BioNumerics®

MICROBIAL DATA ANALYSIS SOFTWARE

BandScoring plugin

PLUGINS

VERSION 7.6

www.applied-maths.com

Contents

NOTES

SUPPORT BY APPLIED MATHS

While the best efforts have been made in preparing this manuscript, no liability is assumed by the authors with respect to the use of the information provided.

Applied Maths will provide support to research laboratories in developing new and highly specialized applications, as well as to diagnostic laboratories where speed, efficiency and continuity are of primary importance. Our software thanks its current status for a part to the response of many customers worldwide. Please contact us if you have any problems or questions concerning the use of BioNumerics®, or suggestions for improvement, refinement or extension of the software to your specific applications:

Applied Maths NV
Keistraat 120
9830 Sint-Martens-Latem
Belgium
PHONE: +32 9 2222 100
FAX: +32 9 2222 102
E-MAIL: info@applied-maths.com
URL: http://www.applied-maths.com

Applied Maths, Inc.
11940 Jollyville Road, Suite 115N
Austin, Texas 78759
U.S.A.
PHONE: +1 512-482-9700
FAX: +1 512-482-9708
E-MAIL: info-US@applied-maths.com

LIMITATIONS ON USE

The BioNumerics® software, its plugin tools and their accompanying guides are subject to the terms and conditions outlined in the License Agreement. The support, entitlement to upgrades and the right to use the software automatically terminate if the user fails to comply with any of the statements of the License Agreement. No part of this guide may be reproduced by any means without prior written permission of the authors.

BioNumerics® uses following third-party software tools and libraries:

- The Python® 2.7.4 release from the Python Software Foundation (http://www.python.org/).
- A library for XML input and output from the Apache Software Foundation (http://www.apache.org).
- NCBI toolkit version 2.2.10 (http://www.ncbi.nlm.nih.gov/BLAST/).
- The Boost c++ libraries (http://www.boost.org/).
- Samtools for interacting with SAM / BAM files (http://www.htslib.org/download/)
- The 7-Zip command line version (7za.exe) from 7-Zip, copyright 1999-2010 Igor Pavlov. http://www.7-zip.org/
- Velvet for Windows, source code can be downloaded from http://www.applied-maths.com/download/open-source.
- Ray for Windows, source code can be downloaded from http://www.applied-maths.com/download/open-source.
- Mothur for Windows, source code can be downloaded from http://www.applied-maths.com/download/open-source.
- Cairo 2D graphics library version 1.12.14 (http://cairographics.org/).
- Crypto++ Library version 5.5.2 (http://www.cryptopp.com/).
- libSVM library for Support Vector Machines (http://www.csie.ntu.edu.tw/~cjlin/libsvm/).
- SQLite version 3.7.17 (http://www.sqlite.org/).
- Gecko engine version 21 (https://developer.mozilla.org/en-US/docs/Mozilla/Gecko).
- pymzML Python® module for high throughput bioinformatics on mass spectrometry data (https://github.com/pymzml/pymzML).
- Numpy Python® library version 1.8.1 (http://www.numpy.org/).
- BioPython Python® library version 1.64 (http://www.biopython.org/).
- PIL Python library® version 1.1.7 (http://www.pythonware.com/products/pil/).
- The SPAdes genome assembler version 3.7.1 (http://bioinf.spbau.ru/spades).

Chapter 1

Starting and setting up BioNumerics

1.1 Introduction

This guide is designed as a tutorial for the *BandScoring plugin* of BioNumerics. In its simplest form, scoring a band is the process of determining whether a band is absent, uncertain or present at a certain position in a fingerprint pattern. The *BandScoring plugin*, however, is more advanced and addresses a common problem in plant breeding, where one wants to know if a genetic marker of interest is either homozygously, heterozygously or not at all occurring in the genome of a hybrid. Such a genetic marker can be a band e.g. on an Amplified Fragment Length Polymorphism (AFLP) gel and corresponds to a (desired or undesired) phenotypic trait.

Consider a mating experiment (see Figure 1.1) with two parent lines (P_1 and P_2), each having a desirable genetic marker that one would like to see homozygously combined in the offspring. The $P_1 \times P_2$ mating will result in an F_1 offspring that is heterozygous for both markers. The subsequent $F_1 \times F_1$ cross generates F_2 offspring, displaying *segregation* of the genetic markers. For each of the heterozygous markers in the F_1 generation, there is 25% chance that this marker will be absent, 50% chance that it will be heterozygous and 25% chance that it will be homozygous.

For the two unlinked genetic markers in our example, there is a 6.25% (or 1 out of 16) chance that both markers are homozygous in the F_2 offspring, i.e. that both desired properties inherited from P_1 and P_2 are ”pure” (highlighted in Figure 1.1).

In practice, however, the band height distribution differs from the situation depicted in Figure 1.1. This is due to differences in intensity between lanes (e.g. caused by variations in DNA concentration applied) and intensity differences along the run length (e.g. the typical ”ski-sloping” observed in chromatograms from an automated sequencer). Therefore, the band heights need to be calibrated within and between lanes before they can be scored as homozygous, heterozygous or absent.

The minimal configuration for the installation of this plugin includes the BioNumerics Fingerprint data and Character data modules.

1.2 Startup program

When BioNumerics is launched from the Windows start panel or when the BioNumerics shortcut () on your computer's desktop is double-clicked, the **Startup program** is run. This program shows the *BioNumerics Startup* window (see Figure 1.2).

A new BioNumerics database is created from the Startup program by pressing the  button.

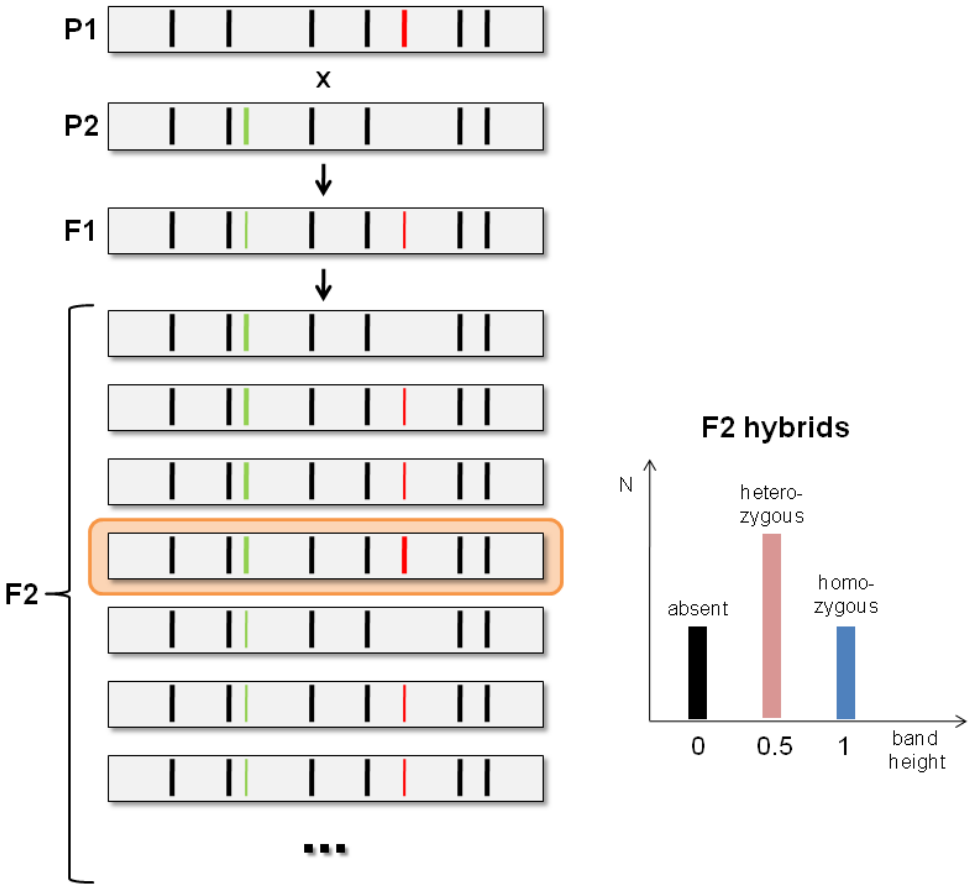

Figure 1.1: Example of a plant breeding experiment. The goal is to select F_2 offspring in which both genetic markers (indicated as colored bands) inherited from the parent lines are homozygously combined.

An existing database is opened in BioNumerics with  or by simply double-clicking on a database name in the list.

1.3 Downloading the BandScoring demonstration database

To illustrate the use of BioNumerics and the *BandScoring plugin*, a demonstration database is available on the Applied Maths website.

3.1 Click the ***Download example databases*** link, located in the lower right corner of the *BioNumerics Startup* window.

This calls the *Tutorial databases* window (see Figure 1.3).

