 Sorting by phenotype . . . 62

25 Select Traits **65**

26 QTL Visualization **67**
26.1 Importing QTL data . . . 67
26.2 QTL summary information . . . 68
26.3 Displaying QTLs . . . 68
26.4 Interacting with QTLs . . . 69
26.5 Filtering visible QTLs . . . 69

27 Filter QTLs **71**

28 Graph Visualization **73**
28.1 Importing graph data . . . 73
28.2 Interacting with a graph . . . 74

29 Find by Name **75**
29.1 Search options . . . 76
29.2 Example regular expressions . . . 77

30 Sort Lines **79**
30.1 By similarity . . . 79
30.2 By trait . . . 79
30.3 By importing an order . . . 79

31 Similarity Matrix Creation **81**

32 Dendrogram Creation **83**

33 Marker Assisted Back Crossing **85**
33.1 MABC QTL format . . . 86
33.2 Recurrent parent proportion (RPP) calculations . . . 87
33.3 Weighted calculation . . . 87
33.4 Unweighted calculation . . . 88
33.5 Genome Coverage . . . 88
33.6 Linkage Drag(LD) Calculations . . . 88

34 Marker Assisted Back Crossing Tutorial **91**
34.1 Tutorial data . . . 91
34.2 Importing data . . . 91
34.3 Exploring and filtering data . . . 96
34.4 Running the analysis . . . 97
34.5 Viewing the analysis results . . . 97
34.6 Filtering the results . . . 99
34.7 Sorting the results . . . 101
34.8 Selecting within results . . . 103
34.9 Ranking results . . . 105
34.10 Exporting results . . . 105
34.11 Viewing results in the genotype view . . . 106
34.12 Conclusion . . . 107

35 Pedigree Verification (F1s Known Parents) **109**
35.1 Options . . . 110
35.2 Understanding the statistics . . . 110

36 Pedigree Verification (F1s Known Parents) Tutorial **113**
36.1 Importing data . . . 113
36.2 Exploring and filtering data . . . 117
36.3 Running the analysis . . . 118

36.4 Viewing the analysis results . . . 118
36.5 Filtering the results . . . 120
36.6 Sorting the results . . . 122
36.7 Selecting within results . . . 124
36.8 Exporting results . . . 126
36.9 Viewing results in the genotype view . . . 126
36.10 Conclusion . . . 128

37 Analysis Results Tables **129**
37.1 Selecting lines . . . 130
37.2 Auto selecting lines . . . 130
37.3 Filtering lines . . . 130
37.4 Sorting lines . . . 131
37.5 Exporting to a file . . . 132
37.6 Ranking lines . . . 133
37.7 Don't sort/filter . . . 133

38 Preferences **135**
38.1 General . . . 136
38.2 Visualization . . . 137
38.3 Warnings . . . 137

39 Tips and Shortcuts **139**

40 Command Line Support **141**
40.1 Advanced Options . . . 141
40.2 createproject.exe (jhi.flapjack.io.CreateProject) . . . 142
40.3 creatematrix.exe (jhi.flapjack.io.CreateMatrix) . . . 142
40.4 mabcstats.exe (jhi.flapjack.io.GenerateMabcStats) . . . 142
40.5 pedverf1stats.exe (jhi.flapjack.io.GeneratePedVerF1sStats) . . . 143
40.6 splitproject.exe (jhi.flapjack.io.SplitProject) . . . 144

For more information on Flapjack, or to download the application, please visit the Flapjack web site.

CHAPTER 1

Quickstart

Here is a brief set of instructions to get you up and running with Flapjack in the shortest possible time.

Flapjack is designed for the visualization and exploration of plant genotype data, on data sets containing just a few to many thousands of lines and markers.

1.1 Importing data

- Import a data set by selecting `File > Import data` from the menu bar, then choosing the option to provide a map and genotype data from files located on disk.
- Import valid map and genotype data files, using the *Import Data* dialog (the link contains details on the formatting of these files).

Flapjack will now import your data and display it on the main canvas. This "default" view is also listed in the navigation panel down the left-hand side, which is used to show you the data sets you have opened and the various views upon them that may exist.

Each chromosome is displayed within its own tab across the top of the display. You can select and view a different chromosome by clicking on the appropriate tab for it.

1.2 Browsing the data

You can move the view around the data by using the scrollbars, or by clicking on the canvas and dragging with the mouse.

Zoom in or out using the slider. If you have the actual genotype data displayed as text on top of the visualization (`Visualization & Overlay genotypes`) then this will only be visible at larger zoom sizes.

An alternative method of moving around the data is to click and drag with the mouse on the overview panel. The red rectangle drawn on the overview represents the area of the data that the main visualization canvas is currently displaying. Dragging with the mouse moves this rectangle and therefore the main view too.

1.3 Interacting with Flapjack

Notice that as you move the mouse over the canvas, information on the genotype under the mouse is displayed in the status bar. This includes the name of the line and the marker, as well as the genotype.

At the top of the main canvas, you'll notice the chromosome map. This shows the actual positions of the markers on the map, with lines linking them from their actual position to their "virtual" position within the visualization. As you move the mouse around, the marker currently under the mouse is highlighted and its name displayed.

The line overview canvas (below the main canvas) shows you a scaled-to-fit overview of the entire line currently under the mouse. As with the main overview, the red rectangle shows the region of the line that is currently visible on the main view.

The marker overview canvas (to the right of the main canvas) shows you a scaled to fit overview of the marker currently under the mouse, along with information on its two adjacent neighbours.

CHAPTER 2

Projects & Data Formats

Everything you do with Flapjack is stored within a project file; imported data, sort orders, trait information, colour schemes, etc. A Flapjack project is active at all times when using the application - even at startup, when a default new project is already created and waiting for data to be imported into it.

A Flapjack project can store zero or more `data sets`.

2.1 Data sets, maps, and genotypes

A data set usually contains information from an imported `map file` and `genotype file`.

Note: Both the map file and the genotype file must be in plain-text, tab-delimited format.

The map file is used to provide details on the chromosomes (name and length; see warning below) and the markers (name, chromosome, and position). Order does not matter as Flapjack will group and sort them by chromosome and distance once they are loaded. A short example is shown below.

```
# fjFile = MAP
1H          125.0          # Only valid for version 1.16.10.x or above
Marker1     1H     32.5
Marker2     1H     45.4
Marker3     2H     23.8
```

The genotype file contains a list of variety lines, with allele data per marker for that line. It also requires a header line specifying the marker information for each column.

```
# fjFile = GENOTYPE
            Marker1   Marker2   Marker3
Line1       A         G         G
Line2       A         -         G/T
Line3       T         A         C
```

Note: You can include additional headers which let Flapjack know the URLs for trying to access additional information about lines and markers held in external *databases*.

Note: In a future release, additional pedigree information can be provided as headers within the genotype file. See *Pedigree Information* for details.

2.2 Flapjack views

Flapjack stores the lines and markers internally in a structure and form that can never be modified. A default `view` upon this data is created whenever an import is successful, and any subsequent operations upon the lines or markers will happen to the view, not to the data set.

Each view (and you can create as many as you like) will hold the set of `chromosomes` for that data set. Each chromosome is displayed independently, but the lines are obviously common to all chromosomes and any modification to the order or display of lines on one chromosome will be reflected across all the others too.

Colour scheme information is generally specific to a view although some settings will be chromosome-specific, such as colouring by marker.

2.3 Phenotypes/Traits

A data set can optionally also store information on one or more `traits` that are associated with the lines. Trait information is imported from a file with the following tab-delimited format:

```
# fjFile = PHENOTYPE
             Trait1       Trait1       Trait2
             Experiment1  Experiment2  Experiment1
Line1        50           High         Short
Line2        2.3          High         Medium
Line3        99.3         Low          Long
```

Trait data for a single trait can be either numerical or categorical. The line containing experiment information for each trait is optional.

2.4 QTLs

A data set can also optionally store information on one or more `QTLs` that are associated with the map. QTL information is imported from a file with the following tab-delimited format:

```
# fjFile = QTL
Name  Chromosome  Position  Pos-Min  Pos-Max  Trait   Experiment  [optional_1] ..␣
↪[optional_n]
QTL1  1H          10        8        12       Height  Exp1        25.5            high
QTL2  1H          20        19       26       Height  Exp1        34.8            low
QTL3  2H          10        8        13.5     Temp    Exp1        99.2            low
```

The **Name** to **Experiment** columns are required and must be included and listed in the order shown. After that, each QTL may have zero or more optional columns of numerical or textual data that can be included too.

2.5 Graphs

A data set can also optionally store information on one or more `graphs` that are associated with the map. Graph information is imported from a file with the following tab-delimited format:

```
# fjFile = GRAPH
SIGNIFICANCE_THRESHOLD   Graph1   5.1
SIGNIFICANCE_THRESHOLD   Graph2   7.5
Marker1                  Graph1   1.3
Marker1                  Graph2   4.3
...
Marker2                  Graph1   1.8
Marker2                  Graph2   3.9
```

Any number of graphs can be stored in a single file with data points per marker. The **SIGNIFICANCE_THRESHOLD** entry is optional (per graph) but defines the significance threshold for that graph if included which will be drawn on Flapjack's display.

CHAPTER 3

Pedigree Information

Flapjack now supports the following pedigree extensions to its genotype file option, which can be imported by including the appropriate headers at the top of a genotype file.

Note: All header lines to Flapjack files begin with **#**

The headers should be a tab-separated list of columns, in the following format (`# fjPedigree` is only space separated):

```
# fjPedigree     <progeny>     <parent-type>     <list of parents of this type>
```

Parent type can currently be one of:

```
RP        (recurrent parent)
DP        (donor parent)
N/A       (not applicable)
```

The `<progeny>` field can either be a specific progeny name, or the special case of *, meaning that this entry applies to all lines in the dataset (a line will never be assigned itself as a parent though). Multiple instances of the same progeny or parent (by name) will be mapped to all instances of that line name.

For example:

```
# fjFile = GENOTYPE
# fjPedigree     *       RP      rpParent
# fjPedigree     line1   DP      dpParent1     dpParent2     dpParent3
# fjPedigree     line2   DP      dpParent1
# fjPedigree     line3   N/A     rndParent1    rndParent2
# fjPedigree     line4   N/A     rndParent1    rndParent2
                 mrkr1   mrkr2   mrkr3
line1            A/T     C       T
line2            G       G/A     C
line3            G       G/A     A
line4            A/T     C       T
```

(continues on next page)

(continued from previous page)

```
rpParent          A/T      C       T
etc...
```

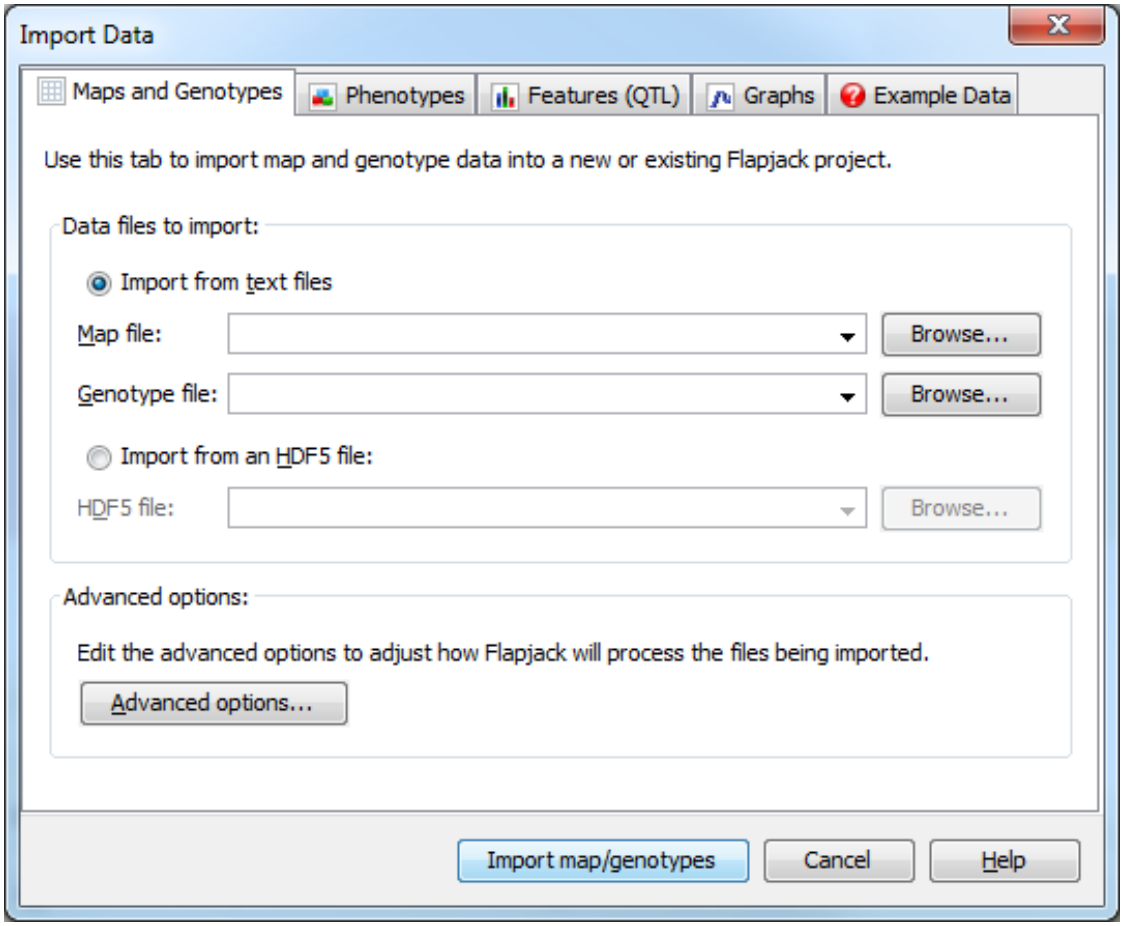
CHAPTER 4

Import Data

The `Import Data` dialog (`File->Import data`) is used to provide information on data files should be used to import data into Flapjack.

4.1 Importing data

The `Maps and Genotypes` tab is used to specify the map file and genotype file to load into Flapjack. With the `Import from text files` radio button selected, use the browse buttons to locate and select the map and genotype files you wish to load into Flapjack.



The **map file** should contain information on the markers, the chromosome they are on, and their position within that chromosome. The markers do not need to be in any particular order as Flapjack will group and sort them by chromosome and distance once they are loaded. A short example is shown below:

```
# fjFile = MAP
Marker1      1H      32.5
Marker2      1H      45.0
Marker3      2H      23.9
```

The **genotype file** should contain a list of variety lines, with allele data per marker for that line. It also requires a header line specifying the marker information for each column.

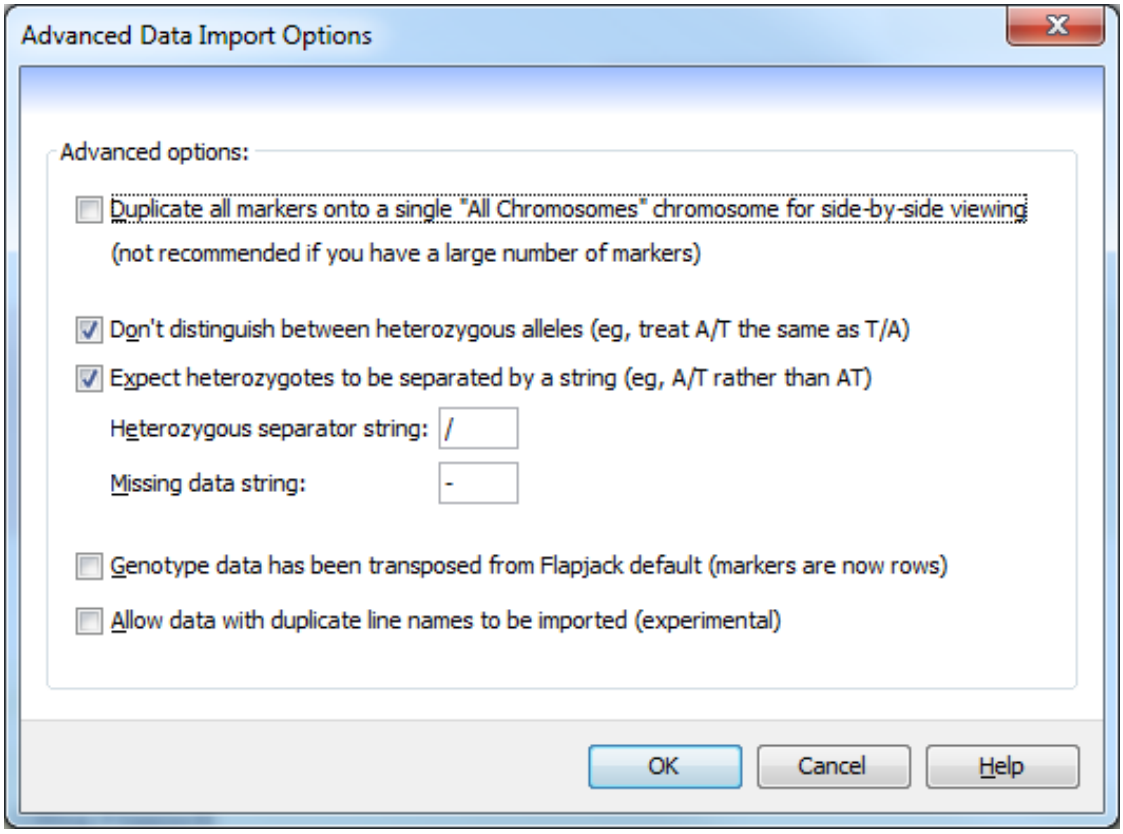
```
# fjFile = GENOTYPE
Marker1   Marker2   Marker3
Line1        A          G          G
Line2        A          -          G/T
Line3        T          A          C
```

Both the map file and the genotype file must be in plain-text, tab-delimited format. The `\# fjFile =` header lines are optional (but recommended) as they allow the files to be loaded into Flapjack via drag and drop. Once you have specified the map and genotype file you wish to load click the `Import map/genotypes` button to import your data.

Clicking the `Advanced options...` button opens the `Advanced Data Import Options` dialog.

4.2 Advanced data import options

This dialog is used to tweak how the basic Flapjack data import behaves.



The advanced data import options are:

- `Duplicate all markers onto a single "All Chromosomes" chromosome for side-by-side viewing` - selecting this option creates an additional "chromosome" which can be selected from the Chromosome dropdown menu on the toolbar. Each chromosome is laid out one after another, with a set of empty dummy markers denoting the end of one chromosome and the beginning of another. This allows you to see the genotypes for the entire dataset together. This is not reccommended when you have very large numbers of markers.
- `Don't distinguish between heterozygous alleles (eg, treat A/T the same as T/A)` - select this option to ensure that duplicate heterozygous alleles are replaced while importing the genotypes. For example, all instances of T/A would be replaced by A/T (or vice versa), so that the final data only contains A/T alleles.
- `Expect heterozygotes to be separated by a string` - this tells Flapjack to look for heterozygotes that are separated by a specific character (defined below). If your data is of the form A/T then this should be used, however, if the data defines AT as a heterozygote then this option should be deselected.
- `Heterozygous separator string` - this specifies the string that is used to define heterozygous alleles within the data.
- `Missing data string` - this specifies the string that is used to define missing genotype data. It can be left blank if no character is used within the genotype file.
- `Genotype data has been transposed from Flapjack default (markers are now rows)` - selecting this option requires effectively swaps the header row with the first column in a genotype file,

i.e. line names in the header row and marker names in the first column.

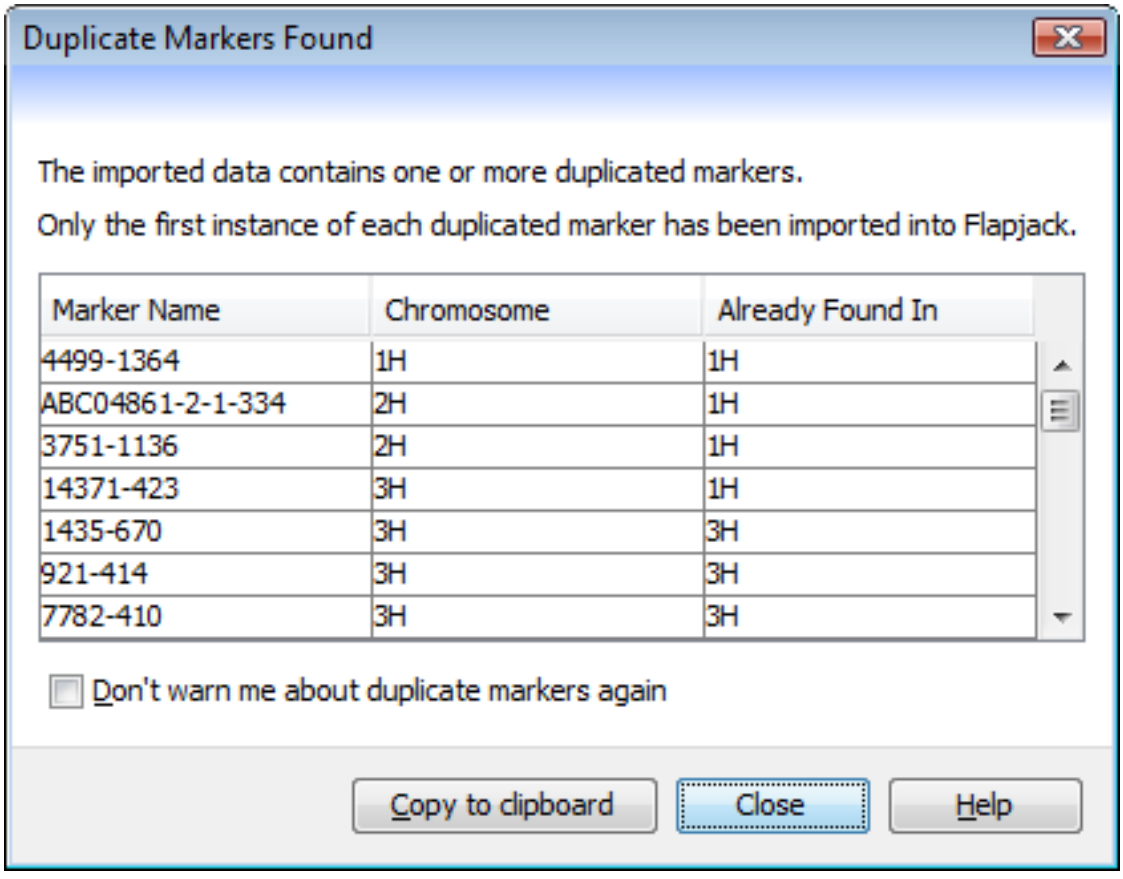
- `Allow data with duplicate line names to be imported (experimental)` - older versions of Flapjack didn't allow the importing of data with duplicate line names. It is now possible - but not reccomended - to import data with duplicate line names.

CHAPTER 5

Duplicate Markers

As data sets get larger, the potential for errors within them increases. Flapjack attempts to compensate for certain situations that may arise due to errors, one of them being detecting multiple instances of the same marker within a chromosome or sets of chromosomes when a data set is first imported.

If and when this happens, Flapjack will warn you via the `Duplicate Markers Found` dialog.



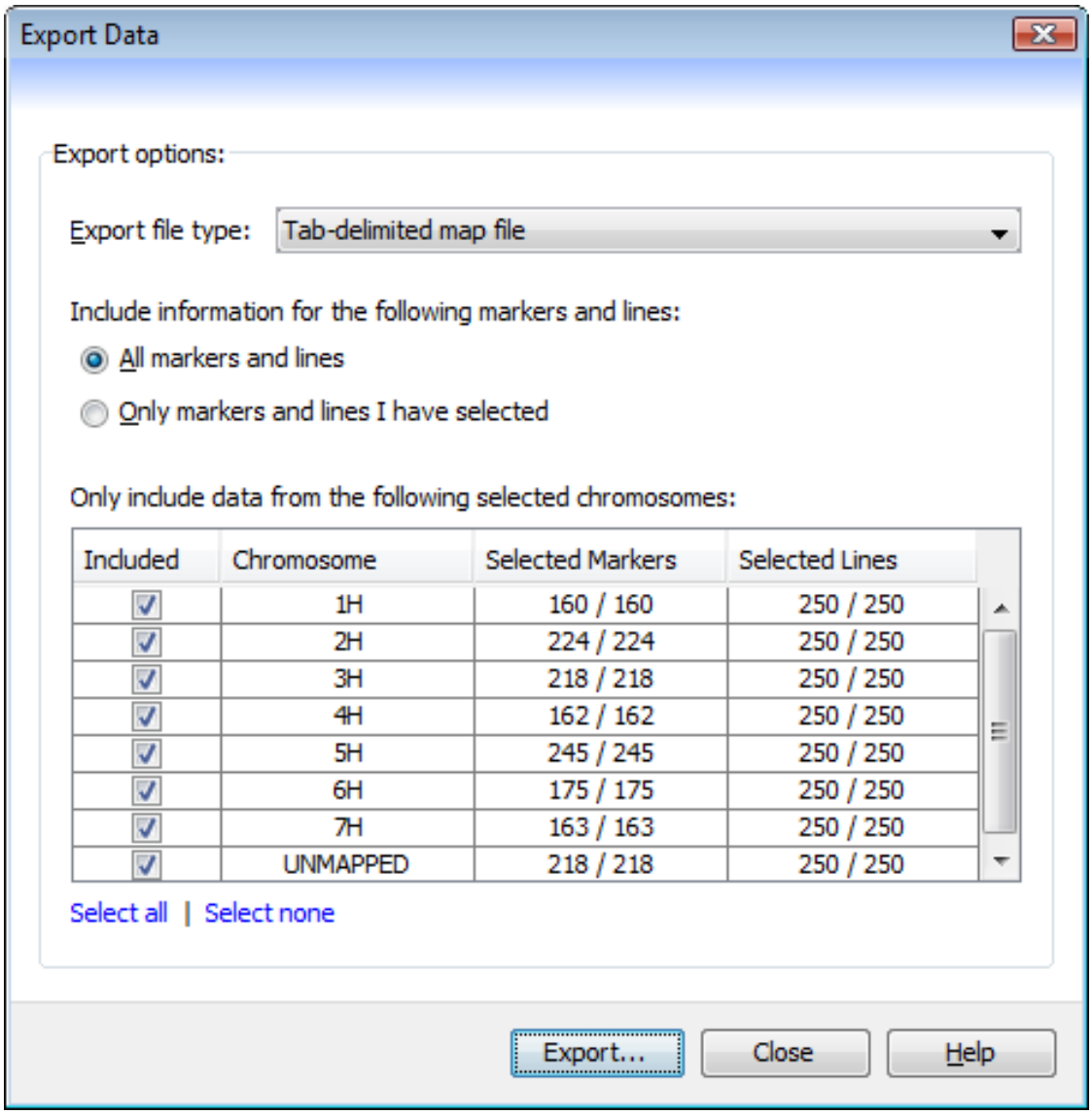
A marker within Flapjack is defined by its (case sensitive) name and only by its name. If another marker is found with the same name - either in the same or in a different chromosome - then Flapjack will ignore all instances of the marker *except* for the first instance it comes across.

Check the `Don't warn me about duplicate markers again` checkbox to have Flapjack import future data sets without warning you when duplicates are found (the duplicates will still be ignored by Flapjack though).

CHAPTER 6

Export Data

The `Export Data` dialog (`Visualization->Export view as text`) can be used to save the state of the current data set (maps, markers, and genotypes) as plain text files. The dialog will create either a .map file (for map data) or a .dat file (for genotype data).



The first step in outputting data is to decide what type of file to create.

- `Tab-delimited map file` - selecting this option creates a .map file, that will contain a tab-delimited list of markers from the dataset. The file contains one marker per line, with each line containing three columns: the marker name, the name of the chromosome it is on, and its position (in cM) within that chromosome.
- `Tab-delimited genotype file` - selecting this option creates a .dat genotype file, that will contain a tab-delimited list of lines from the dataset. A header row references the markers, then each subsequent row contains the name of a line, followed by all the allele scores for that line against each marker.

The second and third steps allow you to filter the exported data, to include either all the information within the current data set, or just a subset of it.

- `All markers and lines` - if selected, data on all markers and all lines will be exported.
- `Only markers and lines I have selected` - selecting this option ensures that only data on currently selected markers and lines will be exported. Any unselected markers or lines will be excluded from the export.
- `Only include data from the following selected chromosomes` - use the table to select which chromosomes should be included in the final output, with only those chromosomes that have their

`Included` field ticked being used. Quickly include or exclude all the chromosomes by using the `Select all` and `Select none` buttons.

Mixing and matching these settings allows you to create files that include either all markers and lines (from either all or just some chromosomes), or just some markers and lines (again, from either all or just some chromosomes).

The exported files can either be re-imported into Flapjack, or can be opened and viewed in an external text editor or spreadsheet.

CHAPTER 7

Quick Export Data

The `Quick Export All Data` dialog (`File->Quick export`) can be used to save the raw data forming the current project to a set of files on disk. The dialog prompts for a location to save to, before writing the following output files:

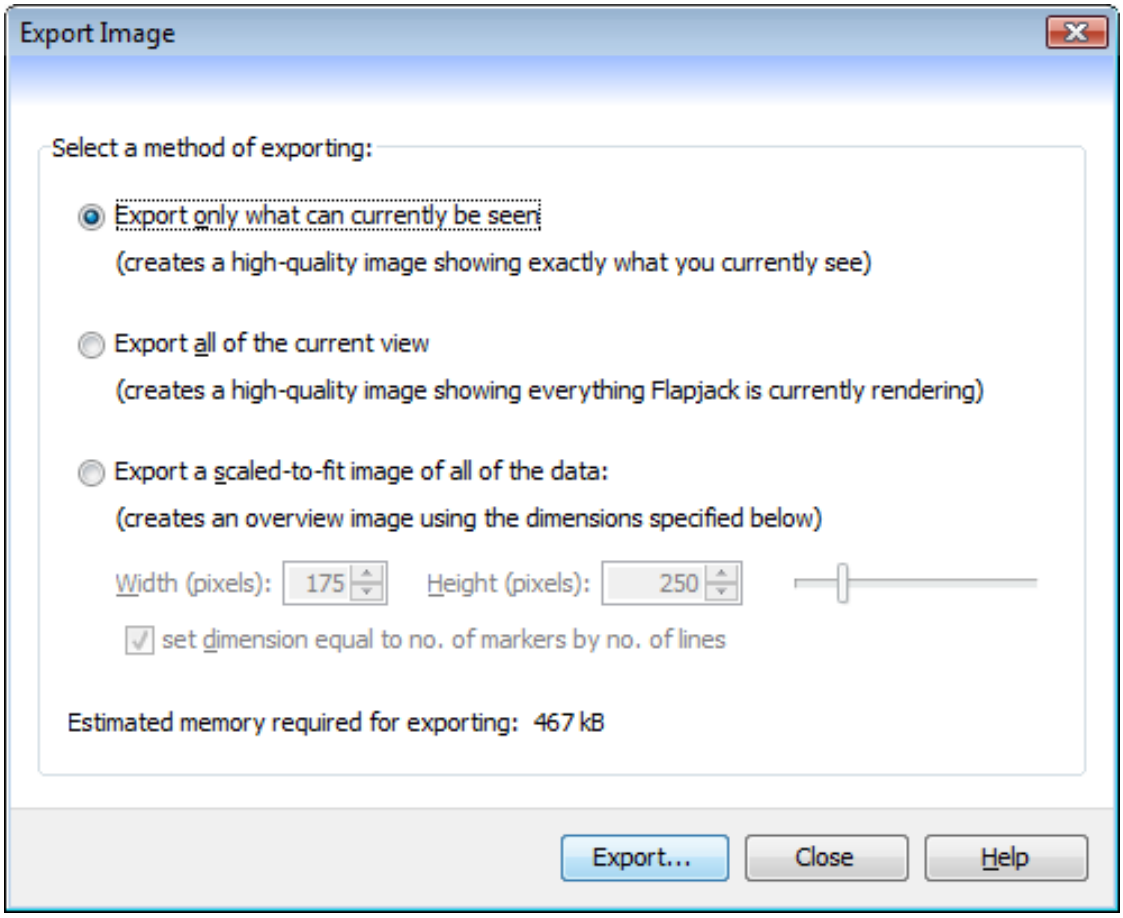
- Map data
- Genotype data
- Phenotype data
- QTL data

Each dataset and/or custom view within the project will have its own set of output files created. All of the files will be in standard Flapjack input format, as defined in the *Projects & Data Formats* topic.

CHAPTER 8

Export Image

The `Export Image` dialog (`Visualization->Export view as image`) can be used to save a PNG format file containing an image of the current chromosome's view.



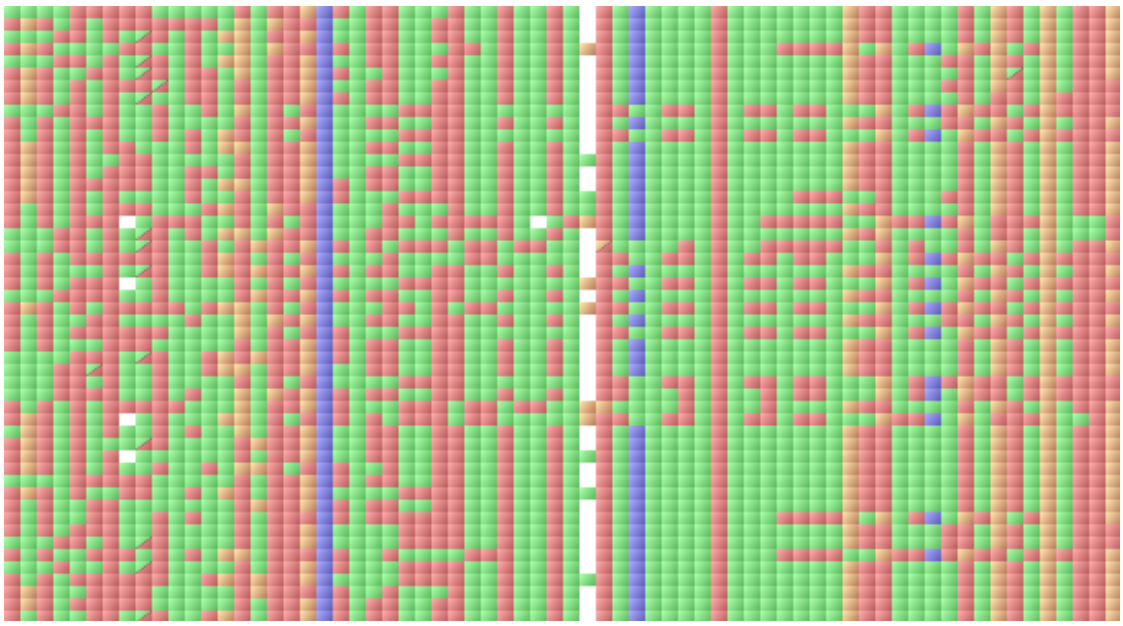
The actual image saved can be one of three possible types:

- `Export only what can currently be seen` - this option will create a high-quality image showing exactly what you currently see. The current chromosome as shown in Flapjack's main window will be replicated exactly to the saved image. Line and map information is 'not' included in this type of export.
- `Export all of the current view` - this option will create a high-quality image showing everything that Flapjack is currently rendering. All of the data within the chromosome, whether visible on screen or not will be replicated exactly to the saved image, along with line and map information.
- `Export a scaled-to-fit image of all of the data` - this option will create an overview image using the dimensions specified, resulting in an image that will show all of the data but scaled to fit the given image dimensions. The quality of the image will depend on this scaling, but you can expect something similar to what is normally shown in a Flapjack overview window. Line and map information is 'not' included in this type of export.

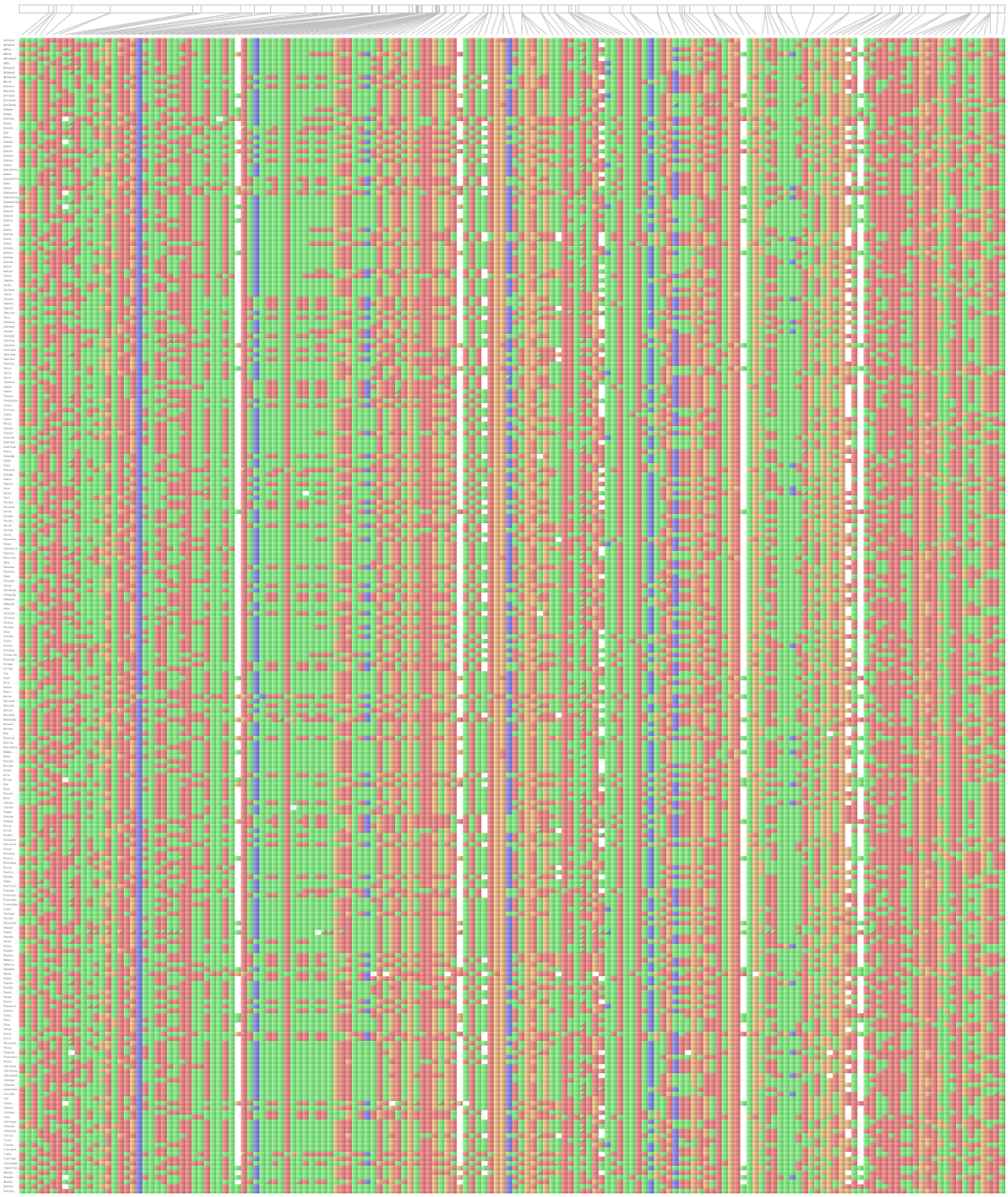
All options are reliant on there being enough free memory for Flapjack to be able to create (and compress to PNG format) the final image. The dialog gives an indication of how much memory will be required but no guarantee can be given that the final image will actually get created.

The following images show some example outputs.

