ViTraM: VIsualization of TRAnscriptional Modules

Version 1.0

June 1st, 2009

Hong Sun, Karen Lemmens, Tim Van den Bulcke, Kristof Engelen,
Bart De Moor and Kathleen Marchal
KULeuven, Belgium

"Quick Start" guide

This "Quick Start" guide will assist the user in exploring the possibilities of ViTraM by means of an example. In this illustrative example, ViTraM was applied on the results obtained by the module detection tool DISTILLER [1]. DISTILLER was applied on *E. coli* data and this resulted in the identification of overlapping regulatory modules and the corresponding regulatory program (in this case the motifs(s) responsible for the observed co-expression = module properties). In addition, the scores of each motif for each gene were obtained from a motif screening (so not automatically assigned by the algorithm). The information on the regulatory modules' composition, the module properties and the gene properties (the motif scores) were included in the module file for ViTraM.

To be able to fully exploit all possibilities of ViTraM, additional information for these modules was included in module file. As DISTILLER assigns only motifs to modules, all functionalities regarding regulatory motifs (module properties) can be used. To allow the user to apply all functionalities regarding regulators on our example data set, the corresponding regulators of these motifs were added to the module file as module properties. As such all functionalities of ViTraM for regulators can also be used.

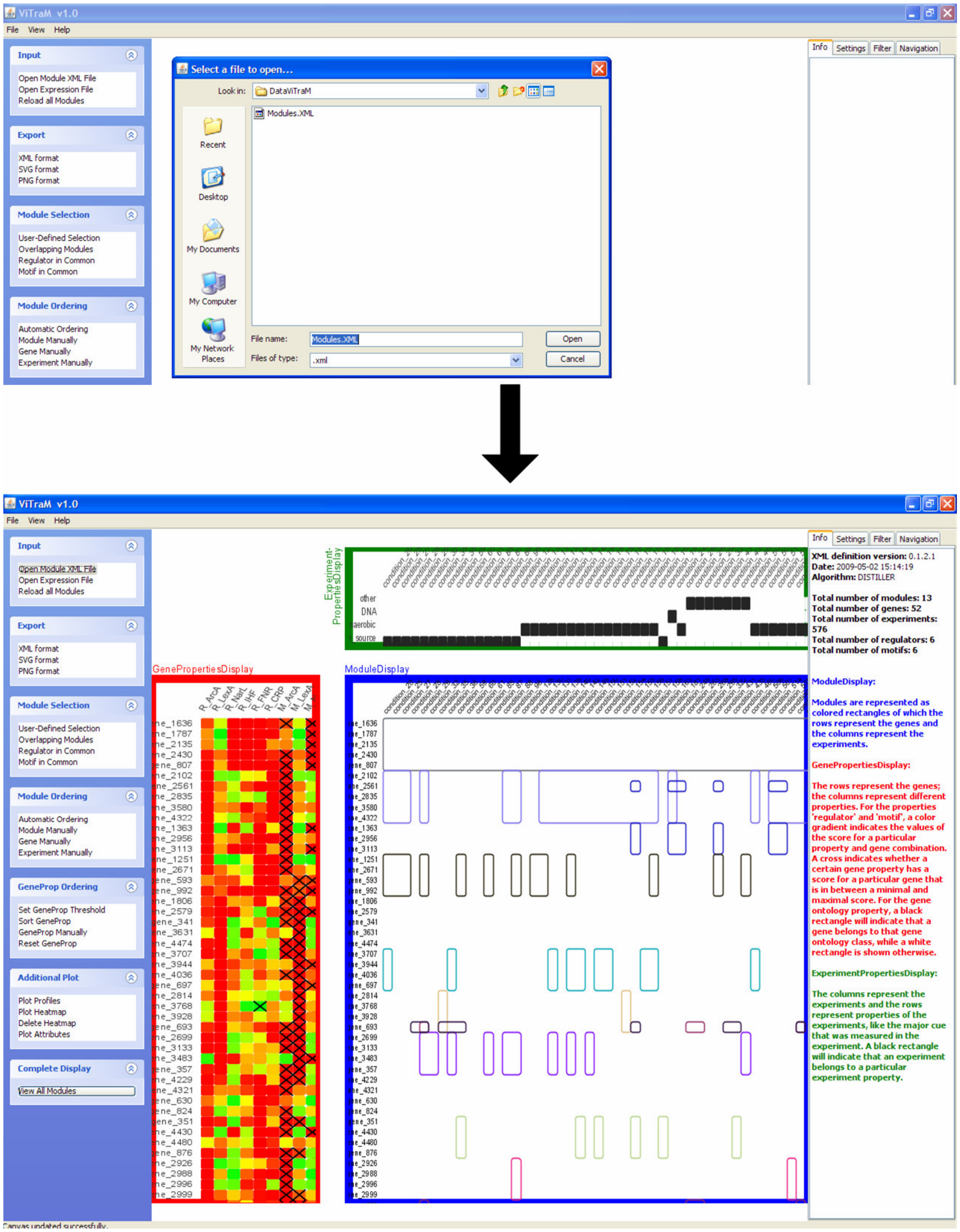
In addition to these module properties, we also included scores that indicate the binding of a regulator to each promoter region. As such, also all functionalities of ViTraM for the gene property 'regulator' can be used. To obtain these scores, the corresponding motif scores, obtained from a motif screening, were multiplied by 100. As such the flexibility of ViTraM in handling all kinds of data can be demonstrated.