3.2 Select the **BandScoring demo database** from the list and select ***Database*** > ***Download*** ().

3.3 Confirm the installation of the demo database and press <***Yes***> after successful installation of the database.

3.4 Close the *Tutorial databases* window with ***File*** > ***Exit***.

The **BandScoring demo database** appears in the *BioNumerics Startup* window.

3.5 Double-click the **BandScoring demo database** in the *BioNumerics Startup* window to open the database.

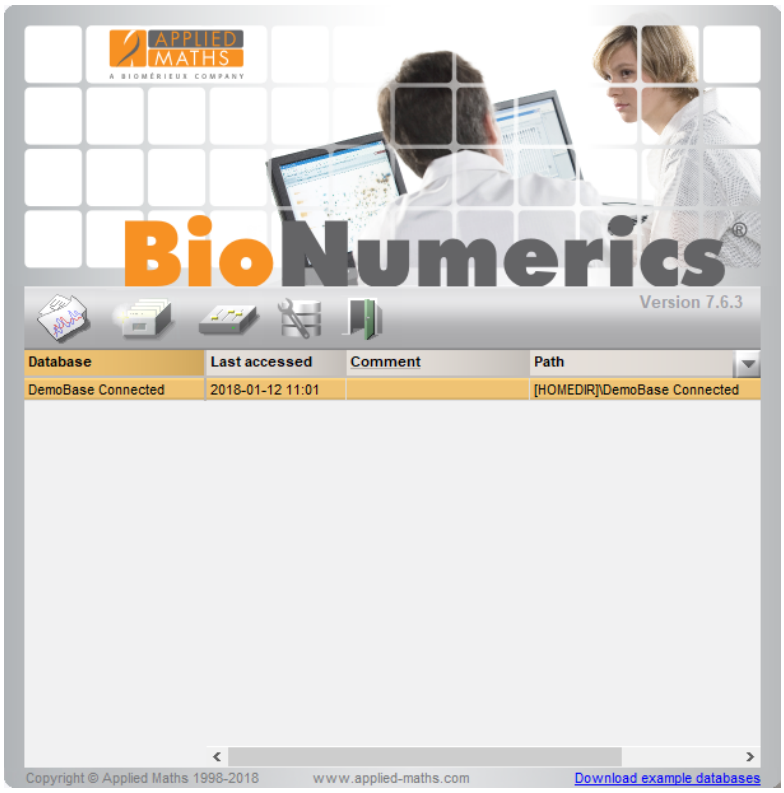

Figure 1.2: The *BioNumerics Startup* window.

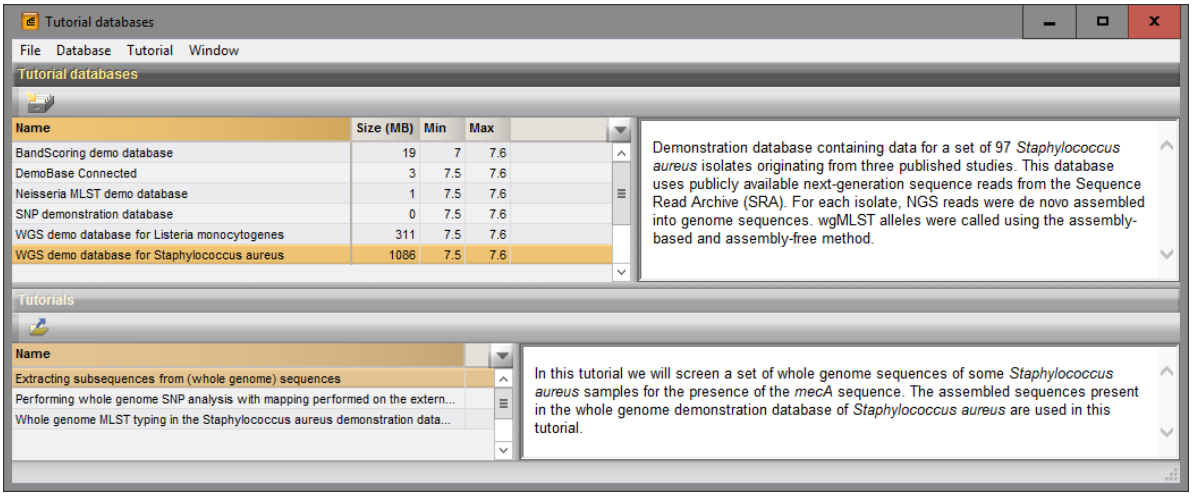

Figure 1.3: The *Tutorial databases* window, used to download the BandScoring demonstration database.

The database that opens should now contain 62 *Lycopersicon lycopersium* entries and one size marker. The database furthermore has one fingerprint type experiment called **AFLP**, with one gel (**Gel1**) linked to it. For more information about the **BandScoring** demonstration database, see 2.1.

1.4 Installing the BandScoring plugin

If a database is opened for the first time, the *Plugins* dialog box will appear by default (see Figure 1.4).

If the database has already been opened previously, the *Plugins* dialog box can be called from the *Main* window by selecting ***File*** > ***Install / remove plugins...*** ().

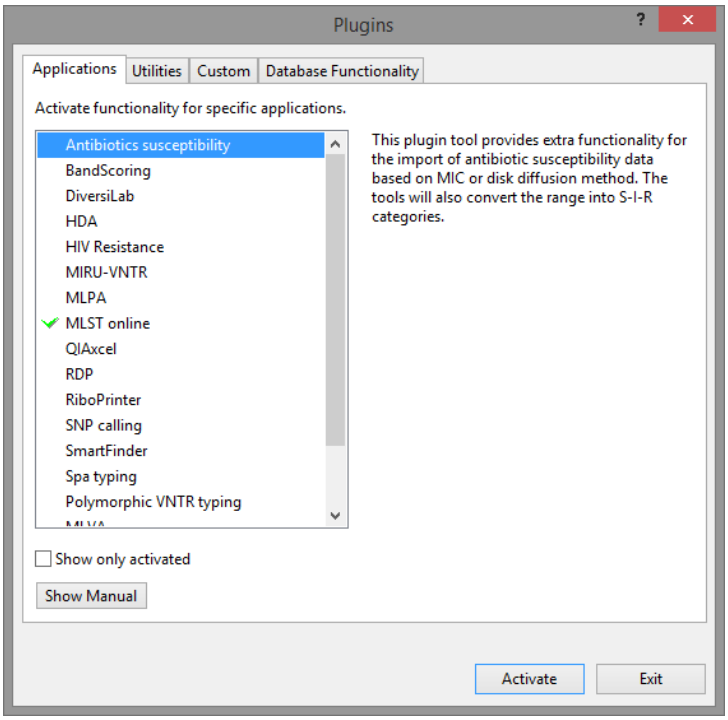

Figure 1.4: The *Plugins* dialog box.

When a particular plugin is selected from the list of plugins, a short description appears in the right panel.

A selected plugin can be installed with the <***Activate***> button. The software will ask for confirmation before installation. Some plugins depend on functionality offered by specific BioNumerics modules. If a required module is missing, the plugin cannot be installed and an error message will be generated.

Once a plugin is installed, it is marked with a green V-sign. It can be removed again with the <***Deactivate***> button.

If the selected plugin is documented, pressing <***Show Manual***> will open its manual in the *Help* window.

4.1 Select the *BandScoring plugin* from the list in the *Applications tab*, press the <***Activate***> button and confirm the installation of the plugin.

The *License string* dialog box pops up (see Figure 1.5). The *BandScoring plugin* can only be installed and activated with a valid *license number*, which needs to be purchased from Applied Maths.

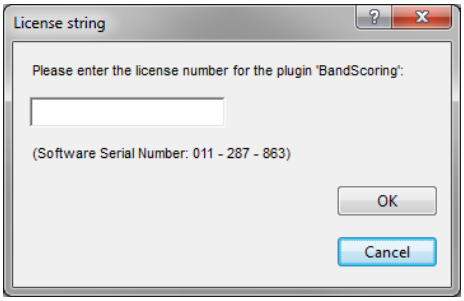

Figure 1.5: The *License string* dialog box.

The *License string* dialog box prompts for a license string that is compatible with the ***Software Serial***

Number listed in the dialog box.

4.2 Enter the six digits license string and press <***OK***>.

4.3 A message box pops up, confirming the installation of the plugin.

4.4 Press <***Proceed***> (or <***Exit***>) to close the *Plugins* dialog box and to continue to the *Main* window.

4.5 Close and reopen the database to activate the features of the *BandScoring plugin*.

The plugin functions can be called from the ***Bandscoring*** menu in the *Main* window and the *Comparison* window (see Figure 1.6).