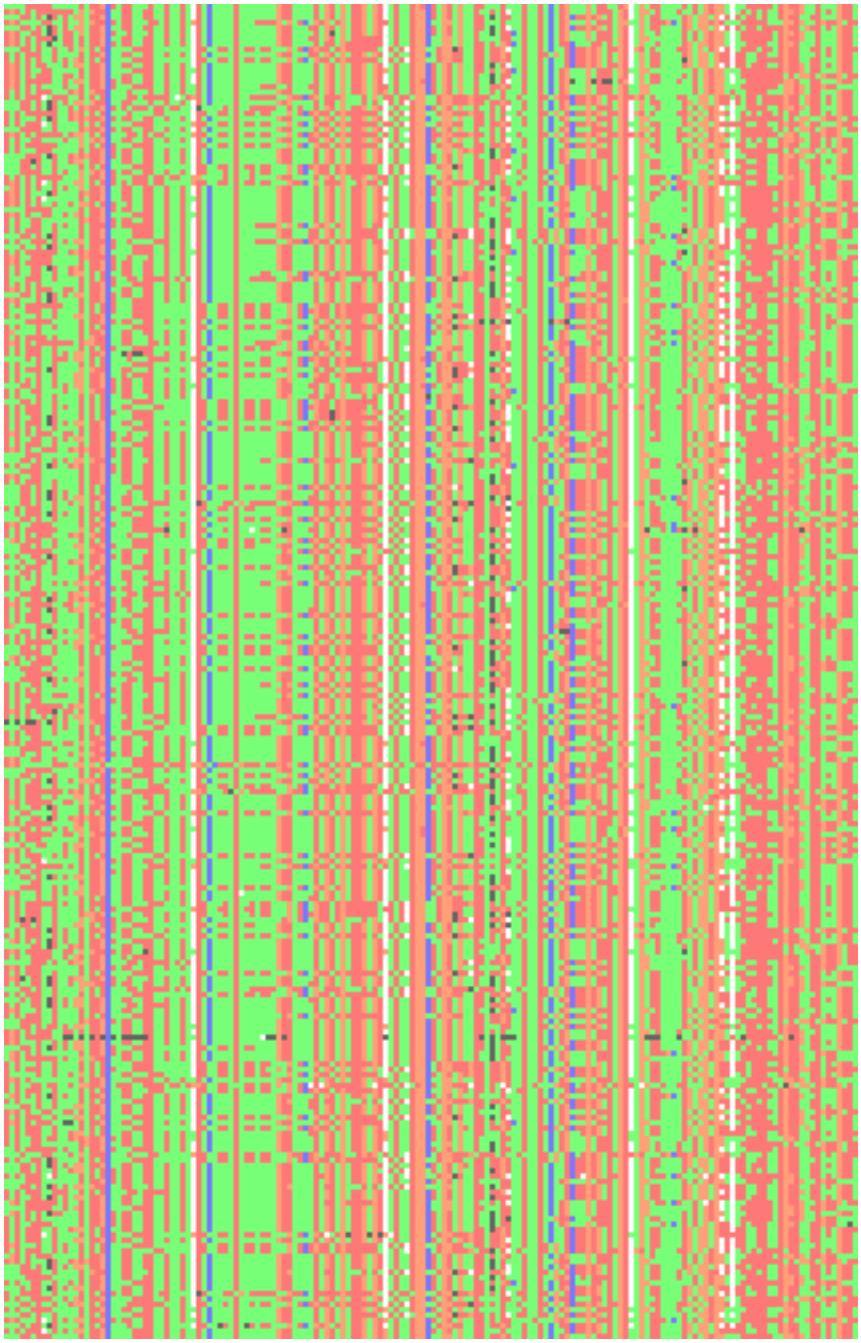
Exporting only what can currently be seen:



Exporting all of the current view:



Exporting a scaled-to-fit image:



CHAPTER 9

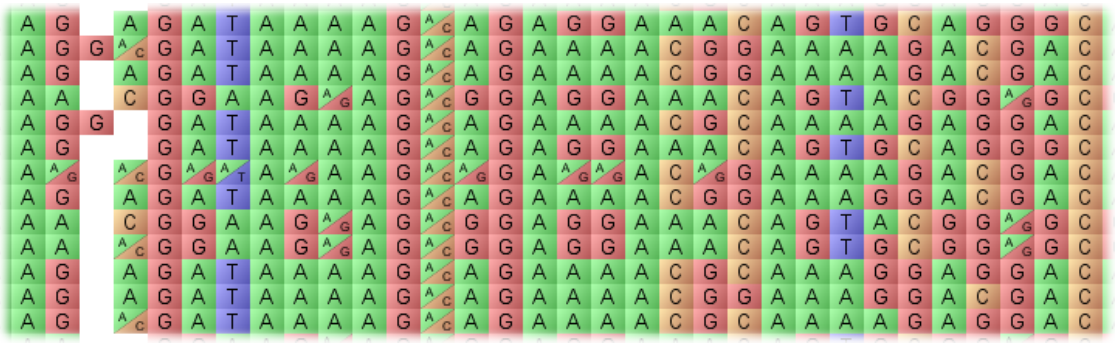
Genotype Visualization

Flapjack's main display is all about genotype visualization. The screen is divided up into several separate areas, each one showing a different component of the various genotype-related visualizations.

Note: This section of help only discusses genotype visualization. Please see the separate entries for help with *Chromosome Visualization*, or *Phenotype Visualization*.

9.1 Genotype display

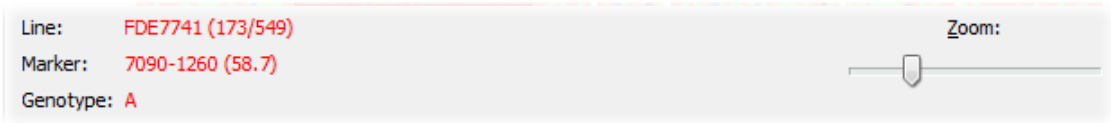
The `genotype display panel` is the primary visualization component, showing a graphical representation of the genotype values for each homozygous (single block) or heterozygous (split diagonal block) allele within the current chromosome. Data on lines is shown horizontally, with data on the markers shown vertically. You can zoom in or out of the view by adjusting the zoom slider.



By default, Flapjack will show a colour per allele type. To overlay the underlying textual value, select `Visualization->Overlay genotypes` from the menu bar. Note that the colours in use are dependant on the *Colour Schemes*.

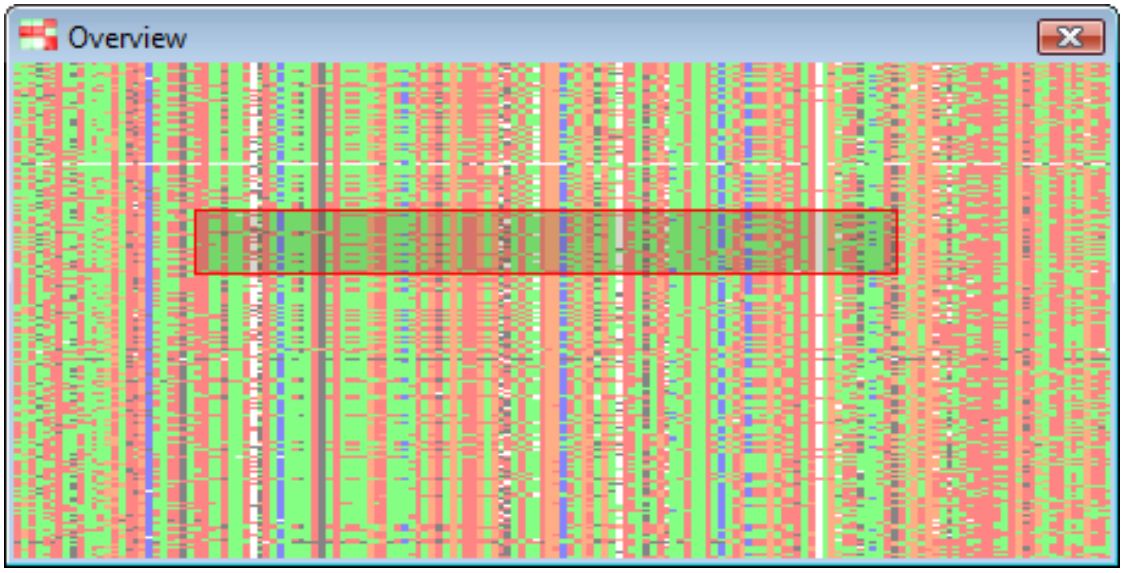
Moving the mouse over any allele will provide more specific details about it in the lower information pane. This includes the name of the line currently under the mouse, the name of the marker currently under the mouse, and the

value of the allele at that position.



9.2 Chromosome overviews

Depending on the zom level, the main display may only be showing a subset of the available data for the current chromosome. A scaled-to-fit overview of the entire chromosome is visible in the bottom left corner of Flapjack's main window. This can also be popped out into its own window by selecting `Visualization->Show overview`.



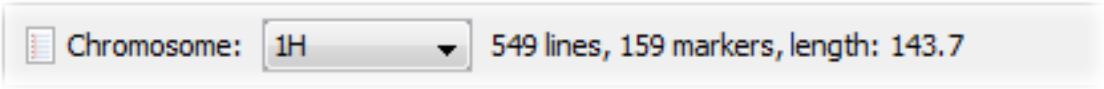
The `overview window` displays all the data, with an inset region (marked in red) that represents how much of the data is currently visible within the main genotype display area.

Note: Although the current colour scheme is maintained within the overview, for some schemes it is impossible for Flapjack to correctly render heterozygotes in an overview. In these cases, *all* such alleles are coloured in grey rather than their split diagonal colouring.

9.3 Line information

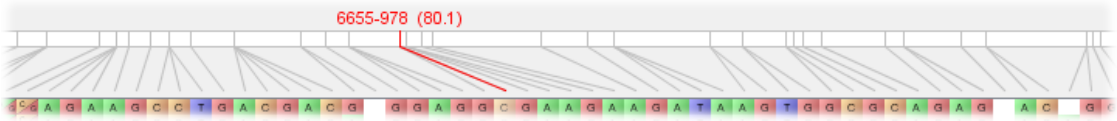
Information on each line is provided by the list of lines names that is displayed to the left-hand side of the genotype display area. As the mouse is moved over the genotypes, the current line's name will be highlighted in red within the list.

9.4 Chromosome display



Flapjack displays one chromosome at a time from within a given data set. A drop down list of all available chromosomes can be found in the chromosome summary panel towards the top of the display area. This gives the name of the currently selected chromosome, along with a count of the number of active lines and markers within it, and its length in centimorgans (cM). (Active lines or markers refers to the number that are currently visible - see *Selecting Lines and Markers* for more information.)

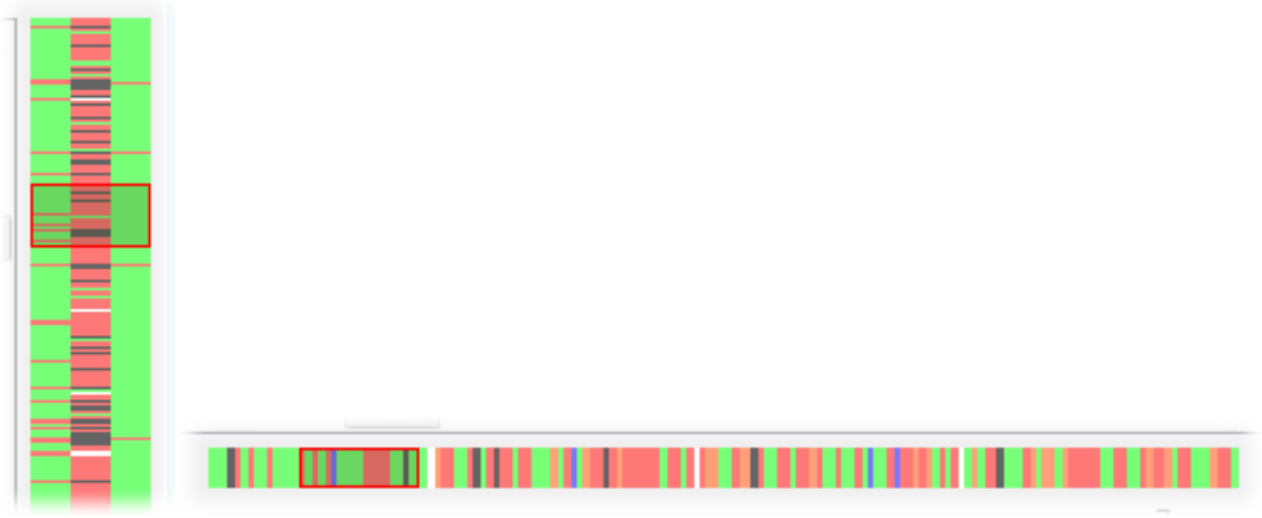
Below the chromosome summary is a graphical representation of the chromosome map. This displays the chromosome itself, along with the positions of all its markers, and how each marker maps from the chromosome to the genotype display area within Flapjack.



As the mouse is moved over the genotypes, the current marker under the mouse is highlighted in red. Its name and position on the chromosome (in brackets) is displayed as well.

9.5 Line and marker overviews

The final elements of the visualization are the two overview displays; one for markers, shown to the right of the main genotype display; and one for lines, shown below the genotype display.



Each of these displays is active whenever the mouse is moved over the genotypes. The markers overview will display a scaled-to-fit visualization of all the data for the marker currently under the mouse, along with its two nearest neighbours. Similarly, the line overview will display a scaled-to-fit visualization of all the data for (just) the current line under the mouse.

As with the main overview display, each of these components will highlight what portion of the (entire) line or marker data is currently visible within the main genotype display by outlining that section within the overview.

CHAPTER 10

Modes and Views

Flapjack has two main visualization modes: *Genotype Visualization* and *Chromosome Visualization*.

Within `Genotype Visualization` there are three further submodes of working.

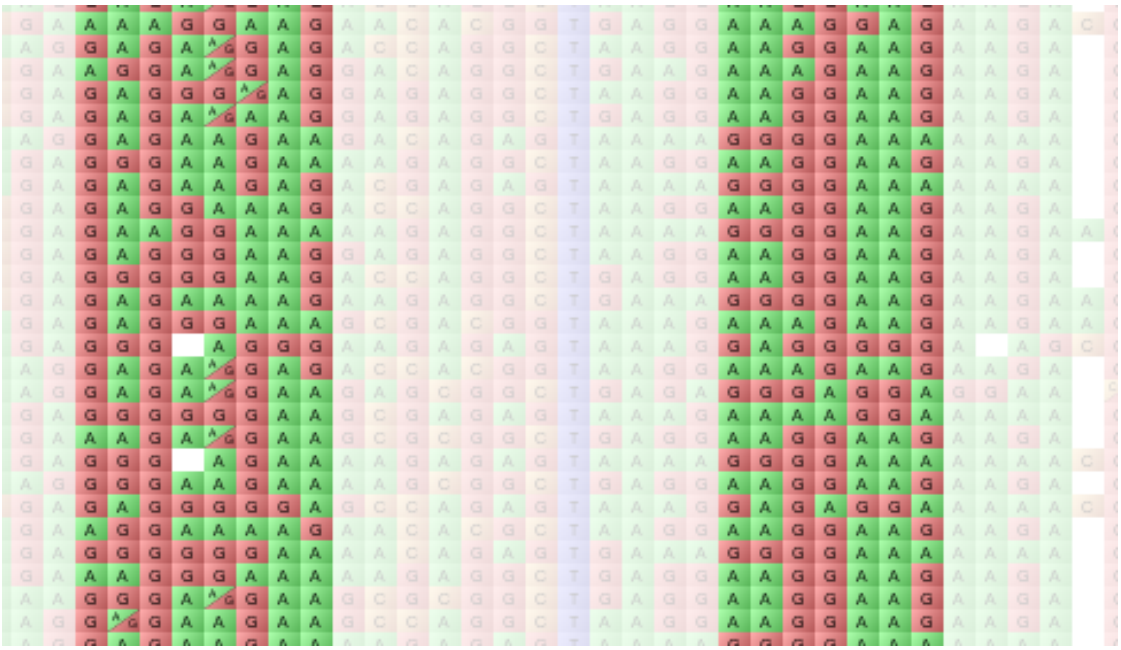
10.1 Navigation mode

In `Navigation Mode` the Flapjack canvas responds to mouse interaction by allowing you to click and drag and canvas around. The selection state of lines and markers is not shown in the this mode.

Select Navigation Mode by using `Edit->Navigation mode` from the menubar or pressing the `Navigation mode` toolbar button.

10.2 Marker mode

In `Marker mode` Flapjack will display the selection state of markers. By default, all markers are selected (painted normally). Any deselected markers are painted in a fainter colour. You can toggle the selection state of one or more markers by clicking on it, or by click-dragging across multiple markers. In Marker Mode, a CTRL+Double Click (or CMD on macOS) can be used to hide an individual marker.



Select Marker Mode by using `Edit->Marker mode` from the menubar or pressing the `Marker mode` toolbar button.

10.3 Line mode

In `Line mode` Flapjack will display the selection state of lines. By default, all lines are selected (painted normally). You can toggle the selection state of one or more lines by clicking on it, or by click-dragging across multiple lines. In Line Mode, a CTRL+Double Click (or CMD on macOS) can be used to hide an individual line.

Select Line Mode by using `Edit->Line mode` from the menubar or pressing the `Line mode` toolbar button.

CHAPTER 11

Selecting Lines and Markers

Flapjack works on the assumption that all lines and markers - when first visualized - are selected; that is, they are drawn in a normal state and are available for use if any of the analysis tasks are run (for example filtering markers from view). Deselected lines or markers are drawn in a fainter colour (see below) and are `always excluded` from any analysis. For more on this see *Modes and Views*.

11.1 Marking selections

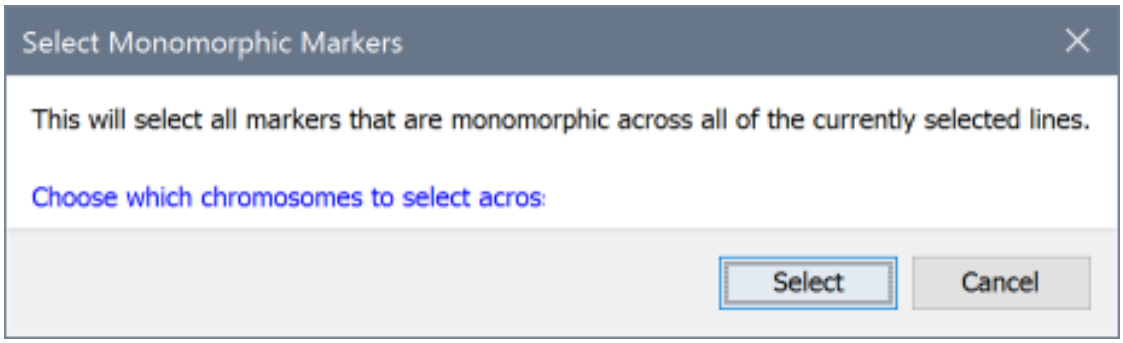
To select or deselect lines or markers in Flapjack, first ensure you are in the correct mode - either `Line mode` for selecting lines or `Marker mode` for selecting markers. Certain operations - for example selecting monomorphic markers - will switch to the correct mode for you.

Clicking on an individual line or marker will toggle its selection state on or off. You can also click-drag across multiple lines or markers at once to set their selection state.

There are also several options within the `Edit->Select lines` and `Edit->Select markers` menu options, allowing you to select all, select none, invert the current selection state, or even to import a custom selection order from a text file. The format of these files is simply one line or marker per line of the file, with any lines or markers found being set to selected. Any lines or markers missing from the file will be left as deselected.

11.2 Selecting monomorphic markers

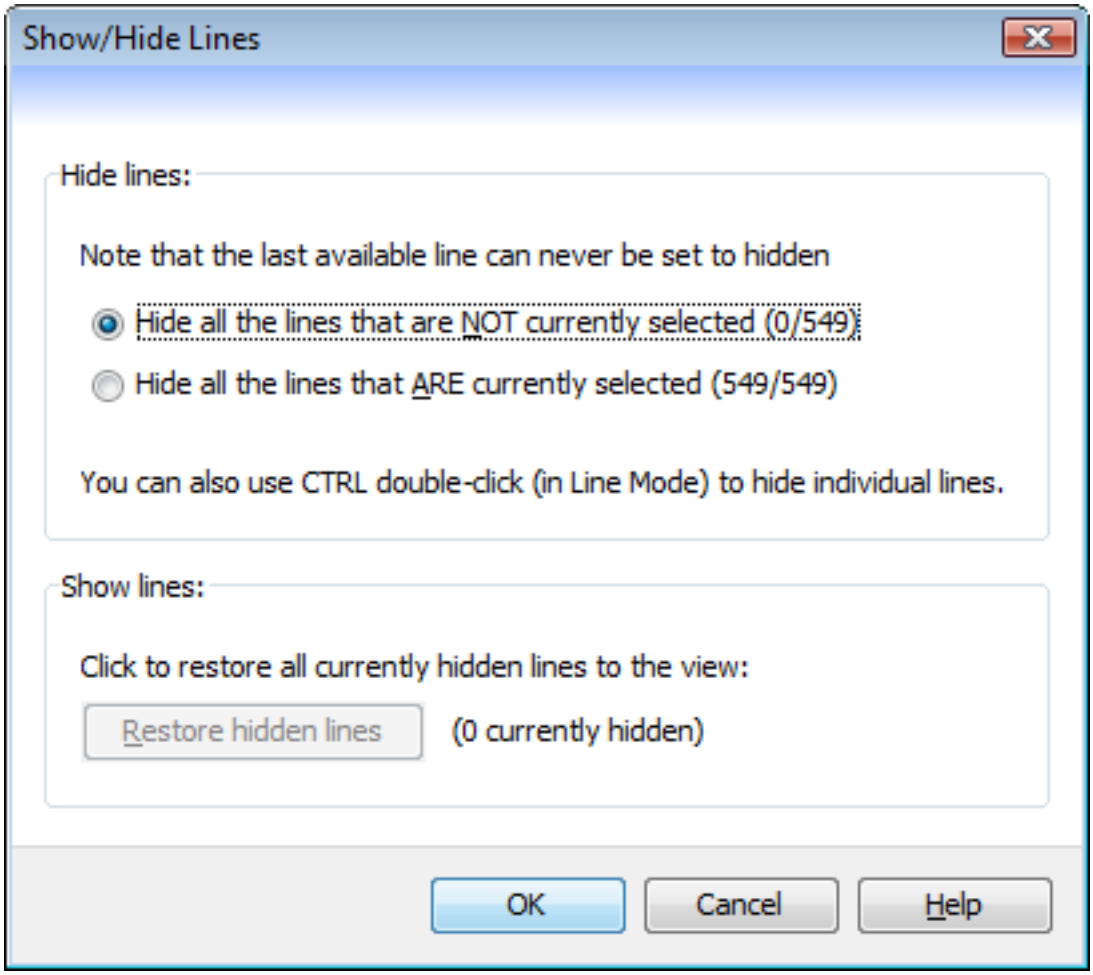
Flapjack contains a special selection dialog that can automatically select all monomorphic markers across all of the currently selected lines. This option can be found within `Edit->Select markers->Select monomorphic...` menu.


Select Monomorphic Markers
This will select all markers that are monomorphic across all of the currently selected lines.
Choose which chromosomes to select acros
Select
Cancel

CHAPTER 12

Show/Hide Lines

You can use the `Show/Hide Lines` dialog (`Edit->Show/hide lines`) to control which lines are visible within the current view. Options are provided that allow you to hide existing lines, or to restore previously hidden lines so that they become visible again.



12.1 Hiding lines

Flapjack offers three methods of hiding lines, two of which are available via this dialog:

- `Hide all the lines that are NOT currently selected` - selecting this method will hide lines that are not part of the currently selected set (these lines will be shown faded on the main display).
- `Hide all the lines that ARE currently selected` - selecting this method will hide lines that are part of the currently selected set.

To visually see which lines are selected or not, ensure Flapjack is in Line Mode before opening the `Show/Hide Lines` dialog.

The third method of hiding markers is available for quickly hiding a single line only. CTRL (or CMD on macOS) double-click a line on the canvas while in Line Mode, and it will be removed from the visible set.

12.2 Restoring lines

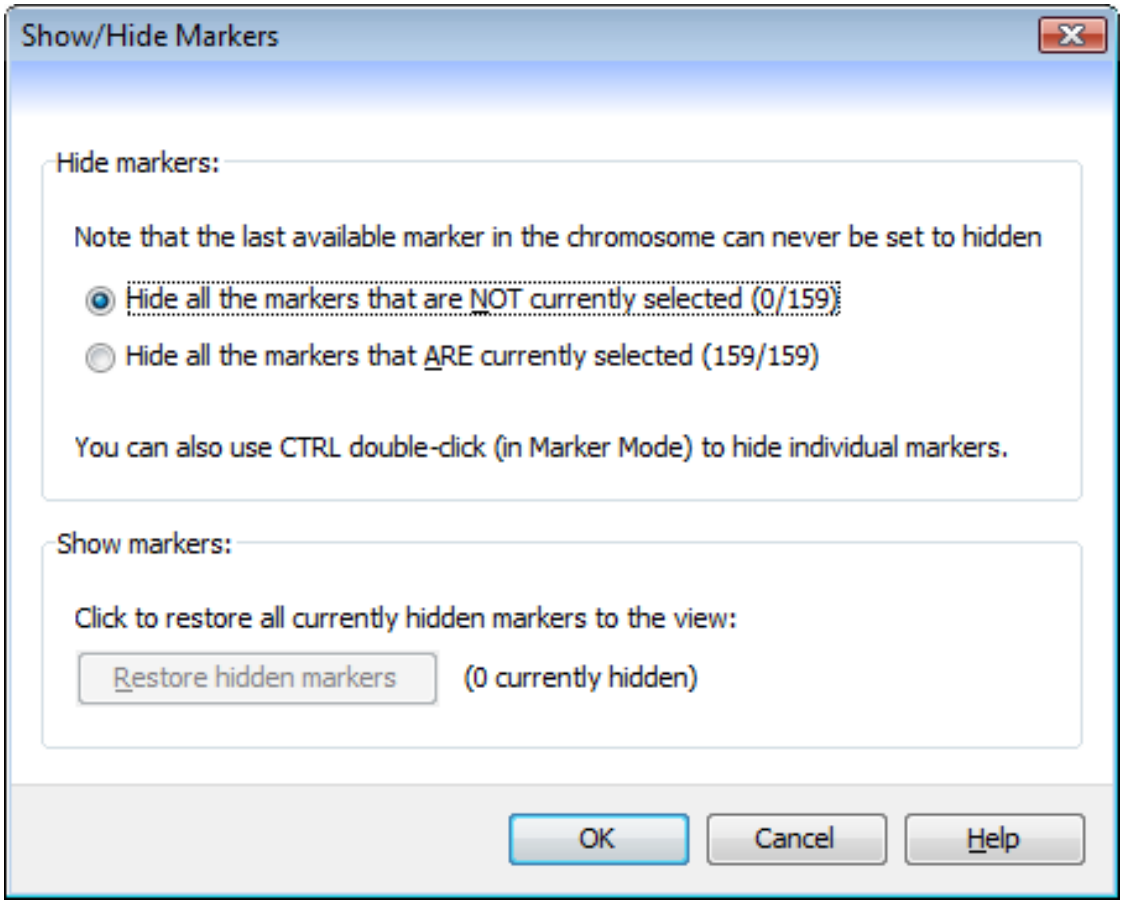
Lines that have been previously hidden can be restored to the view by clicking the `Restore hidden lines` button (which will only be enabled if there are actually lines available to be restored).

Note that all restored lines will be readded to the view **at the end** of the current set of lines.

CHAPTER 13

Show/Hide Markers

You can use the `Show/Hide Markers` dialog (`Edit->Show/hide markers`) to control which markers are visible within the current view. Options are provided that allow you to hide existing markers, or to restore previously hidden markers so that they become visible again.



13.1 Hiding markers

Flapjack offers three methods of hiding markers, two of which are available via this dialog:

- `Hide all the markers that are NOT currently selected` - selecting this method will hide markers that are not part of the currently selected set (these markers will be shown faded on the main display).
- `Hide all the markers that ARE currently selected` - selecting this method will hide markers that are part of the currently selected set.

To visually see which markers are selected or not, ensure Flapjack is in Marker Mode before opening the `Show/Hide Markers` dialog.

The third method of hiding markers is available for quickly hiding a single marker only. CTRL (or CMD on macOS) double-click a marker on the canvas while in Marker Mode, and it will be removed from the visible set.

13.2 Restoring markers

Markers that have been previously hidden can be restored to the view by clicking the `Restore hidden markers` button (which will only be enabled if there are actually markers available to be restored).

Note that when restoring markers, Flapjack will search for the best location to restore each marker to the view, but there will be cases when a marker will not return to the same location it had originally. This can happen for several reasons:

- If the marker shares the exact same chromosome position with one or more other markers. For example, consider the markers A, B, and C, all located at the same position on the chromosome. Marker B is set to hidden, then restored. The marker order after restoration could be ABC, BAC, or ACB - each is equally valid.
- If other (still visible) markers have been moved around the view. Flapjack uses the chromosome position to determine where to reinsert a marker, but if the marker order has been changed, these positions become unusable.

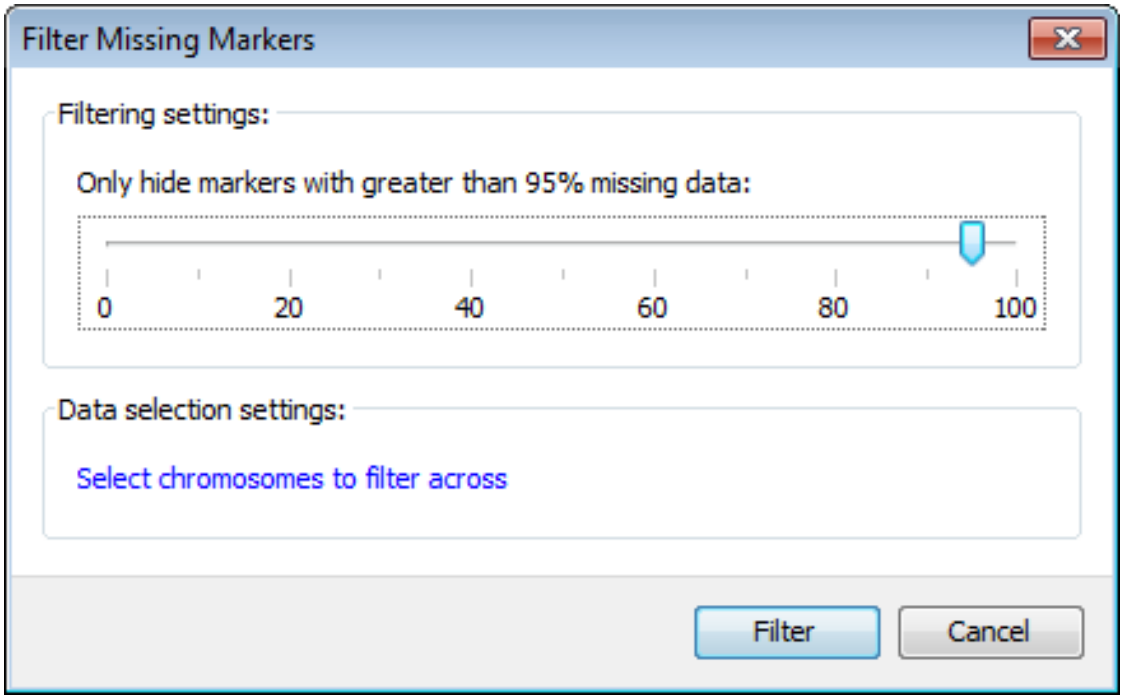
CHAPTER 14

Filtering Markers

You can filter markers in one of four ways, with all methods accessible via the `Edit->Filter markers` menubar option.

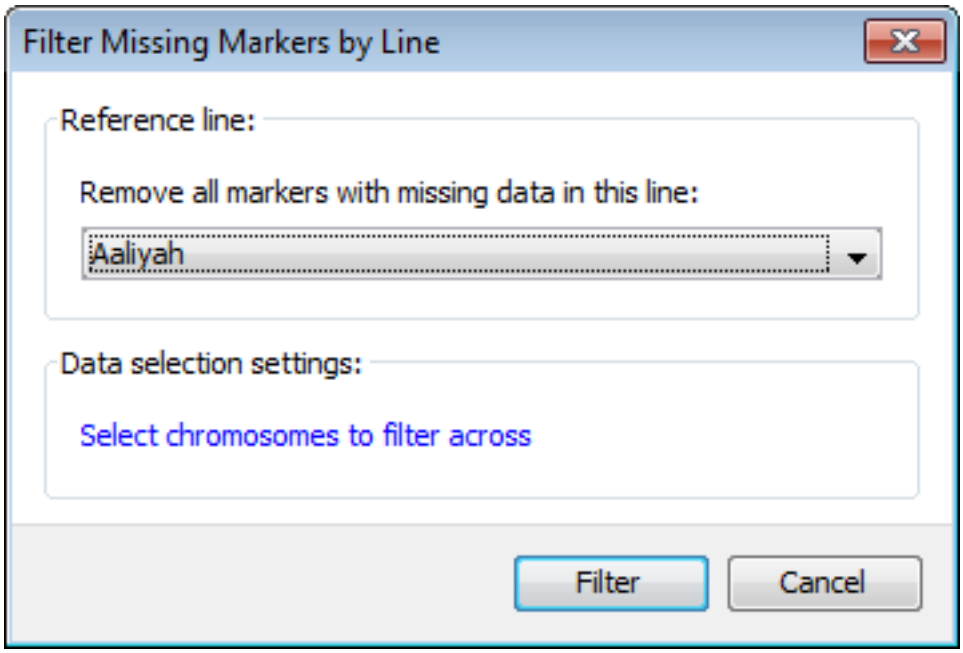
Note that any filter options will only apply to the currently selected set of markers, so you may wish to confirm marker selection before choosing one of these options, and/or modifying the filter by choosing to include or exclude entire chromosomes using the `Select chromosomes to filter across` option, located within each of the filter dialogs.

14.1 Filter missing markers



Use the dialog to select a percentage cutoff to filter markers with missing data.

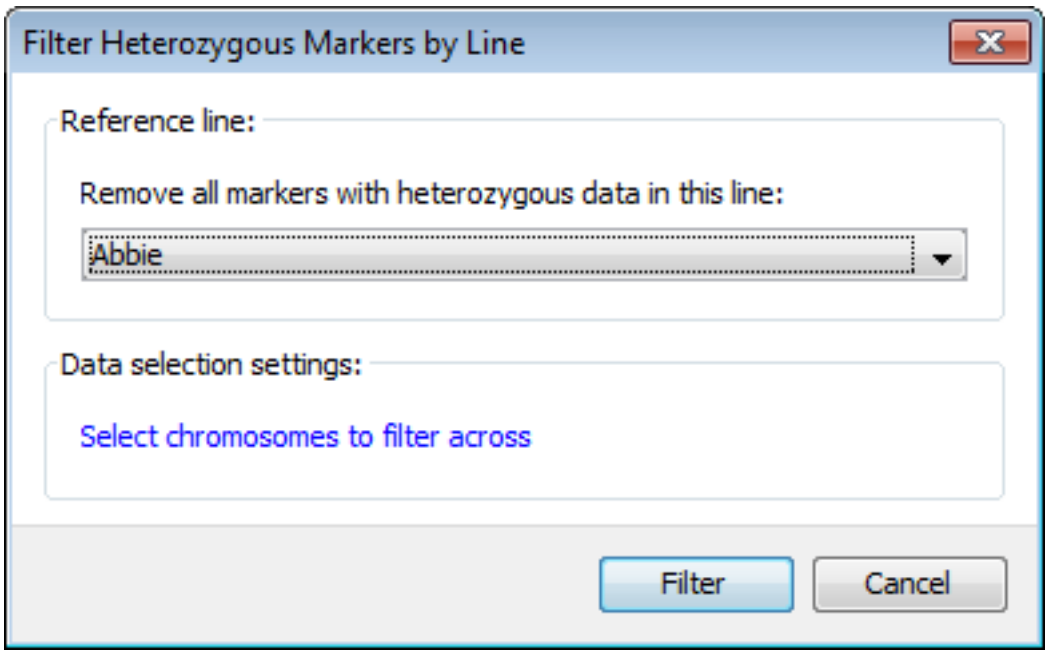
14.2 Filter missing markers by line



Use the dialog to select which line should be used; any markers with missing data for that line (and only that line) will be hidden from view. Note that if you right-click a line on the main visualization view, then select this filter from the

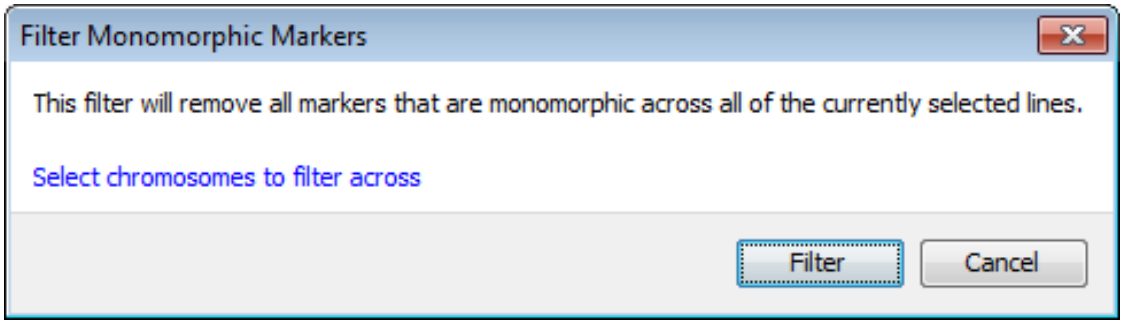
pop up menu, the line will then be preselected in the dropdown.

14.3 Filter heterozygous markers by line



Use the dialog to select which line should be used; any heterozygous markers for that line (and only that line) will be hidden from view. Note that if you right-click a line on the main visualization view, then select this filter from the pop up menu, the line will then be preselected in the dropdown.

14.4 Filter monomorphic markers



This filter will remove all markers that are monomorphic across all of the currently selected lines.

CHAPTER 15

Chromosome Map Scaling

Flapjack's chromosome map (shown above the main graphical genotype area) supports three different display methods. Experiment with the different scaling options to see which ones work best for the data in hand.

15.1 Global map scaling

Global map scaling simply fits the entire chromosome map onto the screen.

15.2 Local map scaling

Local map scaling adjusts the chromosome to display only the area covered by the markers currently on screen. Its left visible edge will be the same as the left-most marker's position, and similarly its right visible edge will be the same as the right-most marker's position.

As you scroll, the visible region of the map scrolls to accommodate all visible markers, however because the scrolling is based on the fixed-width genotype display, the map will appear to expand and contract as you move.

15.3 Classic map scaling

Classic map scaling - so called because it was the only scaling method available in early versions of Flapjack - draws a chromosome that is pixel-mapped to the main genotype display, resulting in a visual chromosome that is as wide as the main display. The area of the map visible at any point in time will not necessarily cover the areas holding the markers on screen, so you may see their 'connecting' lines disappearing off the left and right edges of the display.

CHAPTER 16

Colour Schemes

Flapjack contains several variant colour schemes for displaying genotype data. You can select a colour scheme, either by using the options on the `Visualization->Colour scheme` menu, or by right-clicking on the main genotype display area and selecting from the popup menu.

You can customize any of the colour schemes by using the *Customize Colours* dialog.

16.1 Nucleotide model

The nucleotide colour scheme provides colour information on the assumption that the data contains alleles that are of the form A, C, G, or T. Each base is assigned its own colour, and these colours apply to both homozygous and (diploid) heterozygous alleles. Heterozygous alleles are also given a special separate colour when rendering in any of the overviews to help distinguish them. All other alleles found within the data are rendered using the Other colour.

16.2 Simple 2 colour model

This is a simple two-state colour model that can be used when the data are of the form A/B (or /-, 0/1, etc). The first two homozygous allele states that are found in the imported data are assigned to the two primary colours states listed below. All other alleles found within the data are rendered using the Other colour.

16.3 By similarity to line (2 colour)

Colouring by line similarity will apply a single consistent colour to the selected comparison line. All other lines will then be coloured according to whether the allele at any given locus matches the allele of the comparison line at that locus.

16.4 By similarity to marker (2 colour)

Colouring by marker similarity will apply a single consistent colour to the selected comparison marker. All other markers will then be coloured according to whether the allele on any given line matches the allele of the comparison marker for that line.

16.5 By allele frequency

Colouring by allele frequency will use a specified frequency threshold to display alleles within a marker in one of two colours - those below the threshold (low frequency), and those above it (high frequency). Note that the frequencies are computed per marker, so a common allele in one marker may still be rare in another.

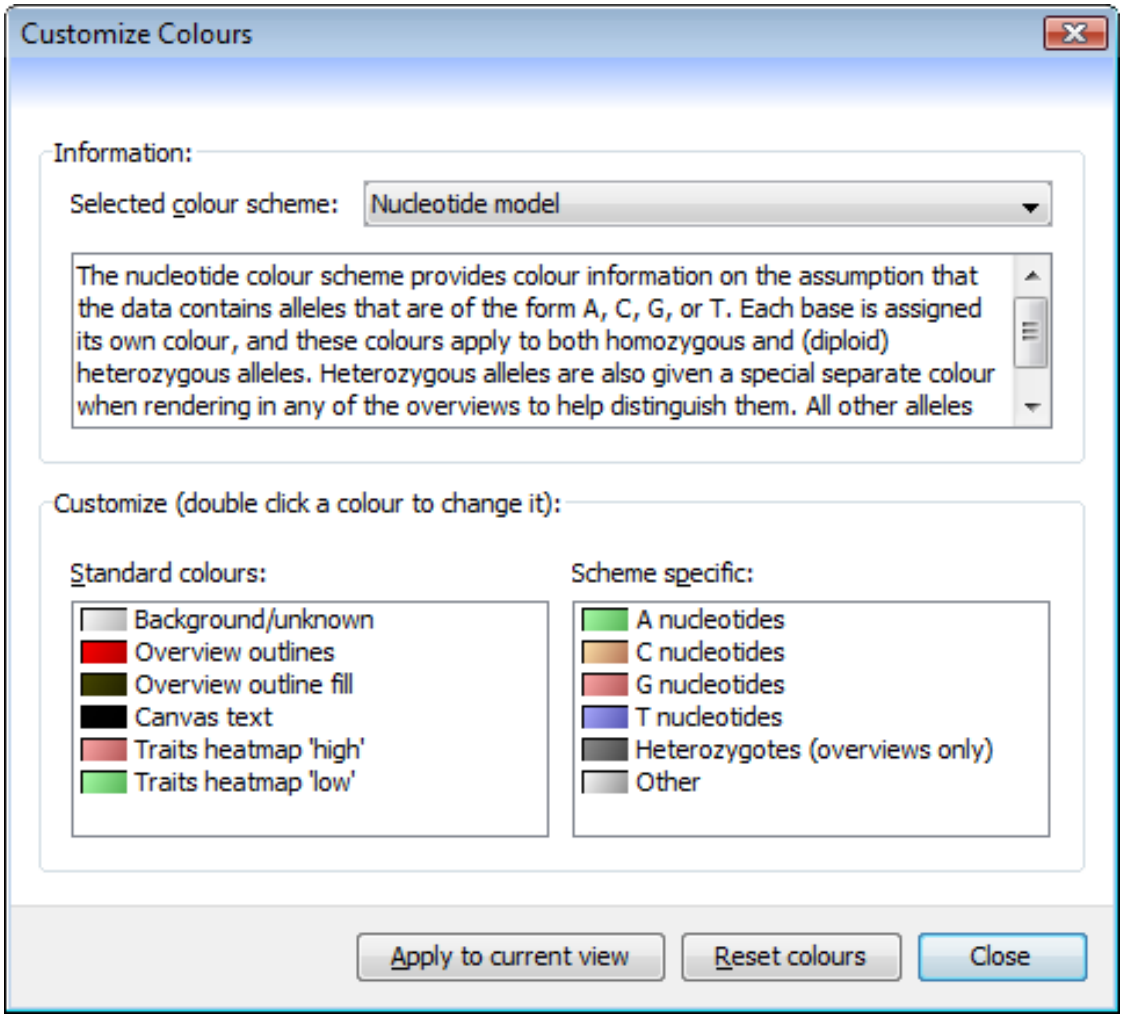
16.6 Random colour schemes

The random colour schemes apply entirely random colours to each allele state found within your data. The colours are randomized using a seed that is regenerated each time a scheme is applied to the data. The random schemes can selects colours from either the Hue Saturation Brightness (HSB) model or from a palette of 216 “web safe” colours.

