

A user guide to the

# PhylochipAnalyzer\*

A program for the analysis of experiments  
with hierarchical probe-sets  
on DNA-mircoarrays

Stephan Frickenhaus<sup>#</sup>

Alfred-Wegener-Institute for Polar and Marine Research  
and  
Hochschule Bremerhaven

May 2007

\* Katja Metfies, Philipp Borsutzki, Christine Gescher, Linda K. Medlin,  
Stephan Frickenhaus. PhylochipAnalyzer - A Program for Analysing Hierarchical Probe-Sets,  
Molecular Ecology Notes (2007), *submitted*

\* <http://www.awi.de/en/go/phylochipalyzer>

# <http://www.awi.de/People/show?sfricken>

## Introduction

The PhylochipAnalyzer [1] is a graphical Windows program that describes a hierarchy of a set of molecular probes. The hierarchy reflects the fact that certain probes are hybridizing against more than a single species (RNA or DNA). Thus, the hierarchy is intrinsic to the probe design, e.g., by inspection of multiple alignments (e.g., software ARB [2]) one will find sequence regions that are species specific and regions that are common maybe to a whole clade in a phylogenetic tree. The species level or clade level is denoted the lowest level in the hierarchy of the probe-set. If a probe is found to be species specific it remains unclear whether there exist other species to hybridize as well, because the dataset of the multiple alignment will always be incomplete.

The PhylochipAnalyzer allows for editing the hierarchy visualized as a tree. It is not implied that this hierarchy reflects absolute taxonomic relations, e.g., relations that are computed from full alignments by phylogenetic inference. Because only an oligo-length fraction of the alignment is used, the user must make cross-checks and possibly shifts against all other known sequences, a task denoted *probe match* within ARB.

The PhylochipAnalyzer is in particular valuable for the analysis of redundant chips, i.e., chips with some or even many copies of the probes. These sets of copies are denoted “blocks” in the following. They allow for an analysis of the mean signals as a quality control.

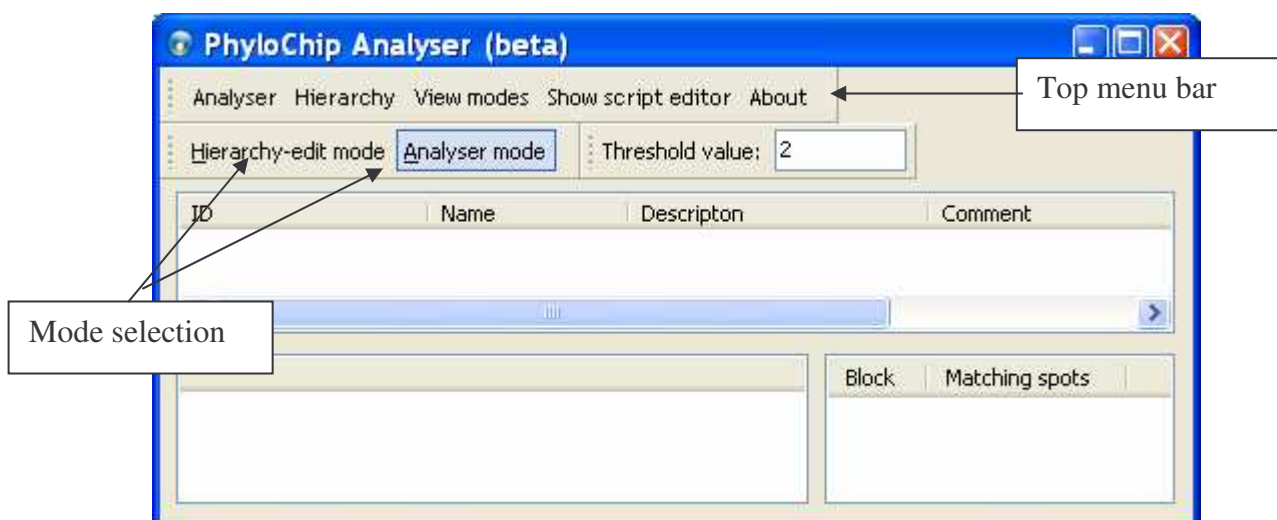
## Step by Step : from Hierarchy Editing to Data Analysis