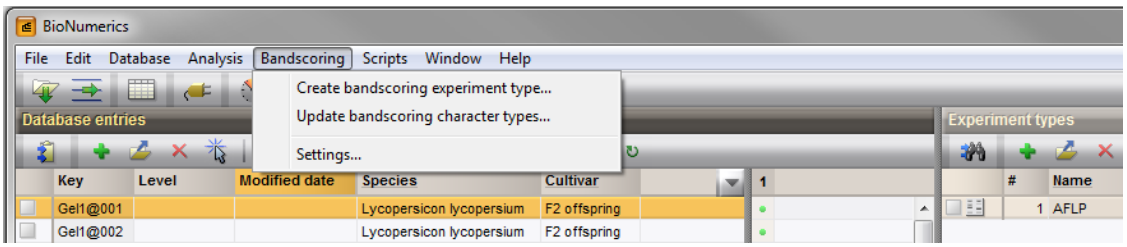

Figure 1.6: Bandscoring menu in the *Main* window.

Chapter 2

Working with the BandScoring plugin

2.1 Workflow for bandscoring in BioNumerics

Before explaining the functionality of the *BandScoring plugin*, the work flow and terminology used in BioNumerics might need some explanation. A schematic overview of the work flow is given in Figure 2.1. Any densitometric record seen as a one-dimensional profile of peaks or bands, such as gel and capillary electrophoresis patterns, is categorized as a *Fingerprint type* in BioNumerics. Data belonging to a fingerprint type are entered as *gels* (pseudo gels in case of capillary electrophoresis), on which *lanes* are defined that are linked to *entries* in the database. Each lane has a densitometric *curve* that is used in the numerical analysis. These curves can be optimized using background subtraction and filtering. Before curves can be compared with each other (intra- or inter gel), they need to be normalized against an external *reference system*. To local maxima on the densitometric curves, *bands* can be assigned. Bands occurring in different lanes can be assigned to *band classes*, with each class representing a certain biological entity. In the context of plant breeding, each band class thereby corresponds to a genetic marker. Bands that belong to a band class can be scored and then additional information becomes available about these bands. Bands are scored to one of the following five states: absent, uncertain, heterozygous, undefined or homozygous. This information is stored in two auxiliary *Character type* experiments, associated with the fingerprint type experiment. Character types are used in BioNumerics to store arrays of named characters. In case of the auxiliary character types created by the *BandScoring plugin*, the names of the characters correspond to the band classes that are saved with the fingerprint type. One character type (with suffix "_bm_pos") is used to store band positions, the other (with suffix _bm_zyg") is used to store the actual band scores and can be used for analysis.

The two auxiliary character types are required by the software to store information in. There will never be any need to manually alter information stored in these character sets. In fact, any manual change to the characters or character values could result in malfunction of the *BandScoring plugin*.

The **BandScoring** demonstration database (for installation instructions, see 1.3) contains the result of a fictitious tomato breeding experiment. In this experiments, the commercially available cultivar "Moneymaker", which produces fine-flavored tomatoes and is resistant against Blossom End Rot, is crossed with an experimental cultivar designated "NemaProof". This (fictitious) second cultivar is resistant against nematodes, but unfortunately produces very small and unpalatable fruits. Both cultivars grow well in humid conditions. The goal of the breeding program is of course to produce a homozygous F_2 offspring that is resistant to nematodes, but has the large, tasty fruit and Blossom End Rot resistance of the "Moneymaker" cultivar.

One AFLP gel is present in the database, on which the two parent lines are loaded and all remaining lanes represent F_2 offspring. This AFLP gel was already processed (lane finding, background subtraction, normalization and automatic band detection). For a detailed description of the fingerprint processing steps, we refer to the Reference manual, Chapter Setting up fingerprint type experiments. Depending on the nature of the data (TIFF or chromatogram files from automated sequencers), different work flows are followed.

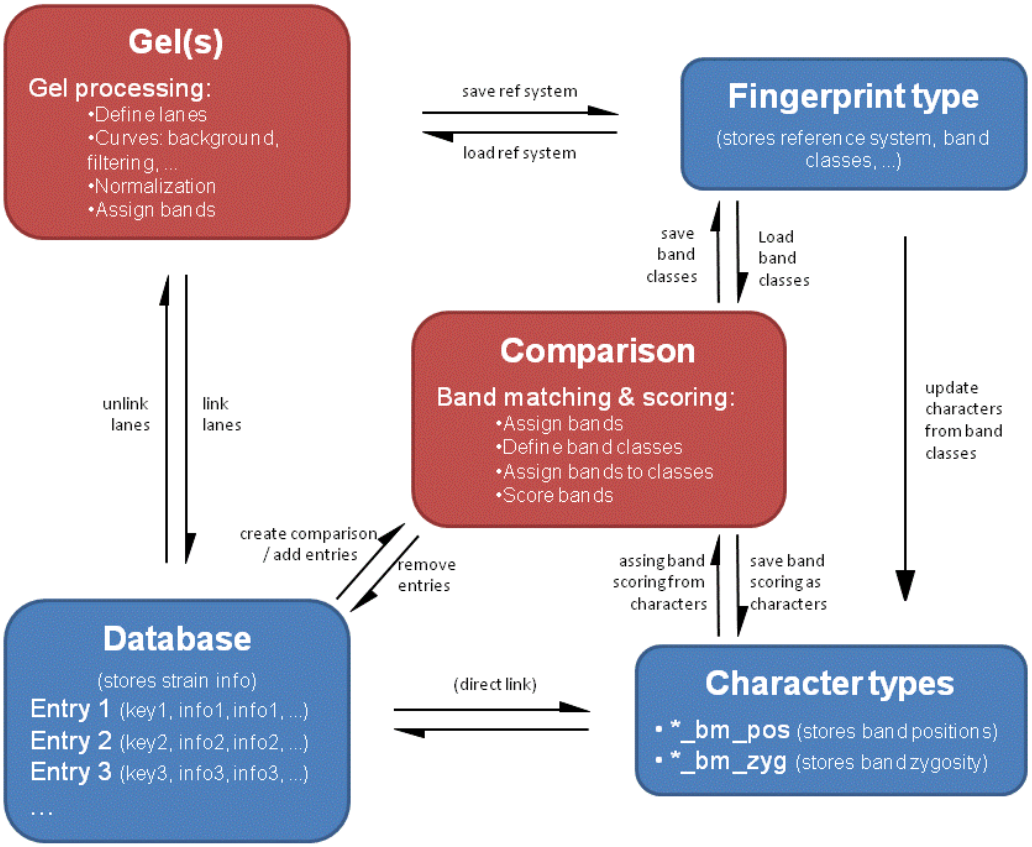

Figure 2.1: Work flow for bandscoring of a fingerprint type in BioNumerics. The elements that require processing are indicated in dark red, the elements that just store information are indicated in blue.

2.2 Creating bandscoring character type experiments

As stated above, BioNumerics uses two auxiliary character type experiments to store bandscoring information in. Therefore, before any bands can be scored, these character type experiments need to be created first.

2.1 In the *Main* window with the **BandScoring** database loaded, select ***Bandscoring*** > ***Create bandscoring experiment type***.

This calls the *Create bandscoring experiment* dialog box (see Figure 2.2).

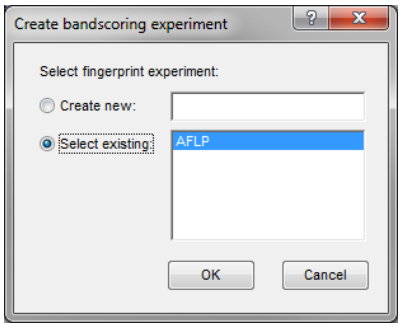

Figure 2.2: The *Create bandscoring experiment* dialog box.

The *Create bandscoring experiment* dialog box offers the option to create auxiliary character type experiments for an existing fingerprint type or to create a new fingerprint type and corresponding auxiliary character types for bandscoring.

2.2 Check ***Select existing***, make sure **AFLP** is selected and press <***OK***>.

Two additional character type experiments, called **AFLP_bm_pos** and **AFLP_bm_zyg**, are now created and appear in the *Experiment types* panel. No characters are present yet in these experiment types.

The colors in which the band scores (Absent, Heterozygous, Undefined, and Homozygous) will be displayed, can be set.

2.3 Select ***Bandscoring*** > ***Settings***.

This pops up the *Bandscoring color settings* dialog box (see Figure 2.3).

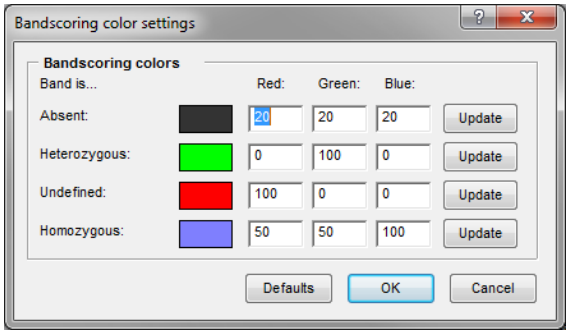

Figure 2.3: The *Bandscoring color settings* dialog box, from which the bandscoring colors can be set.

To modify a display color, type a number for the red, green and blue values (in percent) in the corresponding boxes and press <***Update***>.