CHAPTER 17

Customize Colours

The `Customize Colours` dialog (`Visualizaton->Colour scheme->Customize`) is used to modify (or reset) the colours used by Flapjack for various components of its visualization. The dialog can also be used to select the current colour scheme in use.



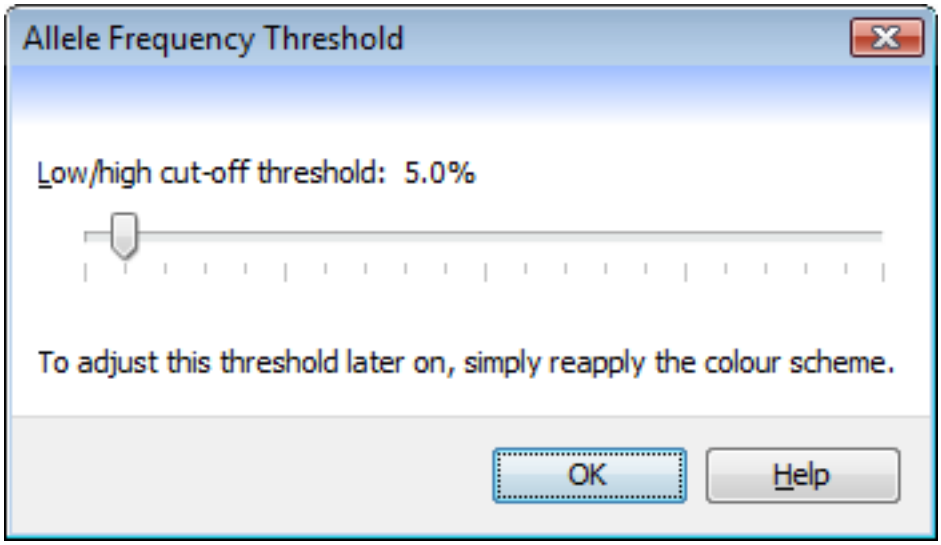
The colours within Flapjack that can be modified are listed in the `Standard colours` list within the dialog. These are the colours that are used regardless of the scheme selected. Scheme-specific colours are shown in the `Scheme specific` list. Changing to another colour scheme will change the list of modifiable colours.

To change a colour, simply double click it with the mouse, and choose a new colour from the dialog that appears. To reset all colours (across all schemes), press the `Reset colours` button.

CHAPTER 18

Allele Frequency Threshold

After selecting the `by allele frequency` colour scheme, the `Allele Frequency Threshold` dialog appears, allowing you to choose the cut-off threshold between low and high frequency alleles.



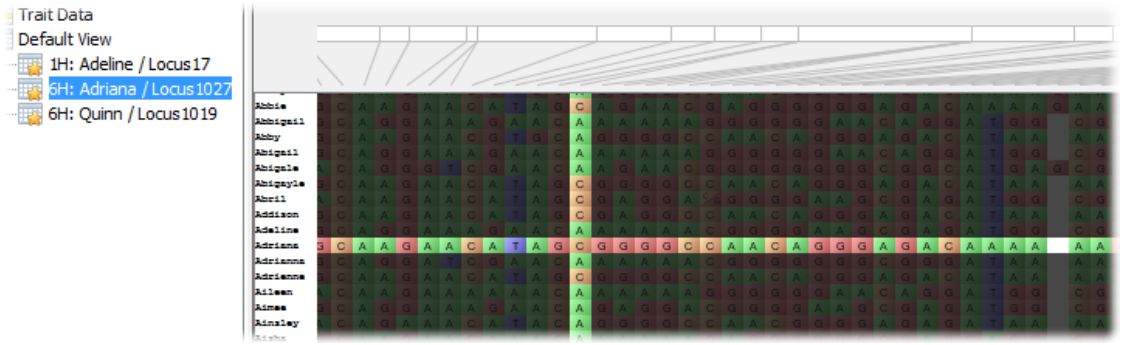
A low frequency allele (blue by default) will be any allele within a given marker that occurs at a percentage less than the chosen threshold. A high frequency allele (green by default) will be any allele within a given marker that occurs at a percentage greater than the chosen threshold.

The underlying method for this scheme calculates a percentage frequency for every allele found across a marker, ignoring unknown genotypes. For example, if you had five lines with the genotypes A/A, A/T, A/A, A/T and A/A at a given locus, then the score for A is 8/10 (80%) and the score for T is 2/10 (20%).

CHAPTER 19

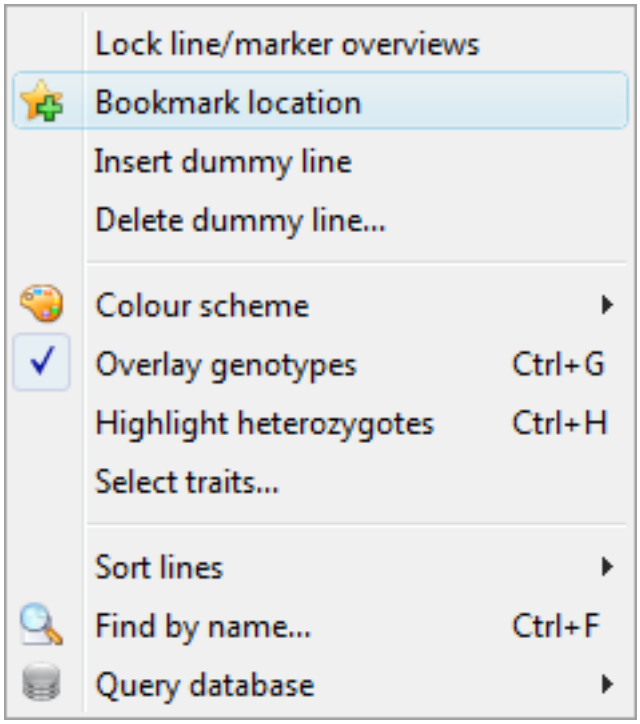
Bookmarks

A bookmark in Flapjack is simply a tracked intersection between a line and a marker (and therefore the allele at that position). You can use bookmarks to track a specific allele of interest, or just a region of the screen in general by picking a specific point within it.



19.1 Creating a bookmark

To bookmark a location, right-click an allele with the mouse, and then select `Bookmark location` from the popup menu that appears. The bookmark will then be added to the navigation tree on the left-hand side of the screen, underneath the node for the current view.



Returning to a previously created bookmark

To return to a bookmark, simply select it from the navigation tree. Flapjack will adjust the current view to show the correct chromosome for that bookmark, and will graphically highlight the line and marker intersection for a few seconds.

19.2 Deleting bookmarks

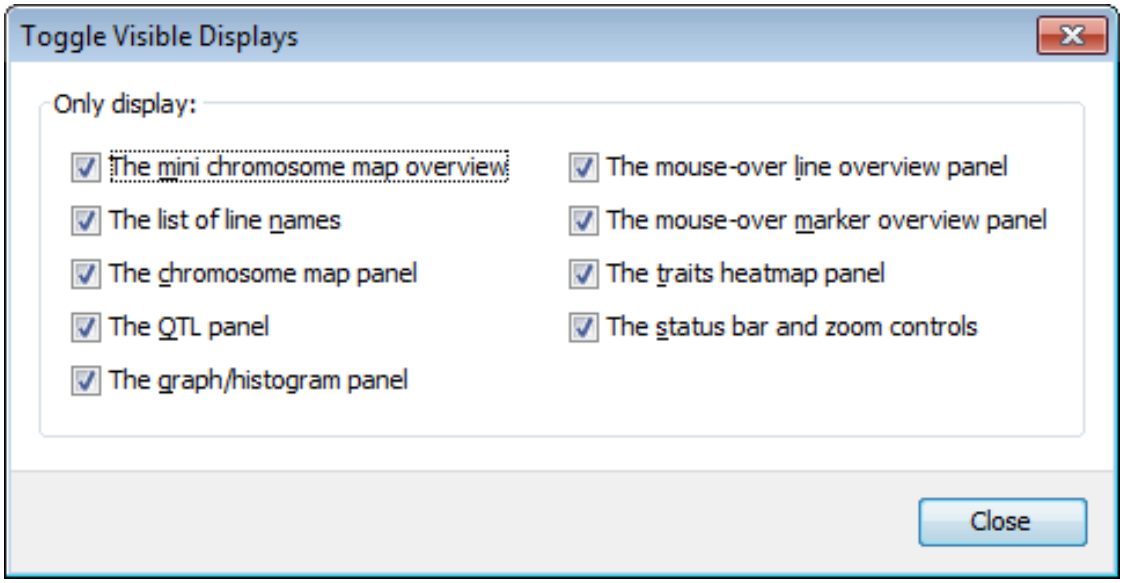
To delete a bookmark, right-click it with the mouse and then select `Delete bookmark` from the popup menu that appears.

CHAPTER 20

Toggle Visible Displays

The `Toggle Visible Displays` dialog can be opened via the `View->Toggle visible displays` menubar option.

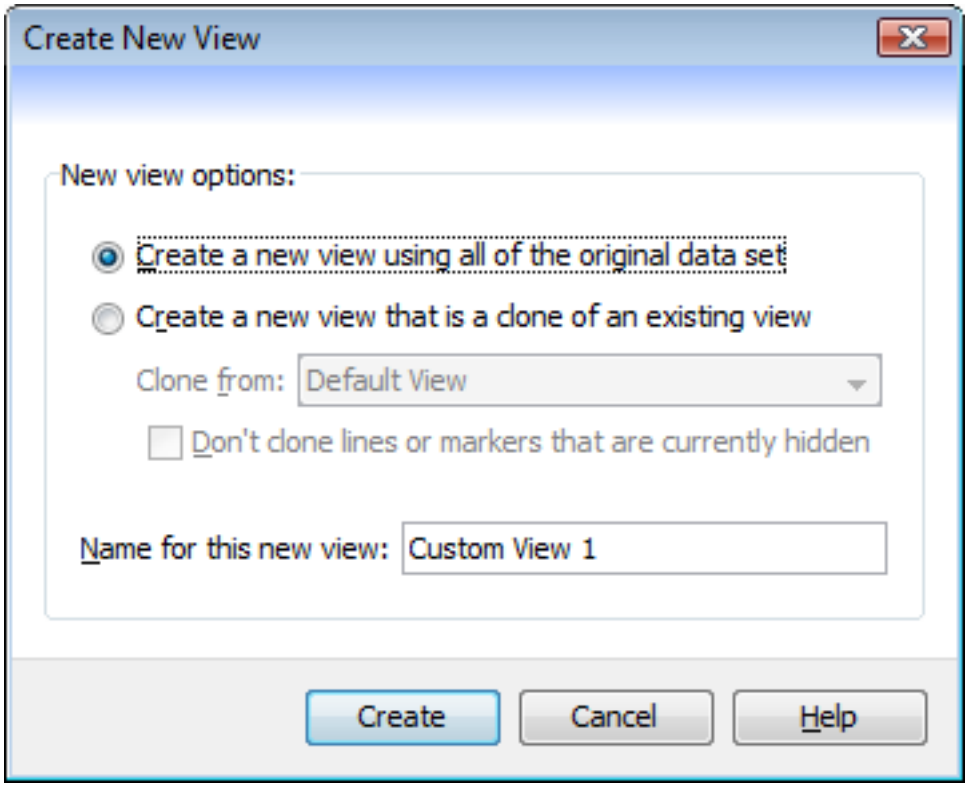
The dialog allows you to select which of Flapjack's display panels should be shown or not, potentially opening up more screen real estate for genotype visualization.



CHAPTER 21

Create New View

There is no set limit (other than memory) on the number of views that Flapjack can contain. To create a new view, use the `Create New View` dialog accessible via the `View->Create new view` menubar option.



To create a new view on your data, choose from one of the following two options:

- `Create a new view using all of the original data set` - selecting this option will create a new view that contains all of the data that was originally imported into Flapjack. The lines and markers will be shown in the same order that the underlying data contains them in.

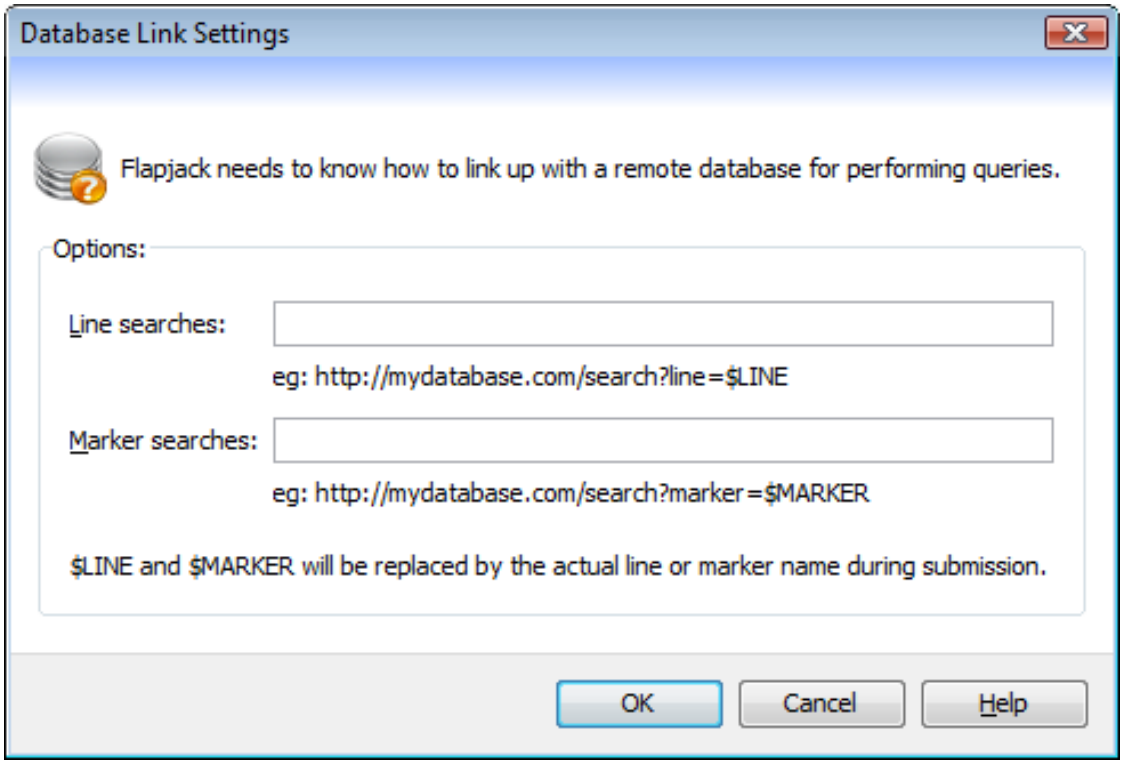
- `Create a new view that is a clone of an existing view` - selecting this option will create a new view that is a clone of one of the existing views that may already exist. It will be an exact copy of that view, with the same colour scheme, line order, marker order, selection states, etc. You also have the option to exclude any hidden lines or markers from the new view - they will **not be cloned**, and therefore cannot be restored to the new view at a later time.

With either option, you can also provide a custom name for the new view.

CHAPTER 22

Database Link Settings

Use the `Database Link Settings` dialog (`Data->query database->Settings`) to set or update the URLs that Flapjack will use when attempting to query a remote database for line or marker information.



- `Line searches` - enter a valid URL containing a location that can be queried for information on a line, where $LINE will be replaced by the actual name of the line.
- `Marker searches` - enter a valid URL containing a location that can be queried for information on a marker, where $MARKER will be replaced by the actual name of the line.

22.1 Genotype File Headers

Flapjack's genotype file format supports headers for specifying `Database Link Settings`.

- **`Line searches`**

```
# fjDatabaseLineSearch = http://mydatabase.com/search?line=$LINE
```

  As above, enter a valid URL containing a location that can be queried for information on a line, where $LINE will be replaced by the actual name of the line.

- **`Marker searches`**

```
# fjDatabaseMarkerSearch = http://mydatabase.com/search?marker=$MARKER
```

  As above, enter a valid URL containing a location that can be queried for information on a marker, where $MARKER will be replaced by the actual name of the line.

CHAPTER 23

Chromosome Visualization

The `Chromosomes Visualization` tab can be switched to by selecting its option on the toolbar, or by using the `View->Chromosomes view` menubar option.

Chromosome visualization presents an overview diagram of all chromosomes in the current dataset, and the location and density of their markers. Each chromosome can be clicked on, which will then render additional information about it using the graph view at the bottom of the window.

CHAPTER 24

Phenotype Visualization

Flapjack can import supplementary phenotype (trait) data that is associated with the genotype data set.

24.1 Importing phenotypic data

To import phenotype data into Flapjack, open up the `Import Data` dialog and select the `Phenotypes` tab where you can choose the file to import.

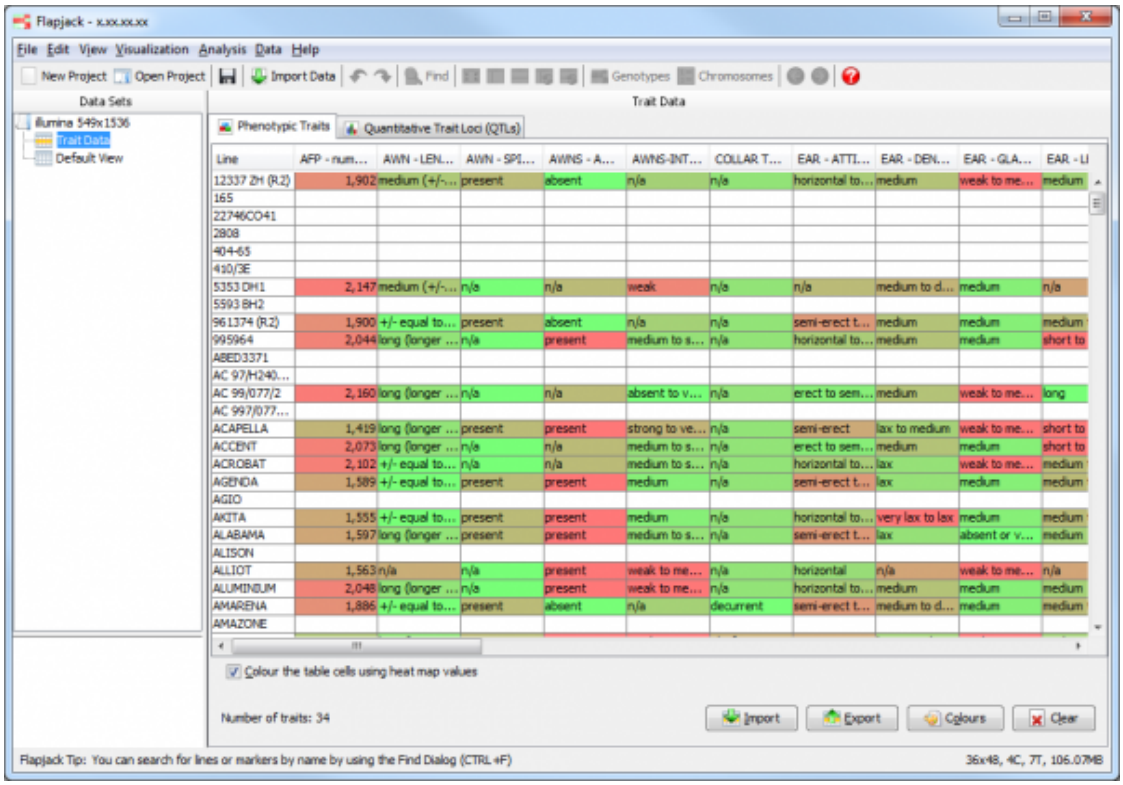
Phenotypes can be imported from files containing the following tab-delimited format:

```
# fjFile = PHENOTYPE
              Trait1        Trait1        Trait2
              Experiment1   Experiment2   Experiment1
Line1         50            High          Short
Line2         2.3           High          Medium
Line3         99.3          Low           Long
```

Trait data for a single trait can be either numerical or categorical. The line containing experiment information for each trait is *optional*.

24.2 Phenotype summary information

One the phenotypes are imported, Flapjack will display a summary table of the data within the `Trait Data` node associated with the main data set. This is a matrix of data with lines down the left hand side and phenotypes across the top of the display. Each intersection of a line and a phenotype represents a phenotype value.



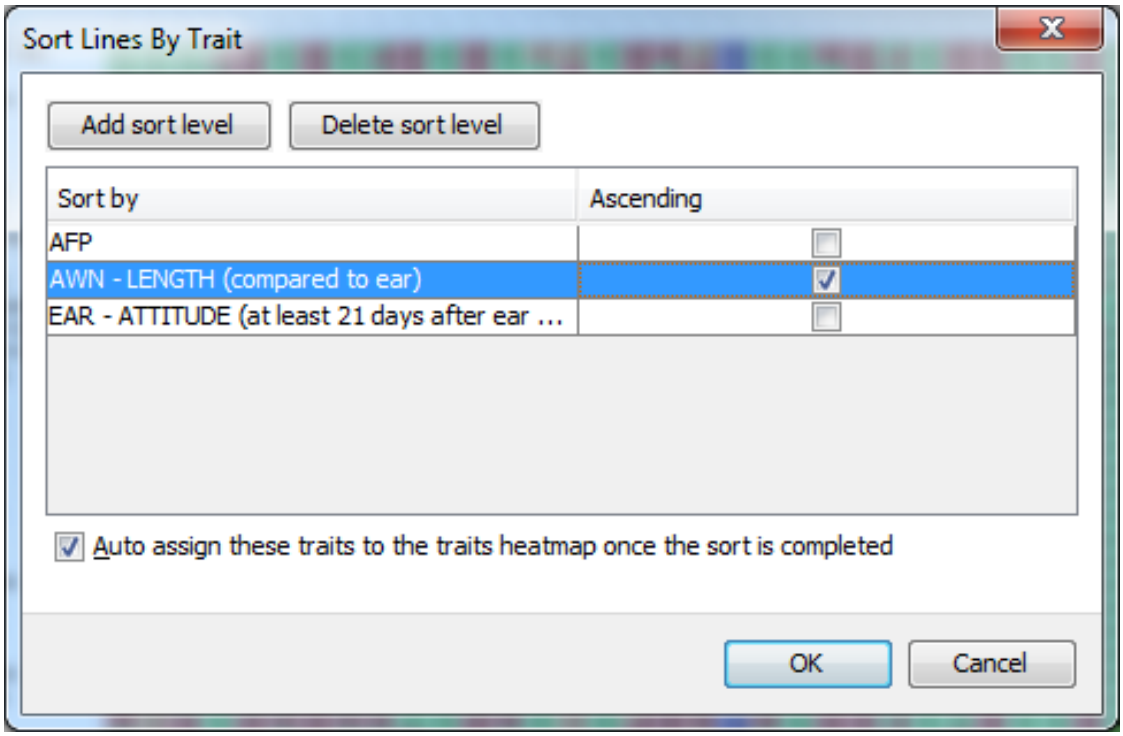
By default the phenotypes values are displayed in a heatmap format, with green representing low values and red representing high values. These colours can be changed either Flapjack wide, or for a specific trait by clicking the `Colours` button on the `Phenotypic Traits` tab of the `Trait Data` view and selecting new colours from the `Trait Colours` dialog.

24.3 Visualizing alongside genotype data

When you load phenotypic data into Flapjack, Flapjack selects the first three traits found in the file and displays them in a heatmap next to the main graphical genotype display. To select which phenotypes are in the heatmap, select `Data->Select traits` to open the `Select Traits` dialog. Select the traits you wish to display next to the lines in the genotype view, then click `OK`.

24.4 Sorting by phenotype

Sort the lines by your phenotypic data by right clicking on a phenotype and choosing `Sort A-Z` or `Sort Z-A` when you have categorical data loaded, or `Sort smallest to largest'` or `Sort largest to smallest` when you have numerical phenotype data loaded. You can also carry out an advanced sort by selecting `Analysis->Sort lines->By trait`.

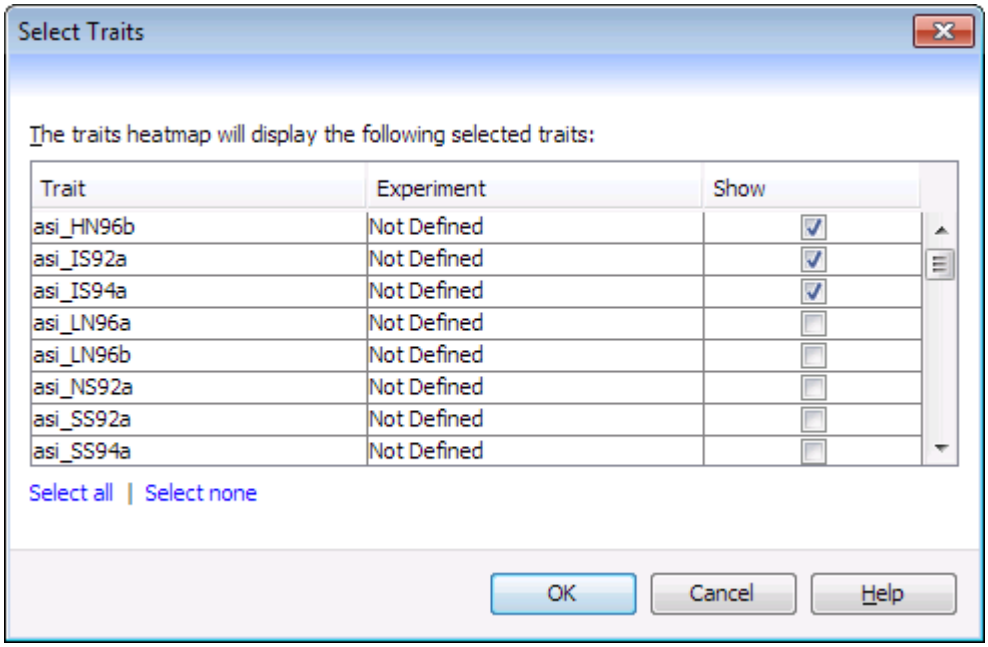


This type of sort allows you to sort by one trait first, then for lines which share values for that trait you can sort within those by a secondary trait and so on. By default the dialog will display a sinle trait to sort by and whether or not the sort will be in ascending order. You can add these secondary (tertiary, n-ary) sorts by clciking the `Add sort level` button. If you decide to remove a sort level, simply click on the row you wish to remove then select `Delete sort level`. Note the `Auto assign these traits to the traits heatmap once the sort is completed` checkbox. If this is checked it will automatically display the traits used to carry out your sort next to the genotypes.

CHAPTER 25

Select Traits

Use the `Select Traits` dialog (`Data->Select traits`) to pick which traits should be displayed on the traits heatmap alongside the list of line names.

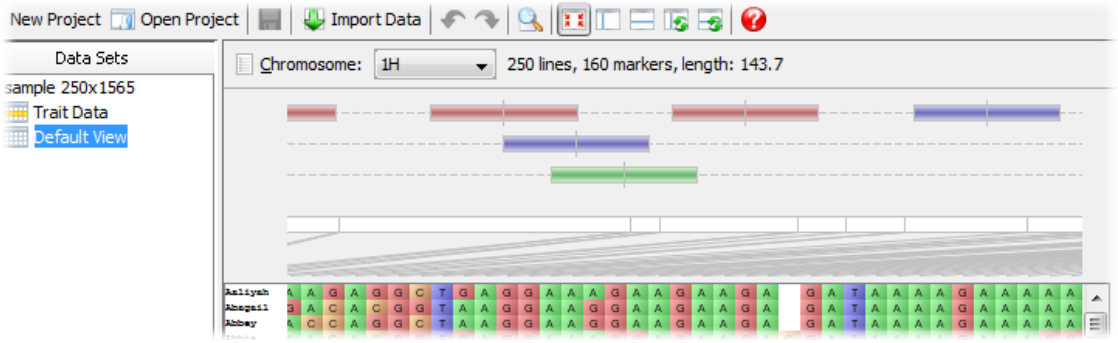


Flapjack can display any number of traits in the heatmap. Each trait that is selected in the table will be made visible.

CHAPTER 26

QTL Visualization

Flapjack can import supplementary QTL (quantitative trait locus) data that is associated with the main map and genotype data set. Flapjack supports multiple QTLs per chromosome, displayed over multiple ‘’tracks” per chromosome.



Each QTL can be colour-coded to the trait it is associated with, and ultimately used to select or de-select groups of markers within the main display.

26.1 Importing QTL data

To import QTL data into Flapjack, open up the `Import Data` dialog and select the `Features (QTL)` tab.

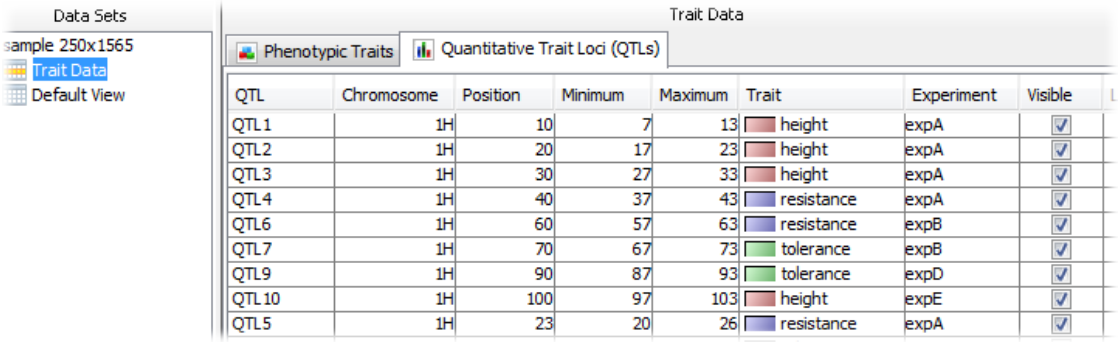
QTLs can be imported from files containing the following tab-delimited format:

```
# fjFile = QTL
Name  Chromosome  Position  Pos-Min  Pos-Max  Trait  Experiment  [optional_1] ..␣
↪[optional_n]
QTL1  1H          10        8        12       Height  Exp1       25.5            high
QTL2  1H          20        19       26       Height  Exp1       34.8            low
QTL3  2H          10        8        13.5     Temp    Exp1       99.2            low
```

The **Name** to **Experiment** columns are required and must be included and listed in the order shown. After that, each QTL may have zero or more optional columns of numerical or textual data that can be included too, for example, LOD scores, r^2 values, etc.

26.2 QTL summary information

Once the QTLs are imported, Flapjack will display a summary table of the data within the `Trait Data` node associated with the main data set. This lists each QTL, along with its values, and also provides a checkbox to set its *visible* state, that is, whether that QTL should be displayed or not.



QTL	Chromosome	Position	Minimum	Maximum	Trait	Experiment	Visible
QTL1	1H	10	7	13	height	expA	☑
QTL2	1H	20	17	23	height	expA	☑
QTL3	1H	30	27	33	height	expA	☑
QTL4	1H	40	37	43	resistance	expA	☑
QTL6	1H	60	57	63	resistance	expB	☑
QTL7	1H	70	67	73	tolerance	expB	☑
QTL9	1H	90	87	93	tolerance	expD	☑
QTL10	1H	100	97	103	height	expE	☑
QTL5	1H	23	20	26	resistance	expA	☑

If any QTLs are not completely on the chromosome map they are associated with, Flapjack will refuse to make them visible and will tag them with a red X in the summary table.

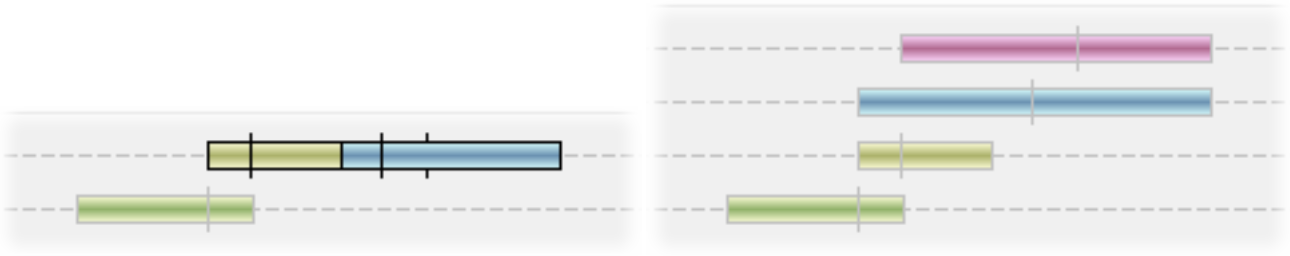
26.3 Displaying QTLs

To ensure the QTLs are displayed on the main visualization area, check the settings in the `Toggle Visible Displays` dialog and check the `QTL panel` option. The display of QTLs is common across all views for a data set - you cannot have a custom set of QTLs per view (unlike lines or markers).

QTLs are displayed along the top of the map with their exact position being marked by a small vertical bar. Their left and right error margins are shown by the overall size of the box itself. Each QTL will be colour-coded by trait.

Mouse-over any QTL to see further information on it, including its name and position values, trait, experiment, and any further optional data that may have been included in the import file.

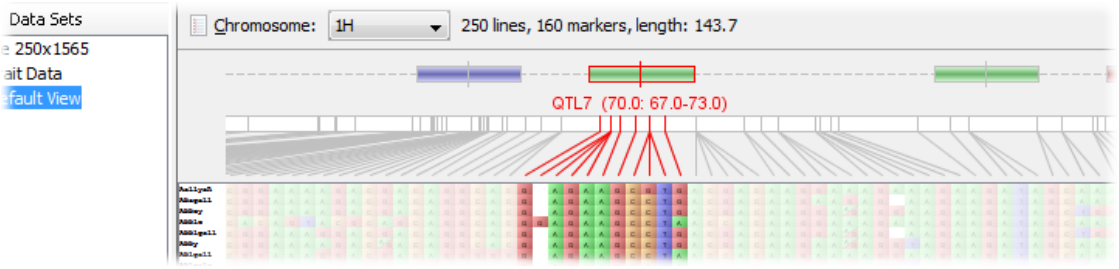
The number of visible tracks can be adjusted by dragging the slider below the QTLs either up or down. If there is is not enough room to display all the QTLs at a given location, Flapjack will collapse them together, and will outline the group in black. To ensure that all QTLs are visible, drag the slider lower to open more tracks until all black outlines have been removed.



26.4 Interacting with QTLs

To move a QTL between tracks, hold down the CTRL key and drag it with the mouse to the new track. Flapjack will remember any custom positioning across the tracks, but only so long as you leave the number of tracks the same. If you change the number of active tracks used for displaying QTLs, then a new layout will be created for them, losing any changes you may have made.

To select or deselect markers that are within the map region of a given QTL, simply click the QTL with the mouse. Flapjack will automatically change to `marker mode` and toggle the selection state of any relevant marker.



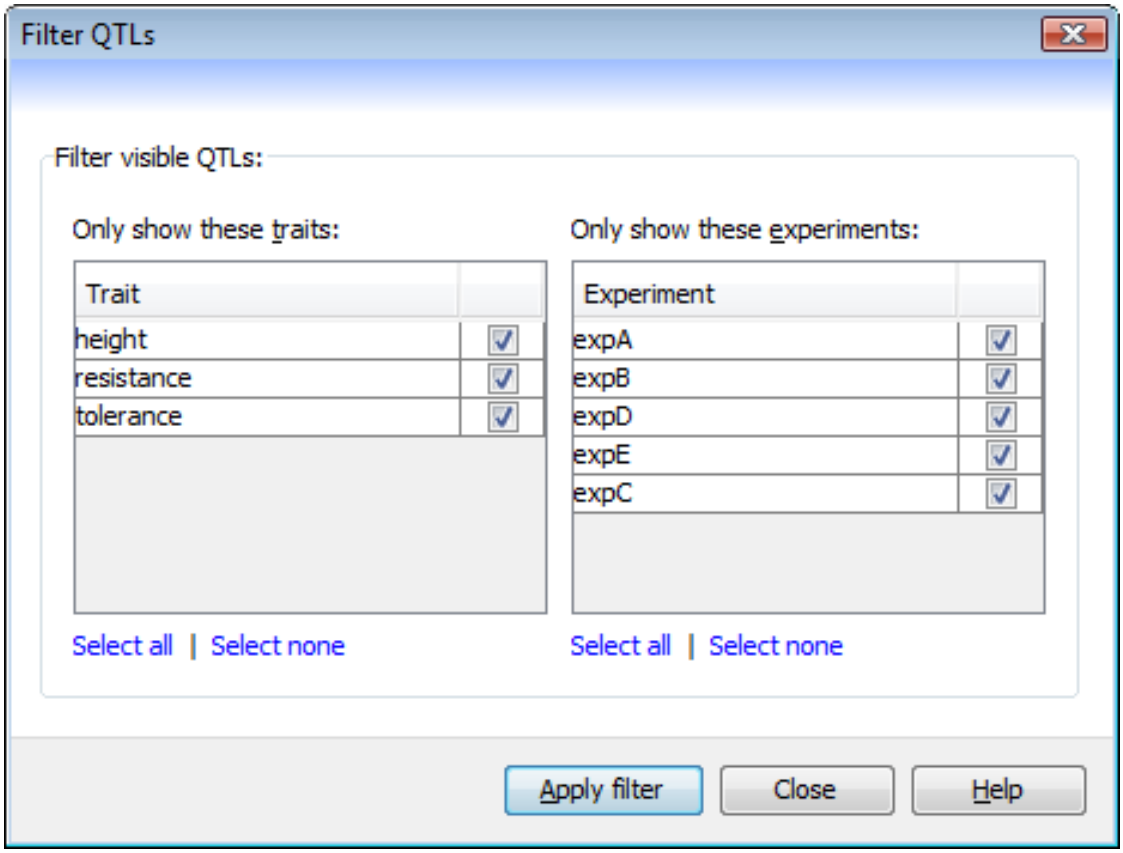
26.5 Filtering visible QTLs

To filter which QTLs are visible, you can either manually select or deselect each QTL using the summary table, or open up the *Filter QTLs* dialog that allows you to perform an automatic filter based on each QTL's associated trait and experiment value.

CHAPTER 27

Filter QTLs

The `Filter QTLs` dialog (`Data->Filter QTLs`) is used to enable or disable the display of QTLs on the map based on a combination of the values assigned to each QTL for its associated trait **and** experiment.



The dialog presents two tables; one showing a list of every trait within the current QTL data set, and another showing a list of every experiment. The QTLs can be filtered by selecting a combination of trait and experiment, so that only

QTLs that match the values of both tables will be left visible.

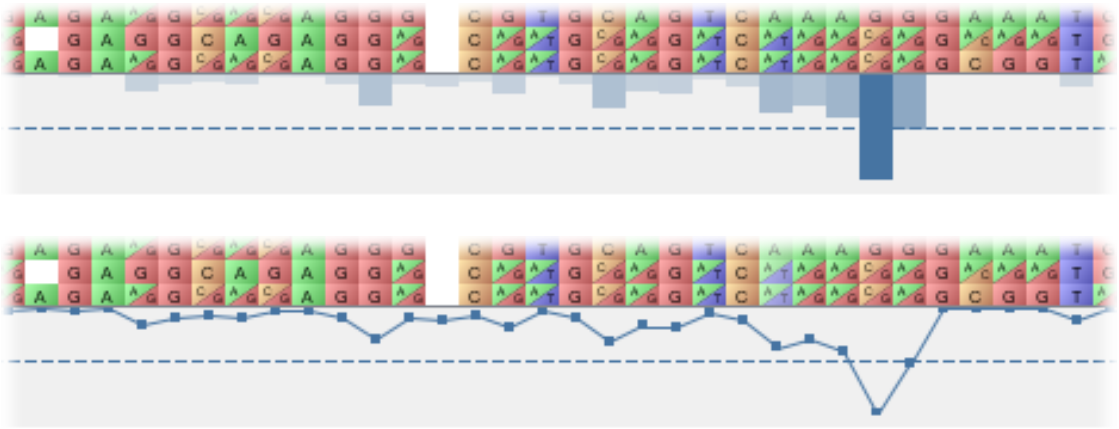
To apply the current filter, click the `Apply filter` button. The effect takes place immeditately and you can check the result in Flapjack without closing the `Filter QTLs` dialog.

Note also, that the checked values within the tables may change dynamically when the filter runs. For example, say you had selected to show only QTLs that were assocaited with the trait **height** across experiments **x1** and **x2**. If **x2** has no QTLs with that trait then its checkbox will be cleared automatically when the filter runs.

CHAPTER 28

Graph Visualization

Flapjack allows you to import "graphs" - which are files that contain one or more values per marker which can be visualized alongside the main display. Flapjack supports two style of graph - histogram and line graph.



28.1 Importing graph data

Import a graph file using the `Import Data` dialog. Note that any graph data that is imported will overwrite any existing graph data held in the current data set, but the file format supports more than one graph anyway.

The structure of a graph file is tab-delimited text. Each line provides details on a marker via three columns in the following format:

```
MarkerName      GraphName      Value
```

For example:

```
11_20479        Height          10.5
11_20480        Height           9.4
11_20479        Resistance     0.234
11_20480        Resistance     0.993
...
```

Each graph can also have an optional significance threshold value associated with it. This is specified by using the reserved "SIGNIFICANCE_THRESHOLD" keyword in the file, for example:

```
SIGNIFICANCE_THRESHOLD    Height     8.0
11_20479                  Height    10.5
...
```

28.2 Interacting with a graph

You can select which graph to display, along with the type of graph (line or histogram) by using the `Select Graph` dialog.