Finally, all genes of a module were assigned to a functional class and all experiments were assigned to some experiment conditional classes. Expression data for the obtained modules were available in the expression data file.

These data can be found on our website:

ftp://ftp.esat.kuleuven.be/sista/klemmens/ViTraM/Index.html

Step 1 After loading the module and expression data file into ViTraM, all modules are initially displayed according to their order in the input file, resulting in a non-optimal representation of the modules.



Look in: DataViTraM

Modules.XML

File name: Modules.XML

Files of type: .xml

XML definition version: 0.1.2.1
Date: 2009-05-02 15:14:19
Algorithm: DISTILLER

Total number of modules: 13
Total number of genes: 52
Total number of experiments: 576
Total number of regulators: 6
Total number of motifs: 6

ModuleDisplay:

Modules are represented as colored rectangles of which the rows represent the genes and the columns represent the experiments.

GenePropertiesDisplay:

The rows represent the genes; the columns represent different properties. For the properties 'regulator' and 'motif', a color gradient indicates the values of the score for a particular property and gene combination. A cross indicates whether a certain gene property has a score for a particular gene that is in between a minimal and maximal score. For the gene ontology property, a black rectangle will indicate that a gene belongs to that gene ontology class, while a white rectangle is shown otherwise.

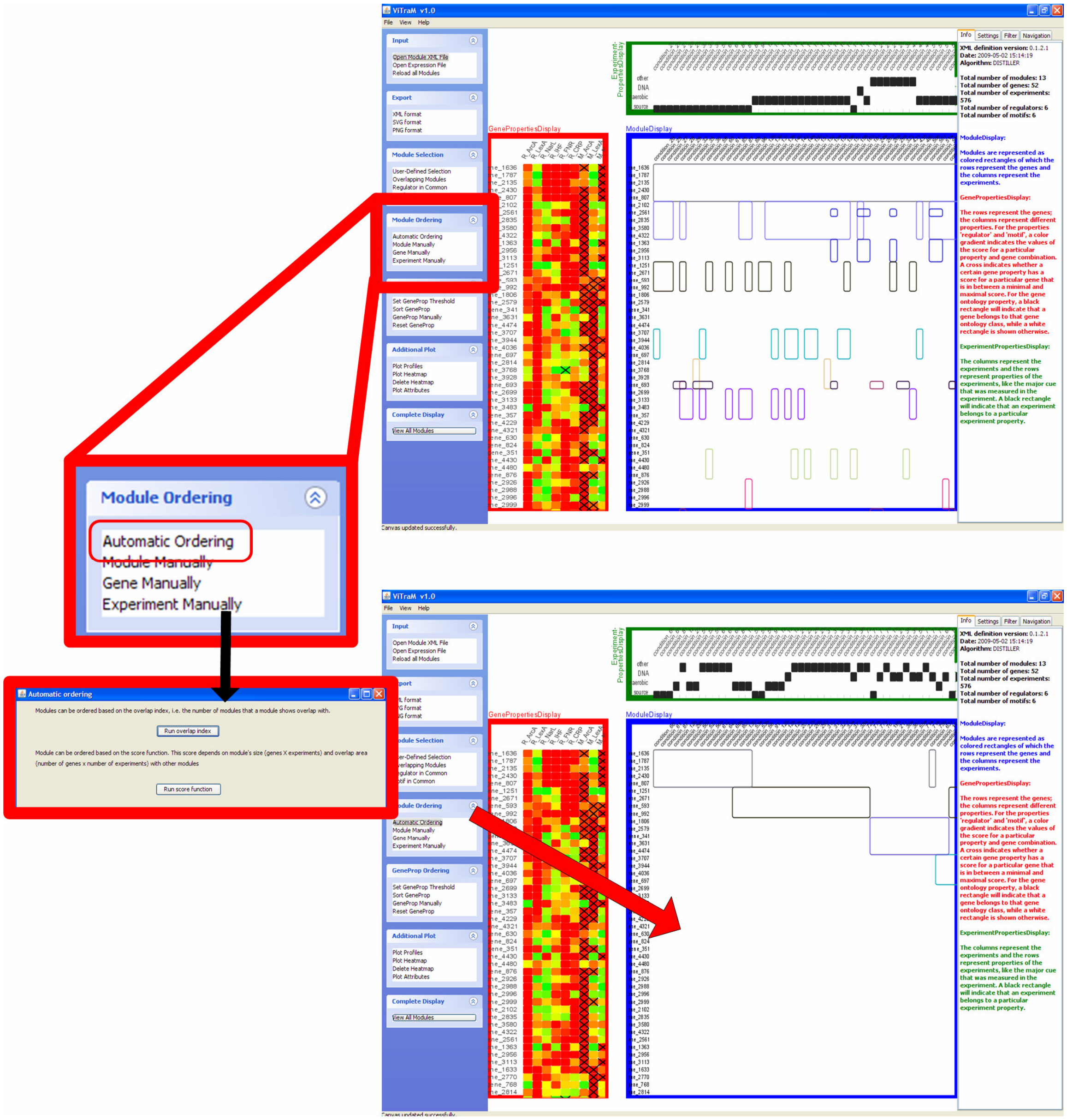
ExperimentPropertiesDisplay:

The columns represent the experiments and the rows represent properties of the experiments, like the major cue that was measured in the experiment. A black rectangle will indicate that an experiment belongs to a particular experiment property.

Canvas updated successfully.

Figure 1: Loading the module file. Initially all modules are displayed according to their order in the module file.

Step 2 By using one of the ordering algorithms of ViTraM, for instance the overlap index, the optimal layout of the modules is determined, resulting in an improved layout. From this ordered ModuleDisplay, it is possible to see which modules show overlap in genes and/ or experiments with other modules.



Canvas updated successfully.

XML definition version: 0.1.2.1
Date: 2009-05-02 15:14:19
Algorithm: DISTILLER

Total number of modules: 13
Total number of genes: 52
Total number of experiments: 576
Total number of regulators: 6
Total number of motifs: 6

ModuleDisplay:

Modules are represented as colored rectangles of which the rows represent the genes and the columns represent the experiments.