To reset the default display colors for all band scores, press <***Defaults***>.

Press <***OK***> when the desired display colors are obtained or <***Cancel***> if you want to leave the bandscoring colors unaltered.

2.3 Creating a comparison to perform a bandscoring

A bandscoring is always done in the *Comparison* window, for a comparison containing any entries you want to score the bands from (see Figure 2.1). These entries could be contained in a single gel or sequencer run (as is the case in the **BandScoring** database), but they could also originate from different gels (runs) if the gels (runs) were properly normalized (see the Reference manual, Chapter Setting up fingerprint type experiments).

In the **BandScoring** database, we want to perform a bandscoring on both parent patterns and all F_2 hybrids.

3.1 In the *Main* window, with the **BandScoring** database loaded, select all entries in the database except the size standard, e.g. use **Ctrl+A** to select all entries and unselect the size marker.

Check boxes for selected entries are indicated as ☑.

3.2 Click on the ➕ button in the *Comparisons* panel to create a new comparison for the selected entries.

3.3 Select ***File*** > ***Save*** (💾, **Ctrl+S**) to save the comparison. When prompted for a name, enter e.g. “BreedingExp1”.

3.4 In the *Experiments* panel of the *Comparison* window, click on the eye icon ( ) that proceeds **AFLP** to display the **AFLP** gel image.

3.5 Select ***Fingerprints*** > ***Perform band matching...*** ( ). This pops up the *Perform band matching* dialog box (see Figure 2.4).

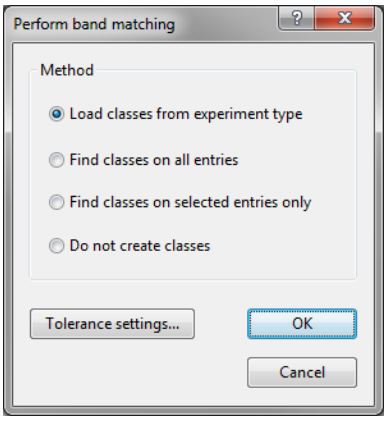

Figure 2.4: The *Perform band matching* dialog box.

3.6 Press <***Tolerance settings***> to open the *Position tolerance settings* dialog box (see Figure 2.5).

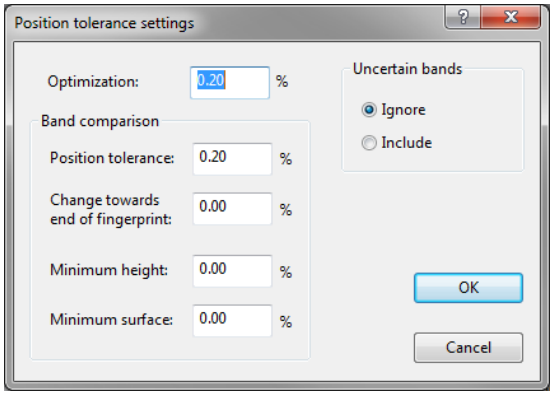

Figure 2.5: The *Position tolerance settings* dialog box.

- The ***Optimization*** is a shift that you allow between any two patterns as a whole and within which the program will look for the best possible matching.
- The ***Position tolerance*** is the maximal shift allowed (in percentage of the pattern length) between two bands allowed to consider them as matching. With ***Change towards end of fingerprint***, you can specify a gradual increase or decrease in tolerance.
- With ***Minimum height*** and ***Minimum surface***, you can exclude weak or irrelevant bands.
- The ***Uncertain bands*** option allows you to either include uncertain bands or ignore them. When ***Ignore*** is chosen, uncertain bands are ignored. This means that in composing a band matching table, the software will omit the uncertain bands, considering them as characters that are unknown. When ***Include*** is chosen, uncertain bands are treated in the same way as certain bands, which means that uncertain bands will contribute to the band classes of a band matching tables in the same way as certain bands.

The settings as specified in the *Position tolerance settings* dialog box will be employed every time an automatic band assignment to band classes is performed.

3.7 For this example, enter a position tolerance of 0.20%, an optimization of 0.20%, a change of 0%, and a minimum height and minimum surface of 0%, and press <***OK***>.

Several methods are available to perform a band matching. The method ***Load classes from experiment type*** will apply the band classes that were saved with the fingerprint type on all entries in the comparison. ***Find classes on all entries*** and ***Find classes on selected entries only*** will automatically search for band classes, respectively on all entries in the comparison or only on the selected entries. In any of the above methods, band classes will be created according to the <***Tolerance settings***> (see above). The last option, ***Do not create classes*** will not perform an automated band matching, but displays the fingerprint data in "band editing mode", ready to create band classes manually and to start scoring bands.

The preferred method will depend on the situation: If the location of the genetic markers is known and saved as band classes with the experiment type, then ***Load classes from experiment type*** will be the obvious choice. If instead the location of the genetic markers is known, but not saved yet as band classes, then ***Do not create classes*** should be selected and band classes created manually. If no information about genetic markers is available, but one just wants to screen for homozygous or heterozygous occurring bands, then ***Find classes on all entries*** or ***Find classes on selected entries only*** can be a convenient option.

3.8 Since we will first illustrate the manual options to create and edit band classes, select ***Do not create classes*** and press <***OK***>.

The **AFLP** image is now in "band editing mode". A large number of bands are shown (indicated as small red rhombs), but no band classes are created yet.

2.4 Adding, deleting and selecting bands in a comparison

In "band editing mode", bands can be added or deleted from within the *Comparison* window. Any changes made are saved on the corresponding gels.

To add a band at a certain position:

4.1 Click on the position where the band should occur. A single selection flag indicates the selected position: .

4.2 Select ***Fingerprints*** > ***Add band*** (**Enter**) or press **Enter** on the keyboard. The band is now added and is displayed with double selection flags:  .

When ***Fingerprints*** > ***Snap to peaks*** is checked (default), the band will be created at the local maximum nearest to the cursor position. When unchecked, bands will be created at the exact cursor position.

To delete a single band:

4.3 Click on the band to select it. Double selection flags now indicate the selected band:  .

4.4 Select ***Fingerprints*** > ***Remove selected band*** (**Del**). The band is now removed and a single selection flag indicates the position where the band used to be: .

A selection of bands can be deleted in a similar way:

4.5 Make a selection of bands using any of the methods described below (see 2.4), e.g. by pressing **Shift** while dragging a rectangle with the mouse.

4.6 Select ***Fingerprints*** > ***Remove selected band*** (**Del**) or press **Delete** on the keyboard.

Since removing bands is easier than adding them (the former can be done for a selection, while the latter is always done one by one), it is often more convenient to have bands searched automatically in step 4 of the gel processing with a low threshold and delete bands assigned in excess later compared to adding a large number of bands manually.

When deleting or adding bands, the Undo/Redo functions are available. To undo a change, select ***Sequence*** > ***Edit alignment*** > ***Undo*** (, **Ctrl+Z**). To redo an undone change, select ***Sequence*** > ***Edit alignment*** > ***Redo*** (, **Ctrl+Y**).

When a comparison is saved using ***File*** > ***Save*** (, **Ctrl+S**) or ***File*** > ***Save as...***, any modified band information (added or deleted bands) is automatically saved to the corresponding gels. Modified band information can also be stored without making any modification to the comparison itself by selecting ***Fingerprints*** > ***Save modified band information...***.

A number of functions in the *BandScoring plugin* work only on the currently selected band(s). Bands can be selected in several ways:

4.7 To select a single band, just click on it with the mouse.

4.8 To select a number of adjacent bands, press the **Shift**-key on the keyboard while dragging a rectangle with the mouse.

4.9 To select all bands belonging to a certain band class, double-click on any band in that band class. A band class needs to be defined first, see 2.5.

4.10 To select all bands in a comparison at once, select ***Bandscoring*** > ***Select all bands*** or press **Shift+A** on the keyboard.

4.11 To select all bands that belong to the currently highlighted entry, select ***Bandscoring*** > ***Select all bands (current entry)*** or press **Shift+E** on the keyboard.

4.12 To select all bands that belong to the currently selected entries, select ***Bandscoring*** > ***Select all bands (selected entries)*** or press **Shift+S** on the keyboard.

Any selection of bands is cleared when clicking on a random position in the gel image.

Bands can be marked as certain via ***Fingerprints*** > ***Mark band(s) as certain*** (**F6**) or uncertain via ***Fingerprints*** > ***Mark band(s) as uncertain*** (**F5**) in a comparison.

When working with bands in the *Comparison* window, it can be useful to hide the band assignments in order to evaluate the band intensity on the underlaying gel image. Band visibility can be toggled via ***Fingerprints*** > ***Settings*** > ***Show / hide bands*** (**F2**) or by pressing **F2** on the keyboard.

Uncertain bands are displayed in a darker color than certain bands.