When you mouse over a graph, the status panel will show you:

- the name of the marker currently under the mouse
- the name of the graph
- the value of the marker (and optionally the graph's threshold value in parenthesis)

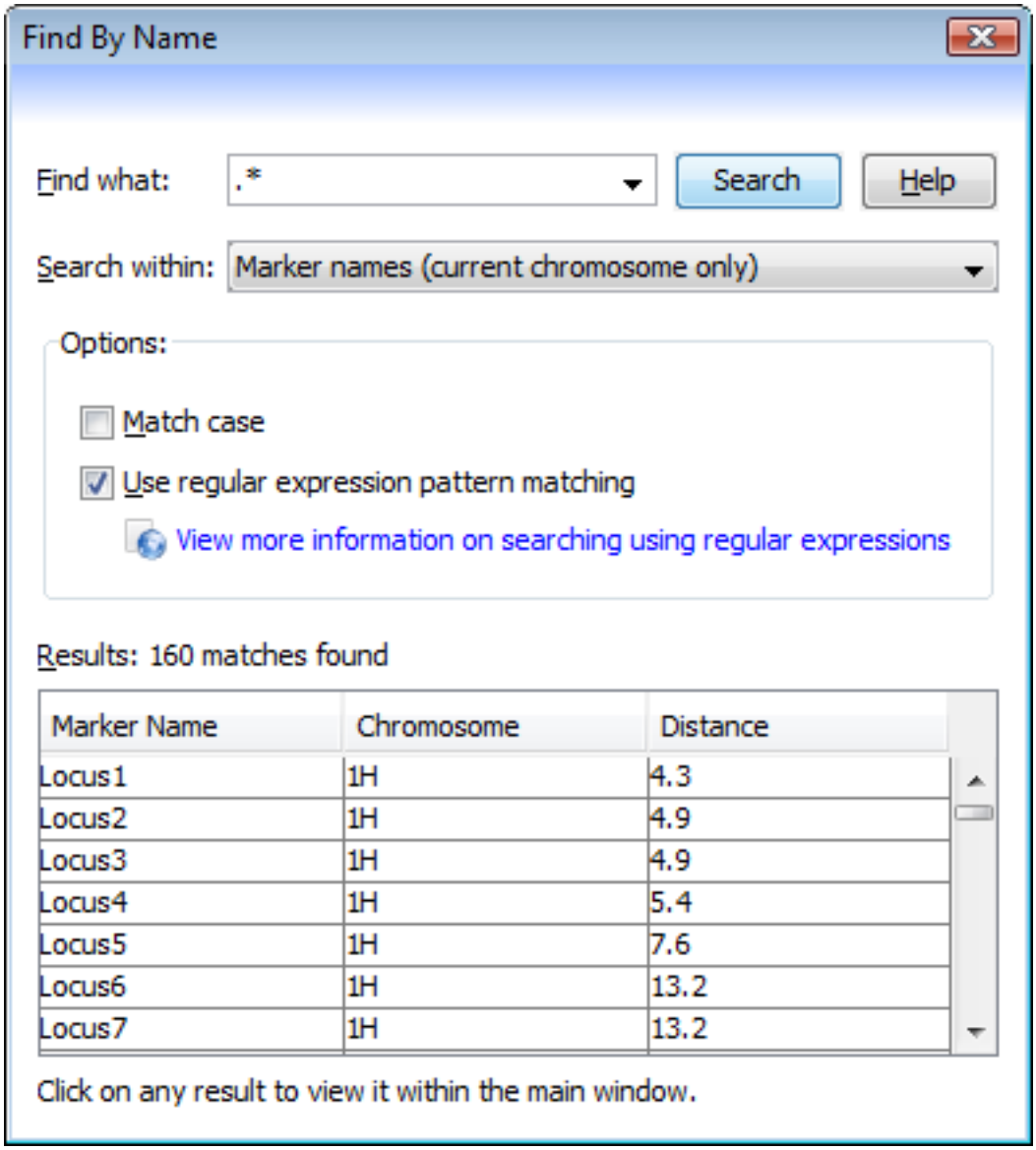
When displaying a histogram, the intensity of the colour is used to represent the value; the stronger the colour, the higher the value.

If a graph has a significance threshold, then this will be displayed on the graph as a dotted line.

CHAPTER 29

Find by Name

The `Find By Name` dialog (`Data->Find by name`) allows you to search for and locate both markers and lines. Searching is possible either via exact matching on a given name or by regular expression matching.



Enter your search parameters into the search box, then click the `Search` button to continue. Any matching results will be displayed in the results table. You can click a result to have Flapjack move the main display to the exact position of the marker or line clicked. It will also be graphically highlighted for a few seconds.

You may also wish to use Flapjack's [bookmarks.shtml bookmark] feature to track results that are of interest without having to search for them again.

29.1 Search options

You can search within three separate areas:

- `Line names` - select this option to search for matching line names.
- `Marker names (current chromosome only)` - select this option to search for matching marker names, with the search limited to markers that are in the currently visible chromosome only.

- `Marker names (across all chromosomes)` - select this option to search for matching marker names across all chromosomes within the current data set.

The dialog also provides two additional options that are available regardless of the search type:

- `Match case` - if this option is checked, then case sensitive matching will be performed. Uncheck the option to ignore case.
- `Use regular expression pattern matching` - if this option is checked, then you can enter a regular expression into the search box.

29.2 Example regular expressions

Here are a few simple regular expressions to help with searching within Flapjack:

- To find all names beginning with the letter 'a' use: `a.*`
- To find all names ending with the letter 'a' use: `.*a`
- To find all names that include the substring 'abc' use `.*abc.*`

For more details on using regular expressions, see the documentation provided at http://java.sun.com/javase/8/docs/api/java/util/regex/Pattern.html.

CHAPTER 30

Sort Lines

30.1 By similarity

[Documentation still to be written]

30.2 By trait

[Documentation still to be written]

30.3 By importing an order

Flapjack allows you to resort a view's ordering of lines by importing a sort order from an external file.

The format of the file is extremely simple; each line of the file should contain the name of a (Flapjack) line, with the ordering being dictated from top to bottom. For example, a file containing the entries:

```
Noelle
Raina
Paloma
Daphne
```

Would resort those four lines in a Flapjack view to be in that order, with Noelle first and Daphne last.

Note that any lines that exist in the view but are not found in the external file will remain in the view, but will be moved to the end/bottom of the new order.

CHAPTER 31

Similarity Matrix Creation

To create a similarity matrix we take two lines and compute a score for the difference between them. We do this for the selected set of lines and across the selected set of markers. The score for the difference between two lines is calculated by comparing each allele of line1 against each allele of line2. If the allele from either line has no value (i.e. missing data) we skip that comparison. If the alleles match, the running score for the line is incremented by 1. If we have a partial match such as in the case of a heterozygous allele like A/T compared with A/G, we increment the running score by 0.5. For every comparison made - except for alleles where either line has missing data - we also increment a count of the number of comparisons carried out. The final value for a line is calculated by taking the running score and dividing it by the number of comparisons made.

Below is an example of the similarity matrix calculation process:

Example lines:

Line 1	A	T		A/T	G	C
Line 2	A	A/T	C	A/C	T	C

Calculations:

Score	1	1.5	1.5	2	2	3
Comparisons	1	2	2	3	4	5

The computed score for Line 1 against Line 2 is 0.6 (3/5).

The resulting similarity matrix is displayed below:

	Line 1	Line 2
Line 1	1	0.6
Line 2	0.6	1

CHAPTER 32

Dendrogram Creation

Dendrograms can be created by querying a basic webservice which we run for the purpose of some data analysis jobs. The webservice runs R jobs on a mini job-scheduling system. Creating a dendrogram in R is just a few simple steps which are outlined below:

- We create a distance matrix from our similarity matrix using the R dist() command using its default arguments.
- We then run a hierarchical cluster analysis on this distance matrix using the R hclust() method passing it the distance matrix created in the previous step as its only parameter.
- Finally we output the dendrogram by creating a png image and using R's plot() passing it the output of hclust and a set of labels which are the line names which were passed to R from Flapjack.

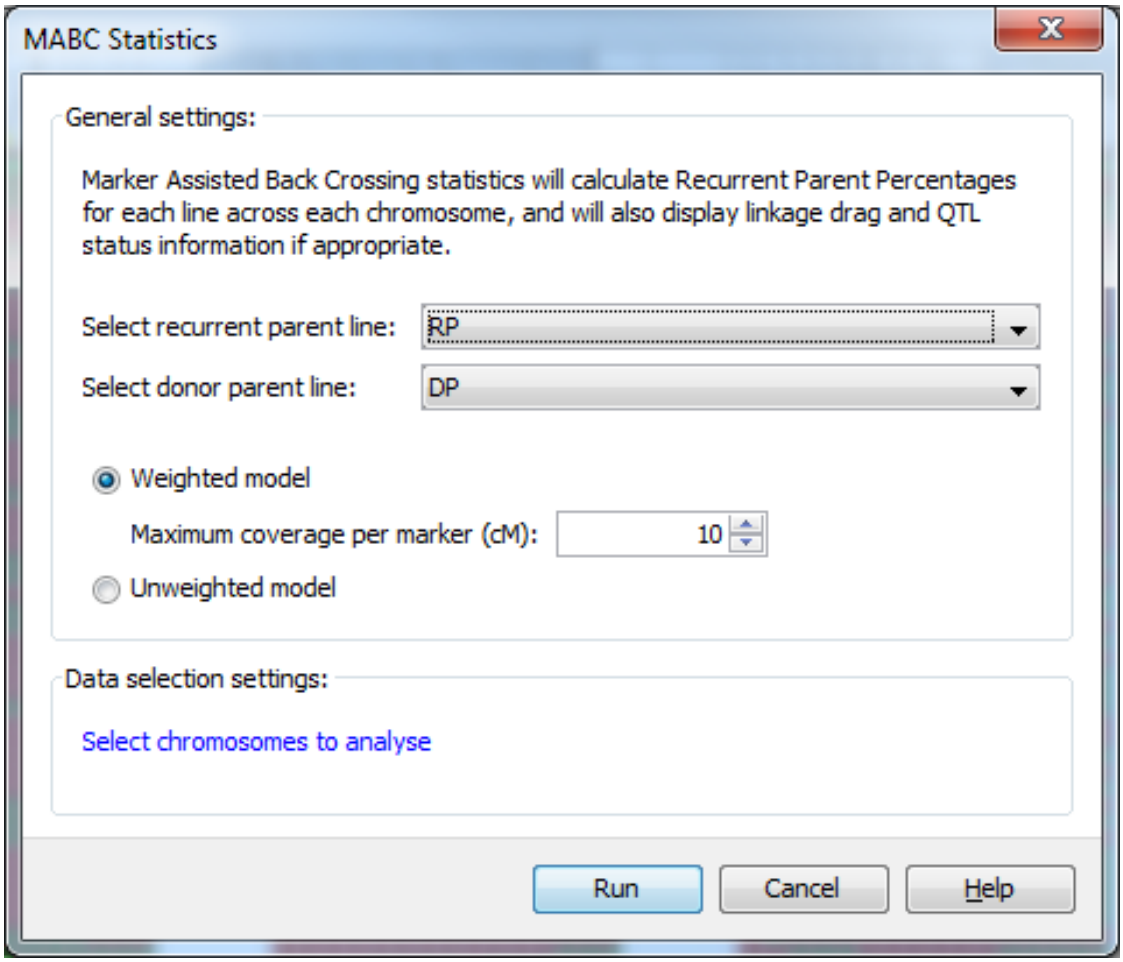
The resulting image is passed back to Flapjack for display.

CHAPTER 33

Marker Assisted Back Crossing

Marker Assisted Back Crossing statistics will calculate Recurrent Parent Percentages for each line across each chromosome, and will also display linkage drag and QTL status information if appropriate.

You can run an MABC analysis by selecting `Analysis->Marker assisted back crossing` from the menubar.



Using the `MABC Statistics` dialog you can select both the recurrent and donor parents with which to run a marker assisted back crossing analysis. You can also choose to use a `Weighted model` which specifies the maximum length of genome a marker can accurately represent in the analysis, or an `Unweighted model` which doesn't take the amount of genome represented by each marker into consideration.

In addition to this help page, you can also read the *Marker Assisted Back Crossing Tutorial*, which runs through the process of running the analysis and viewing the results with a sample dataset.

33.1 MABC QTL format

To generate linkage drag statistics as part of the marker assisted backcrossing analysis, you must provide a QTL file in addition to your genotype and map data. This QTL file differs from the standard Flapjack QTL file as it requires an additional column of data to be able to make use of the data in the analysis.

The format of your QTL file should resemble the example below:

```
# fjFile = QTL
Name  Chromosome  Position  Minimum  Maximum  Trait   Experiment  ``Source``
↪[optional_1] .. [optional_n]
QTL1  1H          10        8        12       Height  Exp1        ``DP``       25.5
↪         high
QTL2  1H          20        19       26       Height  Exp1        ``DP``       34.8
↪         low
```

(continues on next page)

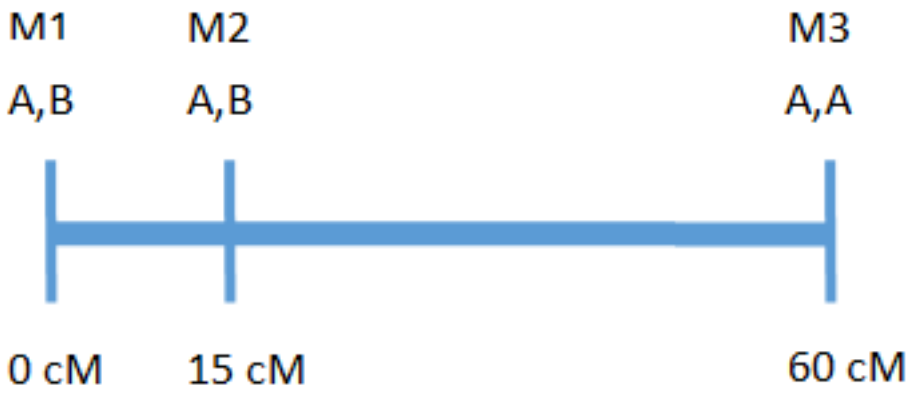
(continued from previous page)

```
QTL3  2H          10        8        13.5     Temp     Exp1        ``DP``        99.2  ␣
↪         low
```

The **Name** to **Experiment** columns are required and must be included and listed in the order shown. After that, each QTL may have zero or more optional columns of numerical or textual data that can be included too. The **Source** column is required for linkage drag statistics to be provided. Flapjack will look for values of DP, and RP, indicating that the alleles we're looking for are coming from either the donor parent, or the recurrent parent.

33.2 Recurrent parent proportion (RPP) calculations

There are two methods for calculating RPP, weighted and unweighted. To illustrate the difference, take the hypothetical chromosome below with three markers:



The above chromosome has a length of 60 cM with 3 markers, where A is the recurrent parent genotype and B is the donor parent genotype.

33.3 Weighted calculation

In the weighted calculation the user specifies the maximum length of genome a marker can accurately represent. For example, if 20 cM is chosen, each marker can accurately represent 10 cM on either side of the marker for a total of 20 cM.

Weights for each marker are determined by taking the minimum of half the distance between the marker and the adjacent marker or half the maximum coverage. Taking the example above:

The distance between M1 and M2 = 15 cM - 0 cM = 15 cM

The distance between M2 and M3 = 60 cM - 15 cM = 45 cM

$$weight_{M1} = \min\left(\frac{15}{2}, \frac{20}{2}\right) = 7.5$$
$$weight_{M2} = \min\left(\frac{15}{2}, \frac{20}{2}\right) + \min\left(\frac{45}{2}, \frac{20}{2}\right) = 17.5$$
$$weight_{M3} = \min\left(\frac{45}{2}, \frac{20}{2}\right) = 10$$

$$RPP_{weighted} = \frac{.5(7.5) + .5(17.5) + 10}{35} = .6428571$$

For the calculation of RPP the weights for markers 1 and two are multiplied by .5 because they are heterozygous and marker 3 is multiplied by 1 because it is homozygous for the recurrent parent. The sum of these values is divided by the total coverage (sum of the weights for each marker).

33.4 Unweighted calculation

In the unweighted calculation the amount of genome being represented by each marker is not taken into consideration. For the above example the RPP would be:

$$RPP_{unweighted} = \frac{.5 + .5 + 1}{3} = .667$$

In this example the RPP calculations do differ as the weights for each marker are not the same. The degree to which the calculations will differ depends on how evenly markers are spaced across the genome. It should be noted that, if no map is provided, the weighted calculation will give effectively the same results as the unweighted calculation.

33.5 Genome Coverage

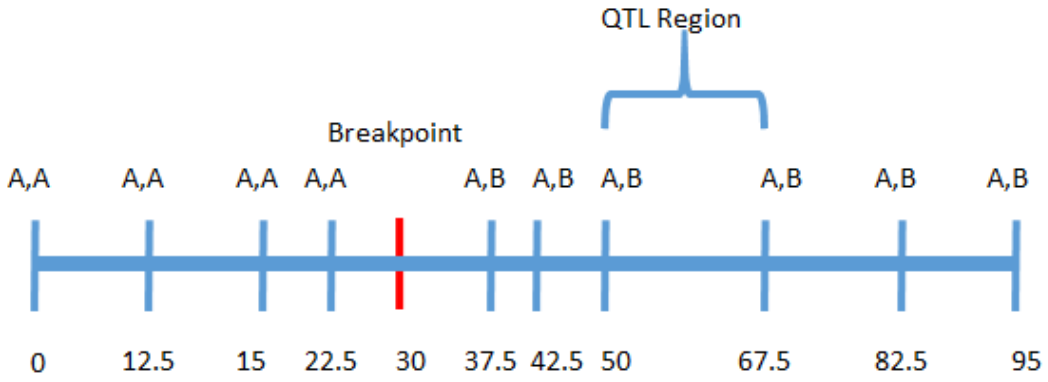
Genome coverage only applies to the weighted RPP calculation. It provides a measure of how much of the genome is covered given the maximum coverage allowed for each marker. In the above example, where maximum length of genome a marker can cover is 20 cM, the genome coverage would be:

$$GenomeCoverage = \frac{35}{60} = .583$$

33.6 Linkage Drag(LD) Calculations

LD is calculated as the length of donor genome linked to the QTL region. This is done by looking for the first recombination on each side of the QTL region and calculating the genome length between the QTL and each breakpoint.

Below is an example to illustrate the calculation:



In this example a break point occurred at approximately 30 cM on the left side of the QTL region and no break point occurred on the right side. In this case linkage drag is calculated by summing the distance between the breakpoint and the left boundary of the QTL region and the distance between the end of the chromosome and the right boundary of the QTL region:

LD = (50 - 30) + (95 - 67.5) = 47.5

CHAPTER 34

Marker Assisted Back Crossing Tutorial

We will assume that you are using the `gobii-test` dataset, which contains simulated data in an A/B format. MABC analysis in Flapjack is not limited to this style of data and works with standard nucleotide data as well. Feel free to follow along either using the dataset as provided, or with your own data.

34.1 Tutorial data

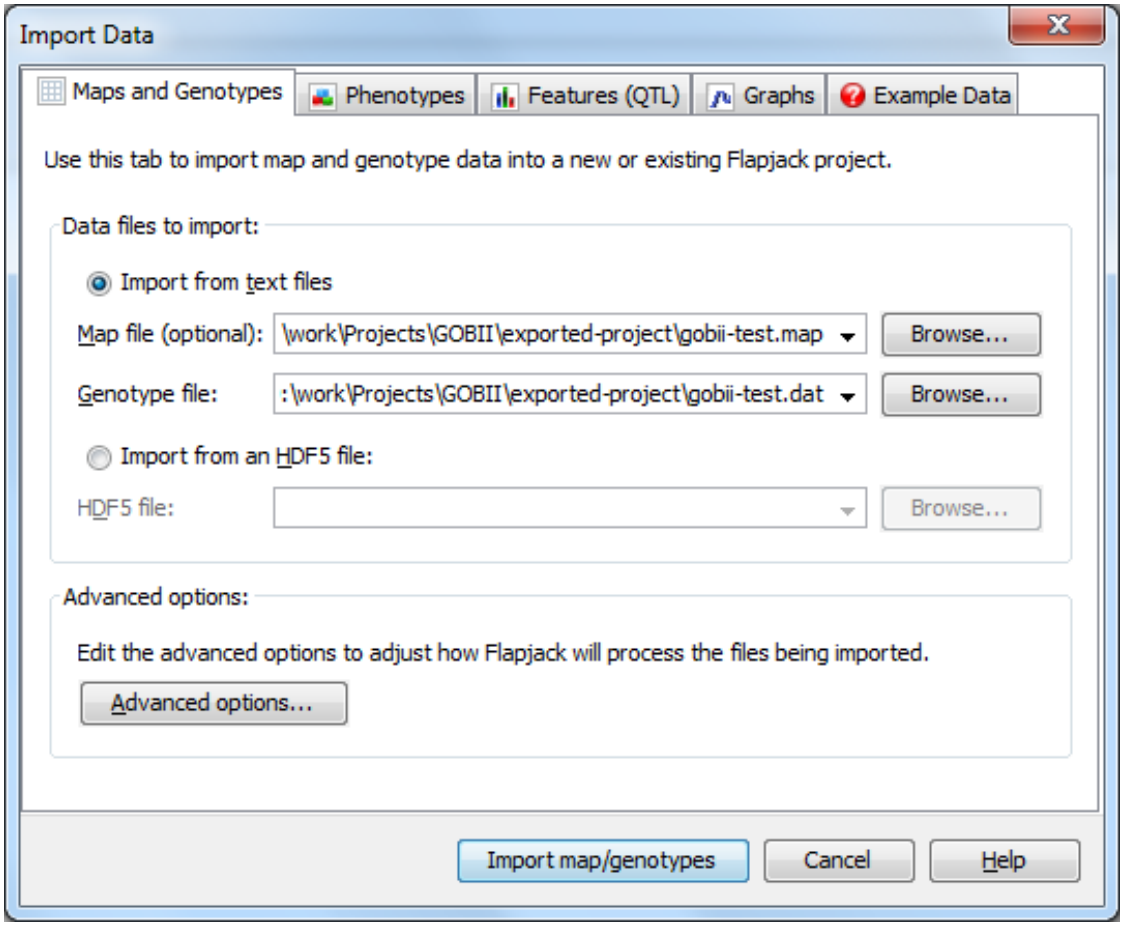
In this tutorial you'll be working with three files, a map file (gobii-test.map) which contains a set of 61 markers across 4 chromosomes, a genotype file (gobii-test.dat) which contains data on a set of 202 lines (one recurrent parent, one donor parent and 200 crosses), and a QTL file (gobii-test.qtl) which contains data on 2 QTL.

Download the data (Right click->Save Link As...):

- gobii-test.map
- gobii-test.dat
- gobii-test.qtl

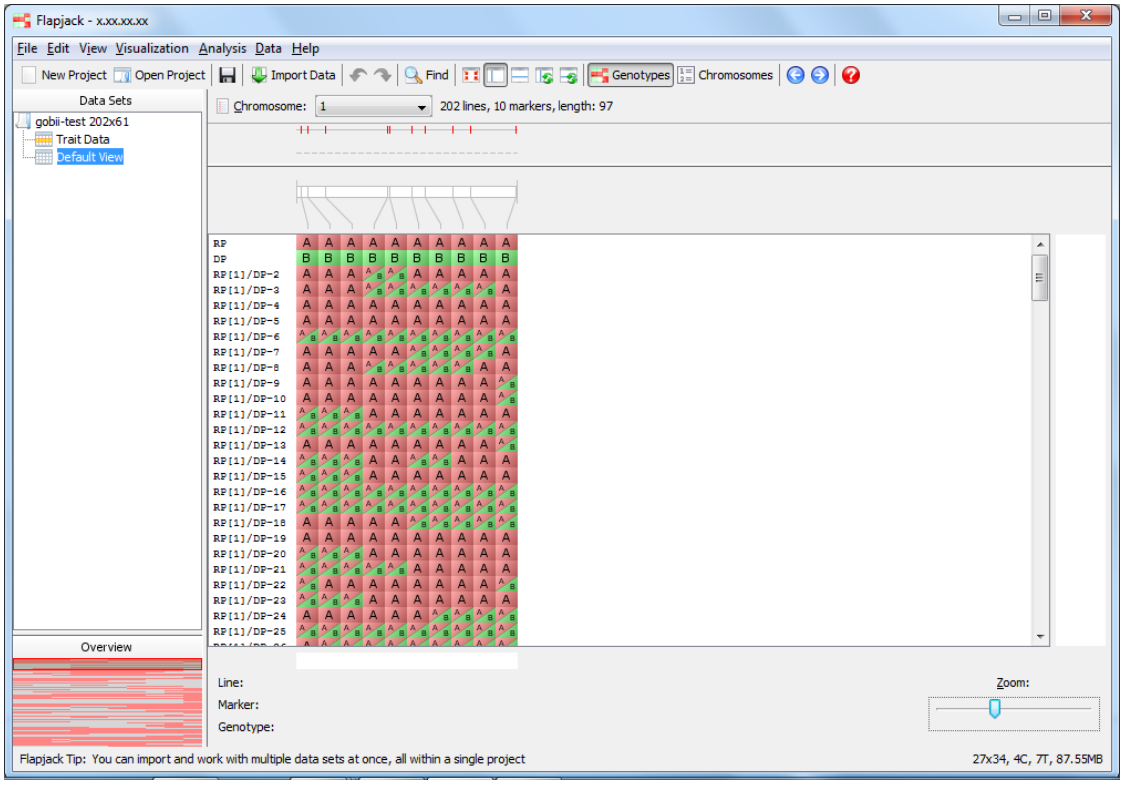
34.2 Importing data

To import data into Flapjack click the `Import Data` button on the toolbar. In the `Import Data` dialog, select the `Maps and Genotypes` tab, then click `Browse` to navigate to and select the map file you wish to import (gobii-test.map), and then do the same for the genotype file you wish to import (gobii-test.dat).

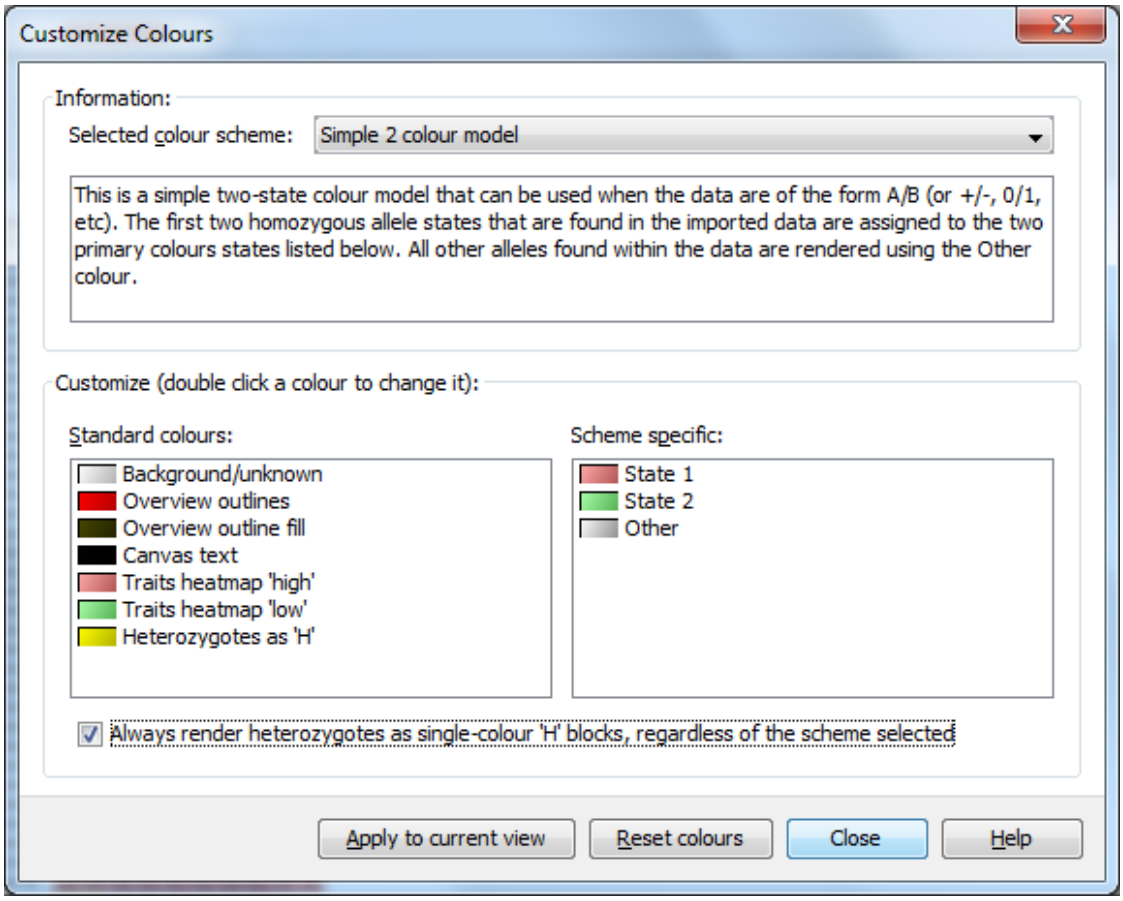


Click the `Advanced options` button to open the `Advanced Data Import Options` dialog and ensure that the `Duplicate all markers onto a single "All Chromosomes" chromosome for side by side viewing` option is selected, then click OK. Finally click the `Import map/genotypes` button to load the data.

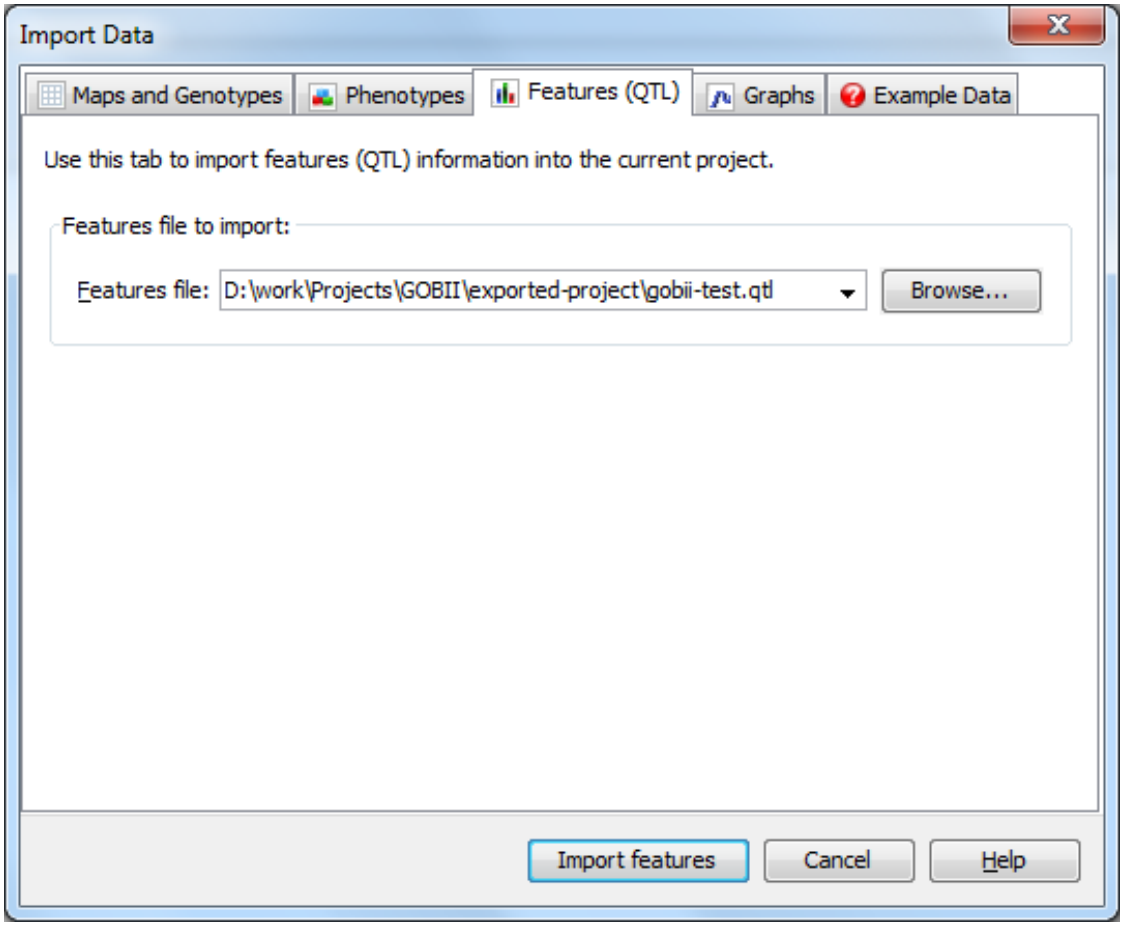
You should now be viewing the Default View on the 1st chromosome of your dataset. This is the main type of visualization in Flapjack and comprises a graphical genotype view of the imported data. Each square represents an allele, found at the cross-section of a line and a marker, lines or varieties can be found down the left hand side of the display and markers can be found along the top (as well as a visualization of the map associated with the markers). Alleles can be either homozygous (in this dataset either an A or a B) or heterozygous, with the latter being rendered as split diagonal blocks, or A/B in the dataset.



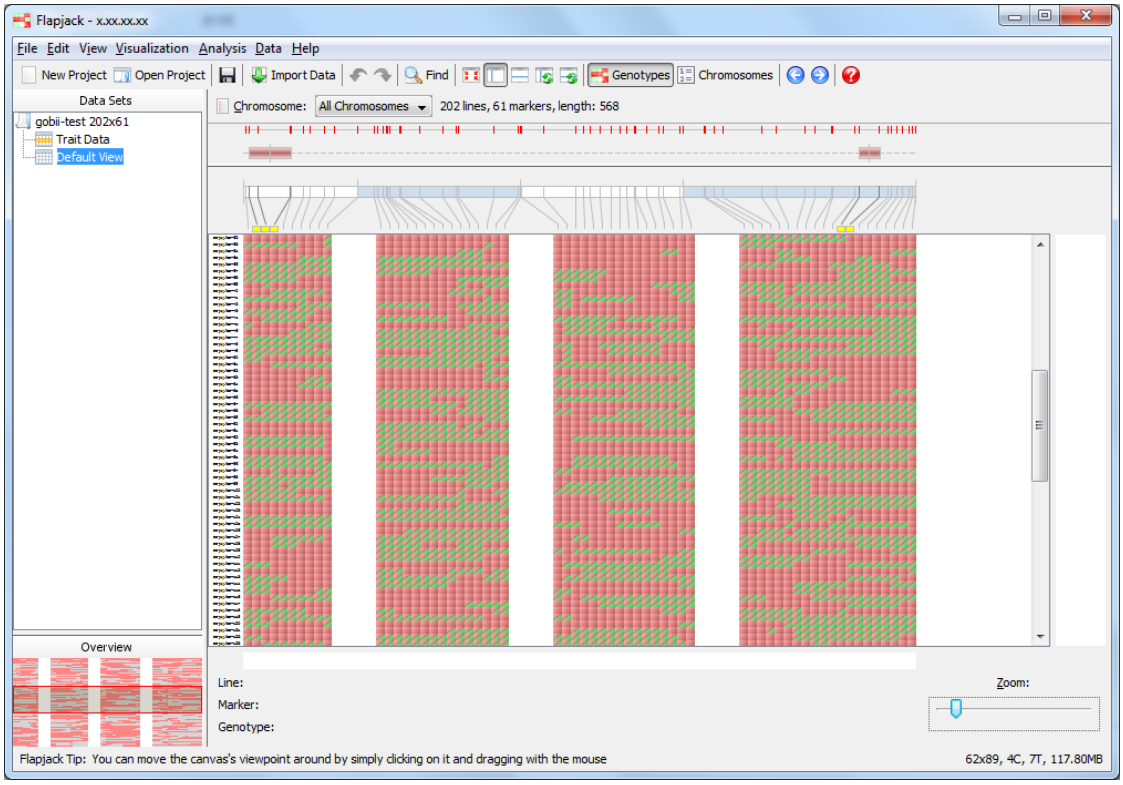
If you're used to viewing heterozygous alleles as simply H instead of a diagonal split view, select `Visualization->Colour scheme->Customize...` from the menubar to open the `Customize Colours` dialog. From there you can select `Always render heterozygotes as single-colour "H" blocks, regardless of the scheme selected` option and click `Apply to current view` - your heterozygotes should now be rendering as a single-colour 'H' block.



Next, click the `Import Data` button again and this time select the `Features (QTL)` tab. Browse for and select the QTL file associated with your dataset (gobii-test.qtl), then click `Import features`. You should now be viewing the Trait Data view on the Quantitative Trait Loci (QTLs) tab which gives a summary view of the imported QTLs.



Click Default View from the navigation tree to return to the visualization of genotype data. Select the All Chromosomes view from the Chromosome dropdown menu underneath the toolbar and use the zoom control at the bottom of the display to zoom out until you can see all of the data. You should now see a representation of all four chromosomes in the dataset, including the two QTL you imported, one on chromosome 1 and the other on chromosome 4.



The QTL in this case are the red rectangles above the maps on the All Chromosomes view. The markers under a QTL are highlighted by yellow boxes just above the genotype display. Now that you have loaded and viewed all of the data required to run a marker assisted backcrossing analysis, it's time to explore the data and filter out any undesirable markers before running the analysis.

34.3 Exploring and filtering data

To navigate the data select `Edit->Navigation mode` from the menubar, then click and drag the main display around with your mouse to examine all of your data. You can also use the scrollbars to navigate the data, as well as clicking and dragging on the overview in the bottom left. To zoom you can either use the zoom slider in the bottom left of the display, double click on the main display to zoom in, or use ctrl /cmd and the mouse's scroll-wheel to zoom in and out.

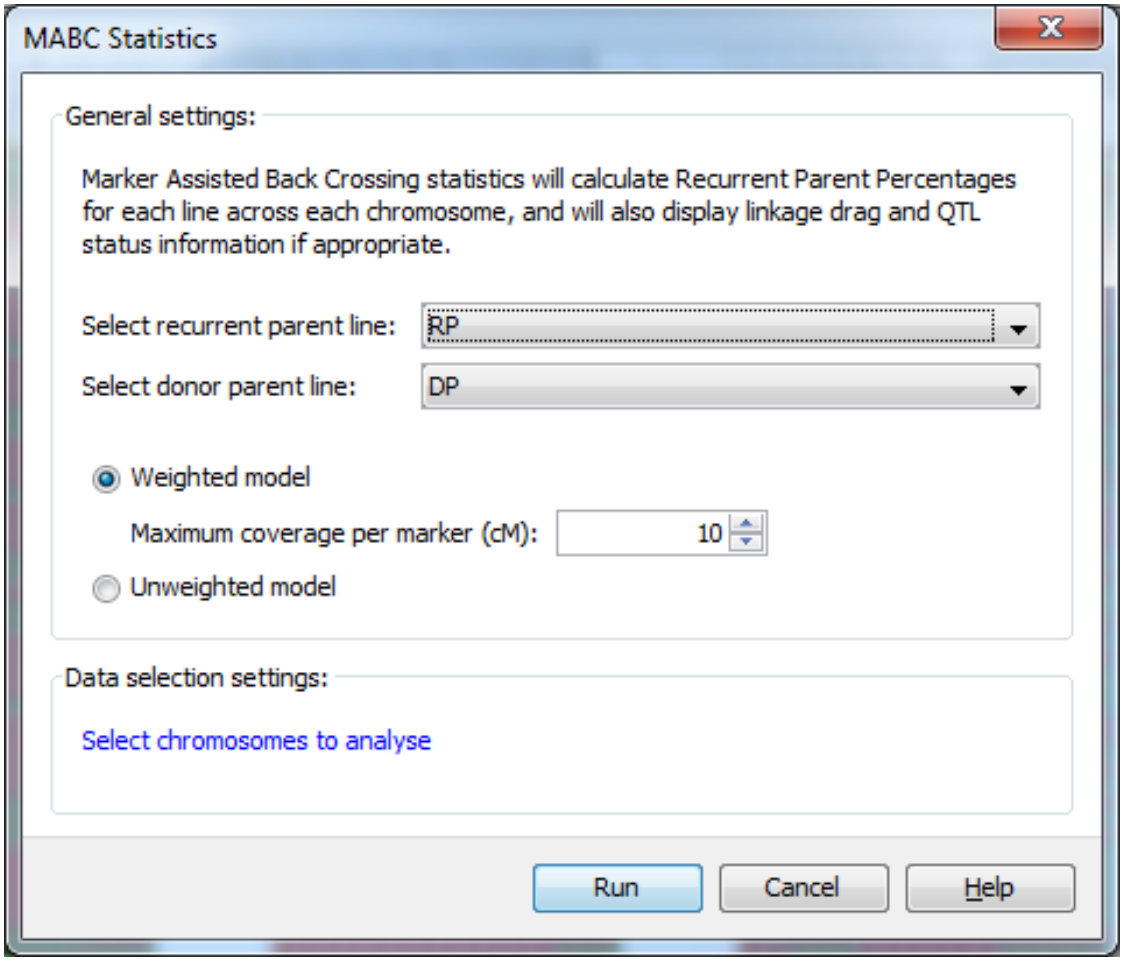
Before you run a marker assisted backcrossing analysis, you may want to filter out markers with lots of missing data, or monomorphic markers. Select `Edit->Filter markers->Missing markers...` to open the `Filter Missing Markers` dialog. Choose the percentage of missing data in a marker required to filter out a marker, then use the `Filter` button to perform the operation. You should see a message detailing the number of markers that were filtered out of the dataset as part of the operation. (Note that in the case of the sample dataset none will be filtered out as no markers have a high percentage of missing data in the simulated dataset.)

The procedure for filtering monomorphic markers is very similar - select `Edit->Filter markers->Monomorphic markers`, and simply apply the filter once the dialog has opened.

It's also possible to filter out markers which have missing (or heterozygous) data in a given line. To do this, right click on a line of interest and and select `Filter markers->Missing markers (by line)` to open the `Filter Missing Markers by Line` dialog. The line should be pre-selected in the dropdown menu. Click `Filter` and Flapjack will remove any markers which have missing data for this line from the display.