GenePropertiesDisplay:

The rows represent the genes; the columns represent different properties. For the properties 'regulator' and 'motif', a color gradient indicates the values of the score for a particular property and gene combination. A cross indicates whether a certain gene property has a score for a particular gene that is in between a minimal and maximal score. For the gene ontology property, a black rectangle will indicate that a gene belongs to that gene ontology class, while a white rectangle is shown otherwise.

ExperimentPropertiesDisplay:

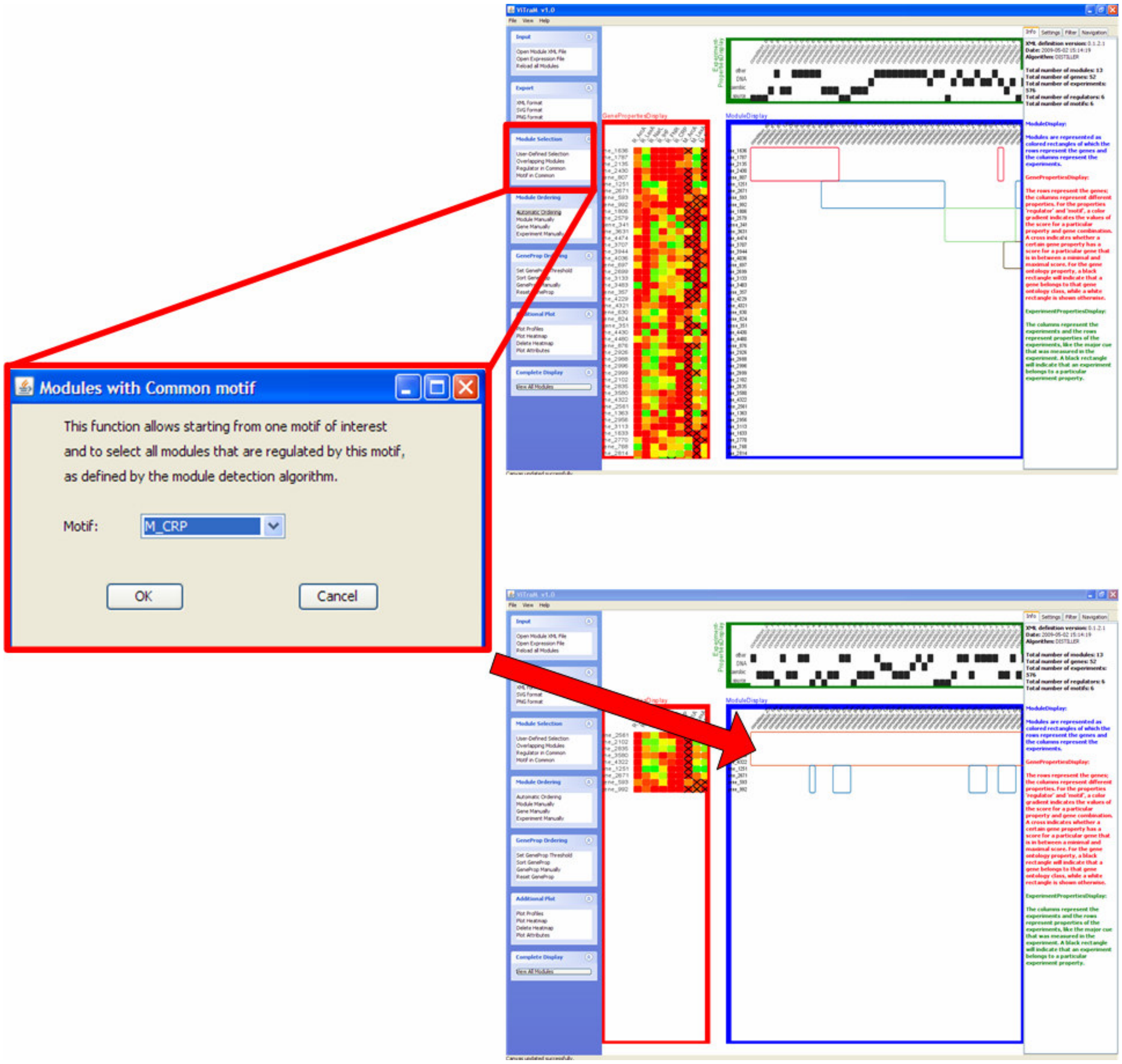
The columns represent the experiments and the rows represent properties of the experiments, like the major cue that was measured in the experiment. A black rectangle will indicate that an experiment belongs to a particular experiment property.