### The Hierarchy Edit Mode

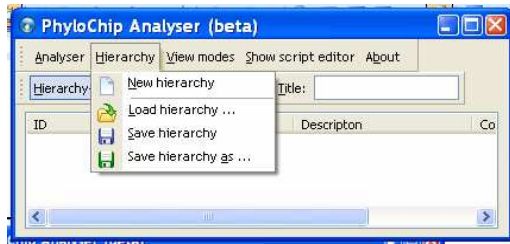
When launching the PhylochipAnalyzer for the first time, the user has three options:

1. Load an existing hierarchy in XML-format, previously saved from the program
2. Construct a hierarchy from a loaded GAL-file by shifting probe descriptions into their correct positions in a tree on screen.
3. Create a hierarchy from scratch.

These options require that the “Hierarchy-edit-mode” is selected in the mode bar. For 1) and 2) the “Load Hierarchy” is chosen in the Hierarchy menu from the menu bar. Usually, the user will start from 2) and restart and continue with 1). Starting from 3 is usefull for planning and documenting the design of a chip without having a chip-describing GAL-file at hand.

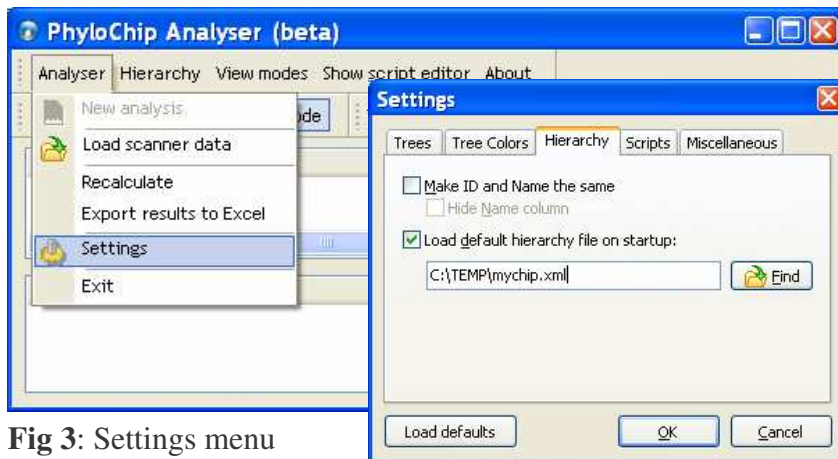


**Fig 1:** Main view of the Phylochip Analyzer



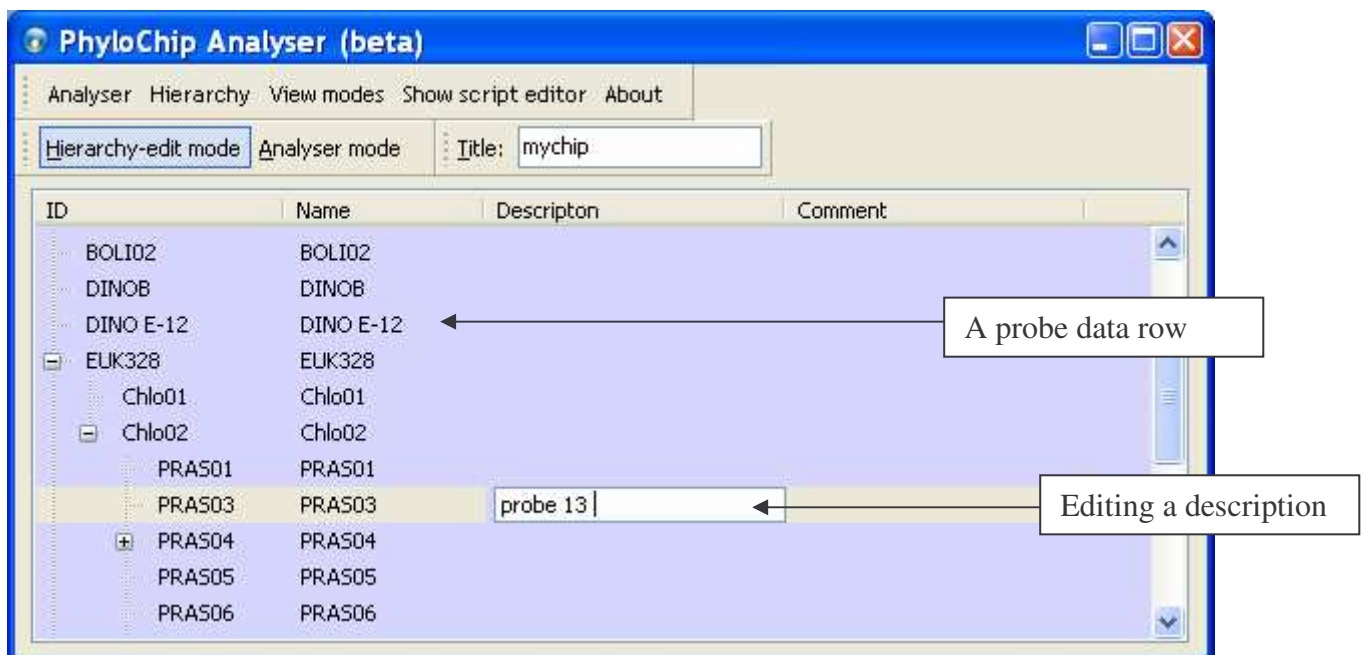
**Fig 2:** Hierarchy pull-down menu

It is possible to load automatically a given hierarchy on start-up of the program. You can select the name of the default hierarchy file in the settings-menu, tab “Hierarchy” (Fig. 3)



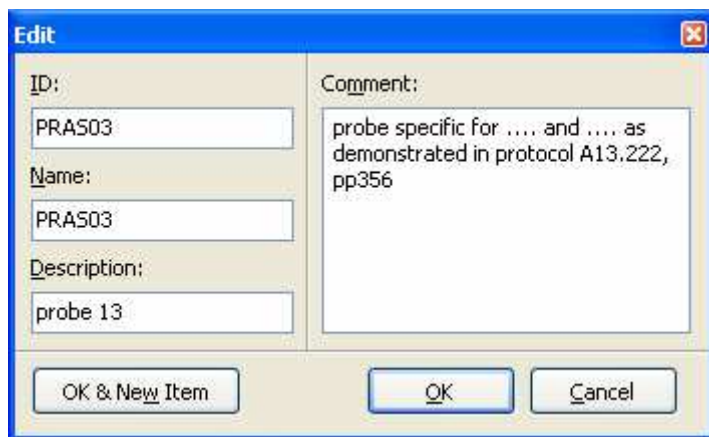
**Fig 3:** Settings menu

The hierarchy editor allows to enter rough description of probes by pressing the left mouse button when the cursor is in the description field, as seen in Fig. 4.



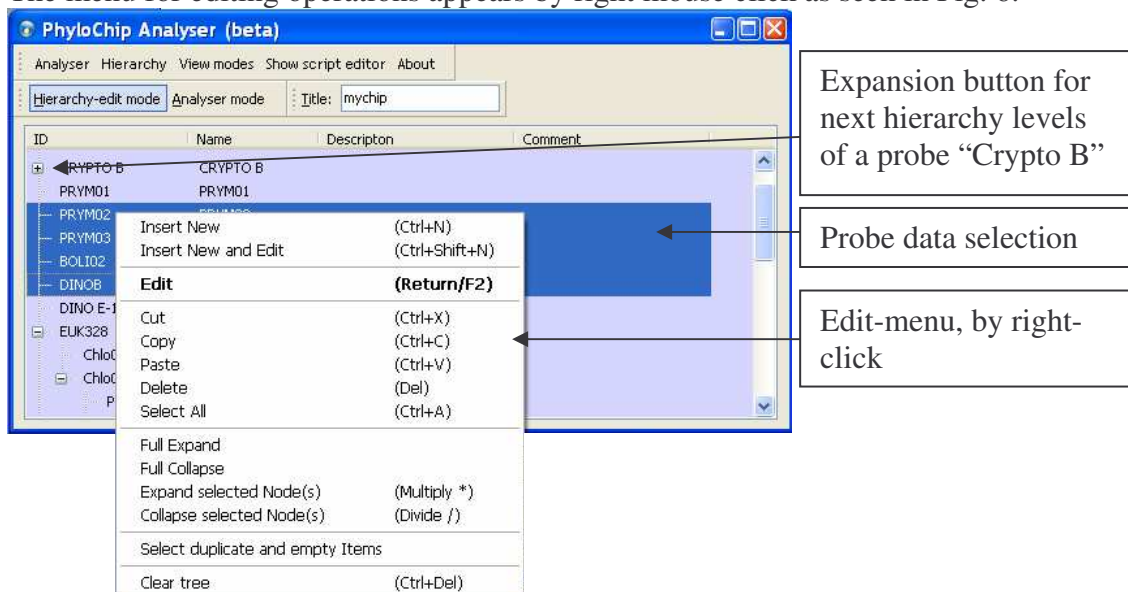
**Fig 4:** Hierarchy editing, entering a description in a probe data row