34.4 Running the analysis

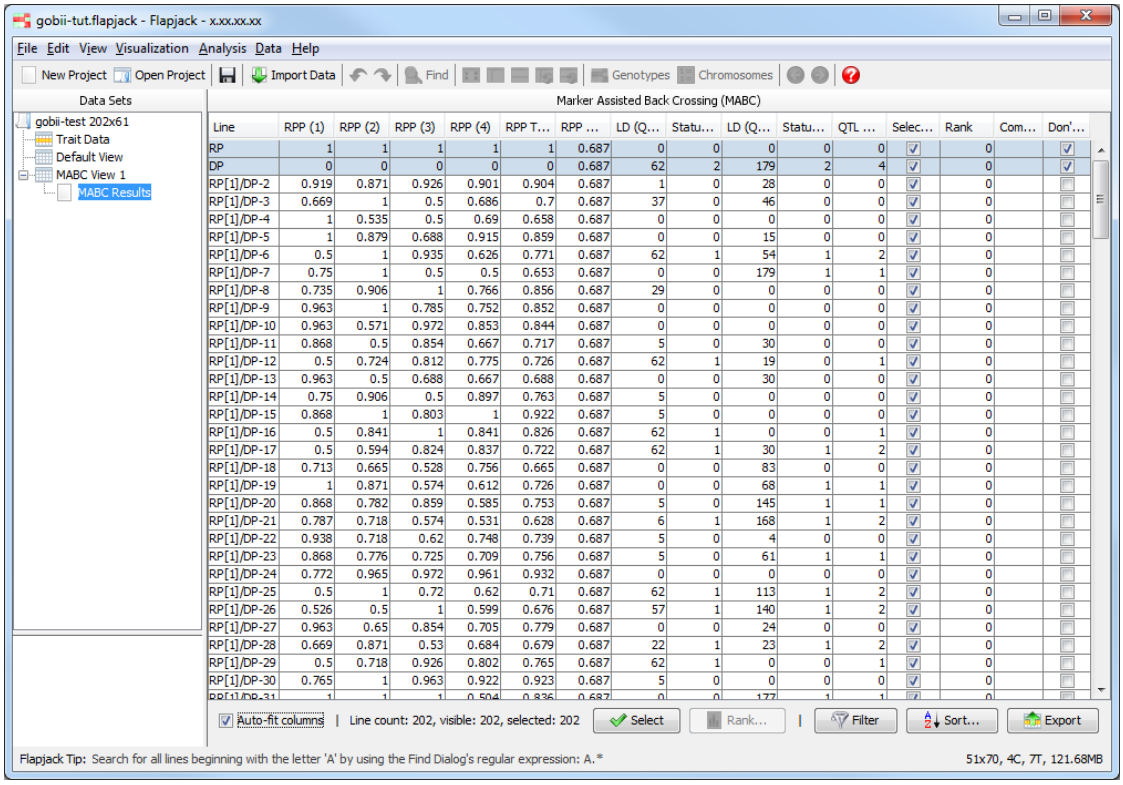
Select `Analysis->Marker assisted backcrossing` to open the `MABC Statistics` dialog. Select the recurrent parent line for your data from the first drop down list and the donor parent line from the second drop down list. The drop down lists automatically select the first and second lines of the genotype input file, so if as in the tutorial data set, your recurrent and donor parent are on the first and second lines of the input file they will be automatically selected. Now select `Weighted model` if it is not already selected, and set the `Maximum coverage per marker` to 10. This means we've specified that each marker can accurately represent 10cM (5 cM either side of the marker) of the genome.



Click `Run` and Flapjack will run the marker assisted backcrossing statistics on your data.

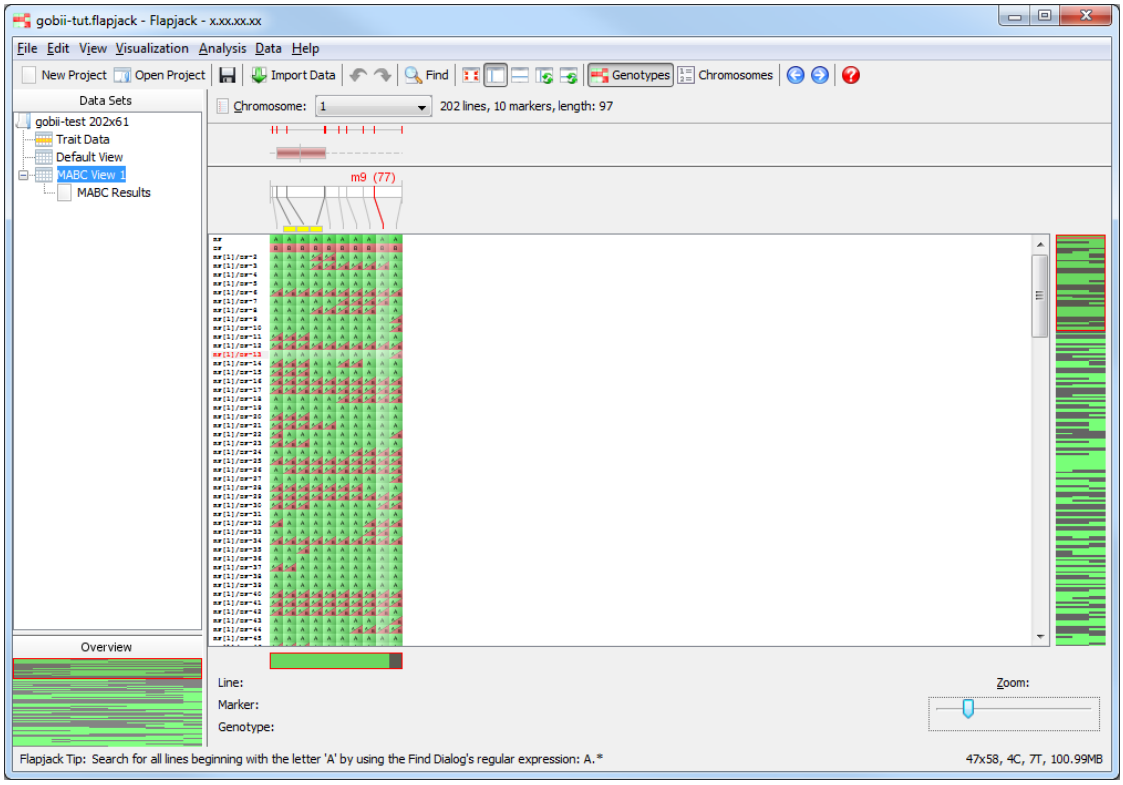
34.5 Viewing the analysis results

Once the analysis has completed you should see a table of results. The results table will contain the lines that you included in the analysis, for each line it will have values for the RPP (Recurrent Parent Proportion) for each chromosome, as well as an RPP total value which measures RPP across all chromsomes, an RPP coverage value, then a linkage drag and status for each QTL, an overall QTL allele count, as well as a selected state, rank and comment. The final column - Don't Sort / Filter - allows you to mark lines that you don't want table sorts and filters to apply to. By default, Flapjack sets both the recurrent and donor parent to neither sort, nor filter. This has the effect of keeping them in the display and always at the top of the table of data.



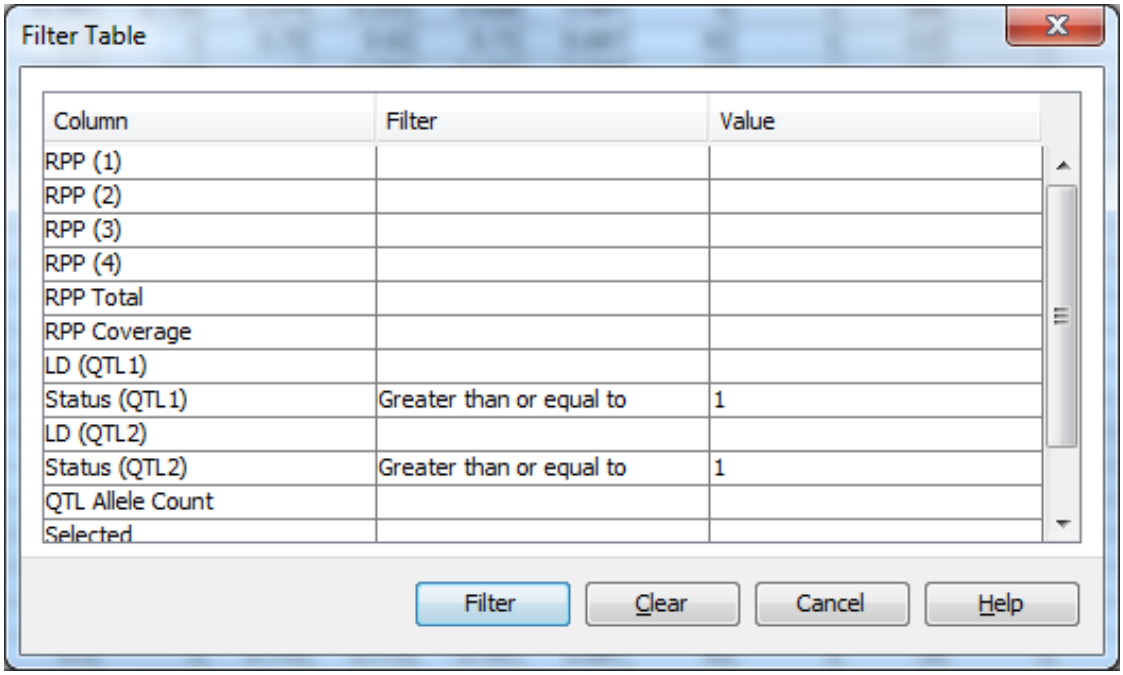
Line	RPP (1)	RPP (2)	RPP (3)	RPP (4)	RPP T…	RPP …	LD (Q…	Statu…	LD (Q…	Statu…	QTL …	Selec…	Rank	Com…	Don'…
RP	1	1	1	1	1	0.687	0	0	0	0	0	☑	0		☑
DP	0	0	0	0	0	0.687	62	2	179	2	4	☑	0		☑
RP[1]/DP-2	0.919	0.871	0.926	0.901	0.904	0.687	1	0	28	0	0	☑	0		☐
RP[1]/DP-3	0.669	1	0.5	0.686	0.7	0.687	37	0	46	0	0	☑	0		☐
RP[1]/DP-4	1	0.535	0.5	0.69	0.658	0.687	0	0	0	0	0	☑	0		☐
RP[1]/DP-5	1	0.879	0.688	0.915	0.859	0.687	0	0	15	0	0	☑	0		☐
RP[1]/DP-6	0.5	1	0.935	0.626	0.771	0.687	62	1	54	1	2	☑	0		☐
RP[1]/DP-7	0.75	1	0.5	0.5	0.653	0.687	0	0	179	1	1	☑	0		☐
RP[1]/DP-8	0.735	0.906	1	0.766	0.856	0.687	29	0	0	0	0	☑	0		☐
RP[1]/DP-9	0.963	1	0.785	0.752	0.852	0.687	0	0	0	0	0	☑	0		☐
RP[1]/DP-10	0.963	0.571	0.972	0.853	0.844	0.687	0	0	0	0	0	☑	0		☐
RP[1]/DP-11	0.868	0.5	0.854	0.667	0.717	0.687	5	0	30	0	0	☑	0		☐
RP[1]/DP-12	0.5	0.724	0.812	0.775	0.726	0.687	62	1	19	0	1	☑	0		☐
RP[1]/DP-13	0.963	0.5	0.688	0.667	0.688	0.687	0	0	30	0	0	☑	0		☐
RP[1]/DP-14	0.75	0.906	0.5	0.897	0.763	0.687	5	0	0	0	0	☑	0		☐
RP[1]/DP-15	0.868	1	0.803	1	0.922	0.687	5	0	0	0	0	☑	0		☐
RP[1]/DP-16	0.5	0.841	1	0.841	0.826	0.687	62	1	0	0	1	☑	0		☐
RP[1]/DP-17	0.5	0.594	0.824	0.837	0.722	0.687	62	1	30	1	2	☑	0		☐
RP[1]/DP-18	0.713	0.665	0.528	0.756	0.665	0.687	0	0	83	0	0	☑	0		☐
RP[1]/DP-19	1	0.871	0.574	0.612	0.726	0.687	0	0	68	1	1	☑	0		☐
RP[1]/DP-20	0.868	0.782	0.859	0.585	0.753	0.687	5	0	145	1	1	☑	0		☐
RP[1]/DP-21	0.787	0.718	0.574	0.531	0.628	0.687	6	1	168	1	2	☑	0		☐
RP[1]/DP-22	0.938	0.718	0.62	0.748	0.739	0.687	5	0	4	0	0	☑	0		☐
RP[1]/DP-23	0.868	0.776	0.725	0.709	0.756	0.687	5	0	61	1	1	☑	0		☐
RP[1]/DP-24	0.772	0.965	0.972	0.961	0.932	0.687	0	0	0	0	0	☑	0		☐
RP[1]/DP-25	0.5	1	0.72	0.62	0.71	0.687	62	1	113	1	2	☑	0		☐
RP[1]/DP-26	0.526	0.5	1	0.599	0.676	0.687	57	1	140	1	2	☑	0		☐
RP[1]/DP-27	0.963	0.65	0.854	0.705	0.779	0.687	0	0	24	0	0	☑	0		☐
RP[1]/DP-28	0.669	0.871	0.53	0.684	0.679	0.687	22	1	23	1	2	☑	0		☐
RP[1]/DP-29	0.5	0.718	0.926	0.802	0.765	0.687	62	1	0	0	1	☑	0		☐
RP[1]/DP-30	0.765	1	0.963	0.922	0.923	0.687	5	0	0	0	0	☑	0		☐

You should see that not only has Flapjack generated this `MABC Results` view, but it has linked this to a new view called `MABC View 1`. Click `MABC View 1` to view it and you should see that it's a clone of the `Default View`, but has the `By similarity to line (2 colour)` colour scheme applied. This colour scheme colours a reference line all green, all other lines have their alleles coloured either green, if they match the reference line, or red, if they don't match the reference line. In this particular case the alleles will be coloured relative to the recurrent parent line you chose when you ran the marker assisted backcrossing analysis. This view `MABC View 1` is linked to the table in the `MABC Results` view. That means moving lines, sorting lines, selecting lines and hiding lines on `MABC View 1` does the same in the linked `MABC Results` view, and sorting lines, selecting lines and filtering lines in the `MABC Results` view does the same in the linked view `MABC View 1`.

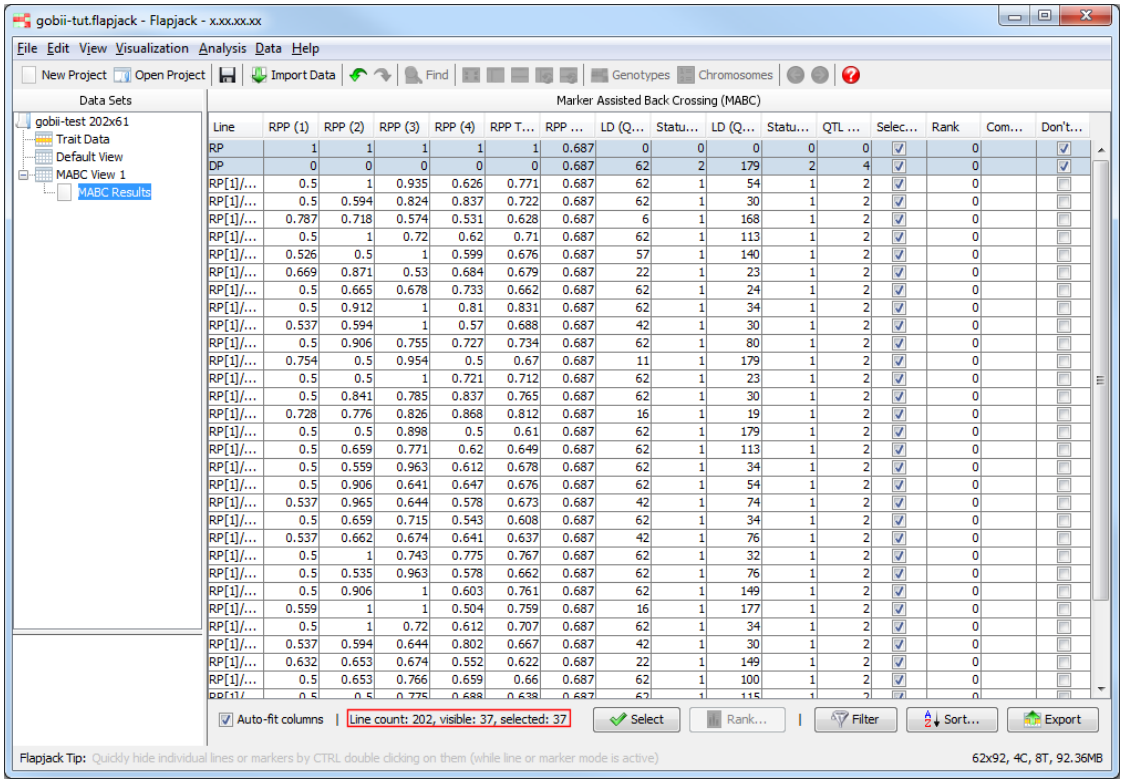


34.6 Filtering the results

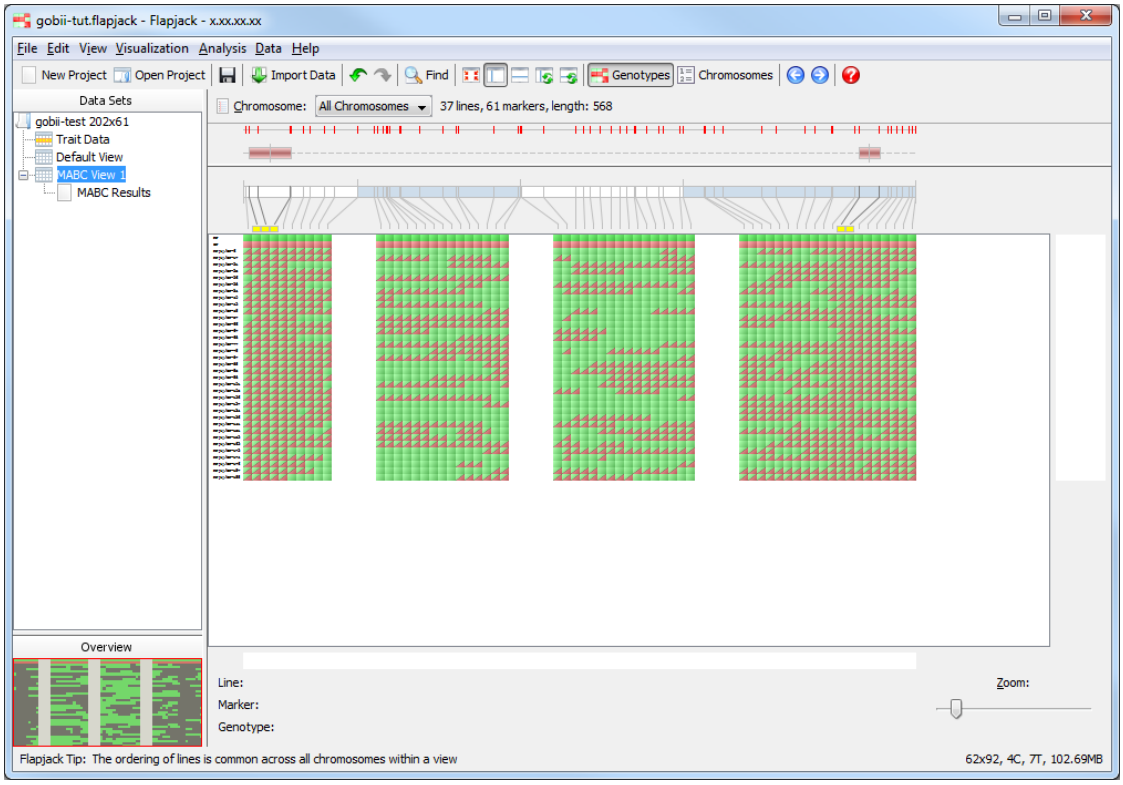
Click on `MABC Results` to return to the results view. Next click `Filter->Filter` to open the `Filter Table` dialog. You should see a table with a list of columns from the table on which you can filter. Click on the filter column for the row called `Status (QTL1)` and select `Greater than or equal to` from the drop down list. Next enter a value of `1` in the adjacent `Value` column. Do the same for `Status (QTL2)`, then click `Filter`.



You should see that the results table has filtered out lines which didn't match the filter criteria. In fact only 37 lines matched the filter criteria.

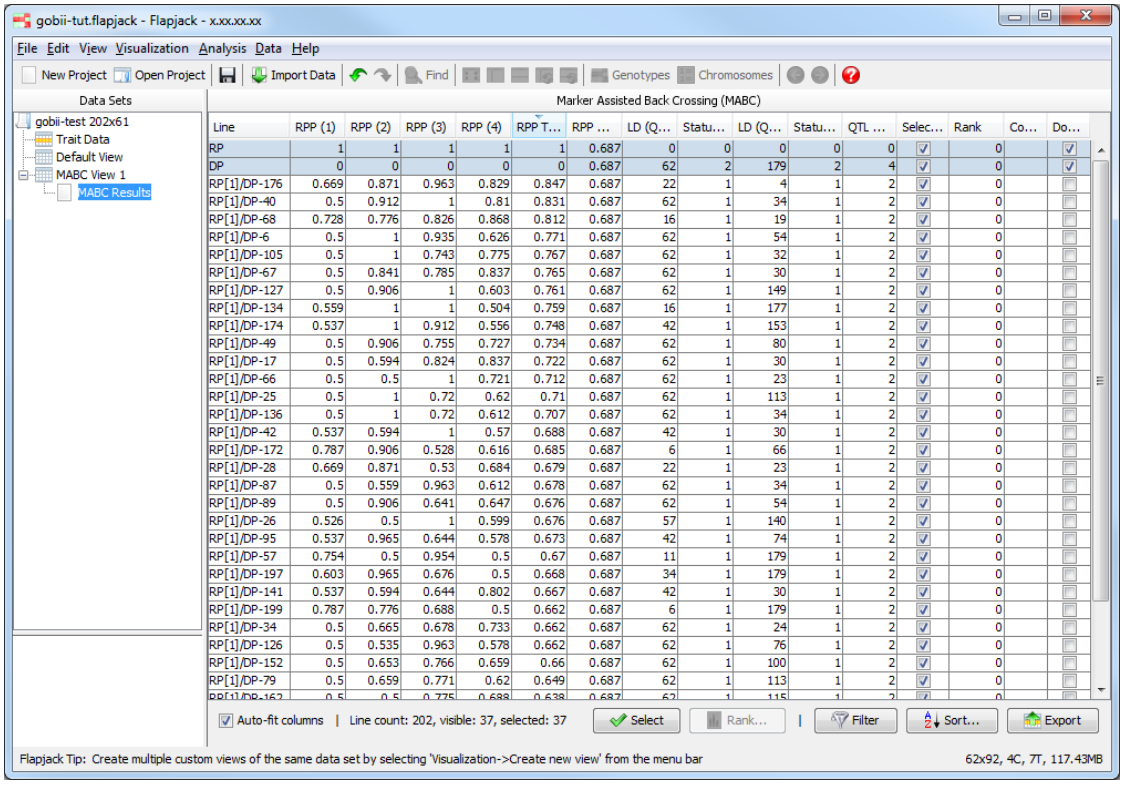


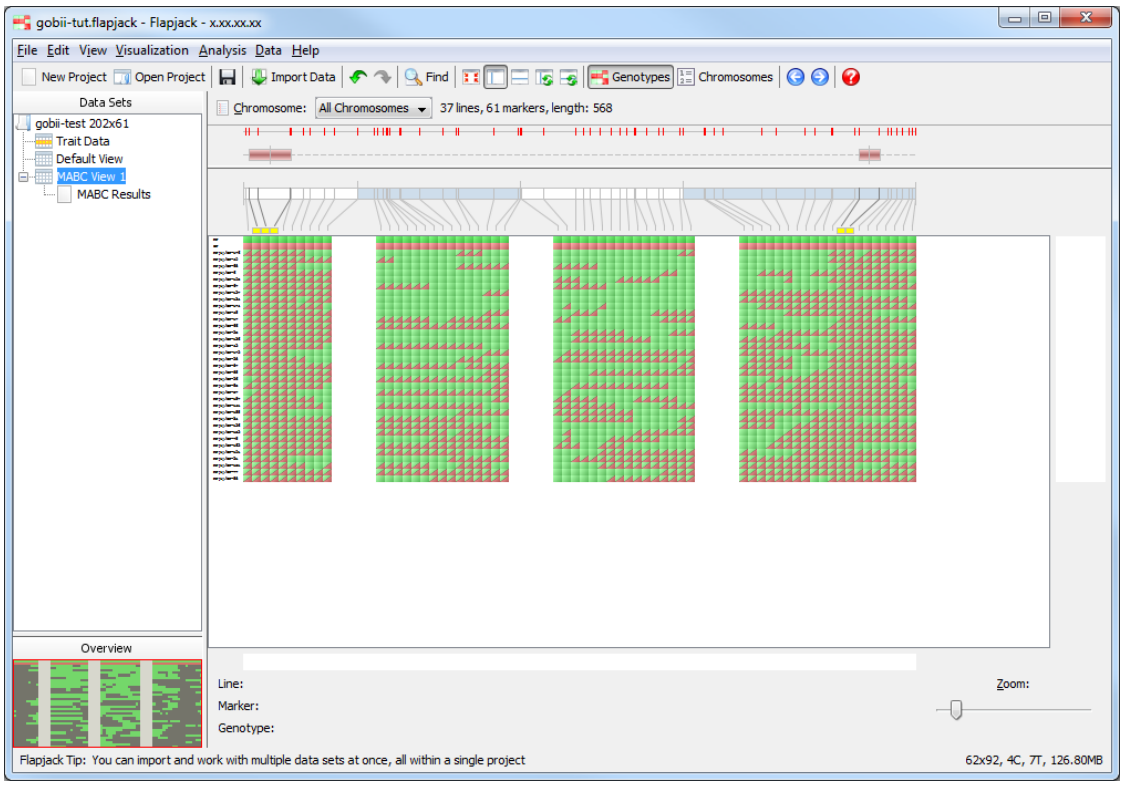
Click on `MABC View 1` if you want to see what the lines which have been kept look like visually.



34.7 Sorting the results

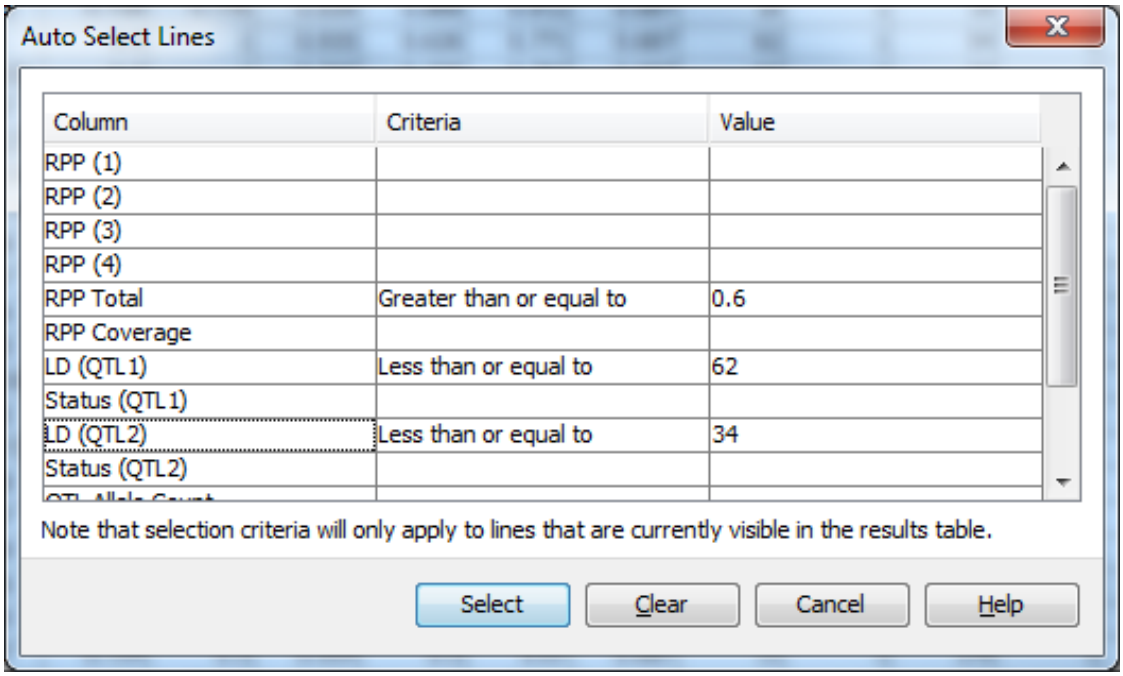
Return to the results table and select `Sort` to use the `Advanced Sort` dialog. Click `Add sort level` twice to add two more entires to the table. Click the first entry in the `Column` column and select `RPP Total` from the drop down list that appears. For the next two entries select LD (QTL1) and LD (QTL) respectively. Finally for the last two entries change their `Order` to be `Smallest to largest`. Click `Sort` to sort the data. You should see that the data in the table and in the genotypes view has been sorted according to your criteria.

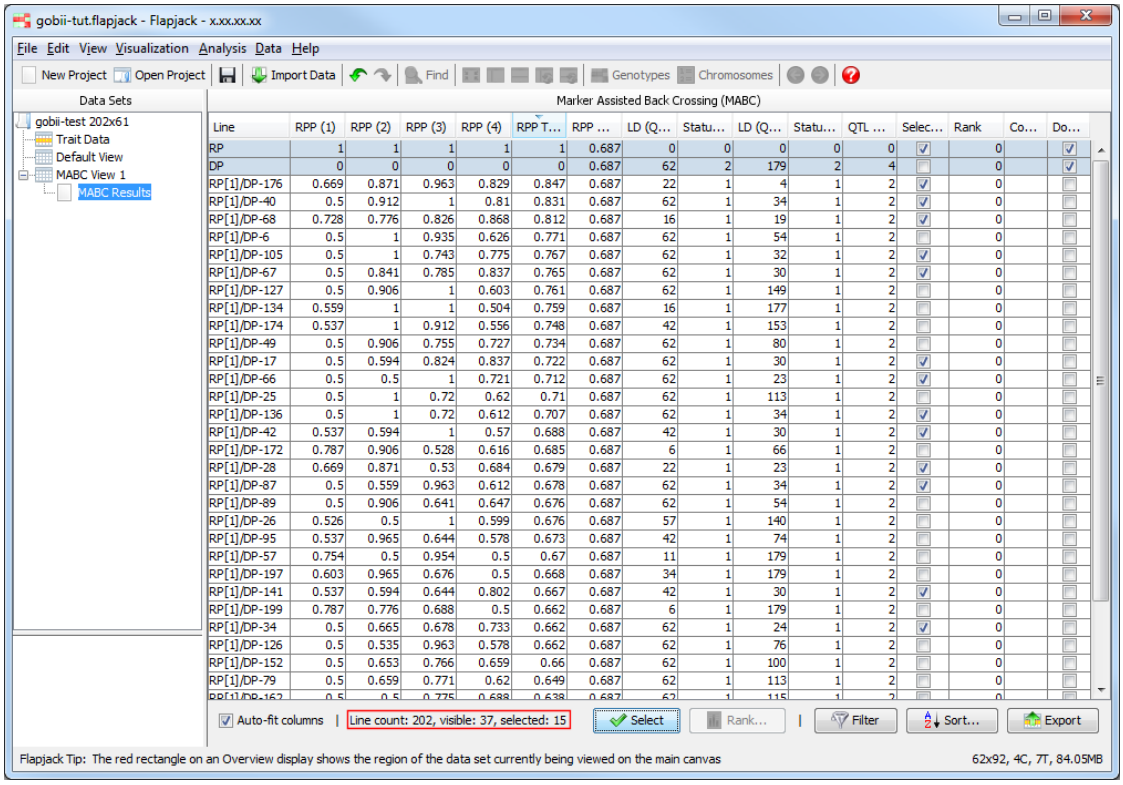

gobii-tut.flapjack - Flapjack - x.xx.xx.xx
File Edit View Visualization Analysis Data Help
New Project Open Project Import Data Find Genotypes Chromosomes
Data Sets
gobii-test 202x61
Trait Data
Default View
MABC View 1
MABC Results
Marker Assisted Back Crossing (MABC)
Line RPP (1) RPP (2) RPP (3) RPP (4) RPP T... RPP ... LD (Q... Statu... LD (Q... Statu... QTL ... Selec... Rank Co... Do...
RP 1 1 1 1 1 0.687 0 0 0 0 0
DP 0 0 0 0 0 0.687 62 2 179 2 4 0
RP[1]/DP-176 0.669 0.871 0.963 0.829 0.847 0.687 22 1 4 1 2 0
RP[1]/DP-40 0.5 0.912 1 0.81 0.831 0.687 62 1 34 1 2 0
RP[1]/DP-68 0.728 0.776 0.826 0.868 0.812 0.687 16 1 19 1 2 0
RP[1]/DP-6 0.5 1 0.935 0.626 0.771 0.687 62 1 54 1 2 0
RP[1]/DP-105 0.5 1 0.743 0.775 0.767 0.687 62 1 32 1 2 0
RP[1]/DP-67 0.5 0.841 0.785 0.837 0.765 0.687 62 1 30 1 2 0
RP[1]/DP-127 0.5 0.906 1 0.603 0.761 0.687 62 1 149 1 2 0
RP[1]/DP-134 0.559 1 1 0.504 0.759 0.687 16 1 177 1 2 0
RP[1]/DP-174 0.537 1 0.912 0.556 0.748 0.687 42 1 153 1 2 0
RP[1]/DP-49 0.5 0.906 0.755 0.727 0.734 0.687 62 1 80 1 2 0
RP[1]/DP-17 0.5 0.594 0.824 0.837 0.722 0.687 62 1 30 1 2 0
RP[1]/DP-66 0.5 0.5 1 0.721 0.712 0.687 62 1 23 1 2 0
RP[1]/DP-25 0.5 1 0.72 0.62 0.71 0.687 62 1 113 1 2 0
RP[1]/DP-136 0.5 1 0.72 0.612 0.707 0.687 62 1 34 1 2 0
RP[1]/DP-42 0.537 0.594 1 0.57 0.688 0.687 42 1 30 1 2 0
RP[1]/DP-172 0.787 0.906 0.528 0.616 0.685 0.687 6 1 66 1 2 0
RP[1]/DP-28 0.669 0.871 0.53 0.684 0.679 0.687 22 1 23 1 2 0
RP[1]/DP-87 0.5 0.559 0.963 0.612 0.678 0.687 62 1 34 1 2 0
RP[1]/DP-89 0.5 0.906 0.641 0.647 0.676 0.687 62 1 54 1 2 0
RP[1]/DP-26 0.526 0.5 1 0.599 0.676 0.687 57 1 140 1 2 0
RP[1]/DP-95 0.537 0.965 0.644 0.578 0.673 0.687 42 1 74 1 2 0
RP[1]/DP-57 0.754 0.5 0.954 0.5 0.67 0.687 11 1 179 1 2 0
RP[1]/DP-197 0.603 0.965 0.676 0.5 0.668 0.687 34 1 179 1 2 0
RP[1]/DP-141 0.537 0.594 0.644 0.802 0.667 0.687 42 1 30 1 2 0
RP[1]/DP-199 0.787 0.776 0.688 0.5 0.662 0.687 6 1 179 1 2 0
RP[1]/DP-34 0.5 0.665 0.678 0.733 0.662 0.687 62 1 24 1 2 0
RP[1]/DP-126 0.5 0.535 0.963 0.578 0.662 0.687 62 1 76 1 2 0
RP[1]/DP-152 0.5 0.653 0.766 0.659 0.66 0.687 62 1 100 1 2 0
RP[1]/DP-79 0.5 0.659 0.771 0.62 0.649 0.687 62 1 113 1 2 0
Auto-fit columns | Line count: 202, visible: 37, selected: 37
Select Rank... Filter Sort... Export
Flapjack Tip: Create multiple custom views of the same data set by selecting 'Visualization->Create new view' from the menu bar
62x92, 4C, 7T, 117.43MB

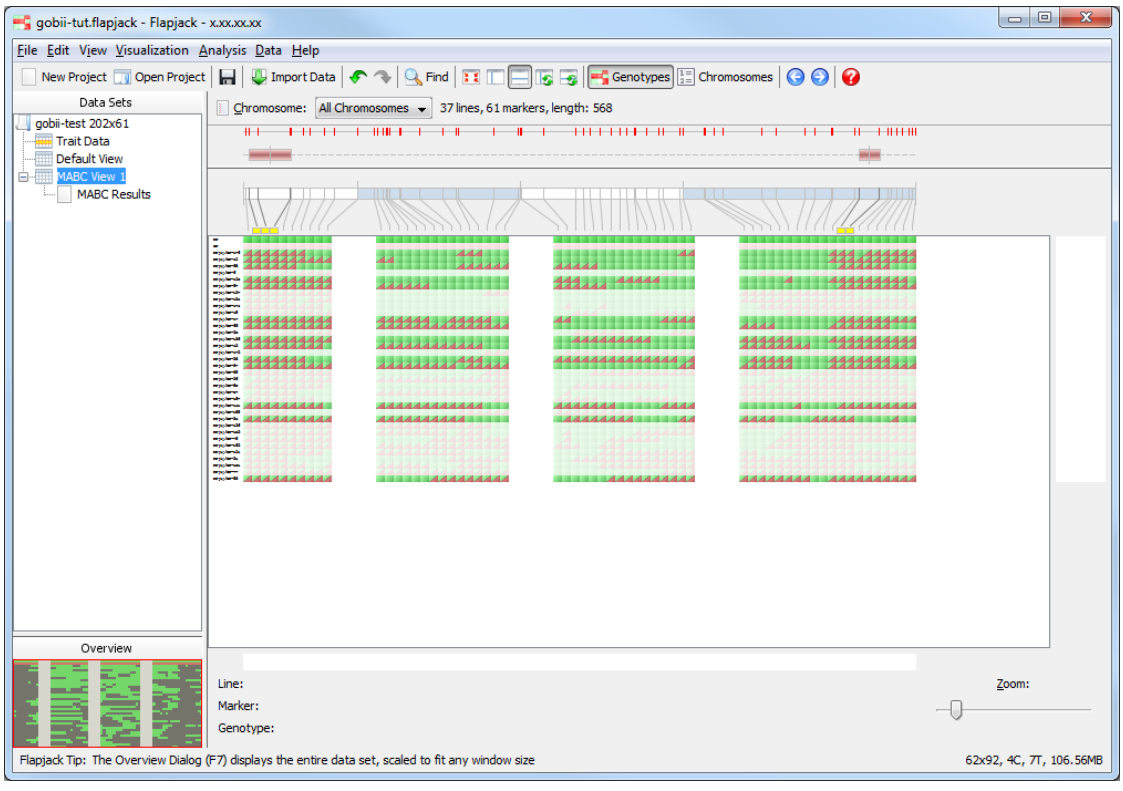

gobii-tut.flapjack - Flapjack - x.xx.xx.xx
File Edit View Visualization Analysis Data Help
New Project Open Project Import Data Find Genotypes Chromosomes
Data Sets
gobii-test 202x61
Trait Data
Default View
MABC View 1
MABC Results
Chromosome: All Chromosomes
37 lines, 61 markers, length: 568
Overview
Line:
Marker:
Genotype:
Zoom:
Flapjack Tip: You can import and work with multiple data sets at once, all within a single project
62x92, 4C, 7T, 126.80MB

34.8 Selecting within results

Click `Select->Auto select` to open the `Auto Select Lines` dialog. Fill the dialog in as with the `Filter` dialog, with values for `RPP Total` of `Greater than or equal to` `0.6`, `LD (QTL1)` of `Less than or equal to` `62`, and `LD (QTL2)` of `Less than or equal to` `34`. Click `Select` to apply the selection criteria to the data in the results table. You may see a dialog informing you that Flapjack has switched to line mode. Click Ok to dismiss this. Then view the results of your selection in both the results and genotype views. You should note that there are now only 15 lines selected. Deselected lines are de-emphasised in the genotype view in Flapjack by ligthening their colour.