2.5 Working with band classes

In this tutorial, we start from a situation in which the position on the gel of the genetic markers of interest is already known, but these positions are not saved yet as band classes with the experiment type. This will be the most common case when analyzing a first AFLP gel containing F_2 hybrid samples. For later gels, the genetic marker positions will already be available as band classes, so the *Load classes from experiment type* method is used when searching for band classes (see Figure 2.4).

Suppose following genetic marker are present on the AFLP gel **Gel1** (percentages are approximate running distances):

- pleasant flavor (59.70%)
- dwarf fruit (65.14%)
- growth in humid conditions (71.16%)
- Blossom End Rot resistance (76.34%)
- nematode resistance (81.36%)

We will now create band classes in comparison **BreedingExp1**, which correspond to the above genetic markers.

Only a brief description of the manual editing functions available for band matching analysis is given in this section, for more details see the Reference manual, Chapter Band matching and polymorphism analysis.

5.1 Select ***Fingerprints*** > ***Settings*** > ***Show metrics scale*** () to display a scale with the percentage running distance on top of the gel image.

5.2 Zoom in on the gel image using the horizontal zoom slider  , if needed.

A new band class can be created as follows:

5.3 Click on a band for which a new band class should be created, e.g. a band occurring at about 59.70% running distance. Double selection flags now indicate the selected band:  .

5.4 Select ***Fingerprints*** > ***Add new band class*** (**Shift+Enter**) or press **Shift+Enter** on the keyboard.

The program asks ”Do you want to automatically search and assign other bands to the new class?”. If you press <***No***>, the new band class will contain only the selected band. If you select <***Yes***>, any band that is within the *Position tolerance settings* (see Figure 2.3) and not already assigned to another band class is automatically assigned to the new class.

5.5 Press <***Yes***> to automatically assign all relevant bands to the new band class.

Instead of its metrics value (the default name), a more meaningful name can be given to a band class.

5.6 With the band class highlighted, select ***Fingerprints*** > ***Band class information...*** (**Ctrl+I**) or press **Ctrl+I** on the keyboard. Alternatively, just double-click the band class label.

5.7 In the *Band class information* dialog box that pops us (see Figure 2.6), enter a more meaningful name, e.g. ”pleasant flavor” for the band class occurring at 59.70% running distance and press <***OK***>.

5.8 Repeat Instruction 5.3 to Instruction 5.7 to create the band classes corresponding to the four remaining genetic markers. The final result should look like depicted in Figure 2.7.

Visually check the automatic band assignment. In case a band was wrongly assigned to a band class, it can be removed from that class.

5.9 Select the band that was wrongly assigned and select ***Fingerprints*** > ***Remove band from class*** (**Ctrl+Del**) or press **Ctrl+Del** on the keyboard. The band is now displayed as a small red rhombus  .

A band that was not automatically assigned to its corresponding class can be manually assigned.

5.10 Select the band class and the band that you want to assign to this class. Select ***Fingerprints*** > ***Assign band to class*** (**Ctrl+Enter**) or press **Ctrl+Enter** on the keyboard. The band is now displayed as a red ”+” sign  .

Removing bands from band classes and (re-)assigning bands to band classes can also be done with a simple drag-and-drop procedure.

5.11 Select a band using the mouse. While pressing the mouse button, drag it away from the band class it was assigned to and release the mouse button. The band is now removed from the band class.

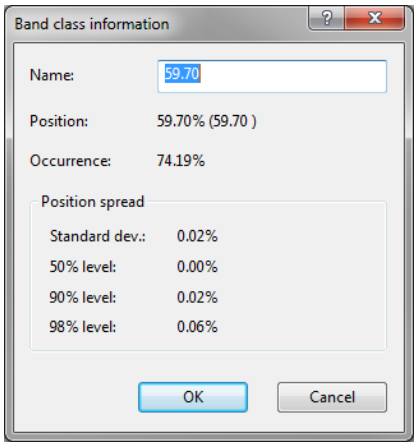

Figure 2.6: The *Band class information* dialog box, displaying the default band class name.

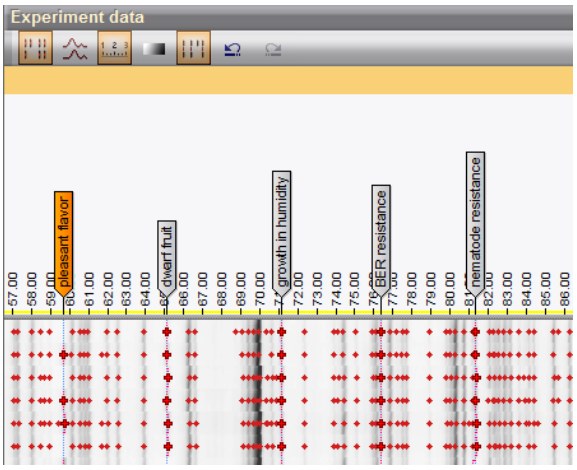

Figure 2.7: Illustration of the band classes created in the **BandScoring** database, corresponding to the fictitious genetic markers.

5.12 Select the band again and drag it towards the first or the second class. You will notice that - when the mouse pointer gets into the proximity of a band class - the band will be assigned to that band class.

A band class that does not correspond to a genetic marker of interest can be deleted as follows:

5.13 Click on a band belonging to the band class.

5.14 Select ***Fingerprints*** > ***Remove band class*** (**Shift+Del**) or press **Shift+Del** on the keyboard to remove the selected band class.

When all band classes are created that correspond to the genetic markers of interest, the band classes can be saved to the experiment type.

5.15 Select ***Fingerprints*** > ***Save band classes to experiment type...*** and press <***Yes***> in the confirmation dialog box.

From the **AFLP** *Fingerprint type* window, you can verify that the band classes are indeed saved with the experiment type.

5.16 In the *Experiment types* panel of the *Main* window, double-click on **AFLP** to open the **AFLP** *Fingerprint type* window and click on the *Band Classes* panel. The band classes should be displayed, similar to Figure 2.8.

When additional gels would be analyzed, having the same genetic markers of interest, it now becomes

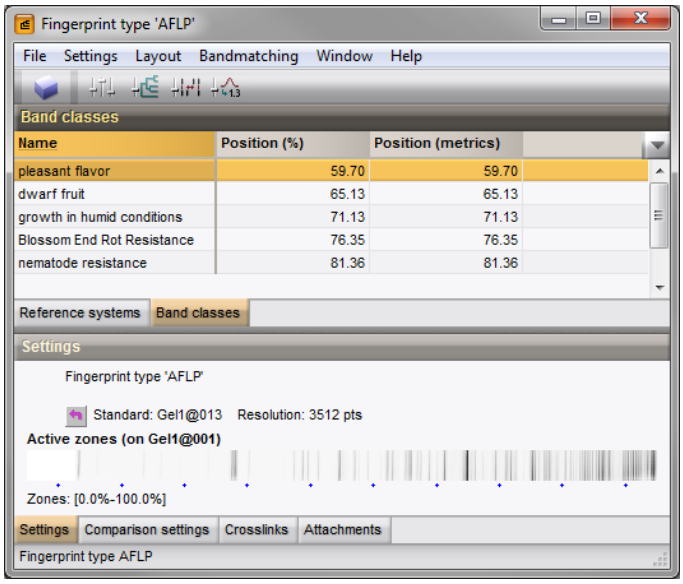

Figure 2.8: The **AFLP** *Fingerprint type* window in the **BandScoring** database, displaying the band classes saved with the experiment.

possible to select ***Load classes from experiment type*** when searching for band classes (see Figure 2.4).

2.6 Updating the bandscoring character type experiments

When band classes are saved with the fingerprint type (here **AFLP**), it becomes possible to update the two auxiliary bandscoring character types (in casu **AFLP_bm_pos** and **AFLP_bm_zyg**) with these band classes.

6.1 First, save and close comparison **BreedingExp1** or any other comparison that might be open.

6.2 In the *Main* window, select ***Bandscoring*** > ***Update bandscoring character types***.

This pops up the *Update bandscoring experiment* dialog box (see Figure 2.9).

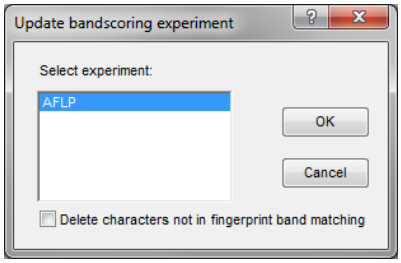

Figure 2.9: The *Update bandscoring experiment* dialog box.

Pressing <***OK***> updates the character data sets with the band classes saved with the fingerprint type. The update is additive, meaning that new band classes are added to the existing character set.

Deleted band classes are only removed from the character sets when ***Delete characters not in fingerprint band matching*** is checked.

6.3 Press <***OK***> to do the update.

If you open any of the two character type experiments, you will notice that now five characters are present. This is illustrated in Figure 2.10 for **AFLP_bm_zyg**.