A complete data set of a single probe can be edited by a double click in the corresponding probe data row, as seen in Fig. 5. The OK&New Item is useful when entering probe by probe from scratch, i.e., when no chip description as a GAL-file is available.



**Fig 5:** Probe editing window, appears after a double-click. New Item inserts a new empty probe data row after the currently edited.

For shifting probe data rows into their correct hierarchical position several mechanisms exist, such as cut and paste, drag and drop and others, all applicable for a selection of probe data. Data selection is done by standard methods, such as Ctrl+Left click or Shift+Arrow up/down. The menu for editing operations appears by right mouse click as seen in Fig. 6.



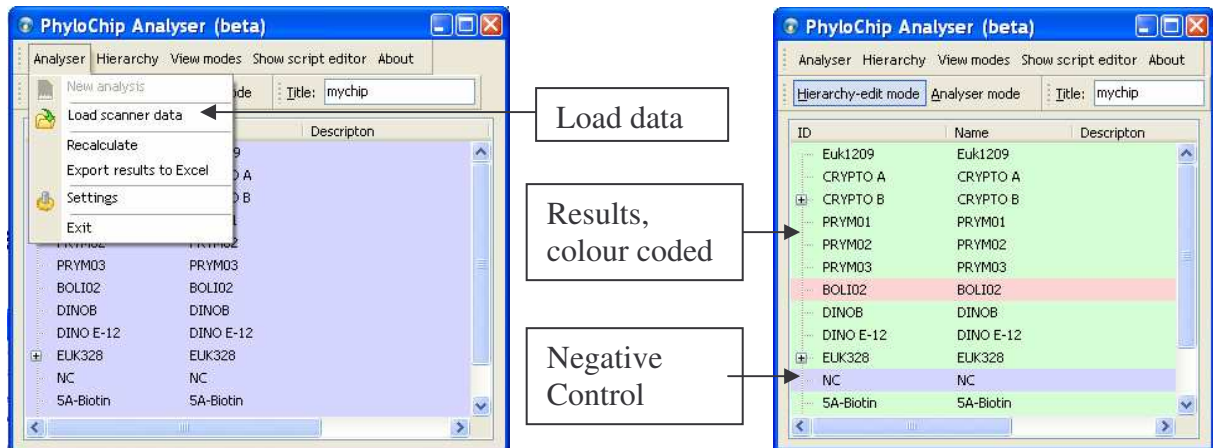
**Fig. 6:** Editing menu by right mouse click, applicable to a selection of probe data rows.

It is possible to collapse hierarchy sub-levels, hiding probe data on lower hierarchy levels, by left-clicking the [-] (expanding by the [+]), as seen in Fig. 6.

At the end of the edit process the user will save the hierarchy in an XML-file, see Fig. 2. The XML-file is a pure text file that can be edited/checked by hand. The user may develop own external software tools to generate appropriate hierarchy files.

## The Analysis Mode

For the analysis, the user will import spot intensity data from a GPR-file that the scanner hardware should produce. It is not necessary to be in analysis mode for the import step.