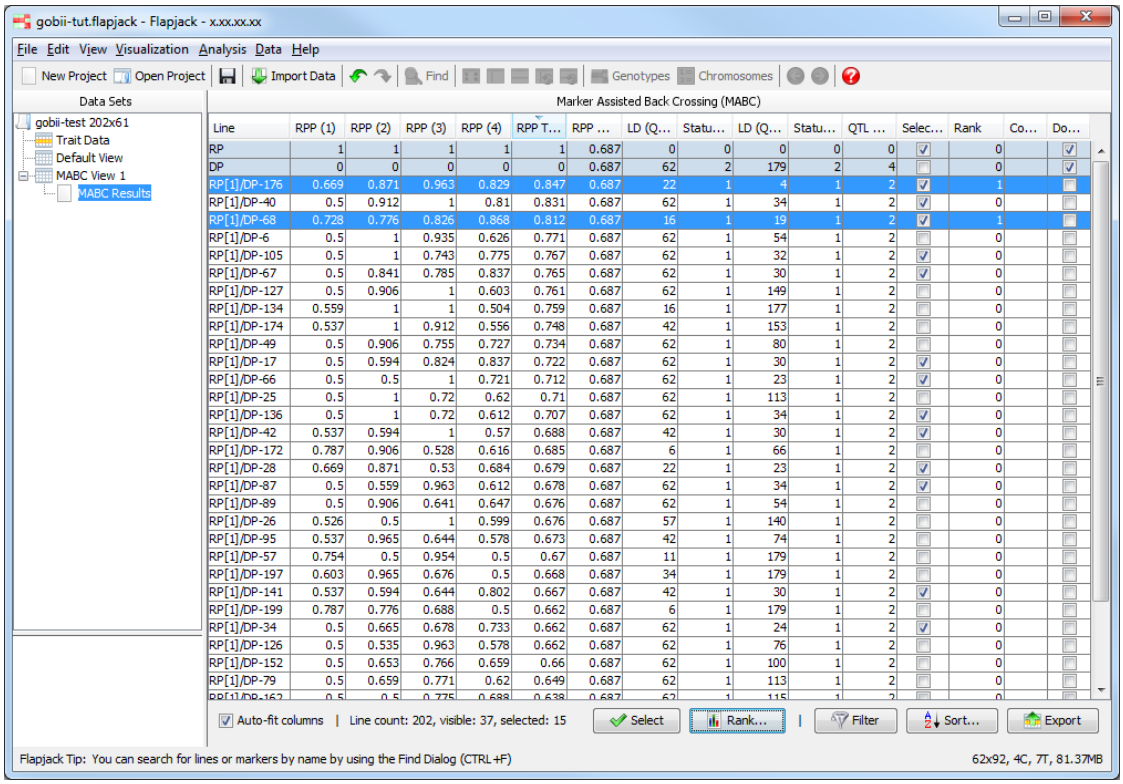



gobii-tut.flapjack - Flapjack - x.xx.xx.xx
File Edit View Visualization Analysis Data Help
New Project Open Project Import Data Find Genotypes Chromosomes
Data Sets
gobii-test 202x61
Trait Data
Default View
MABC View 1
MABC Results
Marker Assisted Back Crossing (MABC)
Line RPP (1) RPP (2) RPP (3) RPP (4) RPP T... RPP ... LD (Q... Statu... LD (Q... Statu... QTL ... Selec... Rank Co... Do...
RP 1 1 1 1 1 0.687 0 0 0 0 0
DP 0 0 0 0 0 0.687 62 2 179 2 4 0
RP[1]/DP-176 0.669 0.871 0.963 0.829 0.847 0.687 22 1 4 1 2 0
RP[1]/DP-40 0.5 0.912 1 0.81 0.831 0.687 62 1 34 1 2 0
RP[1]/DP-68 0.728 0.776 0.826 0.868 0.812 0.687 16 1 19 1 2 0
RP[1]/DP-6 0.5 1 0.935 0.626 0.771 0.687 62 1 54 1 2 0
RP[1]/DP-105 0.5 1 0.743 0.775 0.767 0.687 62 1 32 1 2 0
RP[1]/DP-67 0.5 0.841 0.785 0.837 0.765 0.687 62 1 30 1 2 0
RP[1]/DP-127 0.5 0.906 1 0.603 0.761 0.687 62 1 149 1 2 0
RP[1]/DP-134 0.559 1 1 0.504 0.759 0.687 16 1 177 1 2 0
RP[1]/DP-174 0.537 1 0.912 0.556 0.748 0.687 42 1 153 1 2 0
RP[1]/DP-49 0.5 0.906 0.755 0.727 0.734 0.687 62 1 80 1 2 0
RP[1]/DP-17 0.5 0.594 0.824 0.837 0.722 0.687 62 1 30 1 2 0
RP[1]/DP-66 0.5 0.5 1 0.721 0.712 0.687 62 1 23 1 2 0
RP[1]/DP-25 0.5 1 0.72 0.62 0.71 0.687 62 1 113 1 2 0
RP[1]/DP-136 0.5 1 0.72 0.612 0.707 0.687 62 1 34 1 2 0
RP[1]/DP-42 0.537 0.594 1 0.57 0.688 0.687 42 1 30 1 2 0
RP[1]/DP-172 0.787 0.906 0.528 0.616 0.685 0.687 6 1 66 1 2 0
RP[1]/DP-28 0.669 0.871 0.53 0.684 0.679 0.687 22 1 23 1 2 0
RP[1]/DP-87 0.5 0.559 0.963 0.612 0.678 0.687 62 1 34 1 2 0
RP[1]/DP-89 0.5 0.906 0.641 0.647 0.676 0.687 62 1 54 1 2 0
RP[1]/DP-26 0.526 0.5 1 0.599 0.676 0.687 57 1 140 1 2 0
RP[1]/DP-95 0.537 0.965 0.644 0.578 0.673 0.687 42 1 74 1 2 0
RP[1]/DP-57 0.754 0.5 0.954 0.5 0.67 0.687 11 1 179 1 2 0
RP[1]/DP-197 0.603 0.965 0.676 0.5 0.668 0.687 34 1 179 1 2 0
RP[1]/DP-141 0.537 0.594 0.644 0.802 0.667 0.687 42 1 30 1 2 0
RP[1]/DP-199 0.787 0.776 0.688 0.5 0.662 0.687 6 1 179 1 2 0
RP[1]/DP-34 0.5 0.665 0.678 0.733 0.662 0.687 62 1 24 1 2 0
RP[1]/DP-126 0.5 0.535 0.963 0.578 0.662 0.687 62 1 76 1 2 0
RP[1]/DP-152 0.5 0.653 0.766 0.659 0.66 0.687 62 1 100 1 2 0
RP[1]/DP-79 0.5 0.659 0.771 0.62 0.649 0.687 62 1 113 1 2 0
Auto-fit columns
Line count: 202, visible: 37, selected: 15
Select
Rank...
Filter
Sort...
Export
Flapjack Tip: The red rectangle on an Overview display shows the region of the data set currently being viewed on the main canvas
62x92, 4C, 7T, 84.05MB


gobii-tut.flapjack - Flapjack - x.xx.xx.xx
File Edit View Visualization Analysis Data Help
New Project Open Project Import Data Find Genotypes Chromosomes
Data Sets
gobii-test 202x61
Trait Data
Default View
MABC View 1
MABC Results
Chromosome: All Chromosomes
37 lines, 61 markers, length: 568
Overview
Line:
Marker:
Genotype:
Zoom:
Flapjack Tip: The Overview Dialog (F7) displays the entire data set, scaled to fit any window size
62x92, 4C, 7T, 106.56MB

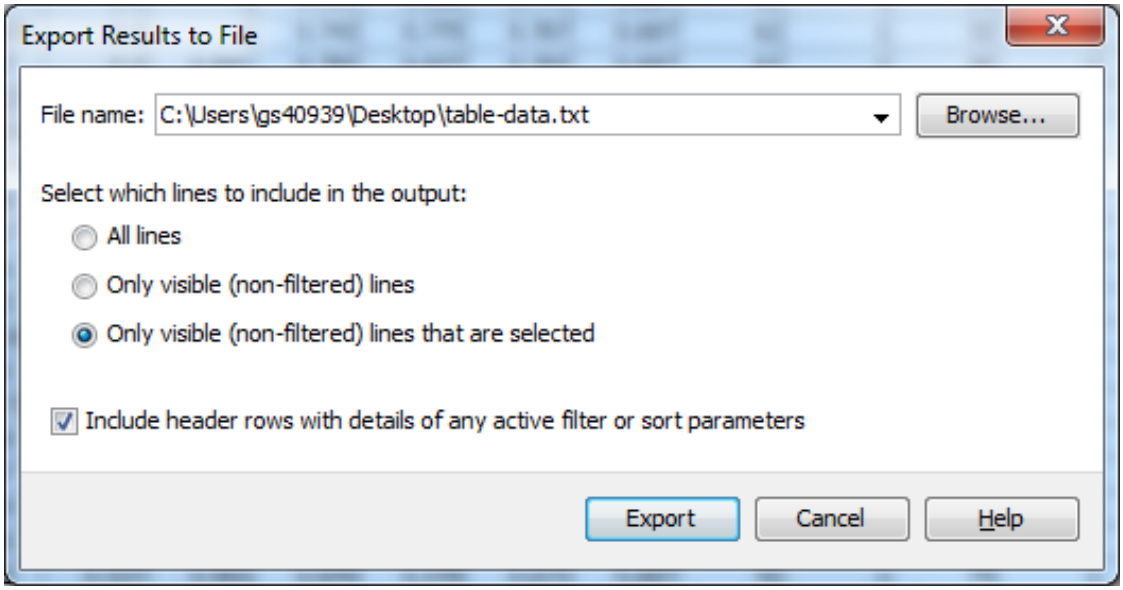
34.9 Ranking results

Looking at the selected lines in the results table a little more closely, two lines in particular stand out for having particularly low LD values for both QTL in the data set. It may be worthwhile marking these up of being particular interest. CTRL/CMD click on the rows for `RP[1]/DP-176` and `RP[1]/DP-68` in the table to highlight these rows. Click the `Rank` button to open the `Rank Lines` dialog. As all other lines in the data set currently have the default rank of 0, leave the value in the `Assign all highlighted lines a rank of` field as 1 and click `OK`. The rank for the highlighted rows will be updated to 1. You can also click on the comment field for each line and type a descriptive comment about the line, perhaps explaining why it's been given a particular rank.



34.10 Exporting results

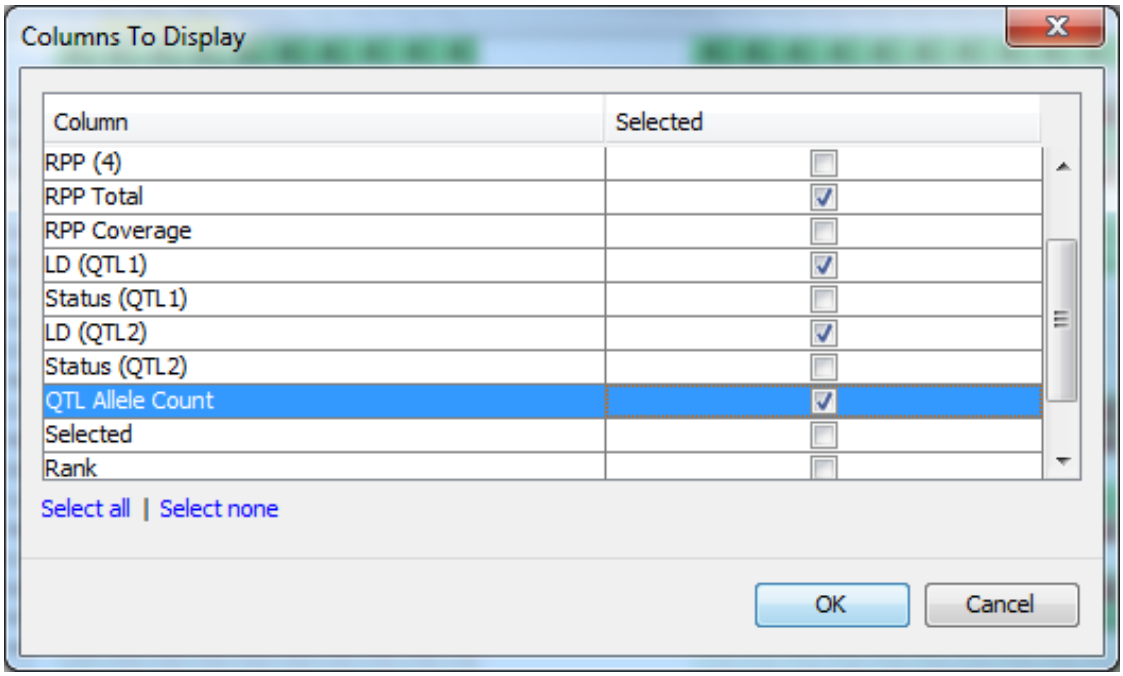
Click `Export` to open the `Export Results to File` dialog. You can select three separate types of export from this dialog, `All lines` which exports all of the data whether it had been filtered or not, `Only visible (non-filtered) lines` which outputs the results table as it appears in Flapjack, and `Only visible (non-filtered) lines that are selected` which outputs only the lines which are visible in the table and are selected. Select the third option: `Only visible (non-filtered) lines that are selected`. Ensure that the `Include header rows with details of any active filter or sort parameters` is selected, this outputs information about the filtering and sort that was applied to reach the current view of the data, which can be useful for the purpose of reproducing the steps at a later date. Click `Browse` to select a location to save your file, as well as a file name. Finally click `Export` to output the data to file.

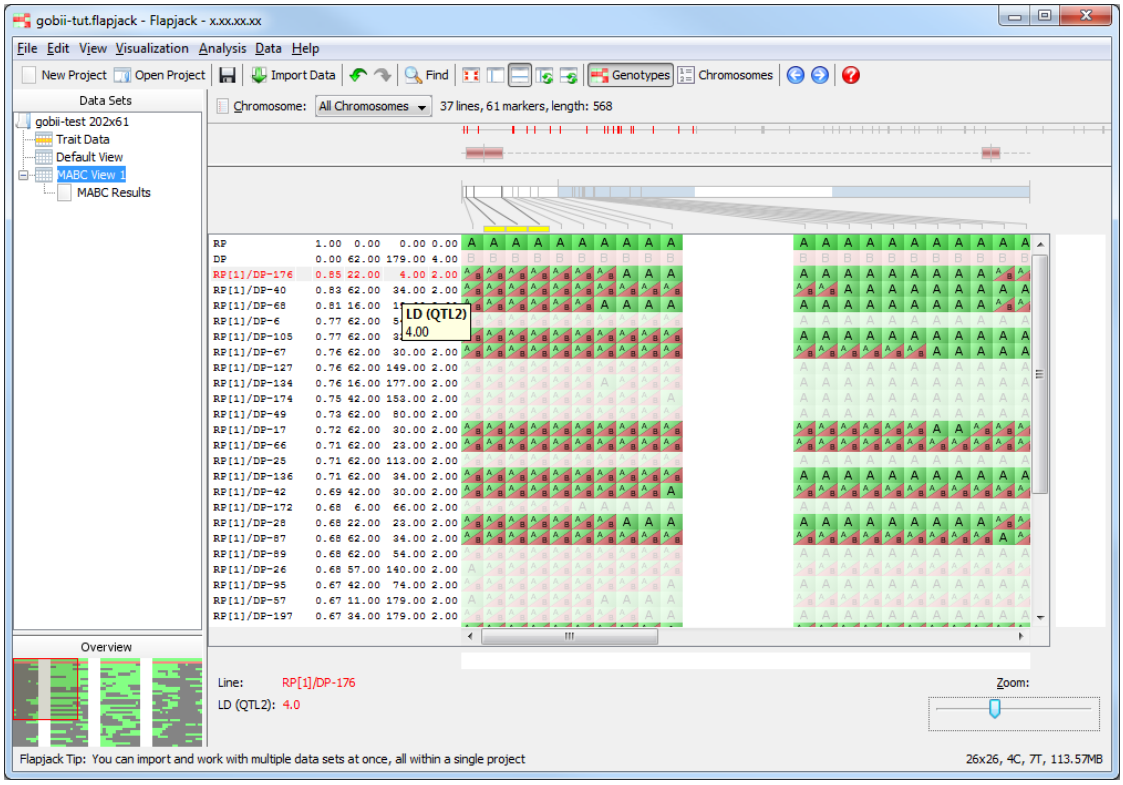


You can view the contents of the file in any text editor.

34.11 Viewing results in the genotype view

Return to the genotype view of the data (`MABC View 1`) and zoom in until you can comfortably read the line names which are just to the left of the genotype visualization. Right-click on the list of line names and select `Show table results` from the menu to open the `Columns To Display` dialog. This dialog allows you to select columns of data from the results table to view side by side with the genotype data. This can be good for a final visual validation of what you're seeing from the analysis results. Select `RPP Total`, `LD (QTL1)`, `LD (QTL2)` and `QTL Allele Count` by clicking the corresponding checkboxes in the dialog, then click `OK`.





You should see that there are four extra columns of data between the lines names and the genotype display. These are the four columns you selected in the `Columns To Display` dialog and represent the data found for those columns for the lines in `MABC Results` view. Mouseover the columns to see tooltips with the column name and value for the line under the mouse, this information is also displayed in the status panel at the bottom of the display.

34.12 Conclusion

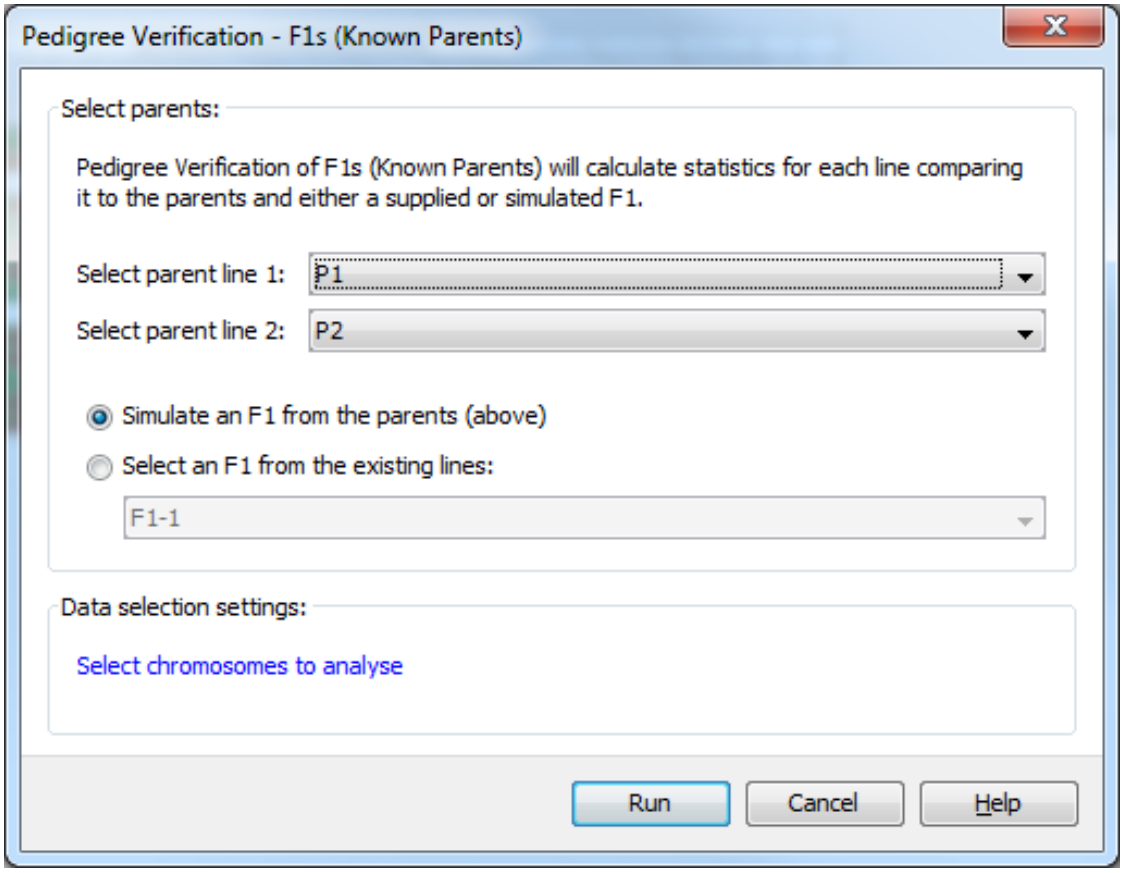
You've now had a chance to experiment with running a markers assisted backcrossing analysis in Flapjack and analysing the results of that analysis. Feel free to experiment more by applying different sorts, filters, selections, or better yet make a start on analysing your own data.

CHAPTER 35

Pedigree Verification (F1s Known Parents)

This analysis provides several key statistics to determine whether a putative F1 sample is a true F1 cross, and that the marker alleles are originating from the expected parents.

In addition to this help page, you can also read the *Pedigree Verification (F1s Known Parents) Tutorial*, which runs through the process of running the analysis and viewing the results with a sample dataset.



Using the `Pedigree Verification - F1s (Known Parents)` dialog you can select the two parental lines which you are comparing putative F1s against. You can also select whether to simulate an expected F1 from the parents you have selected, or to use an expected F1 which was defined in the input data.

35.1 Options

If you choose to simulate an F1, Flapjack will generate an expected F1 line from the two parental lines you selected in the dialog. To do this Flapjack compares the alleles of the two parental lines to determine the alleles for the expected F1. For any allele where a parent has missing data, or a heterozygous allele the allele in the expected F1 is left as missing data. For other alleles if parent 1 has A and parent 2 has A, the expected F1 gets A as the allele for that marker. If parent 1 has A and parent 2 has T the expected F1 get A/T as the allele at that marker.

- `% allele match to expected` - the simplest way to understand whether an F1 is a true cross from the expected parents is first to simulate the expected F1 alleles based on those parents, and then compare the expected F1 alleles to the putative F1 samples. A 100% match indicates the putative F1 sample is a true F1 from the expected parents. Note, for a match to occur, the putative F1 sample has to EXACTLY match the allele pattern of the simulated F1 ie a match occurs if the putative F1 sample is A/A and the simulated F1 is A/A, but NOT if the putative F1 sample has A/A and the simulated F1 is A/T.

Note: Expected F1 alleles can only be simulated if both parents have homozygous alleles. If either parent has a heterozygous call or missing data for a marker, then the expected F1 allele data cannot be simulated and will have a missing value.

- `Marker Count and % Missing` - since for many F1 tests a small number of markers are used, and any small amount of marker failure can result in misleading analyses results, statistics are first provided for `Marker Count` and `% Missing`. Results can therefore first be filtered by `Marker Count` or `% Missing` to eliminate potentially skewed results.
- `% het and % Deviation from Expected` - an F1 made from a cross between two inbred lines will have elevated % heterozygosity compared to parents. Therefore, the two statistics provided; % of markers that are heterozygous (% het), and deviation from the expected level of heterozygosity compared to the simulated F1 (% Deviation from Expected), will indicate whether a cross has been successfully made. In the case that a putative F1 sample has low % hets and high deviation from expected, then the sample is most likely an inbred line and not a true cross.
- `% P1 Contained and % P2 Contained` - to determine whether each parent has contributed alleles to the putative F1 sample, the analysis provides `% contained` results. If a single parent is 100% contained in the F1 this means that 100% of the marker alleles from the parent are contributing to the F1 and is a true parent. A lower % contained results can mean genotyping error or that the parent was not involved in the cross. In the case that the parent has a heterozygous call (or missing data) for a marker, then that marker datapoint is not used in the % contained analysis.

35.2 Understanding the statistics

These 3 sets of analyses can help determine whether your putative F1 sample is a true F1 from a cross of the expected parents.

Warning: No attempt has been made to define exact analysis values that result in the likely scenarios as this will depend on the level of genotyping error in your experiment as well as the shared alleles, or genetic similarity,

between the parents being crossed as well as the alternate inbreds that could be resulting in seed mix ups or outcrossing.

Below are various combinations of values for the statistics.

A true F1 from the expected parents:

% Allele match to expected	**High**
% het	**As expected**
% Deviation from expected	**Low**
% contained parent 1	**High**
% contained parent 2	**High**

A self of parent 1, or sample is parent 1 (seed mix-up):

% Allele match to expected	**Low**
% het	**Low**
% Deviation from expected	**High**
% contained parent 1	**High**
% contained parent 2	**Low**

An F1 cross between parent 1 and an unknown inbred:

% Allele match to expected	**Low**
% het	**Approximately as expected**
% Deviation from expected	**Medium-high**
% contained parent 1	**High**
% contained parent 2	**Low**

An F1 cross between two unknown inbreds:

% Allele match to expected	**Low**
% het	**Approximately as expected**
% Deviation from expected	**Medium-high**
% contained parent 1	**Low**
% contained parent 2	**Low**

CHAPTER 36

Pedigree Verification (F1s Known Parents) Tutorial

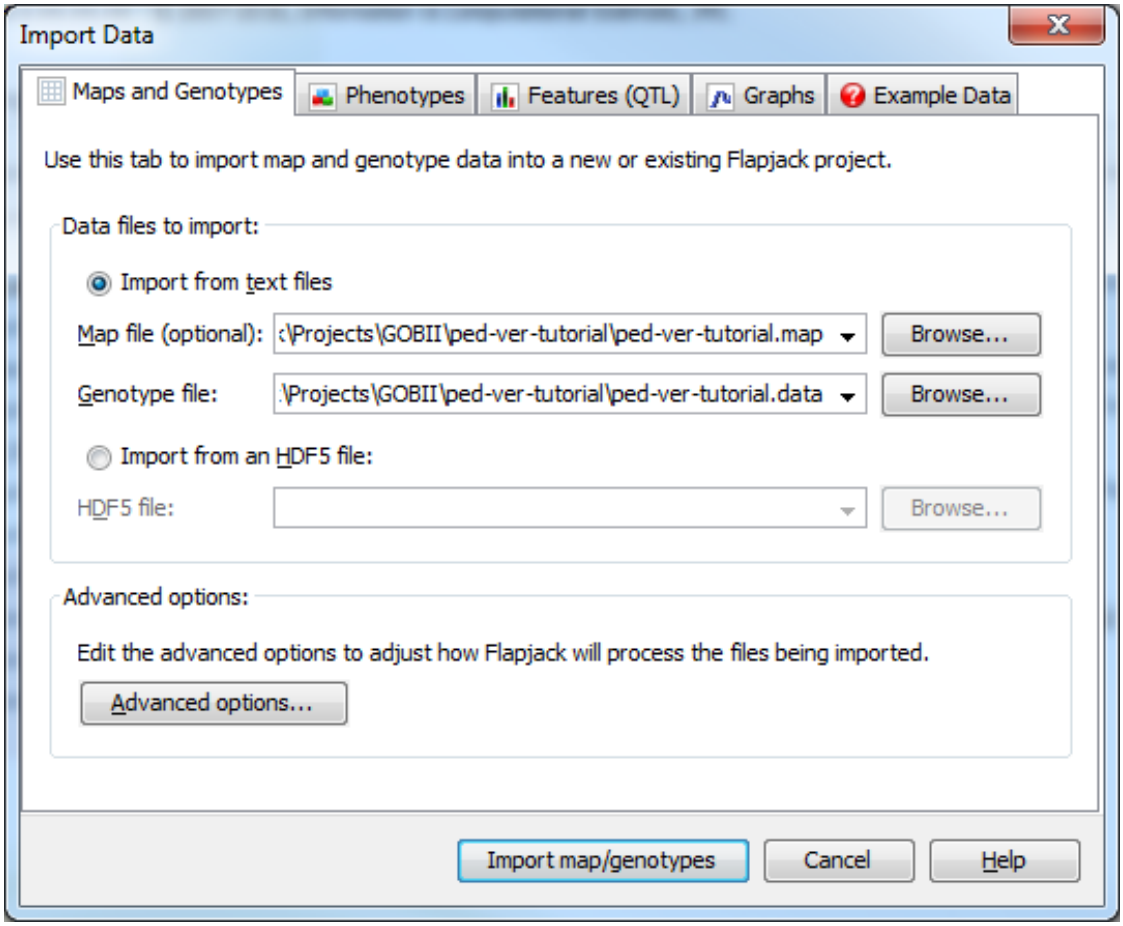
In this tutorial we will assume that you are using the `ped-ver-tutorial` dataset, which contains simulated data with two parent lines and twelve F1s that we'd like to test against these parental lines to see if they are likely to be derived from the parental lines. You'll be working with two files, a map file (ped-ver-tutorial.map) which contains a set of 20 markers across 3 chromosomes and a genotype file (ped-ver-tutorial.dat) which contains data on a set of 14 lines (two of which are parents, 12 of which are putative F1s).

Download the data (Right click->Save Link As):

- ped-ver-tutorial.map
- ped-ver-tutorial.dat

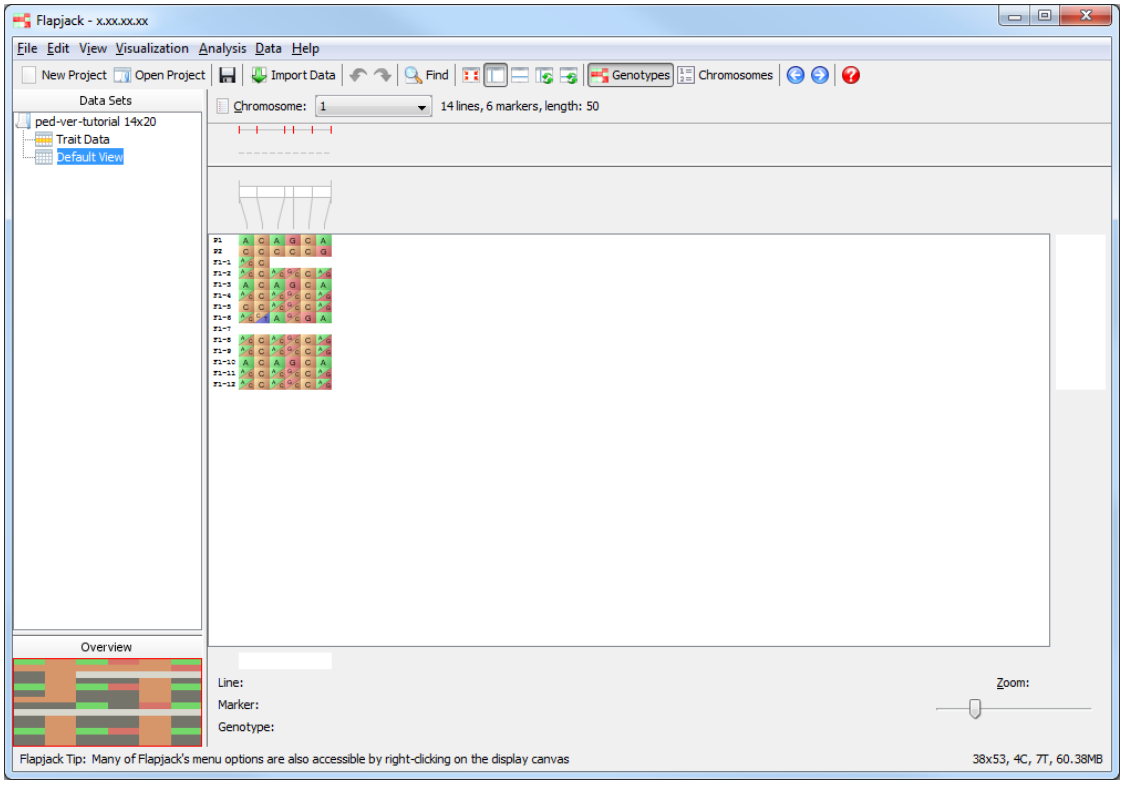
36.1 Importing data

To import data into Flapjack click the `Import Data` button on the toolbar. In the `Import Data` dialog, select the `Maps and Genotypes` tab, then click `Browse` to navigate to and select the map file you wish to import (ped-ver-tutorial.map), and then do the same for the genotype file you wish to import (ped-ver-tutorial.data).

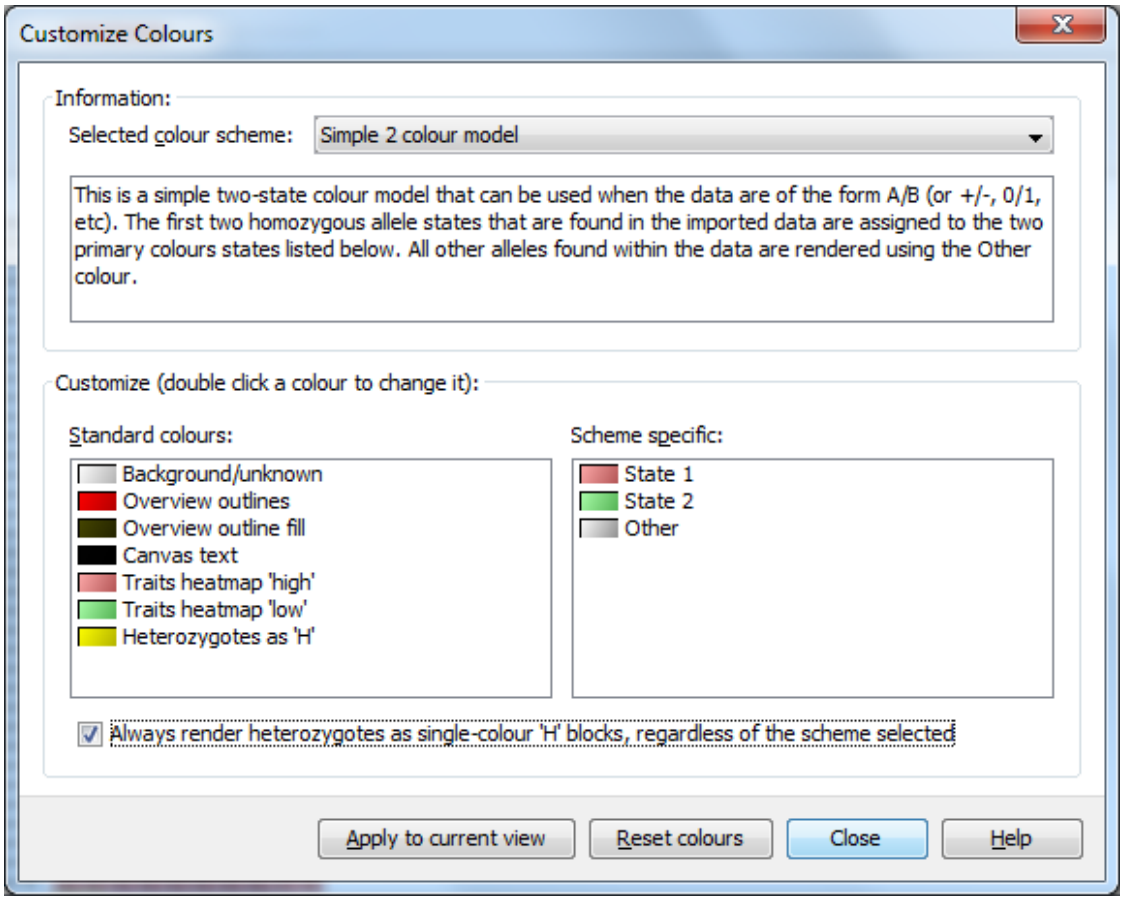


Click the `Advanced options` button to open the `Advanced Data Import Options` dialog and ensure that the `Duplicate all markers onto a single "All Chromosomes" chromosome for side by side viewing` option is selected, then click OK. Finally click the `Import map/genotypes` button to load the data.

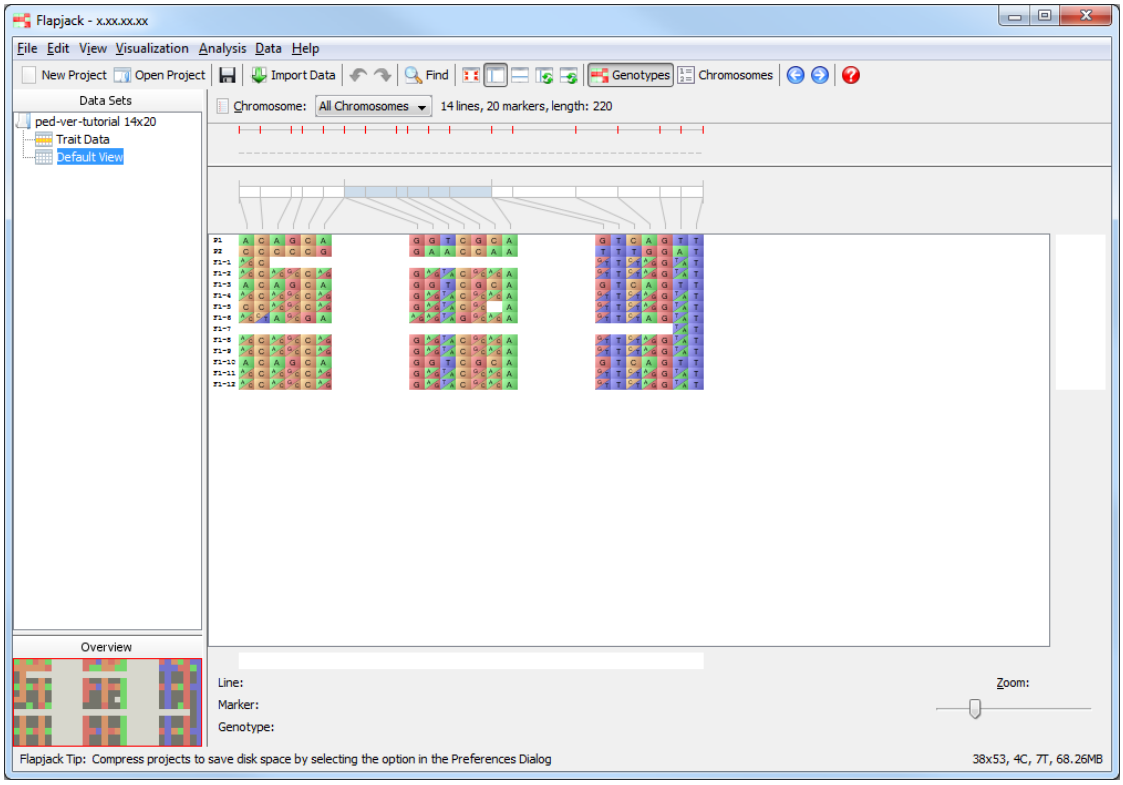
You should now be viewing the Default View on the 1st chromosome of your dataset. This is the main type of visualization in Flapjack and comprises a graphical genotype view of the imported data. Each square represents an allele, found at the cross-section of a line and a marker, lines or varieties can be found down the left hand side of the display and markers can be found along the top (as well as a visualization of the map associated with the markers). Alleles can be either homozygous (in this dataset either an A or a B) or heterozygous, with the latter being rendered as split diagonal blocks, or A/B in the dataset.



If you're used to viewing heterozygous alleles as simply H instead of a diagonal split view, select `Visualization->Colour scheme->Customize` from the menubar to open the `Customize Colours` dialog. From there you can select `Always render heterozygotes as single-colour 'H' blocks, regardless of the scheme selected` option and click `Apply to current view` - your heterozygotes should now be rendering as a single-colour 'H' block.



Select the All Chromosomes view from the Chromosome dropdown menu underneath the toolbar and use the zoom control at the bottom of the display to zoom out until you can see all of the data. You should now see a representation of all three chromosomes in the dataset.



Note the colour of the "map" alternating between white and blue. The "map" for the `All Chromosomes` view is a combination of the other maps with ticks which extend above the map to show the start and end of each real map within the "map" of the `All Chromosomes` view. Every second map is also coloured light blue to make it easier to differentiate between maps on the `All Chromosomes` view.

36.2 Exploring and filtering data

To navigate the data select `Edit->Navigation mode` from the menubar, then click and drag the main display around with your mouse to examine all of your data. You can also use the scrollbars to navigate the data, as well as clicking and dragging on the overview in the bottom left. To zoom you can either use the zoom slider in the bottom left of the display, double click on the main display to zoom in, or use ctrl /cmd and the mouse's scroll-wheel to zoom in and out.

Before you run a marker assisted backcrossing analysis, you may want to filter out markers with lots of missing data, or monomorphic markers. Select `Edit->Filter markers->Missing markers` to open the `Filter Missing Markers` dialog. Choose the percentage of missing data in a marker required to filter out a marker, then use the `Filter` button to perform the operation. You should see a message detailing the number of markers that were filtered out of the dataset as part of the operation.

Note: In the case of the sample dataset none will be filtered out as no markers have a high percentage of missing data in the simulated dataset.

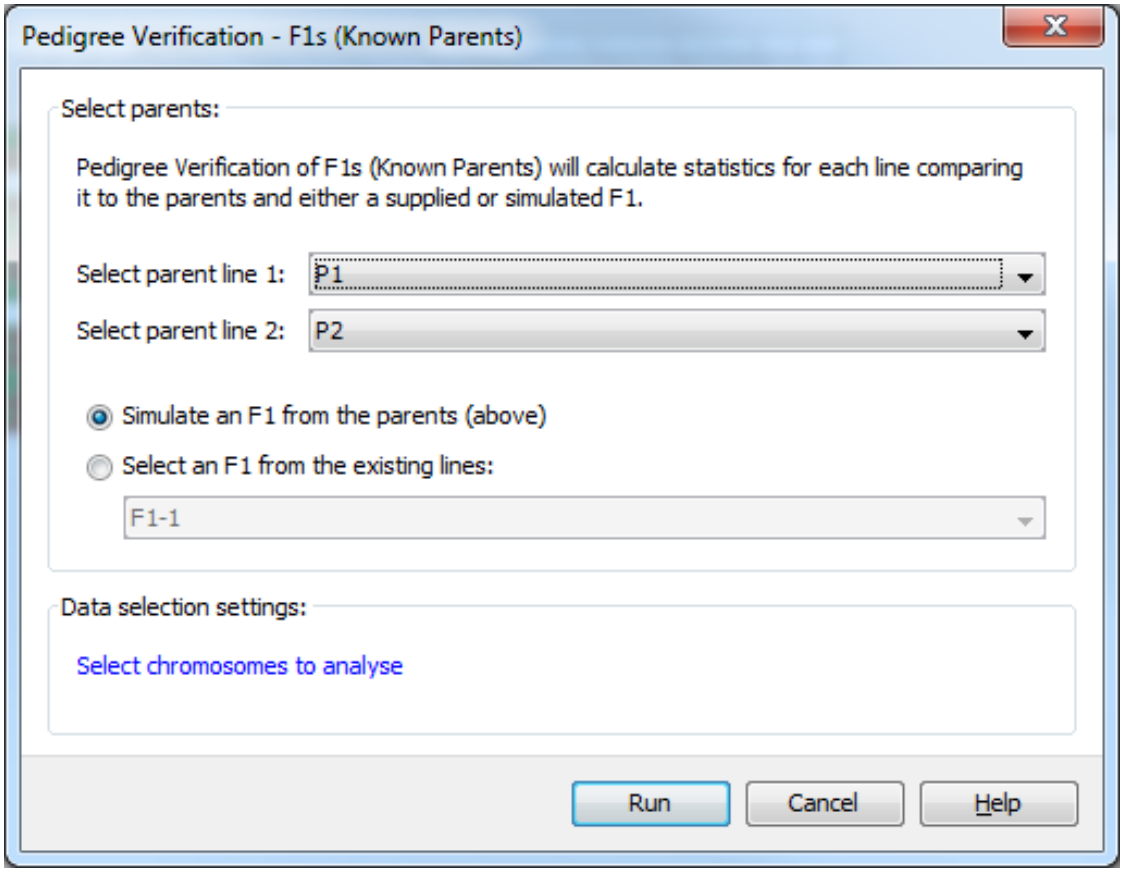
The procedure for filtering monomorphic markers is very similar - select `Edit->Filter markers->Monomorphic markers`, and simply apply the filter once the dialog has opened.