Modules can be ordered based on the overlap index, i.e. the number of modules that a module shows overlap with.

Module can be ordered based on the score function. This score depends on module's size (genes X experiments) and overlap area (number of genes x number of experiments) with other modules

Figure 2: ViTraM includes several ways for obtaining an optimal layout of the modules. After using, for instance, the "Overlap Index" function the modules are ordered such that genes and experiments of a module are grouped together as much as possible, bringing into account overlap with other modules. Ordering the modules therefore allows for a clearer overview on the modules and makes the biological analysis of the modules more easy.

Step 3 In addition to viewing all modules, ViTraM also has several options for selecting modules the user wants to display. These functionalities will always operate on all modules in the module file. ViTraM allows selecting modules based on a common regulator or motif, such as for instance CRP. It is possible to select all overlapping modules of a particular module of interest. A user-defined selection is also possible: manually selecting modules 9 and 20 for instance will give the same result as selecting modules that have the CRP motif in common.



Modules with Common motif

This function allows starting from one motif of interest and to select all modules that are regulated by this motif, as defined by the module detection algorithm.

Motif: M_CRP

Figure 3: A subset of modules can be selected in several ways. Selecting all modules with a common motif M_CRP, for instance, results in the selection of two modules out of 13. Only these two modules will be displayed.

Step 4 After selecting modules, an optimal layout can again be obtained by ordering the modules by using the overlap index.