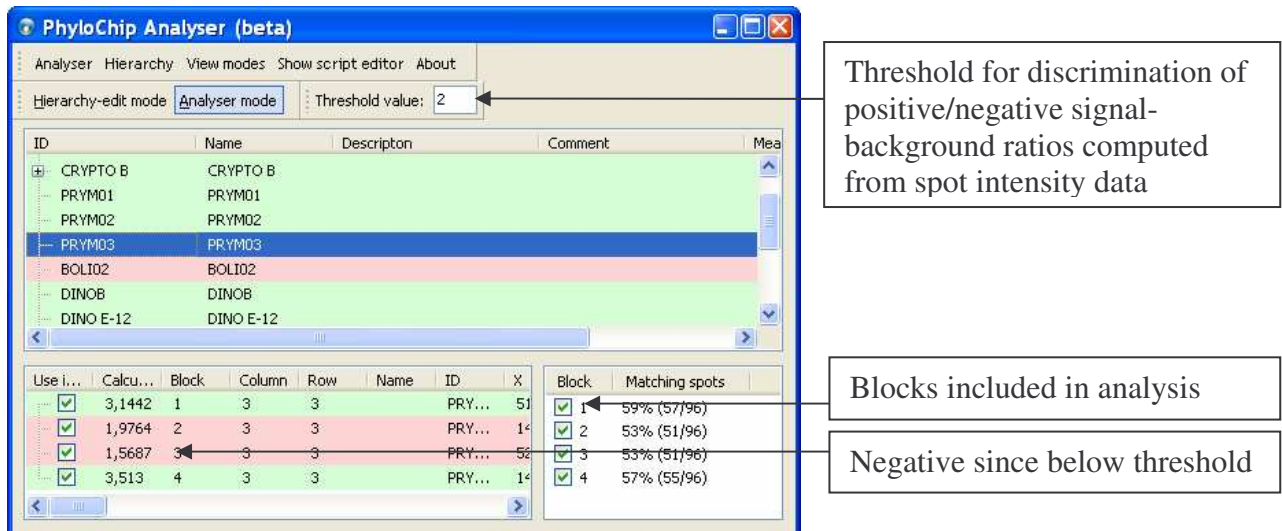


**Fig. 7:** Data import for analysis by “Load scanner data” (left). View of the results within “Hierarchy-edit mode” mode (right) after loading the GPR-file.

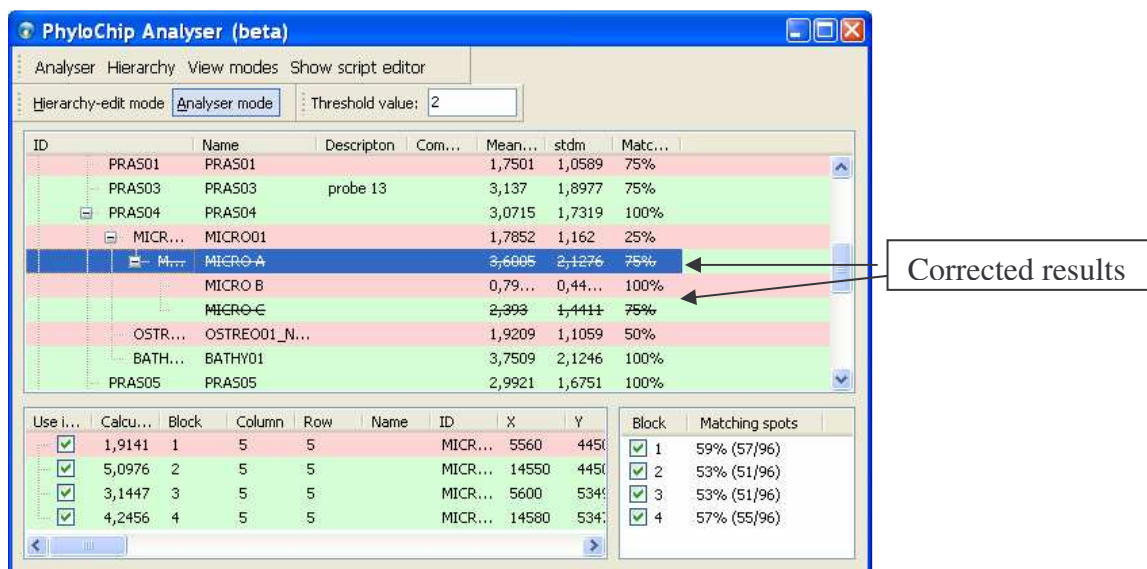
After the import, the spot data is automatically analysed and signal results are visible as red bars for negative and green bars for positive, see Fig. 7, right.

Spot data for probes not yet present in the hierarchy are printed in *italics*. They are easily identified, as their NAME field is empty. It is noteworthy, that such spots data is also included in the analysis. In this way, it is possible to have several other probes on the chip that might serve as additional positive or negative controls.