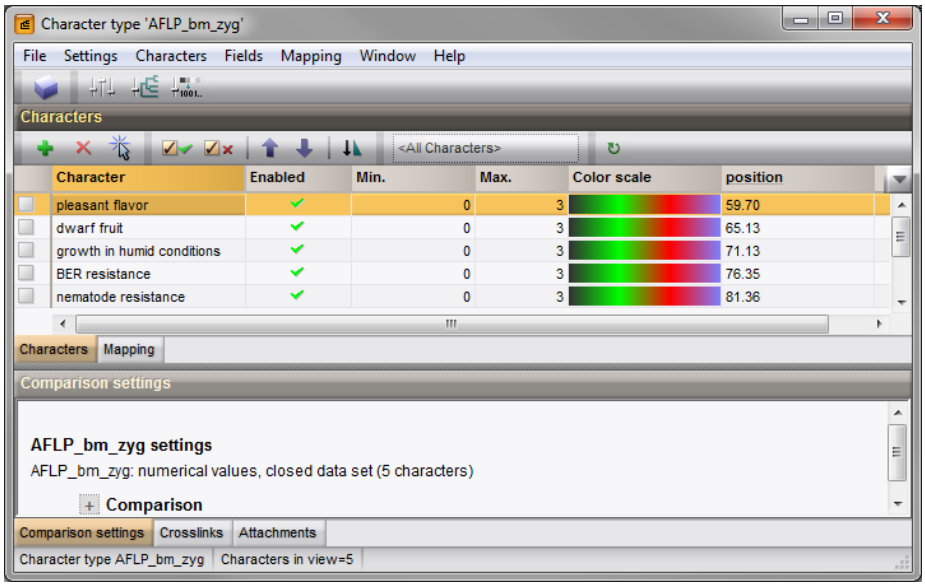

Figure 2.10: The auxiliary bandscoring character type AFLP_bm_zyg, displaying characters that correspond to band classes.

2.7 Band class filtering

In the previous section, band classes were created for the four genetic markers and the correct bands were assigned to these classes. However, the comparison still contains a large number of irrelevant bands, not assigned to any band class (see Figure 2.7). These bands can be automatically removed by the software.

7.1 Open comparison **BreedingExp1** by double-clicking it in the *Comparisons* panel of the *Main* window.

7.2 In the *Experiments* panel of the *Comparison* window, click on the eye button ( ) that proceeds **AFLP** to display the **AFLP** gel image.

7.3 Select ***Fingerprints*** > ***Show bands*** ( ) to display the bands and band classes of the **AFLP** fingerprint type.

7.4 Select ***Bandscoring*** > ***Remove bands not associated with a band class***. This will remove the uninformative bands in the *Comparison* window.

7.5 Save the comparison with ***File*** > ***Save*** (, **Ctrl+S**) button. The band information will be updated at the gel level as well.

Sometimes, e.g. when using an automated band class search, a number of band classes only correspond to one or a few band. These classes are uninformative for plant breeding purposes and can be removed automatically by selecting ***Bandscoring*** > ***Remove band classes with few bands***. This calls the *Remove band classes* dialog box (see Figure 2.11).

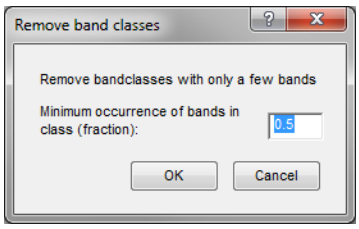

Figure 2.11: The *Remove band classes* dialog box.

The dialog prompt for the ***Minimum occurrence of bands in class***. This ratio is the number of band occurrences over the total number of entries in the comparison. Band classes with lower occurrence (and their corresponding bands) will be removed.

2.8 User–defined and automated band assignment

In the **BandScoring** database, the occurrence (and zygosity) of the genetic markers in the two parent lines is known. We also know that the ”growth in humid conditions” marker occurs homozygously for both parent lines and their offspring. Therefore, these assignments can be made manually.

8.1 For the ”Moneymaker” cultivar parent line (see ’Cultivar’ information field), select the band belonging to the ”pleasant flavor” band class and select ***Bandscoring*** > ***Assign band(s) as homozygous*** or press **Ctrl+F8** on the keyboard. Repeat the same action for the band belonging to the ”Blossom End Rot resistance” band class.

8.2 For the ”NemaProof” cultivar (see ’Cultivar’ information field), assign the bands corresponding to ”dwarf fruit” and ”nematode resistance” as homozygous, similar as described in the previous step.

8.3 Double-click on any band in the ”growth in humid conditions” band class to select all bands that belong to this class and select ***Bandscoring*** > ***Assign band(s) as homozygous*** or press **Ctrl+F8** on the keyboard.

Bands can also be user-assigned as heterozygous via ***Bandscoring*** > ***Assign band(s) as heterozygous*** (shortcut: **Ctrl+F6**) or as undefined via ***Bandscoring*** > ***Assign band(s) as undefined*** (shortcut: **Ctrl+F7**).

If a mistake was made during band assignment, the assignment can be reset.

8.4 Select the band(s) for which you want to reset the assignment, using any of the procedures described in 2.4.

8.5 Select ***Bandscoring*** > ***Reset band(s) assignment*** or press **Ctrl+F5** on the keyboard.

The selected bands will be reset to undefined and automatically assigned, ready for manual and/or automatic band assignment.

When all bands for which the zygosity is known are manually assigned, the software can assign the zygosity of the remaining bands automatically.

8.6 Select ***Bandscoring*** > ***Assign bands automatically***.

The *Band calibration window* pops up (see Figure 2.12).

The *Band calibration window* shows the band distributions for each of the band classes.

In the *Comparison* window, unassigned bands are now automatically assigned by the algorithm. User-assigned bands are indicated with a rhombus, colored according to the *Bandscoring colors* (see Figure 2.3), e.g. . Automatically assigned bands are indicated with a ”+” sign and also colored according to the bandscoring colors, e.g. . User-assigned bands will NOT be altered by the automatic assignment.

The algorithm tries to automatically assign hetero- or homozygosity to all bands in all entries. First, an iterative calibration over the entries and band classes is performed. This calibration is stored with the band matching. After the calibration, the bands are scored into one of the five states (absent, uncertain, heterozygous, undefined, or homozygous). The calibration and scoring settings can be adjusted (see 2.12). The automatic band assignment can be assessed and further refined using the band class calibration plots (see 2.9).

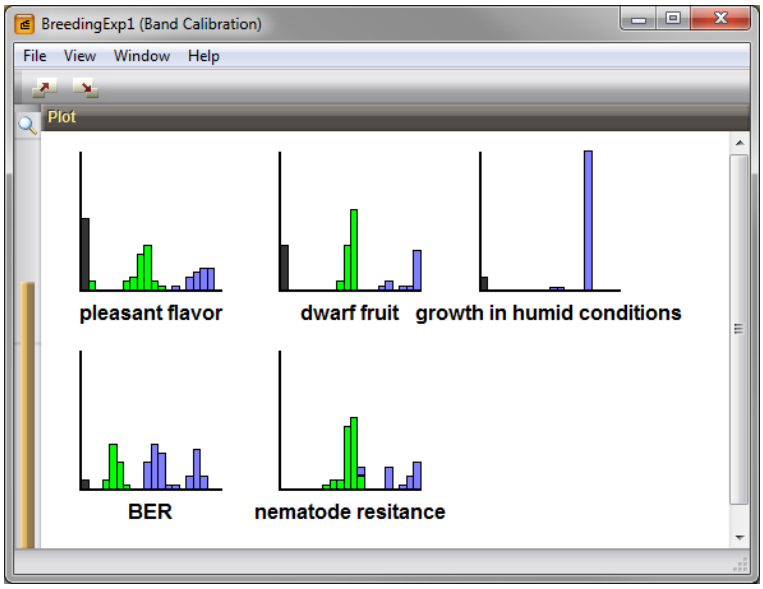

Figure 2.12: The *Band calibration window*, showing the calibrated band height distributions for the band classes defined in the **BreedingExp1** comparison.

2.9 Band class calibration plots

Two types of band class calibration plots are available: an overview plot in the *Band calibration window* and a detailed plot for a selected band class in the *Detailed band calibration window*.

The *Band calibration window* displays an overview of the calibrated band height distributions for all band classes in the comparison. It is displayed after an automatic band assignment or can be called via ***Bandscoring*** > ***Plot heights*** in the *Comparison* window.

If no calibration (automatic assignment) has been performed, the raw height distribution is shown in both the overview and detailed band calibration plot.

Display options for the *Band calibration window* can be set via ***View*** > ***Settings*** (see Figure 2.13).

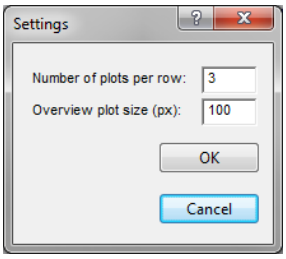

Figure 2.13: The *Settings* dialog box.