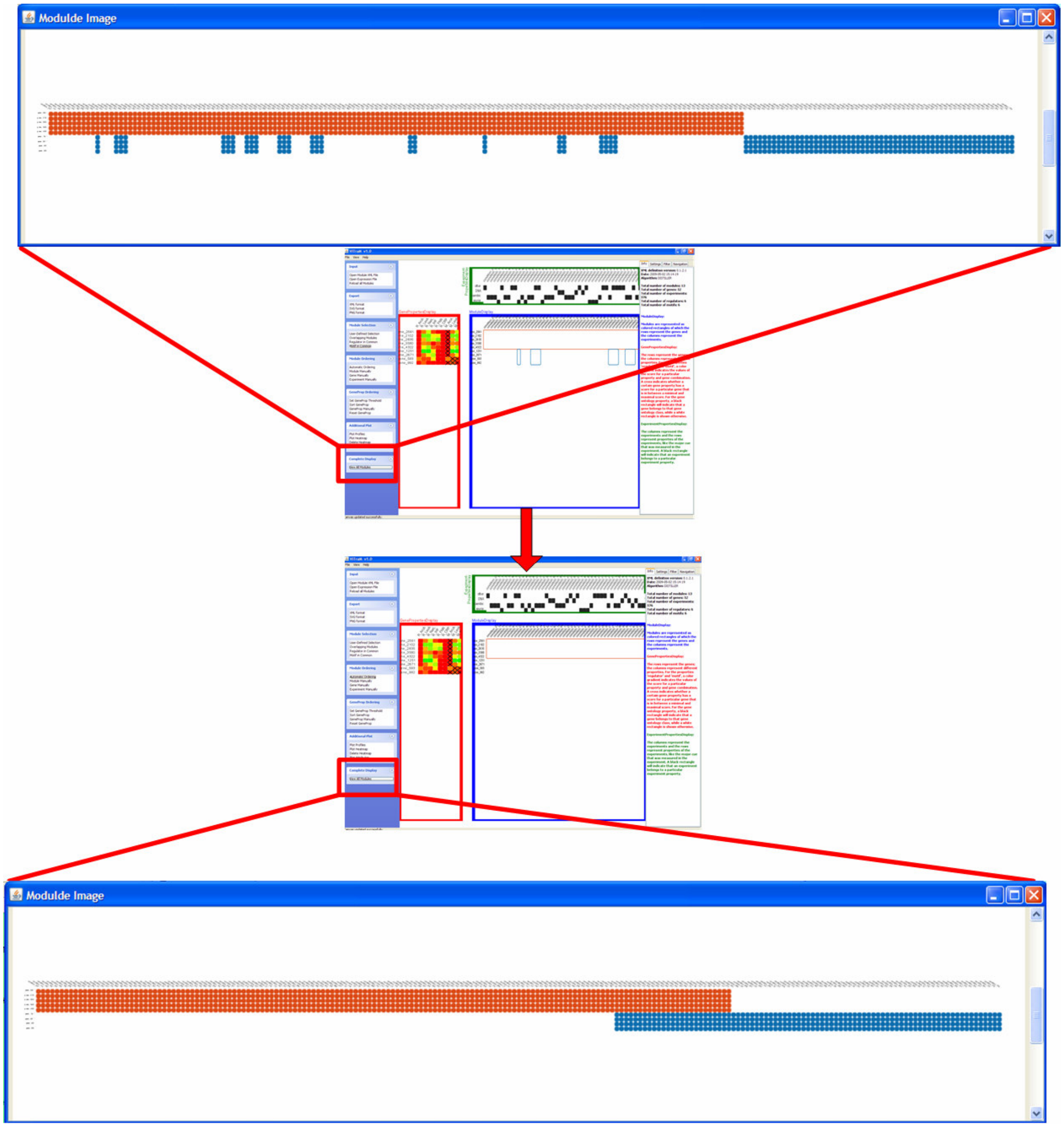