When switching to “Analyser mode” two sub windows appear listing the data (Fig. 8). The left part shows all copies of spots of a probe on the chip, for which data is available from file. In the example, two copies show negative results and two show positive results. In the second column, the “Calculated Value” is listed, which is basically the signal-to-background or signal-to-negative-control ratio. Spots are evaluated “negative” when the ratio is below the threshold value entered in the upper field (see legend in Fig. 8). The threshold value is only seen/modified in the analysis mode.

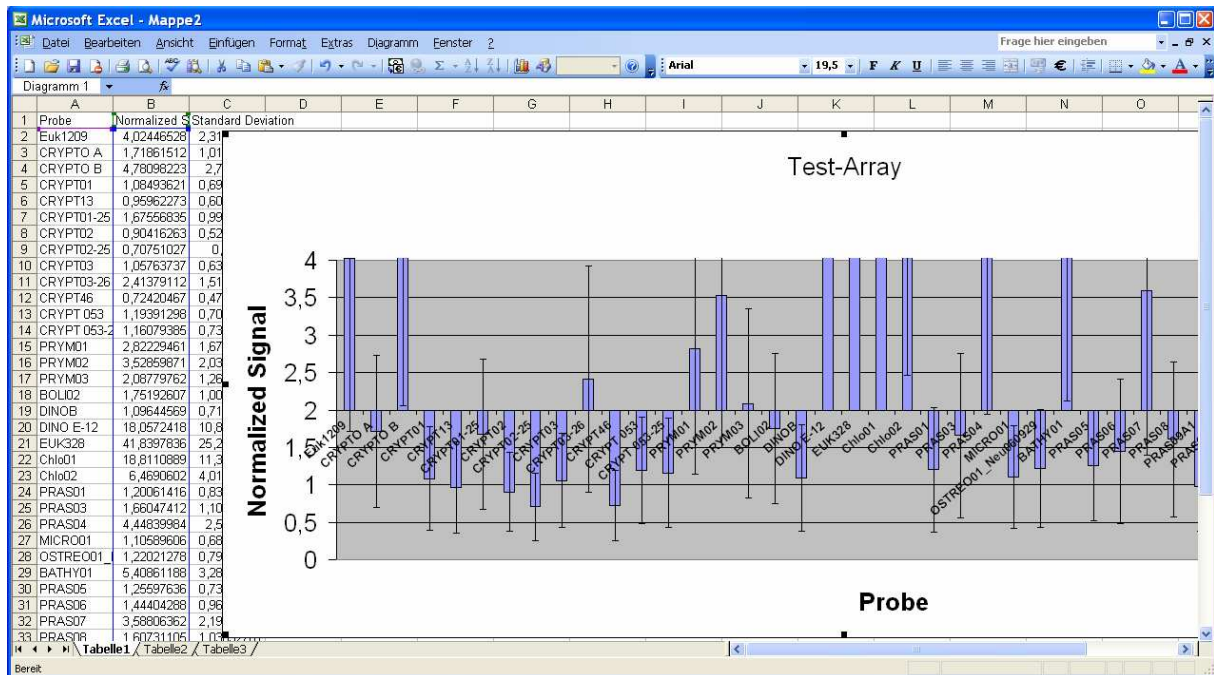


**Fig. 8:** Analyser mode view showing the overall results for a probe in the upper part. Lower left: spot signals for each block for the probe selected in the upper part, i.e., data for 4 copies of this probe were found in the data file. Bottom-right: whole blocks may be excluded from the analysis by deselection of the check box.