The number of plots per row can be selected, as well as the pixel size of the overview plots.

The same settings can also be accessed by selecting ***Bandscoring*** > ***Settings*** in the *Comparison* window (see 2.12).

The plots can be printed via ***File*** > ***Print*** or copied to the clipboard via ***File*** > ***Copy to clipboard***. From the clipboard, the copied information can then be pasted in applications such as Adobe Photoshop, Microsoft PowerPoint, Microsoft Word, etc.

Double-clicking an overview plot in the *Band calibration window* selects the corresponding band class in the *Comparison* window and opens a detailed plot for that band class. From the *Comparison* window, a

detailed band calibration plot can be displayed for a selected band class by selecting ***Bandscoring*** > ***Plot heights (selected band class)***.

9.1 For example, double-click on the ”nematode resistance” plot.

This opens the *Detailed band calibration window* for this band class (see Figure 2.14).

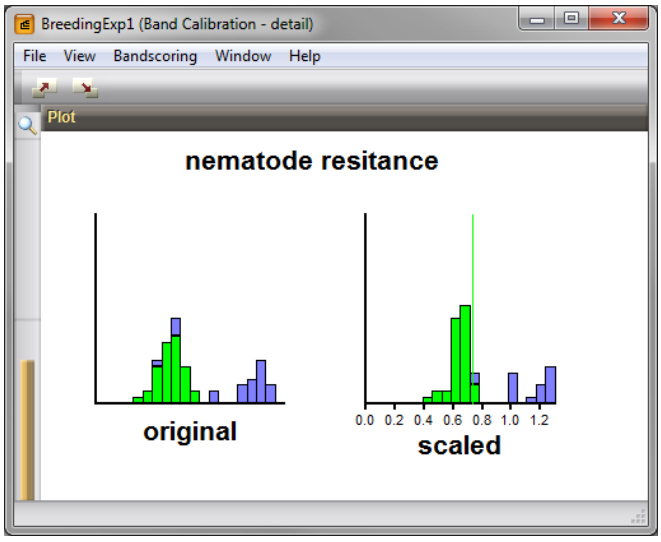

Figure 2.14: The *Detailed band calibration window* for the ”nematode resistance” band class.

Similar as for the *Band calibration window*, display options for the *Detailed band calibration window* can be set by selecting ***View*** > ***Settings***. This calls the *Settings* dialog box (see Figure 2.15).

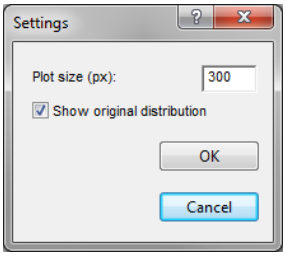

Figure 2.15: The *Settings* dialog box.

The original distribution can be shown or hidden and the pixel size of the plots can be entered.

The detailed plot(s) can be printed via ***File*** > ***Print*** or copied to the clipboard via ***File*** > ***Copy to clipboard***. From the clipboard, the copied information can then be pasted in applications such as Adobe Photoshop, Microsoft PowerPoint, Microsoft Word, etc.

The *Detailed band calibration window* contains some extra functionality to help adjusting the automatic bandscoring.

Selecting ***Bandscoring*** > ***Set assignment limits*** calls the *Set assignment limits* dialog box (see Figure 2.16).

In this dialog the assignment limits between uncertain/heterozygous, heterozygous/undefined and undefined/homozygous scoring can be entered.

Press <***Preview***> for a preview of the entered limits on the scaled plot.

The entered limits can be set by pressing <***Set limits***>. Band assignments are updated simultaneously in

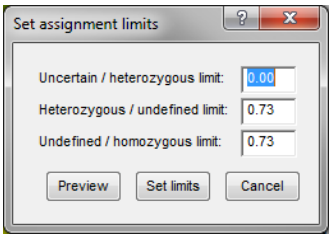

Figure 2.16: The *Set assignment limits* dialog box.

the plot windows and *Comparison* window.

9.2 For the ”nematode resistance” band class, you might want to increase the heterozygous/undefined and undefined/homozygous limits to 0.80.

9.3 Press <***Preview***> for a preview of the entered limits on the scaled plot.

The assignment limits can also be interactively set using the cursor and the available menu items or shortcuts.

9.4 For example, open the *Detailed band calibration window* for the ”Blossom End Rot resistance” band class e.g. by double-clicking on this plot in the *Band calibration window*.

Since the band with a height of about 0.8 should probably still be considered as heterozygous, we will change the heterozygous/undefined limit.

9.5 Click with the mouse on the position where you want to define the heterozygous/undefined limit, e.g. at about 0.85.

9.6 Select ***Bandscoring*** > ***Set heterozygous/undefined limit*** or press **2** on the keyboard.

Now the band with a height of about 0.8 is set to heterozygous. Other limits can be set accordingly.

The assignment limits are only applicable to automatically assigned bands, they do not affect user-assigned bands!

In the *Detailed band calibration window*, bands can be filtered according to their calibrated height.

9.7 Select ***Bandscoring*** > ***Filter band class***.

The *Filter band class* dialog box pops up (see Figure 2.17).

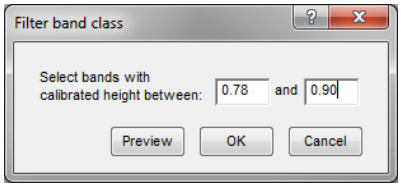

Figure 2.17: The *Filter band class* dialog box to select bands with a specific calibrated height.

This dialog allows you to select bands that have a calibrated height to filter out specific bands.

Press <***Preview***> to display the filtered range as a shaded area.

Press <***OK***> to select the bands with calibrated heights between the specified values in the *Comparison* window.

9.8 In our example enter a calibrated height range between 0.78 and 0.90.

9.9 Press <***Preview***> and press <***OK***>.

In this example, the single band for which the score was altered as a consequence of Instruction 9.5 to Instruction 9.6, is selected in the *Comparison* window.

2.10 Output and analysis of band scoring data

Once the bandscoring has been checked and approved, it can be outputted in different ways.

10.1 In the *Comparison* window, select ***Bandscoring*** > ***Save assignment to file***.

This exports the band assignments to a tab-delimited text file, which can be imported in other programs. The content and format of the text file can be controlled via the settings (see 2.12).

The bandscoring information can be saved to the auxiliary character type experiments.

10.2 Select ***Bandscoring*** > ***Save assignment to character data***.

This function saves the band assignments to the character data sets. As soon as the update is completed, the **AFLP_bm_zyg** character type is displayed automatically in the *Comparison* window.

For **AFLP_bm_zyg** - as for any character type - a whole range of analysis functions is available in BioNumerics, e.g. cluster analysis (using a categorical coefficient), sorting, charts and statistics analysis, etc. (requires the Tree and Network Inference module). For a detailed description of the above functions, see the Reference manual, Part Basic cluster analysis.

The values of the AFLP_bm_zyg character type (visible when ***Characters*** > ***Show values*** () is selected in the *Comparison* window) are 0, 1, 2 and 3 for absent, heterozygous, undefined and homozygous, respectively. Uncertain bands correspond with an absent character value (”null”).

For the fictitious breeding experiment of the **BandScoring** database, we are looking for the F_2 hybrid that is homozygous for ”pleasant flavor”, ”Blossom End Rot resistance” and ”nematode resistance”, but for which the ”dwarf fruit” genetic marker is absent. A simple sort by character value greatly facilitates this search:

10.3 In the *Experiment data* panel of the *Comparison* window, click on the ”dwarf fruit” character. Select ***Characters*** > ***Sort by character value*** ().

The entries in the comparison are now sorted according to the value of the character ”dwarf fruit”, similar as depicted in Figure 2.18. It is obvious that only the entry with key **Gel1@023** has the desired combination of genetic markers.

2.11 Assigning bands from character data

Assume that a second **AFLP** gel was run, containing only F_2 hybrids from the same breeding experiment as **Gel1**. Once this second gel is processed and the lanes linked to database entries (see work flow in Figure 2.1), a new comparison can be created, e.g. containing all entries from the second gel and both parent lines that were run on **Gel1**. Since the band assignments for the parent lines are saved in the auxiliary character set, they can be called again by selecting ***Bandscoring*** > ***Assign bands from saved character data***. When the band classes as defined in the comparison and the characters in the auxiliary character set correspond completely, the band assignment is imported directly.

When band classes and characters do not map, a warning dialog box is displayed (see Figure 2.19 and Figure 2.20).

The *Warning* dialog box is displayed when there are band classes that do not map with the characters in the auxiliary character set. This dialog ask what should be done with the extra band classes: ***Keep present assignment*** or ***Set to undefined***.