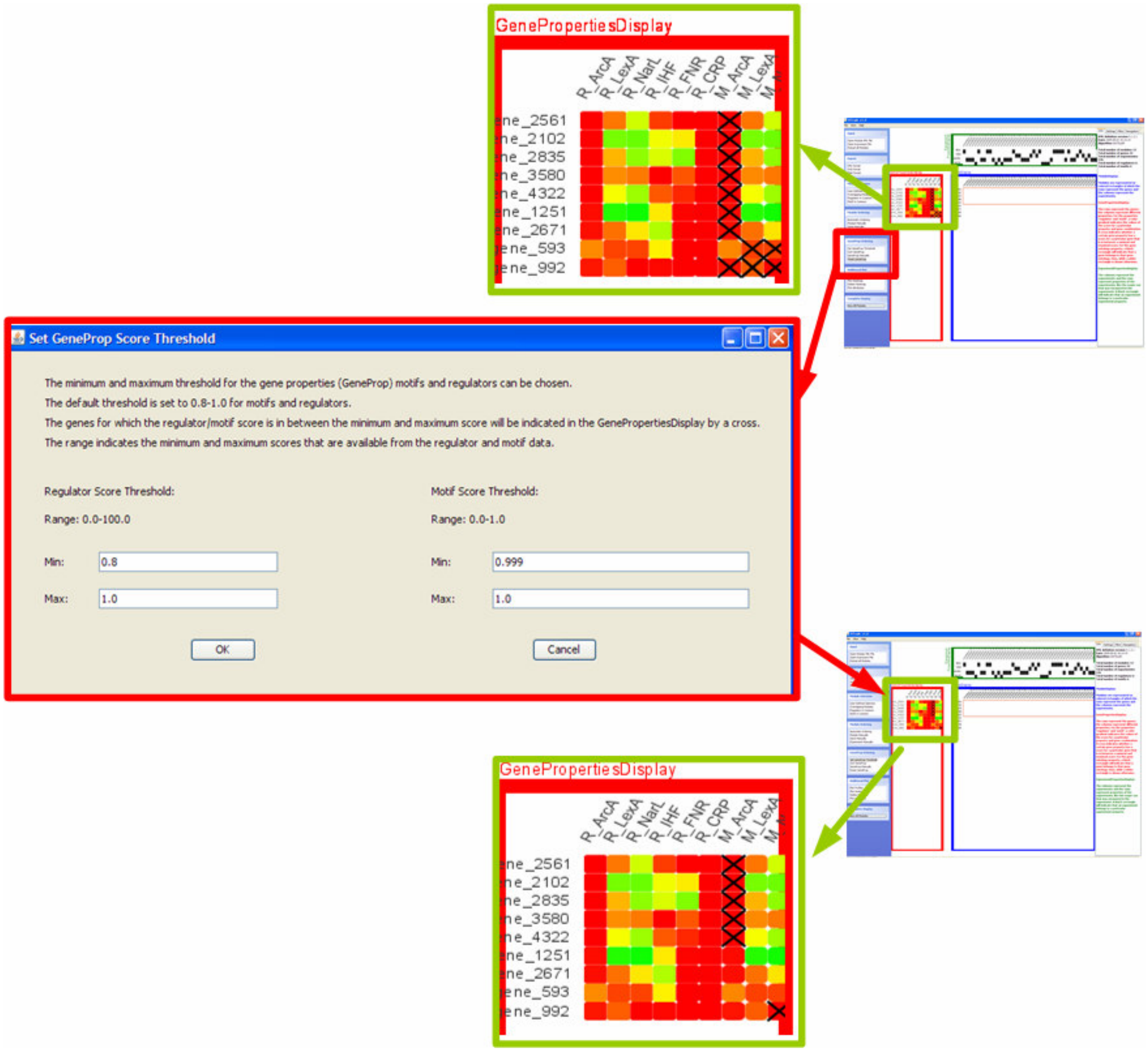