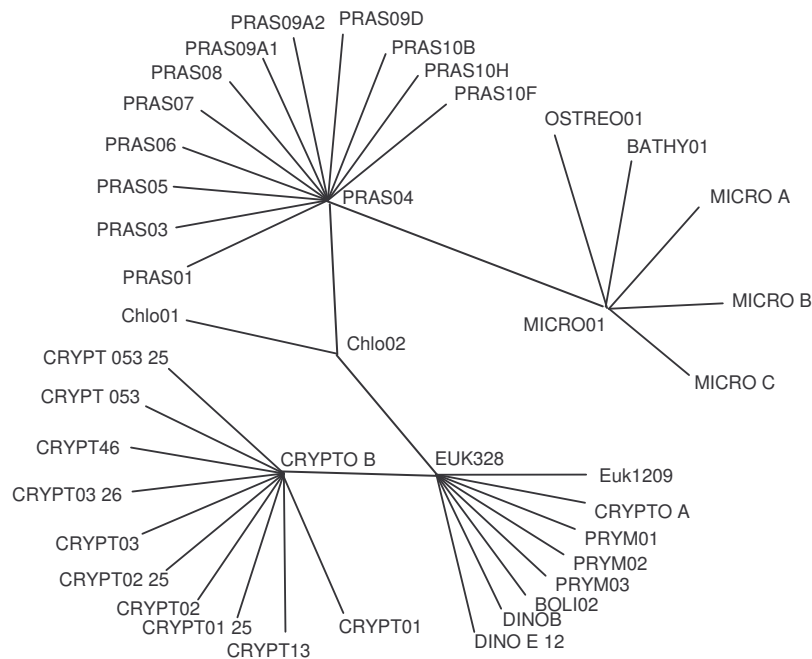


**Fig. 9:** Two false positives corrected because probe MICRO01 on the higher level showed a negative signal.

The analysis results may be exported to Excel directly (see menu in Fig. 7). A view of such an export with manually modified scaling is seen in Fig. 10. Furthermore, the hierarchy may be saved in a tree-file in Newick-format. Such a file can be drawn e.g., by the Treeview program [3] (results shown in Fig. 11).



**Fig. 10:** Zoom into a direct import to Excel, showing mean signal-background ratios (over all copies included in the analysis). The error bars indicate the standard deviation of the mean. Negative mean signals bars point downwards, e.g., probe CRYPT01 shows a negative signal.



**Fig. 11:** A tree from a hierarchy saved in Newick-Format (“Hierachy menu– Save as...”) drawn with the program Treeview.

## Refinement of the Analysis

The analysis may be refined by inspection of the block qualities, e.g., observations of a block showing much less consistency with other blocks can serve as a hint to exclude the questionable block from the analysis. The overall block statistics is always seen in the lower left part of the Analysis mode view (see Fig. 9). It shows how many spot signals of an individual block are matching the overall consensus. A block with large deviation can easily

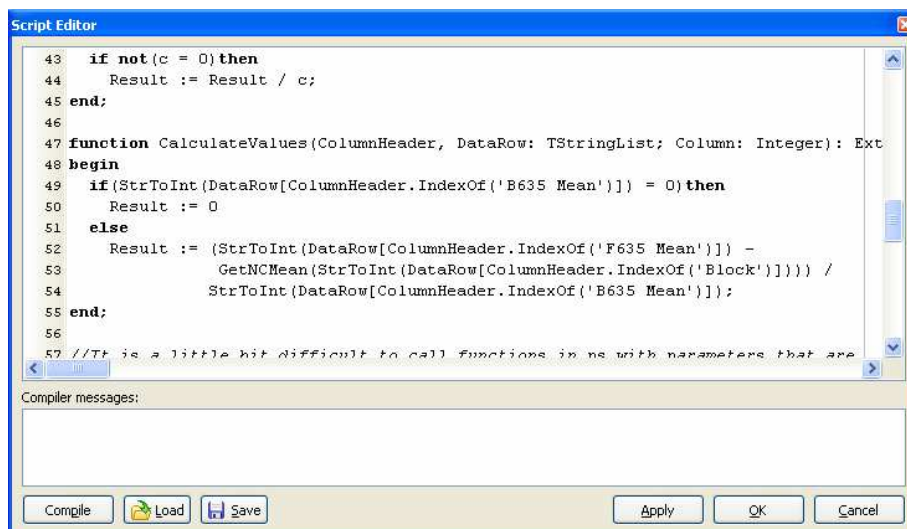
be excluded from the analysis, i.e., all spots of this block are invalidated and not considered in recalculation of mean values.

Furthermore, the user is free to exclude individual spots from the analysis. This may make sense if one can prove that cross-hybridization has occurred, or that an unknown species was present in the sample. When only a few blocks are on the chip, it may be problematic to exclude individual spot data. However, it may also be necessary to exclude a probe from the analysis when it demonstrated to be of too low specificity.

## A Note on Negative Controls

Negative control probes should always be included on the chip and in the analysis, as they allow for an alternative way to calculate signal-background ratios. Negative controls should have ID and Name “NC” to be correctly identified for the computational analysis. A chip without negative controls can only be evaluated by looking at local or global background signals and the according standard deviation.