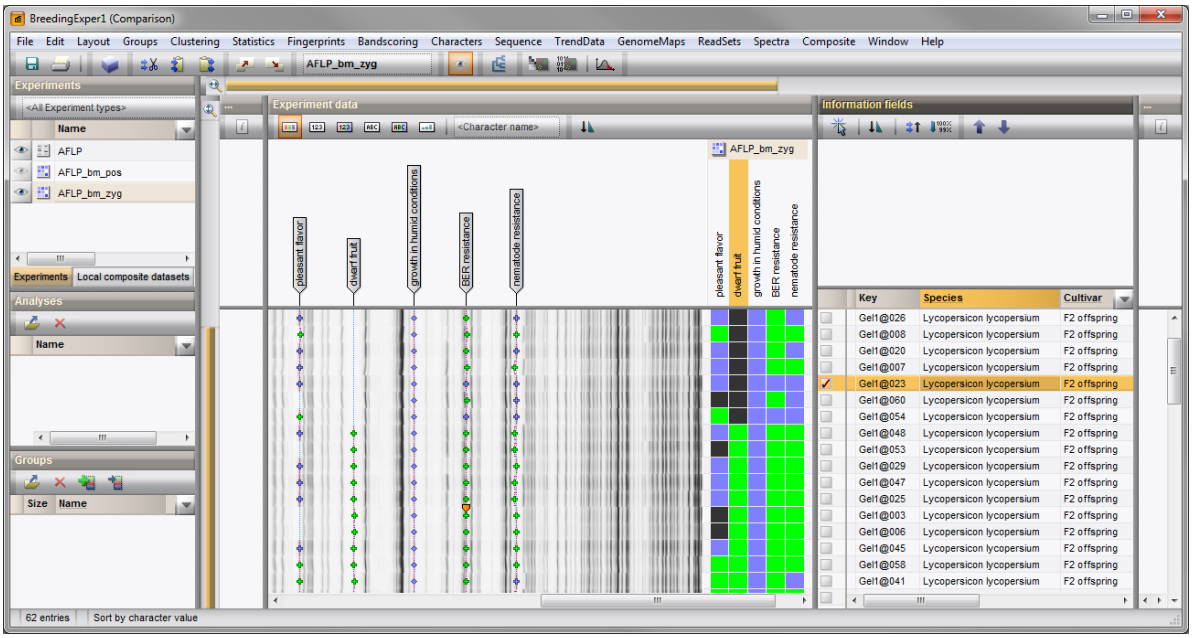

Figure 2.18: Result of the bandscoring analysis in comparison **BreedingExp1**: the selected entry has the desired combinations of genetic markers.

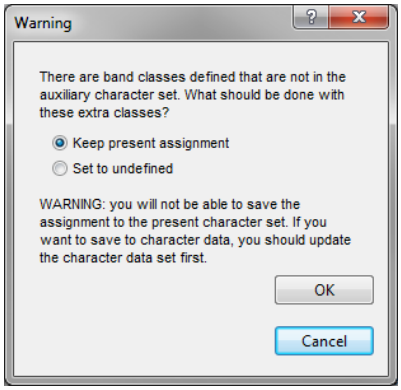

Figure 2.19: The *Warning* dialog box.

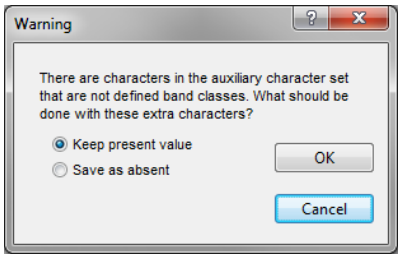

Figure 2.20: The *Warning* dialog box.

The *Warning* dialog box is displayed when there are characters in the auxiliary character set that do not map with the band classes. This dialog ask what should be done with the extra characters: ***Keep present value*** or ***Save as absent***.

2.12 Bandscoring settings

All relevant settings for bandscoring can be accessed from the *Comparison* window.

12.1 In the *Comparison* window, select ***Bandscoring*** > ***Settings***.

This pops up the *Bandscoring settings* dialog box (see Figure 2.21).

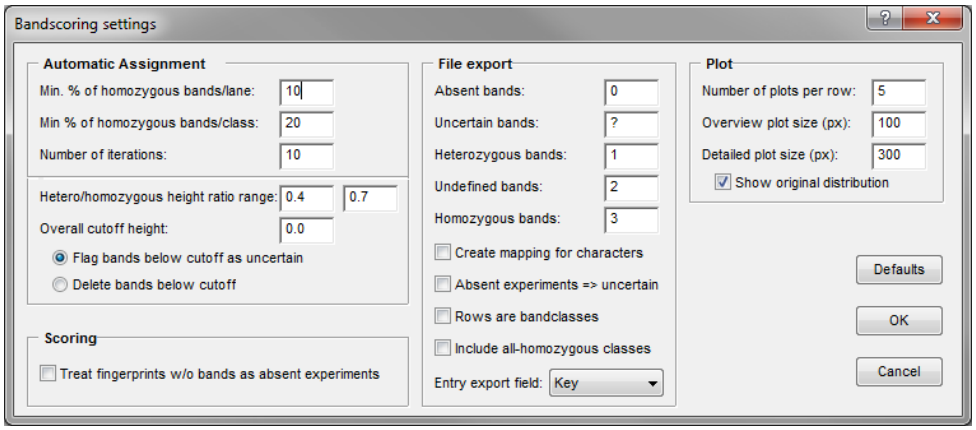

Figure 2.21: The *Bandscoring settings* dialog box, grouping the main settings related to bandscoring.

The bandscoring settings are organized in four groups:

Automatic assignment options

For the calibration to make sense, at least some of the bands in a lane or in a band class need to be homozygous. The ***Min.% of homozygous bands/lane*** and ***Min. % of homozygous bands/class*** can be specified here. The default values of 10, respectively 20% mean that at least 10% of the bands per lane, respectively at least 20% of the bands per class are homozygous. User-defined assignments overrule those settings.

The calibration is an iterative process and the ***Number of iterations*** can be specified. The default value of 10 iterations is usually more than sufficient for convergence.

In principle, the band height of heterozygous bands is half that of homozygous bands. In practice this is not always the case. The program will therefore test a range of heterozygous band fractions (specified as ***Hetero/homozygous height ratio range***) to obtain the best possible separation between homo- and heterozygous bands.

In some cases, bands were assigned that actually should not exist. These bands generally have low heights and can cause problems when trying to separate homo- from heterozygous bands. To solve this issue, an ***Overall cutoff height*** can be specified. Bands with a calibrated height lower than this value are not considered in the homo/heterozygous separation process. One can choose to set these bands to uncertain (***Flag bands below cutoff as uncertain***) or to remove those bands altogether (***Delete bands below cutoff***).

Scoring

By default, if no bands were detected in a lane, they will all be scored as absent when saved to character data. However, in case the option ***Treat fingerprints w/o bands as absent experiments*** is checked and not a single band detected in the lane, all corresponding values in the character type with suffix "_bm_zyg" will be set to "Null", i.e. a missing value. This is equivalent to an absent experiment; when BioNumerics is closed and opened again, the colored dots in the *Experiment presence* panel for the two character types will be removed. The ***Treat fingerprints w/o bands as absent experiments*** option should be checked if you do not wish to systematically unlink lanes of insufficient quality.

File export options

The names or values, representing each of the five band states (***Absent***, ***Uncertain***, ***Heterozygous***, ***Undefined*** and ***Homozygous*** bands) in the exported text file, can be specified here.

If ***Create mapping for characters*** is checked, a mapping will automatically be created for the character type containing the band zygosity (suffix "_bm_zyg"), using the names as specified under the ***File export*** settings.

If ***Absent experiments => uncertain*** is checked, absent experiments (lanes) will be mapped to uncertain when the bandscoring is exported to a text file, i.e. the name or value defined for ***Uncertain bands*** will appear for each of the band classes in the absent lane. If unchecked, no information will be exported for absent experiments.

The exported text file by default has the entries organized in rows and the band classes in columns. The data matrix can be transposed by checking ***Rows are band classes***: this will export band classes as rows and entries a columns.

By default, band classes containing only homozygous bands are not exported. Check ***Include all-homozygous classes*** to include band classes that contain only homozygous bands in the exported text file.

Plot options

The ***Number of plots per row*** and the ***Overview plot size (px)*** can be set for the *Band calibration window*.

For the *Detailed band calibration window*, the ***Detailed plot size (px)*** and whether or not the original distribution is displayed left from the calibrated distribution (***Show original distribution***) can be determined.

Pressing the <***Defaults***> button reverts all settings in the *Bandscoring settings dialog box* to their default value.

APPLIED MATHS

A BIOMÉRIEUX COMPANY

Please contact us for any additional information you might require, we will gladly help you!

Headquarters
Keistraat 120 • 9830 Sint-Martens-Latem • Belgium
+32 922 22 100 info@applied-maths.com

USA and Canada
11940 Jollyville Rd., Suite 115N • Austin, TX 78750 USA
+1 512 482 9700 info-us@applied-maths.com