It's also possible to filter out markers which have missing (or heterozygous) data in a given line. To do this, right click on a line of interest and and select `Filter markers->Missing markers (by line)` to open the `Filter`

`Missing Markers by Line` dialog. The line should be pre-selected in the dropdown menu. Click `Filter` and Flapjack will remove any markers which have missing data for this line from the display.

36.3 Running the analysis

Select `Analysis->Pedigree verification->F1s (known parents)` to open the `Pedigree Verification - F1s (Known Parents)` dialog. Select the first parent line for your data from the first drop down list and the second parent line from the second drop down list. The drop down lists automatically select the first and second lines of the genotype input file, so if as in the tutorial data set, your two parents are on the first and second lines of the input file they will be automatically selected. Select `Simulate an F1 from the parents (above)` as with the tutorial data set we don't have an expected F1 in our data. The simulated F1 is generated by looking at each allele of the parents and deciding what an F1 created from these two lines would have at that allele. We use a simple model such that if either parent has missing or heterozygous data for an allele that allele has missing data in the simulated F1. If we did have an expected F1 in our data set we would select `Select an F1 from the existing lines` and choose our expected F1 from that dropdown list.

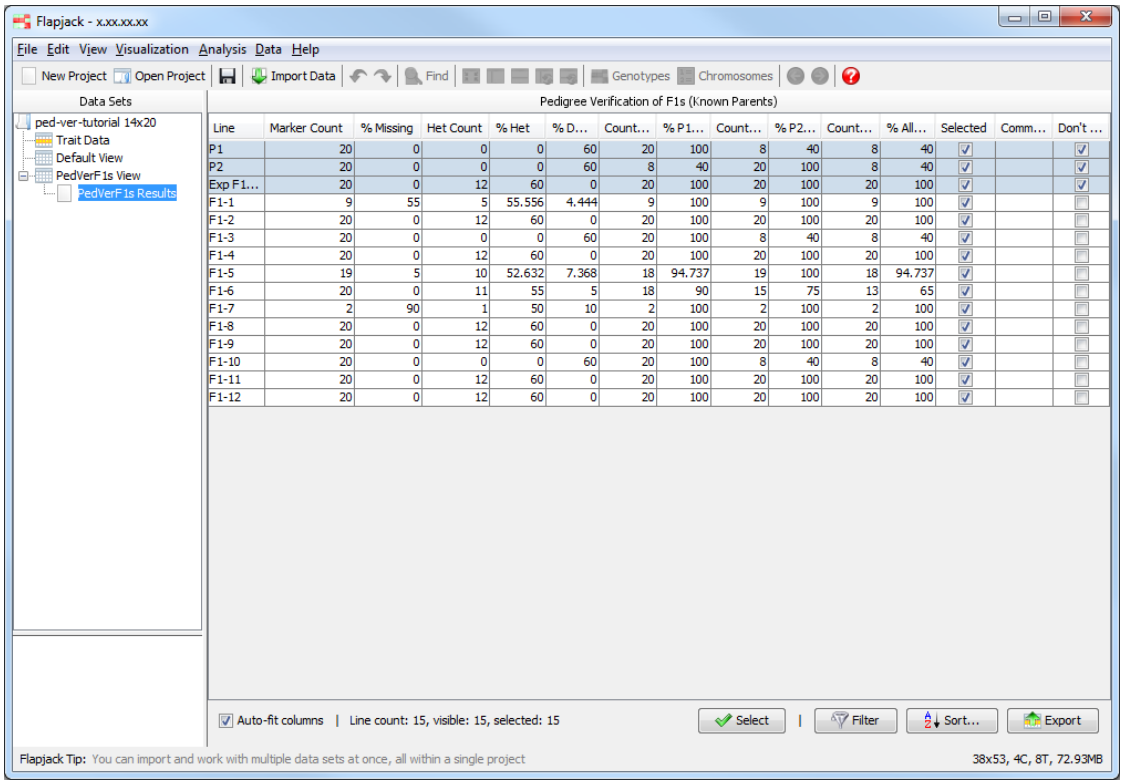


Click `Run` and Flapjack will run the analysis.

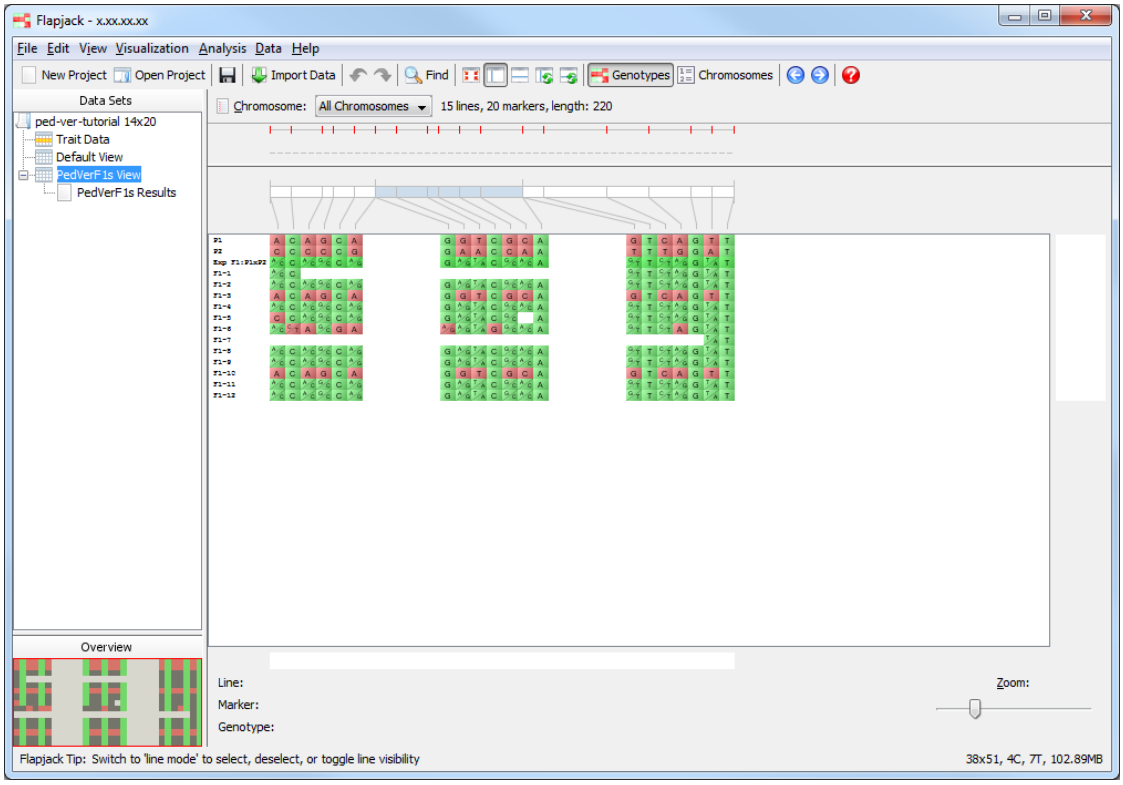
36.4 Viewing the analysis results

Once the analysis has completed you should see a table of results. The results table will contain the lines that you included in the analysis, for each line it will have a marker count, percentage of missing alleles, count of heterozygous alleles, percentage of heterozygous alleles, percentage deviation from the expected F1, count of match to parent 1

alleles, percentage match to parent 1 alleles, count of match to parent 2 alleles, percentage match to parent 2 alleles, count of allele match to the expected F1, percentage of allele match to the expected F1, as well as a selected state, and comment. The final column - Don't Sort / Filter - allows you to mark lines that you don't want table sorts and filters to apply to. By default, Flapjack sets both of the parents, and the expected F1 to neither sort, nor filter. This has the effect of keeping them in the display and always at the top of the table of data.

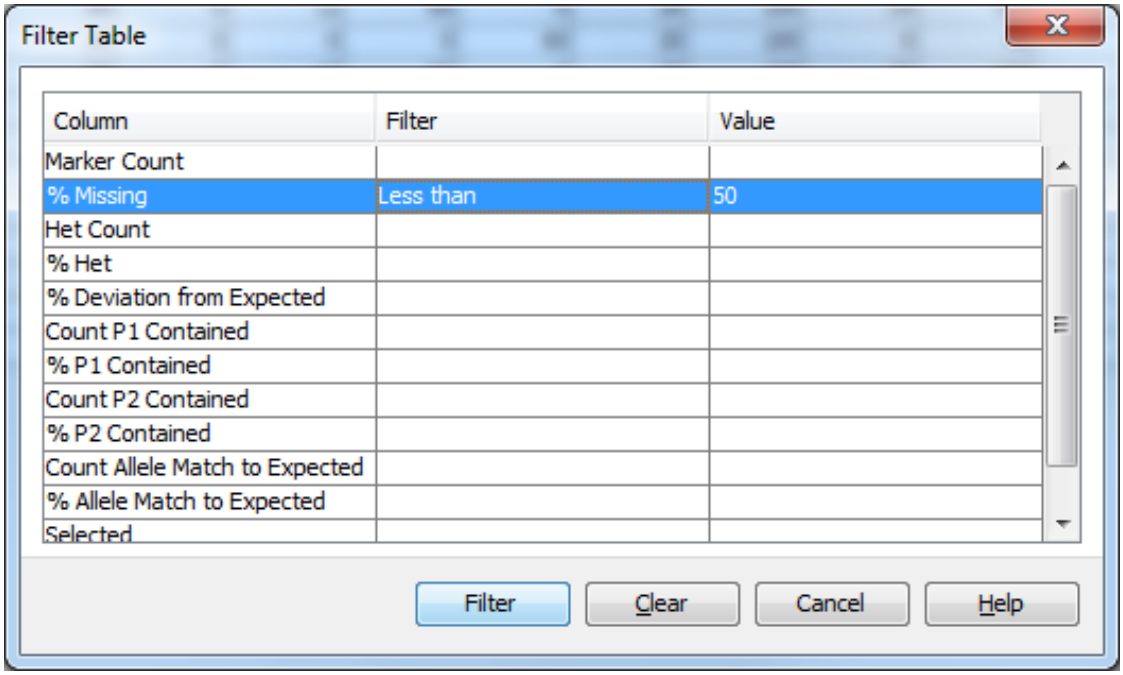


You should see that not only has Flapjack generated this `PedVerF1s Results` view, but it has linked this to a new view called `PedVerF1s View`. Click `PedVerF1s View` to view it and you should see that it's a clone of the `Default View`, but has the `By similarity to line (2 colour)` colour scheme applied. This colour scheme colours a reference line all green, all other lines have their alleles coloured either green, if they match the reference line, or red, if they don't match the reference line. In this particular case the alleles will be coloured relative to the simulated F1 you created when you ran the pedigree verification F1s (known parents) analysis. This view `PedVerF1s View` is linked to the table in the `PedVerF1s Results` view. That means moving lines, sorting lines, selecting lines and hiding lines on `PedVerF1s View` does the same in the linked `PedVerF1s Results` view, and sorting lines, selecting lines and filtering lines in the `PedVerF1s Results` view does the same in the linked view `PedVerF1s View`.

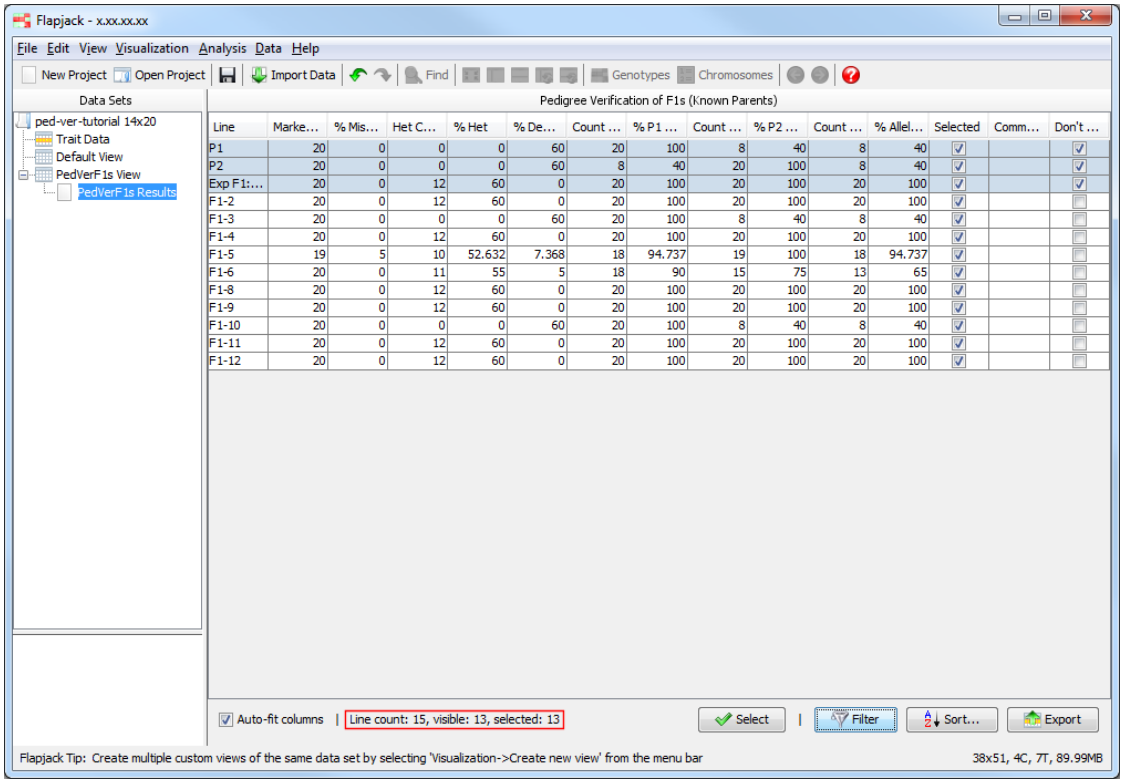


36.5 Filtering the results

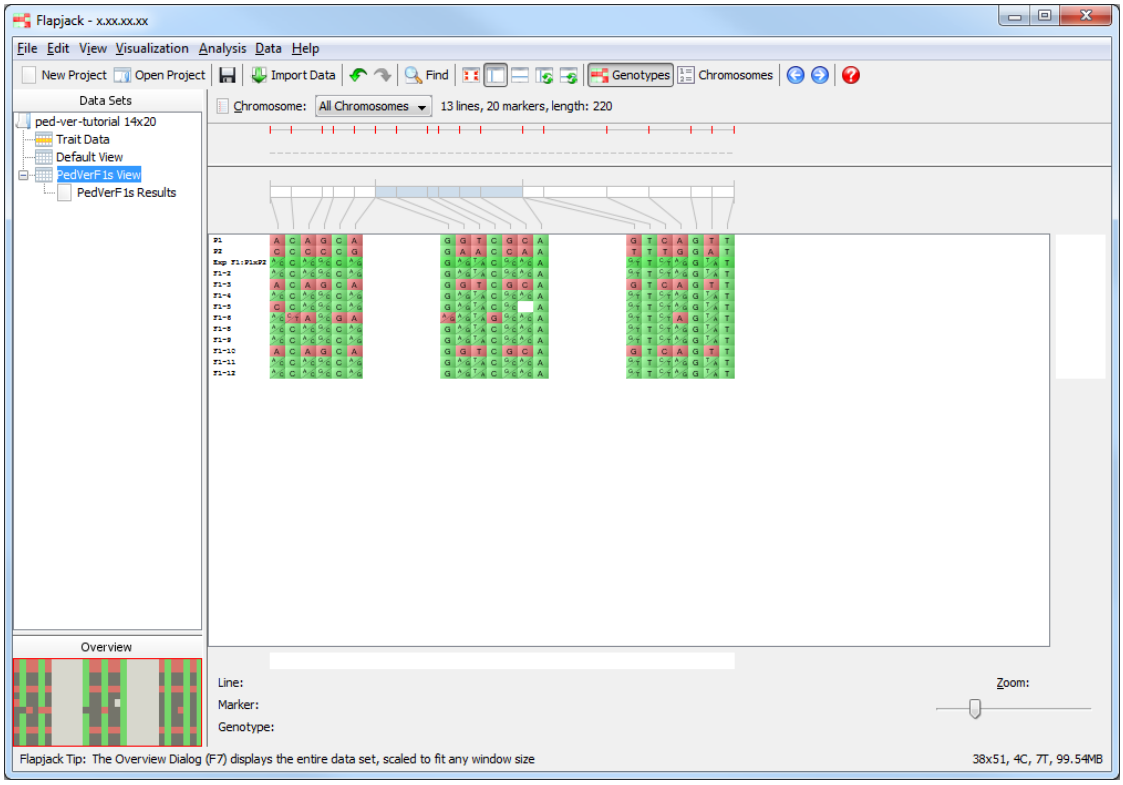
Click on `PedVerF1s Results` to return to the results view. Next click `Filter->Filter` to open the `Filter Table` dialog. You should see a table with a list of columns from the table on which you can filter. Click on the filter column for the row called `% Missing` and select `Less than` from the drop down list. Next enter a value of `50` in the adjacent `Value` column, then click `Filter`.



You should see that the results table has filtered out lines which didn't match the filter criteria. 13 of the 15 lines matched the criteria.

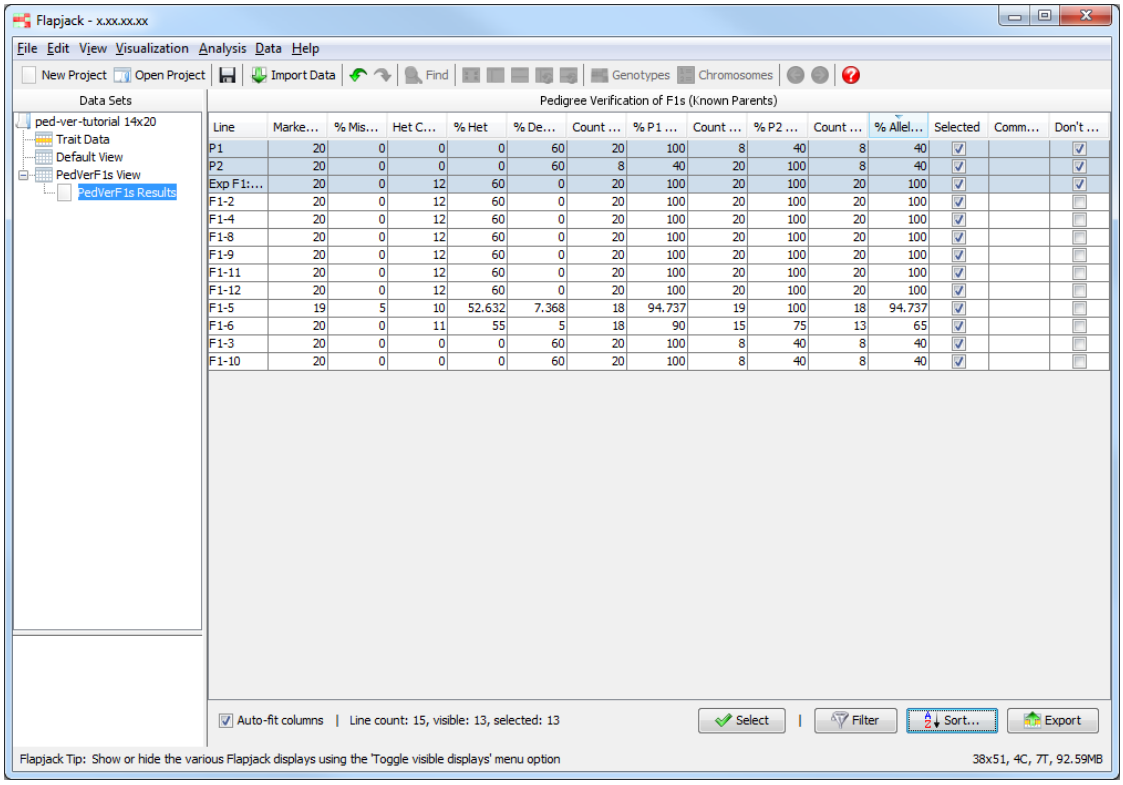


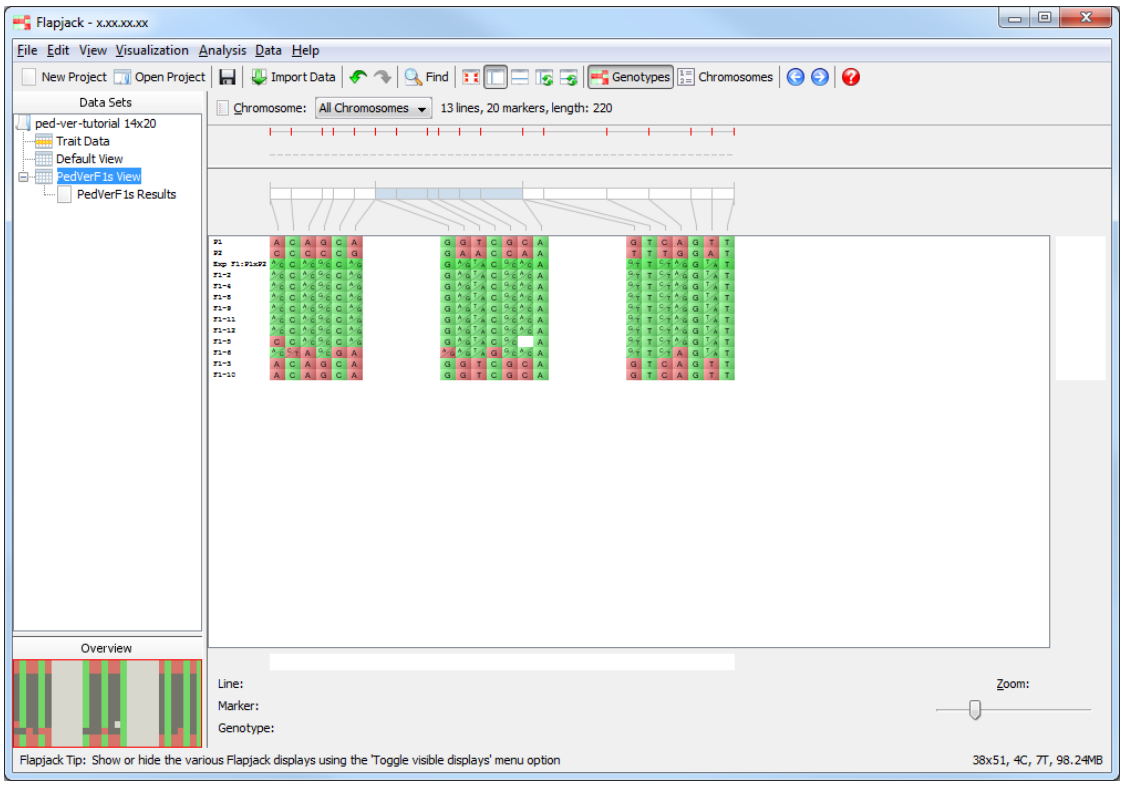
Click on `MABC View 1` if you want to see what the lines which have been kept look like visually.



36.6 Sorting the results

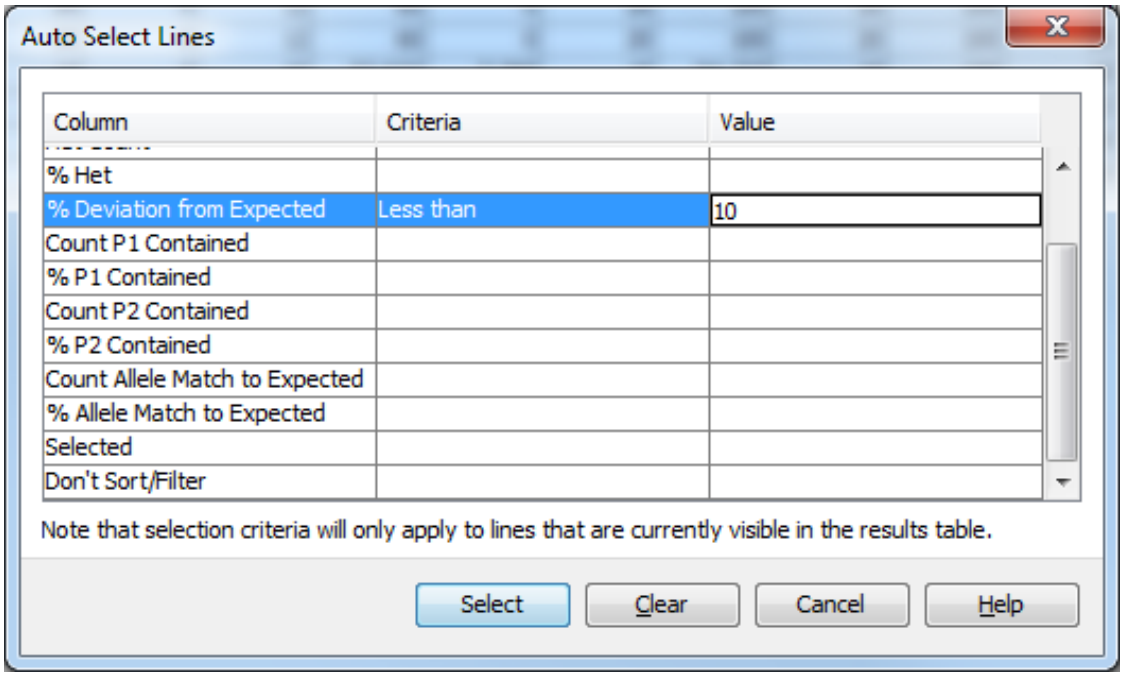
Return to the results table and select `Sort` to use the `Advanced Sort` dialog. Click `Add sort level` to add another entry to the table. Click the first entry in the `Column` column and select `% Allele Match to Expected` from the drop down list that appears. For the next entry select `% P" Contained`. For the last entry change its `Order` to be `Smallest to largest`. Click `Sort` to sort the data. You should see that the data in the table and in the genotypes view has been sorted according to your criteria.

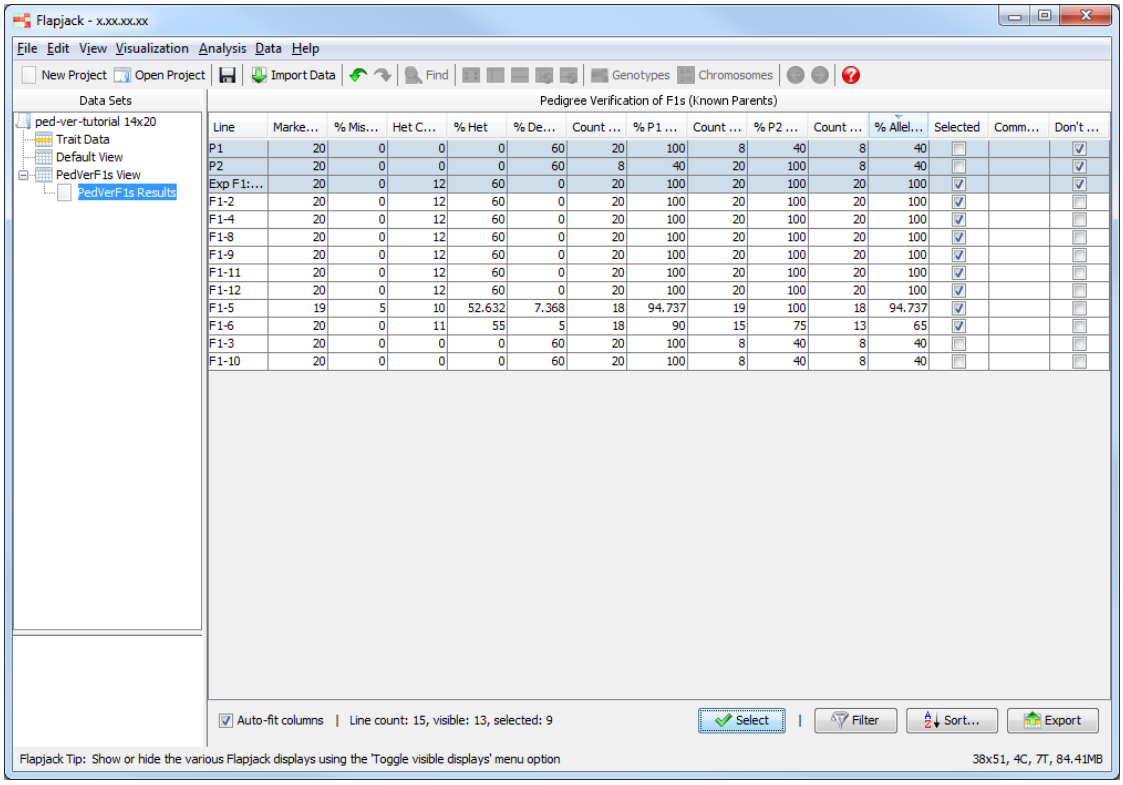

Flapjack - x.xx.xx.xx
File Edit View Visualization Analysis Data Help
New Project Open Project Import Data Find Genotypes Chromosomes
Data Sets
ped-ver-tutorial 14x20
Trait Data
Default View
PedVerF1s View
PedVerF1s Results
Pedigree Verification of F1s (known Parents)
Line Marke... % Mis... Het C... % Het % De... Count ... % P1 ... Count ... % P2 ... Count ... % Allel... Selected Comm... Don't ...
P1 20 0 0 0 60 20 100 8 40 8 40
P2 20 0 0 0 60 8 40 20 100 8 40
Exp F1:... 20 0 12 60 0 20 100 20 100 20 100
F1-2 20 0 12 60 0 20 100 20 100 20 100
F1-4 20 0 12 60 0 20 100 20 100 20 100
F1-8 20 0 12 60 0 20 100 20 100 20 100
F1-9 20 0 12 60 0 20 100 20 100 20 100
F1-11 20 0 12 60 0 20 100 20 100 20 100
F1-12 20 0 12 60 0 20 100 20 100 20 100
F1-5 19 5 10 52.632 7.368 18 94.737 19 100 18 94.737
F1-6 20 0 11 55 5 18 90 15 75 13 65
F1-3 20 0 0 0 60 20 100 8 40 8 40
F1-10 20 0 0 0 60 20 100 8 40 8 40
Auto-fit columns | Line count: 15, visible: 13, selected: 13
Select | Filter Sort... Export
Flapjack Tip: Show or hide the various Flapjack displays using the 'Toggle visible displays' menu option
38x51, 4C, 7T, 92.59MB

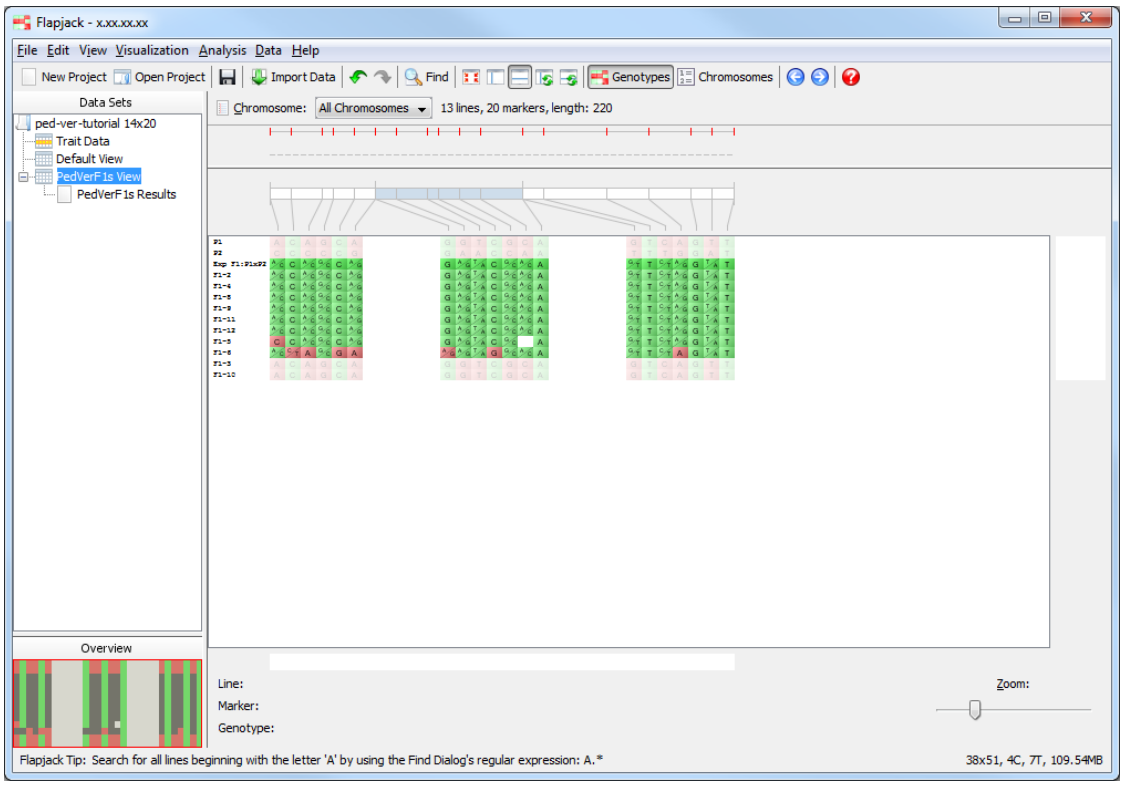

Flapjack - x.xx.xx.xx
File Edit View Visualization Analysis Data Help
New Project Open Project Import Data Find Genotypes Chromosomes
Data Sets
ped-ver-tutorial 14x20
Trait Data
Default View
PedVerF1s View
PedVerF1s Results
Chromosome: All Chromosomes
13 lines, 20 markers, length: 220
Overview
Line:
Marker:
Genotype:
Zoom:
Flapjack Tip: Show or hide the various Flapjack displays using the 'Toggle visible displays' menu option
38x51, 4C, 7T, 98.24MB

36.7 Selecting within results

Click `Select->Auto select` to open the `Auto Select Lines` dialog. Fill the dialog in as with the `Filter` dialog, with values for `% Deviation from Expected` of `Less than` `10`. Click `Select` to apply the selection criteria to the data in the results table. You may see a dialog informing you that Flapjack has switched to line mode. Click Ok to dismiss this. Then view the results of your selection in both the results and genotype views. You should note that there are now only 15 lines selected. Deselected lines are de-emphasised in the genotype view in Flapjack by ligthening their colour.

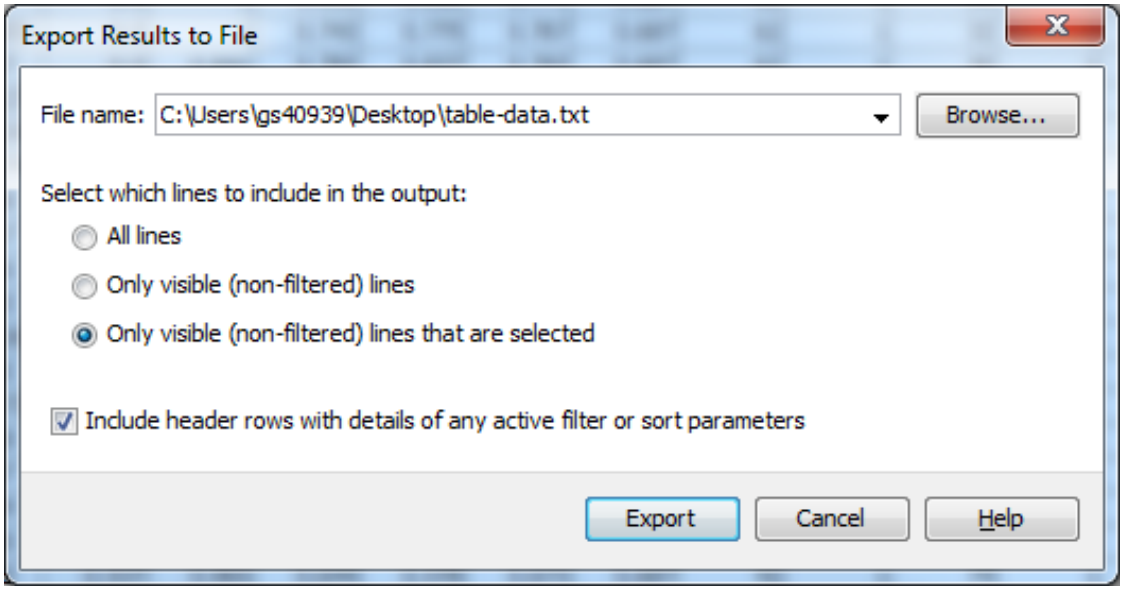



Flapjack - x.xx.xx.xx
File Edit View Visualization Analysis Data Help
New Project Open Project Import Data Find Genotypes Chromosomes
Data Sets
ped-ver-tutorial 14x20
Trait Data
Default View
PedVerF1s View
PedVerF1s Results
Pedigree Verification of F1s (known Parents)
Line Marke... % Mis... Het C... % Het % De... Count ... % P1 ... Count ... % P2 ... Count ... % Allel... Selected Comm... Don't ...
P1 20 0 0 0 60 20 100 8 40 8 40
P2 20 0 0 0 60 8 40 20 100 8 40
Exp F1:... 20 0 12 60 0 20 100 20 100 20 100
F1-2 20 0 12 60 0 20 100 20 100 20 100
F1-4 20 0 12 60 0 20 100 20 100 20 100
F1-8 20 0 12 60 0 20 100 20 100 20 100
F1-9 20 0 12 60 0 20 100 20 100 20 100
F1-11 20 0 12 60 0 20 100 20 100 20 100
F1-12 20 0 12 60 0 20 100 20 100 20 100
F1-5 19 5 10 52.632 7.368 18 94.737 19 100 18 94.737
F1-6 20 0 11 55 5 18 90 15 75 13 65
F1-3 20 0 0 0 60 20 100 8 40 8 40
F1-10 20 0 0 0 60 20 100 8 40 8 40
Auto-fit columns | Line count: 15, visible: 13, selected: 9
Select | Filter Sort... Export
Flapjack Tip: Show or hide the various Flapjack displays using the 'Toggle visible displays' menu option
38x51, 4C, 7T, 84.41MB


Flapjack - x.xx.xx.xx
File Edit View Visualization Analysis Data Help
New Project Open Project Import Data Find Genotypes Chromosomes
Data Sets
ped-ver-tutorial 14x20
Trait Data
Default View
PedVerF1s View
PedVerF1s Results
Chromosome: All Chromosomes
13 lines, 20 markers, length: 220
Overview
Line:
Marker:
Genotype:
Zoom:
Flapjack Tip: Search for all lines beginning with the letter 'A' by using the Find Dialog's regular expression: A.*
38x51, 4C, 7T, 109.54MB

36.8 Exporting results

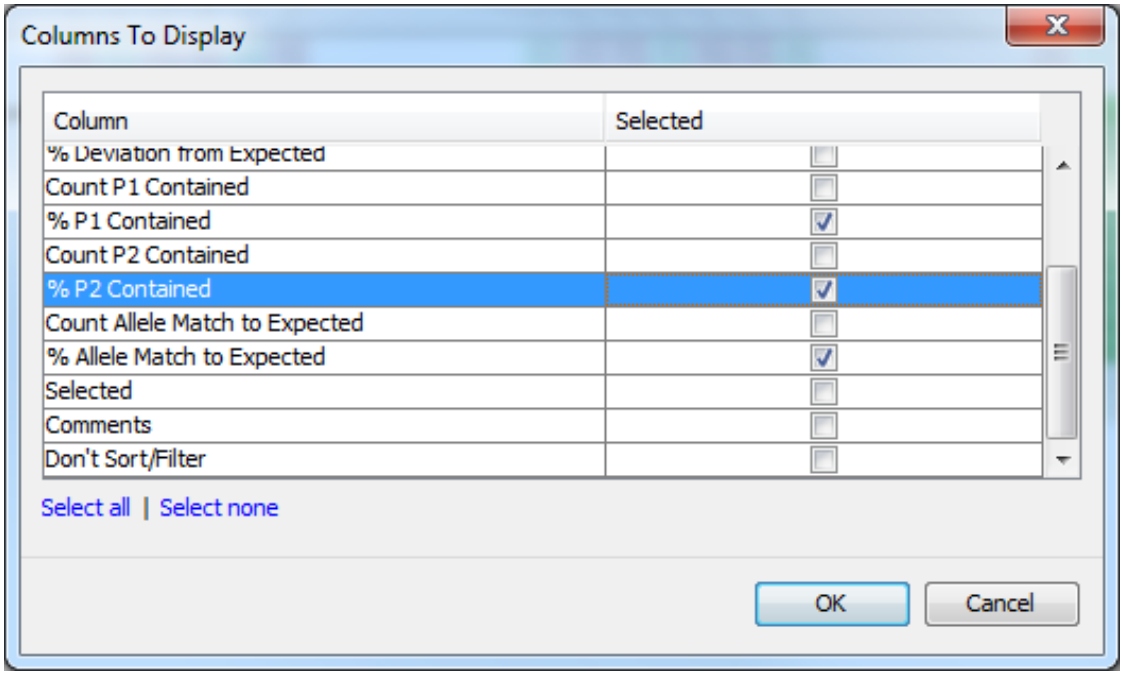
Click `Export` to open the `Export Results to File` dialog. You can select three separate types of export from this dialog, `All lines` which exports all of the data whether it had been filtered or not, `Only visible (non-filtered) lines` which outputs the results table as it appears in Flapjack, and `Only visible (non-filtered) lines that are selected` which outputs only the lines which are visible in the table and are selected. Select the third option: `Only visible (non-filtered) lines that are selected`. Ensure that the `Include header rows with details of any active filter or sort parameters` is selected, this outputs information about the filtering and sort that was applied to reach the current view of the data, which can be useful for the purpose of reproducing the steps at a later date. Click `Browse` to select a location to save your file, as well as a file name. Finally click `Export` to output the data to file.