The evaluation method of the PhylochipAnalyzer is coded in a Delphi-Script, which can be edited (Top-menu “Show Script editor”, Fig. 12). An alternative script “script-alternative.psf” is included in the sample-data folder. It allows for an analysis with respect to spot background mean and standard deviation. In the script the name of the negative control “NC” can also be modified.

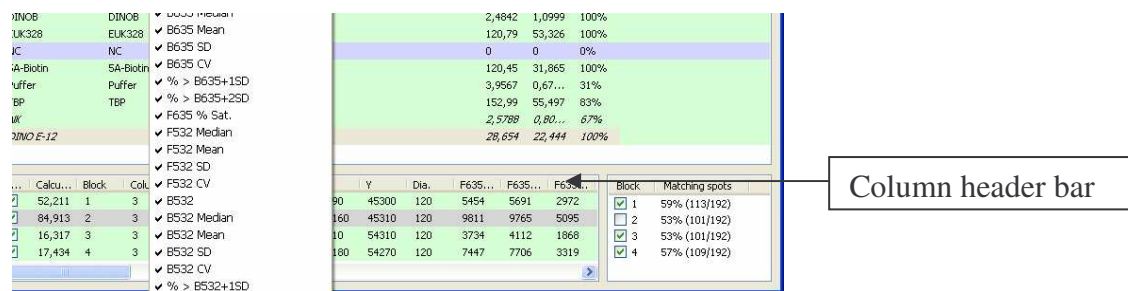


```
43 if not (c = 0) then
44     Result := Result / c;
45 end;
46
47 function CalculateValues(ColumnHeader, DataRow: TStringList; Column: Integer): Ext
48 begin
49     if (StrToInt(DataRow[ColumnHeader.IndexOf('B635 Mean')]) = 0) then
50         Result := 0
51     else
52         Result := (StrToInt(DataRow[ColumnHeader.IndexOf('F635 Mean')]) -
53             GetNCMean(StrToInt(DataRow[ColumnHeader.IndexOf('Block')])) /
54             StrToInt(DataRow[ColumnHeader.IndexOf('B635 Mean')]));
55     end;
56
57 //It is a little bit difficult to call functions in ps with parameters that are
```

**Fig. 12:** Script editor, showing the formula to evaluate signal-noise-ratios from the alternative script.

## Other Features

The view of the data in the lower left part in the analyser mode can be modified in a very flexible way. The user may shift any column to other positions by clicking on the column header for drag and drop. He is also free to deselect quantities from the view by right-clicking on the column header bar, as seen in Fig. 13.



**Fig. 13:** Right-clicking the column header bar lets you select/deselect data columns to display

## To Further Develop

The PhylochipAnalyser is far from being a perfect tool. In particular, it is not possible to save the current state of an analysis session with all its selections (thresholds, check-boxes). A view on more than one chip at the same time is also strongly needed for larger comparative studies or time series. For these, extensions could be thought of, enabling cluster-analysis over probes/species, locations or times. The rather simple elimination procedure for false positives can also be put onto a more quantitative ground by analysing correlations of signals over the different taxonomic levels. This would require an in-depth understanding of the quantitative relations of (cross-) hybridization efficiencies and unspecific background signals. Recently, it was demonstrated that a rather limited set of probes yields consistent quantitative estimates [4].

[1] Katja Metfies, Philipp Borsutzki, Christine Gescher, Linda K. Medlin, Stephan Frickenhaus. (2007). PhylochipAnalyzer - A Program for Analysing Hierarchical Probe-Sets, *Molecular Ecology Notes*, submitted

[2] Kumar Y, Westram R, Behrens S, Fuchs B, Glockner FO, Amann R, Meier H, Ludwig W. (2005). Graphical representation of ribosomal RNA probe accessibility data using ARB software package. *BMC Bioinformatics*. Mar 21;6:61.

[3] Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.

[4] Alex E. Pozhitkov, Kyle D. Bailey, Peter A. Noble (2007). Development of a statistically robust quantification method for microorganisms in mixtures using oligonucleotide microarrays. To appear in: *Journal of Microbiological Methods*. doi:10.1016/j.mimet.2007.05.001