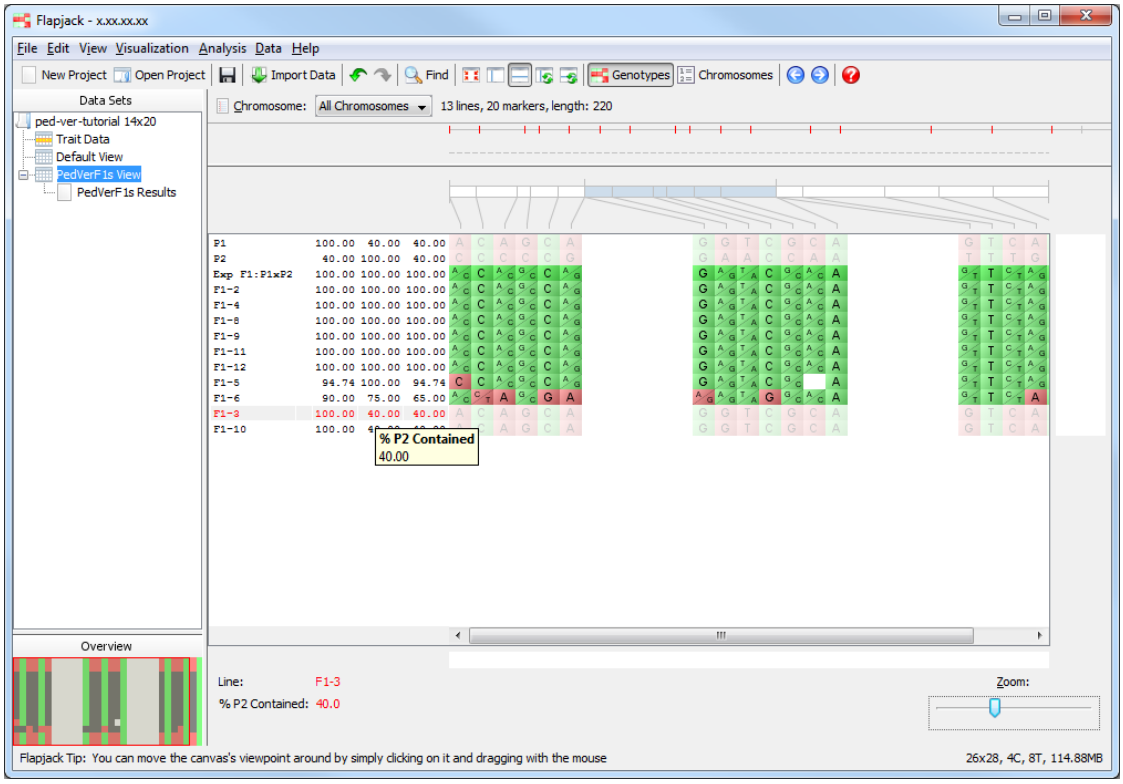


You can view the contents of the file in any text editor.

36.9 Viewing results in the genotype view

Return to the genotype view of the data (`PedVerF1s View`) and zoom in until you can comfortably read the line names which are just to the left of the genotype visualization. Right-click on the list of line names and select `Show table results` from the menu to open the `Columns To Display` dialog. This dialog allows you to select columns of data from the results table to view side by side with the genotype data. This can be good for a final visual validation of what you're seeing from the analysis results. Select `% P1 Contained`, `% P2 Contained`, and `% Allele Match to Expedted` by clicking the corresponding checkboxes in the dialog, then click `OK`.





You should see that there are three extra columns of data between the lines names and the genotype display. These are the four columns you selected in the `Columns To Display` dialog and represent the data found for those columns for the lines in `PedVerF1s Results` view. Mouseover the columns to see tooltips with the column name and value for the line under the mouse, this information is also displayed in the status panel at the bottom of the display. Here we can see that one of the lines we've deselected has a value of % P1 Contained of 100%, but only 40% for % P2 Contained and % Allele Match to Expected. This suggests that the line under the mouse is a self of parent 1.

36.10 Conclusion

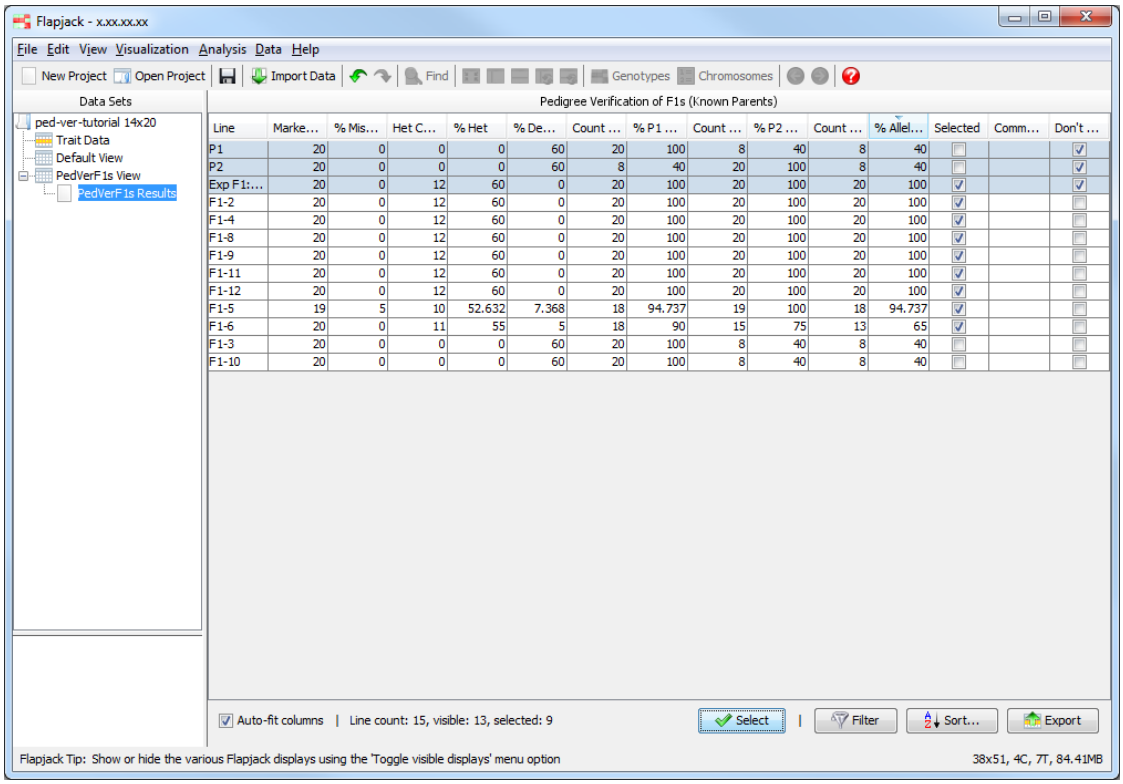
You've now had a chance to experiment with running a markers assisted backcrossing analysis in Flapjack and analysing the results of that analysis. Feel free to experiment more by applying different sorts, filters, selections, or better yet make a start on analysing your own data.

CHAPTER 37

Analysis Results Tables

For certain types of analysis, such as *Marker Assisted Back Crossing* or *Pedigree Verification (F1s Known Parents)* Flapjack will present you with a results table. While the content of this table differs between analyses, you can interact with the table in the same way regardless of the analysis you have run. You can select, filter and sort lines, directly in the table, or through dialogs (for more complex options) and also export the data to a file.

This is a very powerful view because all lines shown in the table will reflect their sort order, filtered state, and selection state with the main visualization. This means you could, for example, sort the lines using the table and have that order reflected in the main view, or sort the main visualization and have that order automatically apply to the table.



Line	Marke…	% Mis…	Het C…	% Het	% De…	Count …	% P1 …	Count …	% P2 …	Count …	% Allel…	Selected	Comm…	Don't …
P1	20	0	0	0	60	20	100	8	40	8	40	☐		☑
P2	20	0	0	0	60	8	40	20	100	8	40	☐		☑
Exp F1:…	20	0	12	60	0	20	100	20	100	20	100	☑		☑
F1-2	20	0	12	60	0	20	100	20	100	20	100	☑		☐
F1-4	20	0	12	60	0	20	100	20	100	20	100	☑		☐
F1-8	20	0	12	60	0	20	100	20	100	20	100	☑		☐
F1-9	20	0	12	60	0	20	100	20	100	20	100	☑		☐
F1-11	20	0	12	60	0	20	100	20	100	20	100	☑		☐
F1-12	20	0	12	60	0	20	100	20	100	20	100	☑		☐
F1-5	19	5	10	52.632	7.368	18	94.737	19	100	18	94.737	☑		☐
F1-6	20	0	11	55	5	18	90	15	75	13	65	☑		☐
F1-3	20	0	0	0	60	20	100	8	40	8	40	☐		☐
F1-10	20	0	0	0	60	20	100	8	40	8	40	☐		☐

Note that you can right click anywhere on the table to access most of the options listed below.

37.1 Selecting lines

Select or deselect a single line by clicking the checkbox in the `Selected` column for that line. To select all the lines in the results table, click the `Select` button at the bottom of the display and pick `Select all`. To deselect all of the lines or invert the selection state of the lines use the `Select none` or `Invert selection` options respectively.

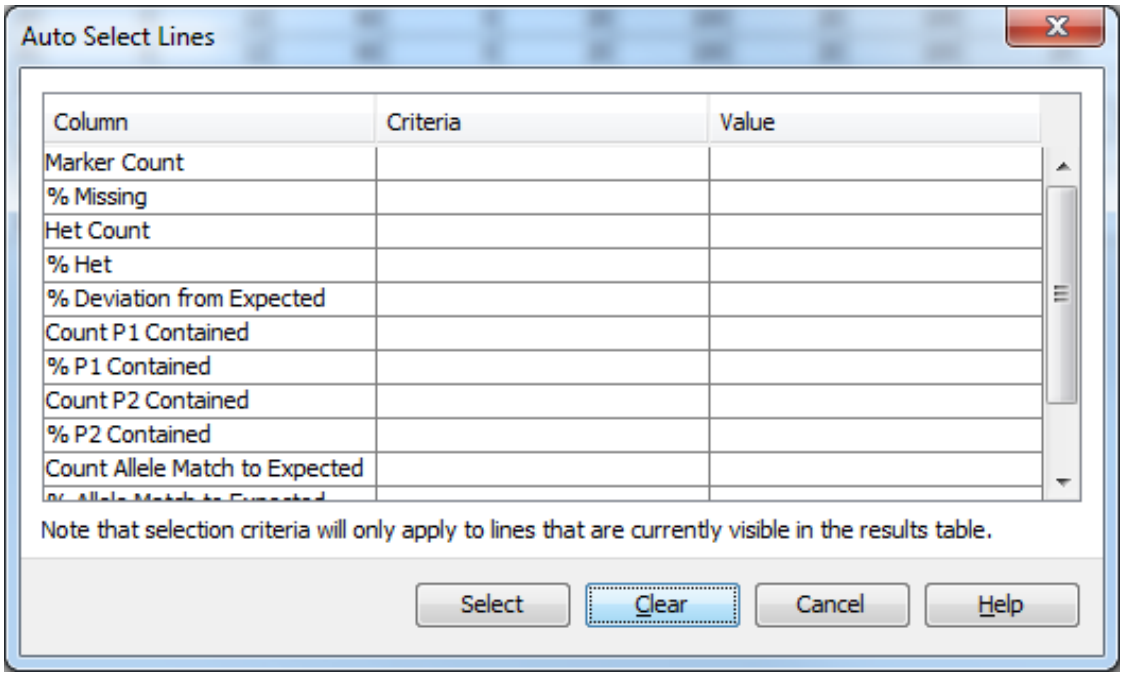
37.2 Auto selecting lines

To select based on a set of criteria click the `Select` button and choose `Auto select` from the popup. This opens the `Auto Select` dialog, where you can specify criteria and values for the results table columns you wish to select on.

For numerical columns pick a criteria from the dropdown menu (eg less than or greater than or equal to) and type a value for that logic operation in the adjoining textbox.

For boolean (true/false) columns (shown as checkboxes in the main table), pick either true or false from the criteria dropdown menu - there is no need to provide a value in the adjoining textbox.

Finally, click `Select` to auto-select lines in the results table based on the options you have provided, or `Clear` to reset the options and start again.

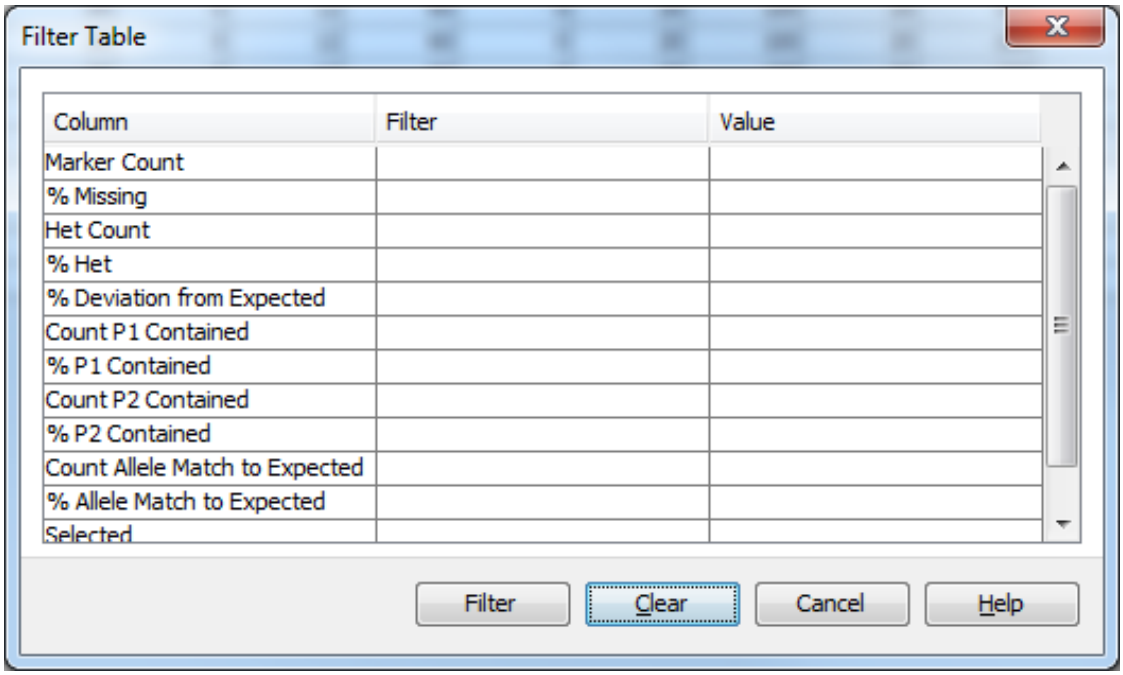


37.3 Filtering lines

To filter lines based on a set of criteria, click the `Filter` button and pick `Filter` from the dropdown menu that appears. For each table column you wish to filter by, specify a filter type and a value. Filter types can be picked from the dropdown menu for each table column and comprise Less than, Less than or equal to, equal to, greater than or

equal to, greater than, or, Does not equal filters. For each column you have specified a filter type you must also specify a value in the adjacent text field, unless the column you are filtering on contains a checkbox, in which case you can only pick from the filter types True and False and can't supply a value.

Once you've set up your filters, you can click `Filter` which will apply the filters to the results table. You can also click `Clear` to reset the dialog to its default state, or `Cancel` to close without doing anything. If you decide you'd like to remove your filters and see all of the data again click `Filter` and select `Reset filters` from the dropdown that appears.

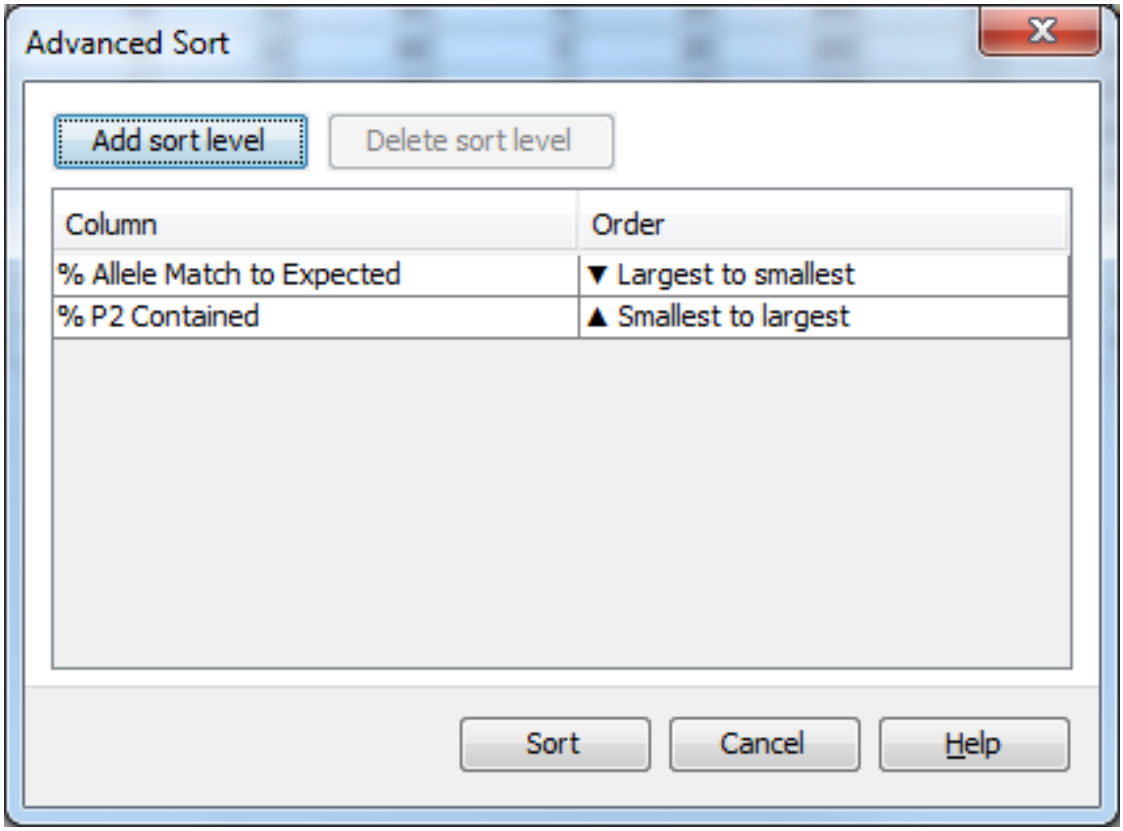


37.4 Sorting lines

To sort a results table you can click the column header of the column you wish to sort by to apply a sort. In the first instance this will sort the table by ascending order of the values in that column. Click on the column header again to toggle the sort to order the table by descending order of the values in your chosen column.

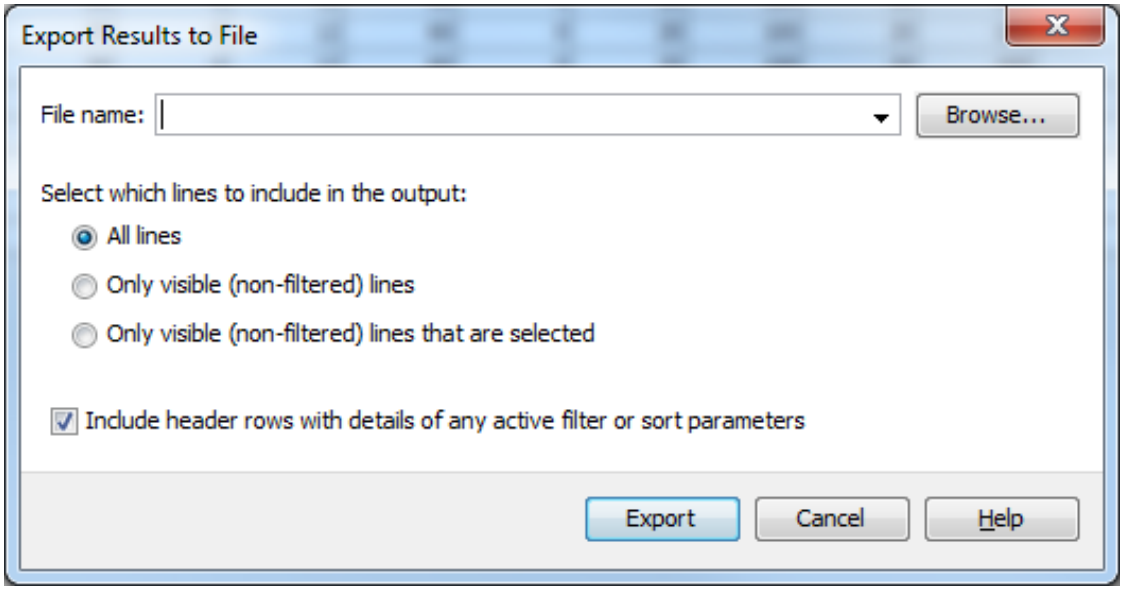
To sort by multiple columns at once click the `Sort` button on the results table view. In the `Advanced Sort` dialog pick the column you wish the primary sort to be on from the column dropdown and the order for the primary sort (either Largest to smallest or Smallest to largest). Click the `Add sort level` button and another set of dropdowns for column and sort order are added to the table. This allows you to sort by multiple columns, with any rows which contain the same value for the primary sort, are then ordered within themselves by the secondary (tertiary, n-ary...) sort order. As an example, in marker assisted backcrossing, if you're primarily interested in lines which have a high QTL allele count value and secondarily in columns with a high RPP total value, you can add two sort levels select QTL allele count and Largest to smallest for the first sort level, then RPP total and Largest to smallest for the second sort level.

Highlight a sort level and click `Delete sort level` to remove any unwanted sort levels. Click `Sort` to apply the sort to the table, or `Cancel` to close the dialog without carrying out the sort.



37.5 Exporting to a file

Click the `Export` button to open the `Export Results to File` dialog. Click `Browse` to select the location to save your file and your desired filename. Select the `All lines` option to export all of the table data (including lines which have been filtered out from the table), with the current sort order applied. Select the `Only visible (non-filtered) lines` option to export the table as you see it in the display. Select the `Only visible (non-filtered) lines that are selected` to export only the data which is visible in the table and has a tick in the selected column. Pick if you wish to include details of any filtering and sorting which is active on the results in the table by ensuring the `Include header rows with details of any active filter or sort parameters` is selected. Click `Export` to export the table results to your chosen file, or click `Cancel` to close the dialog without exporting the data.



37.6 Ranking lines

Click and drag on the table to highlight the lines you wish to rank. Once highlighted, click the `Rank` button to open the `Rank Lines` dialog. Enter an integer value as the rank you wish to give the highlighted lines and click `Ok` to apply that rank to those results, or `Cancel` to close the dialog without ranking the results.

37.7 Don't sort/filter

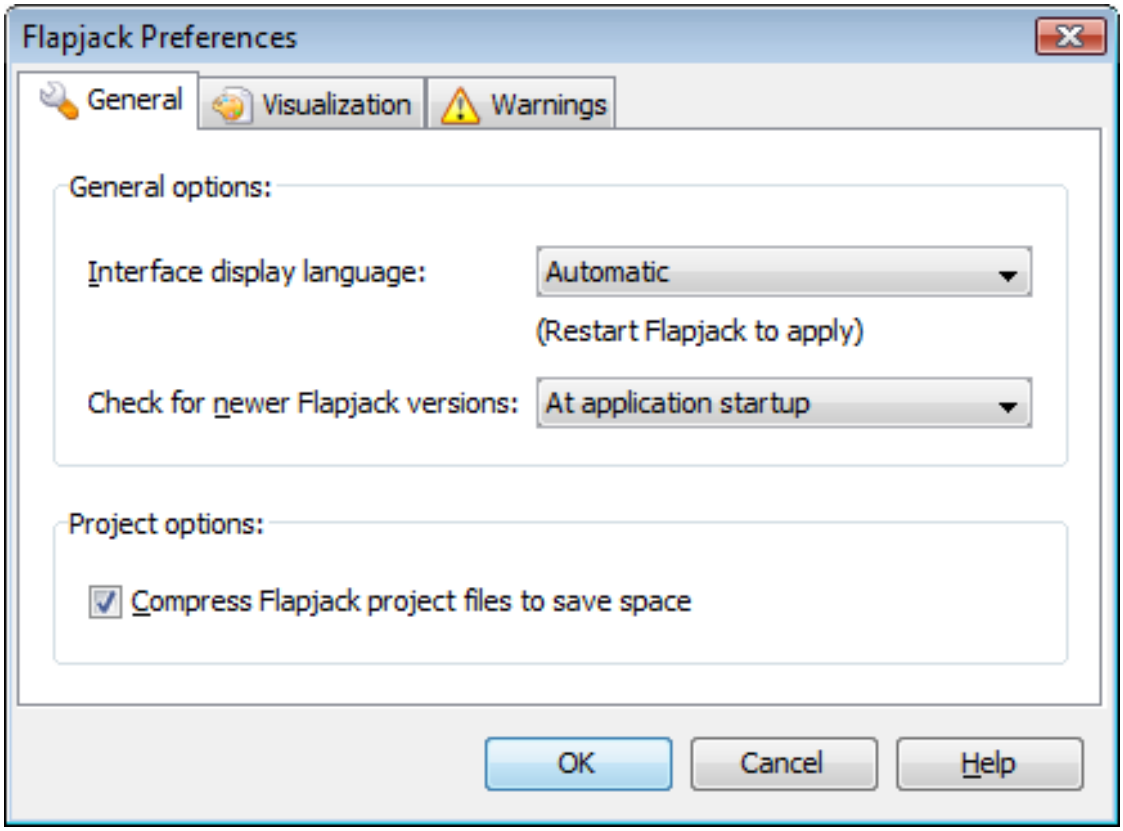
When you run an analysis certain lines may be automatically marked as `Don't Sort/Filter` lines. This means that any table based sorting or filtering won't apply to those lines, in the case of sorting keeping these lines at the top of the table and in the case of filtering keeping them in the table even if they don't match the filter criteria. To mark more lines that shouldn't be sorted, or filtered, click the checkbox for those lines in the `Don't Sort/Filter` column of the table. They won't immediately move to the top of the table, but will do so the next time any sorting, or filtering occurs.

CHAPTER 38

Preferences

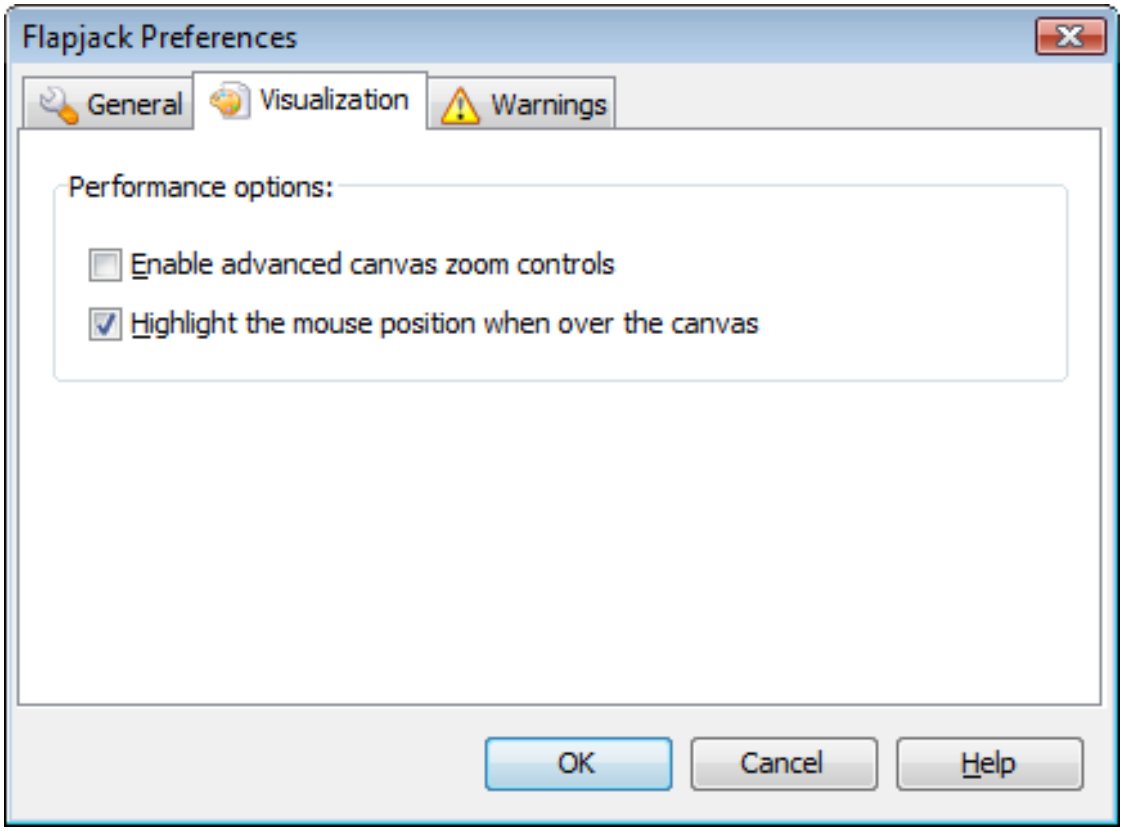
The `Preferences` dialog (`Help->Preferences`) can be used to modify various settings that affect the way Flapjack is used and is displayed.

38.1 General



- `Interface display language` - this setting determines what language Flapjack will display its user interface in. `Automatic` will attempt to pick the most suitable language based on your operating system's settings. If this is not correct, or you wish to use another language, you can also select from `English (UK)`, `English (US)` , or `German`. Flapjack will need to be restarted after any changes are made.
- `Check for newer Flapjack versions` - this setting determines how often Flapjack will attempt to connect back to its download server to see if a newer version is available. The options available are `Never`, `At application startup`, `Once a day`, `Once a week`, and `Once a month`.
- `Compress Flapjack project files to save space` - toggles on or off project compression. By default, project files are zip-compressed which significantly reduces the final file size, however, with very large data sets this may increase the time taken to save and/or load projects.

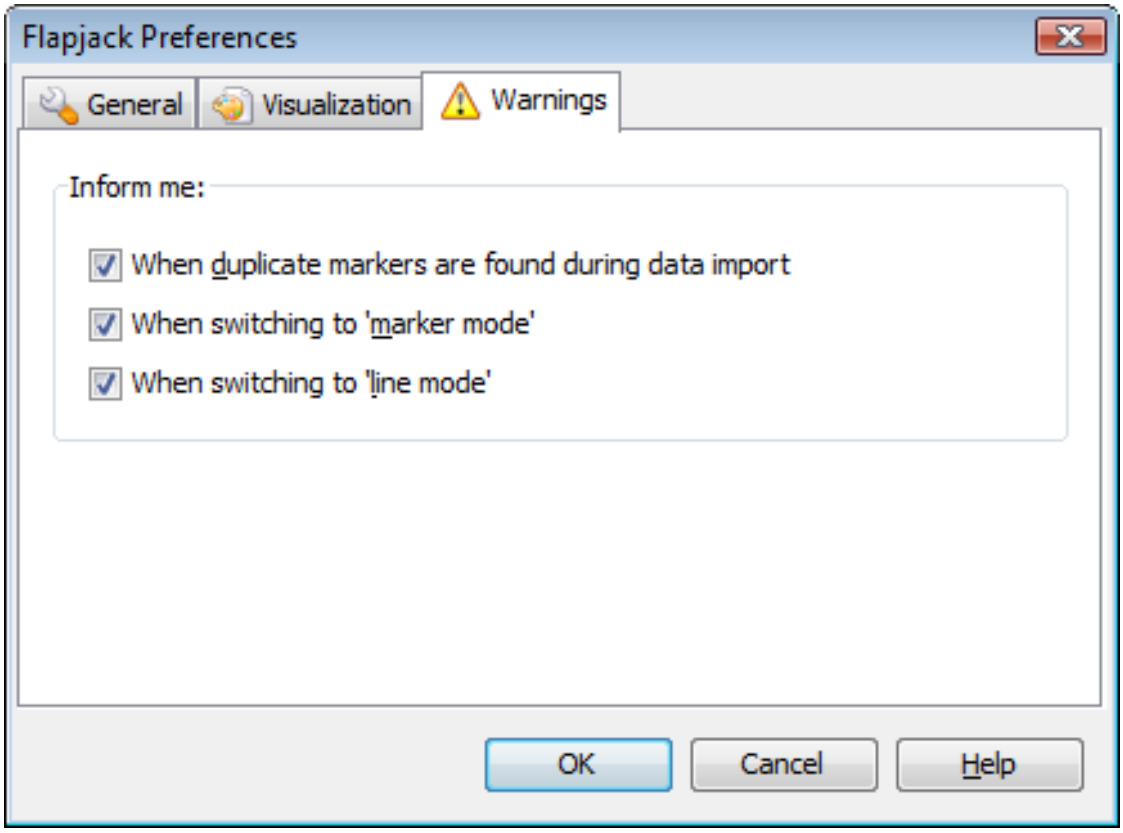
38.2 Visualization



- `Enable advanced canvas zoom controls` - selecting this option will change the standard "zoom" control option for zooming in and out of the display into separate horizontal and vertical zoom controls. This gives more control of the display but may reduce the effect of some visualizations. Use with caution.
- `Highlight the mouse position when over the canvas` - selecting this option will graphically highlight the line and the marker currently under the mouse cursor's position as it is moved over the main display area.

38.3 Warnings

Flapjack can inform you when certain events take place. Once you are used to these events, you may wish to stop further reminders. These options allow for this.



- `When duplicate markers are found during data import` - every marker (and line) within a data set imported into Flapjack must be unique. In the case of markers, Flapjack will continue to import the data even after a duplicate marker has been found, however it will only use the first instance that it came upon. Selecting this option will cause Flapjack to display the `Duplicate Markers` dialog that informs you if duplicates are found.
- `When switching to 'marker mode'` - selecting this option will cause Flapjack to remind you that you are now working in marker mode (along with brief instructions) if a switch is made to marker mode.
- `When switching to 'line mode'` - selecting this option will cause Flapjack to remind you that you are now working in line mode (along with brief instructions) if a switch is made to line mode.

CHAPTER 39

Tips and Shortcuts

Hints and tips on how to use Flapjack are shown periodically in its status bar running along the bottom edge of its main window. For easy reference, the following is a list of all the tips that may appear:

- Hold CTRL while clicking and dragging lines or markers to move them to new positions
- You can import and work with multiple data sets at once, all within a single project
- Many of Flapjack's menu options are also accessible by right-clicking on the display canvas
- Compress projects to save disk space by selecting the option in the Preferences Dialog
- The ordering of lines is common across all chromosomes within a view
- The Overview Dialog (F7) displays the entire data set, scaled to fit any window size
- Navigate around a view quickly by clicking and dragging on one of the Overview displays
- Flapjack will periodically check for new versions at startup
- The red rectangle on an Overview display shows the region of the data set currently being viewed on the main canvas
- Mouse over any allele position to view further information on it
- Toggle the overlaying of actual genotype data onto the canvas by pressing CTRL G
- You can search for lines or markers by name by using the Find Dialog (CTRL F)
- Search for all lines beginning with the letter 'A' by using the Find Dialog's regular expression: A.*
- You can move the canvas's viewpoint around by simply clicking on it and dragging with the mouse
- Zoom in or out by using the slider, or by holding CTRL and scrolling the mousewheel
- Create multiple custom views of the same data set by selecting 'Visualization->Create new view' from the menu bar
- Please send any comments, feature requests, bug reports, etc, to flapjack@hutton.ac.uk
- Customize the various colour schemes by selecting the 'Customize' option from within the 'Colour schemes' sub-menu

- Show or hide the various Flapjack displays using the 'Toggle visible displays' menu option
- Quickly hide individual lines or markers by CTRL double clicking on them (while line or marker mode is active)
- Switch to 'marker mode' to select, deselect, or toggle marker visibility
- Switch to 'line mode' to select, deselect, or toggle line visibility
- Quickly track locations of interest by right-clicking on the display and selecting 'Bookmark location'
- Move QTLs between tracks by holding CTRL and dragging them with the mouse
- Change which QTLs are visible by selecting Data->Filter QTLs from the menu bar

CHAPTER 40

Command Line Support

Flapjack is provided along with various command-line utilities.

The utilities are run in a different way depending on the platform used. For all systems excluding macOS, you can, for example, find the `createproject` executable (.exe or .sh) located in the root folder where Flapjack is installed. For macOS, you must manually run it using:

```
java -cp lib/flapjack.jar jhi.flapjack.io.CreateProject <options>
```

A description of the input file formats accepted by Flapjack is given *Projects & Data Formats*.

40.1 Advanced Options

These options can be used (where they make sense) with the command line programs specified below where the programs accept Flapjack files as input.

```
-A, --all-chromosomes          duplicate all markers onto a single All Chromosomes␣
↪chromosome for side-by-side viewing
-C, --collapse-heteozygotes    don't distinguish between heterozygous alleles (eg␣
↪treat A/T the same as T/A)
-S, --heterozygous-separator   the string used to separate heterozygous alleles␣
↪(default is / or use "" for no separator
-M, --missing-data             the string used to represent missing data (default is -
↪ or use "" for empty string
-T, --transposed               genotype data is transposed compared to Flapjack's␣
↪default
-E, --decimal-english          override locale default and use '.' as the decimal␣
↪separator
-D, --allow-duplicates         allow duplicate line names in input files
```

40.2 createproject.exe (jhi.flapjack.io.CreateProject)

This program can be used to pre-create .flapjack project files from existing tab-delimited text files. This ability allows for the creation of project files outwith the Flapjack environment, for instance, to allow a web server (that links to a database) to make Flapjack project files available for download.

The following options are available:

```
-g, --genotypes <genotypes_file>       the location of the file containing genotype␣
↪data (required)
-p, --project <project_file>           the name of the project file that will be␣
↪created (required)
-m, --map <map_file>                   the location of the file containing map data␣
↪(optional)
-t, --traits <traits_file>             the location of the file containing trait data␣
↪(optional)
-q, --qtls <qtls_file>                 the location of the file containing QTL data␣
↪(optional)
-n, --name <string>                    a name for the dataset to be created in the␣
↪Flapjack project (optional)
```

For example:

```
createproject.exe -m input.map -g input.dat -p output.flapjack
```

40.3 creatematrix.exe (jhi.flapjack.io.CreateMatrix)

This program will take input data files and run Flapjack's *Similarity Matrix Creation* module upon them, outputting a matrix file for use elsewhere (eg in R).

The following options are available:

```
-g, --genotypes <genotypes_file>       the location of the file containing genotype␣
↪data (required)
-o, --output <output_file>             the name of the matrix file that will be␣
↪created (required)
-m, --map <map_file>                   the location of the file containing map data␣
↪(optional)
-p, --project <project_file>           the name of the project file that will be␣
↪created (optional)
```

For example:

```
creatematrix.exe -g input.dat -o output.txt
```

40.4 mabcstats.exe (jhi.flapjack.io.GenerateMabcStats)

This program will take input data files and run Flapjack's *Marker Assisted Back Crossing* statistics module upon them, outputting a tab-delimited text file with results similar to those shown directly in Flapjack's table view had the UI been used.

The following options are available:

```
-m, --map <map_file>                              the location of the file containing␣
↪map data (required).
-g, --genotypes <genotypes_file>                  the location of the file containing␣
↪genotype data (required).
-q, --qtls <qtls_file>                            the location of the file containing␣
↪QTL data (required).
-r, --recurrent-parent <index_of_line>            the index (1-based) of the recurrent␣
↪parent in the file (required).
-d, --donor-parent <index_of_line>                the index (1-based) of the donor␣
↪parent in the file (required).
--model weighted|unweighted                       the model to run (required).
-o, --output <file_name>                          the name of the output file that will␣
↪be created (required).
-c, --max-marker-coverage <coverage_value>        the maximum coverage per marker in cM␣
↪(optional)
-p, --project <project_file>                      the name of the project file that will␣
↪be created (optional)
```

For example:

```
mabcstats.exe -m input.map -g input.dat -q input.qtl -r 1 -d 2 --model weighted -o␣
↪mabc.txt
```

40.5 pedverf1stats.exe (jhi.flapjack.io.GeneratePedVerF1sStats)

This program will take input data files and run Flapjack's *Pedigree Verification (F1s Known Parents)* statistics module upon them, outputting a tab-delimited text file with results similar to those shown directly in Flapjack's table view had the UI been used.

The following options are available:

```
-m, --map <map_file>                      the location of the file containing map data␣
↪(required)
-g, --genotypes <genotypes_file>          the location of the file containing genotype␣
↪data (required)
-f, --parent1 <index_of_line>             the index (1-based) of the first parent in the␣
↪file (required)
-s, --parent2 <index_of_line>             the index (1-based) of the second parent in␣
↪the file (required)
-o, --output=<file_name>                  the name of the output file that will be␣
↪created (required)
-e, --expectedf1 <index_of_line>          the index (1-based) of a line to use as the␣
↪expected F1 (optional)
-p, --project <project_file>              the name of the project file that will be␣
↪created (optional)
```

For example:

```
pedverf1stats.exe -m input.map -g input.dat -f 1 -s 2 -o pedver.txt
```

40.6 splitproject.exe (jhi.flapjack.io.SplitProject)

This program can be used to take an existing .flapjack project file and filter out the raw data again as a collection of tab-delimited plain text files.

Note: This program uses an older style of command line argument parsing and will be updated in a future release.

The following options are available:

```
-project=<project_file>      the location of the project to process (required)
-dir=<directory>             the location to write the output files to (required)
-datasetin=<dataset_name>    the name of a dataset within the project file to process.
↪ If no names are specified, then all datasets will be extracted (optional)
-datasetout=<dataset_name>   overrides the given datasetin name with a new name to␣
↪use when outputting that dataset's files (optional)
-decimalEnglish              whether to always parse numbers assuming they contain␣
↪the English decimal separator, dot rather than comma (optional)
```

For example:

```
splitproject.exe -project=input.flapjack -dir=outputdir
```