Figure 4: The optimal layout for the selected modules can be determined by using one of ViTraM's ordering techniques, for instance *"Overlap Index"*. As can be seen from the Overview Display, the genes and conditions are indeed grouped together such that no module is split up anymore.

Step 5 In the GenePropertiesDisplay, the scores of the regulators and motifs for the genes in the CRP modules are shown. The scores that satisfy the default thresholds are indicated by a cross. Changing the default score settings is possible by clicking “Set GeneProp Threshold” in the “GeneProp Ordering” panel. We changed the default minimum score for the motifs into 0.999 instead of 0.8.



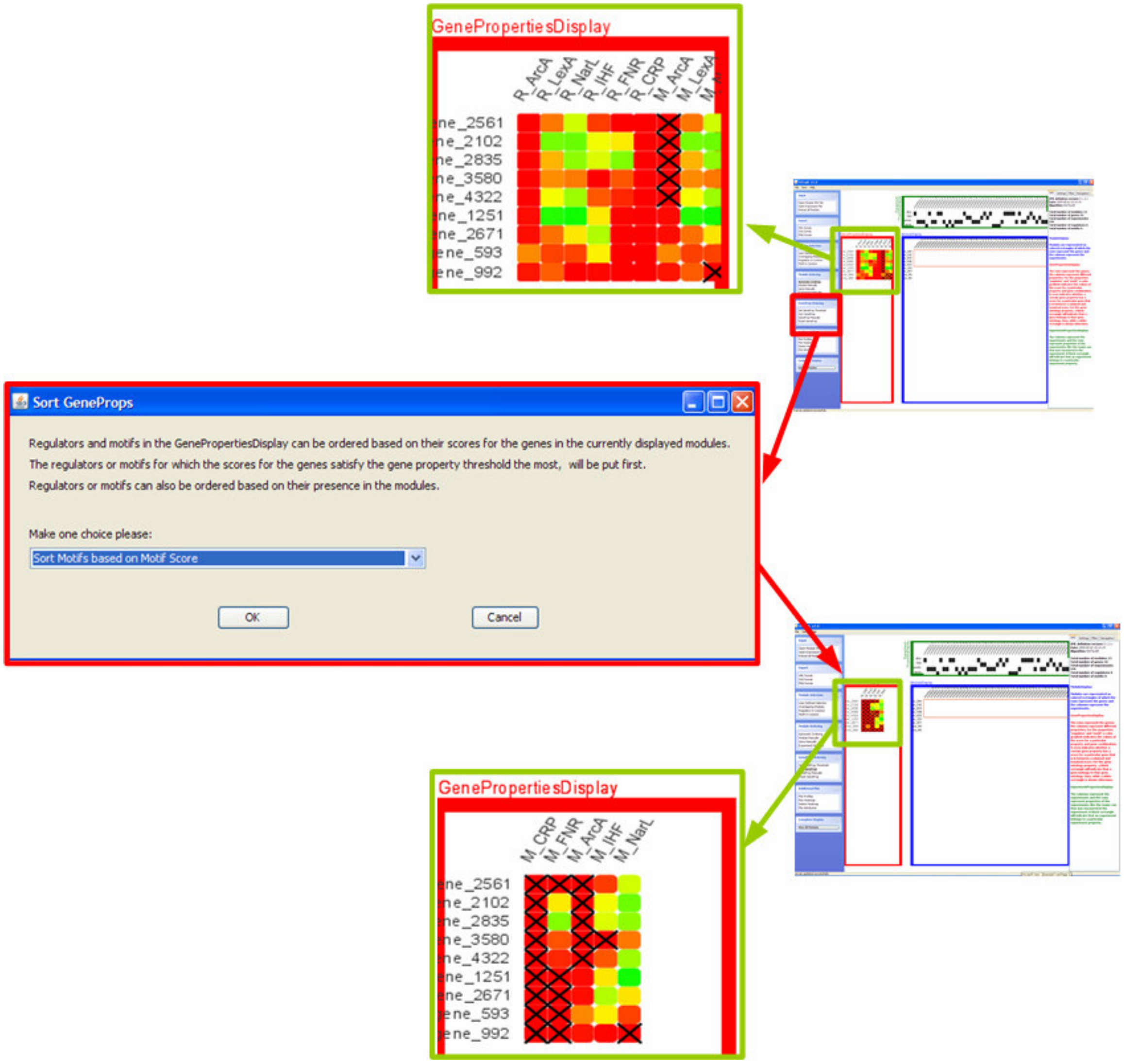
The minimum and maximum threshold for the gene properties (GeneProp) motifs and regulators can be chosen.
The default threshold is set to 0.8-1.0 for motifs and regulators.
The genes for which the regulator/motif score is in between the minimum and maximum score will be indicated in the GenePropertiesDisplay by a cross.
The range indicates the minimum and maximum scores that are available from the regulator and motif data.

Regulator Score Threshold:
Range: 0.0-100.0
Min: 0.8
Max: 1.0

Motif Score Threshold:
Range: 0.0-1.0
Min: 0.999
Max: 1.0

Figure 5: The gene properties are shown in the GenePropertiesDisplay. Scores for regulators and motifs are represented by a color. In red are the highest values, in green are the lowest values. A cross indicates that a particular score is in between the score thresholds. The minimum score threshold is default set to 0.8, while the default maximum score is 1. These thresholds can be changed by using the “Set GeneProp Threshold” button. In this example, we changed the minimum motif score to 0.999.

Step 6 When sorting the motifs in this display according to the scores, it is clear that in addition to the CRP motif, also the ArcA motif is important for one module, as this motif has high scores for all genes in this module. On the other hand, the FNR motif has high scores for the genes in the second module.

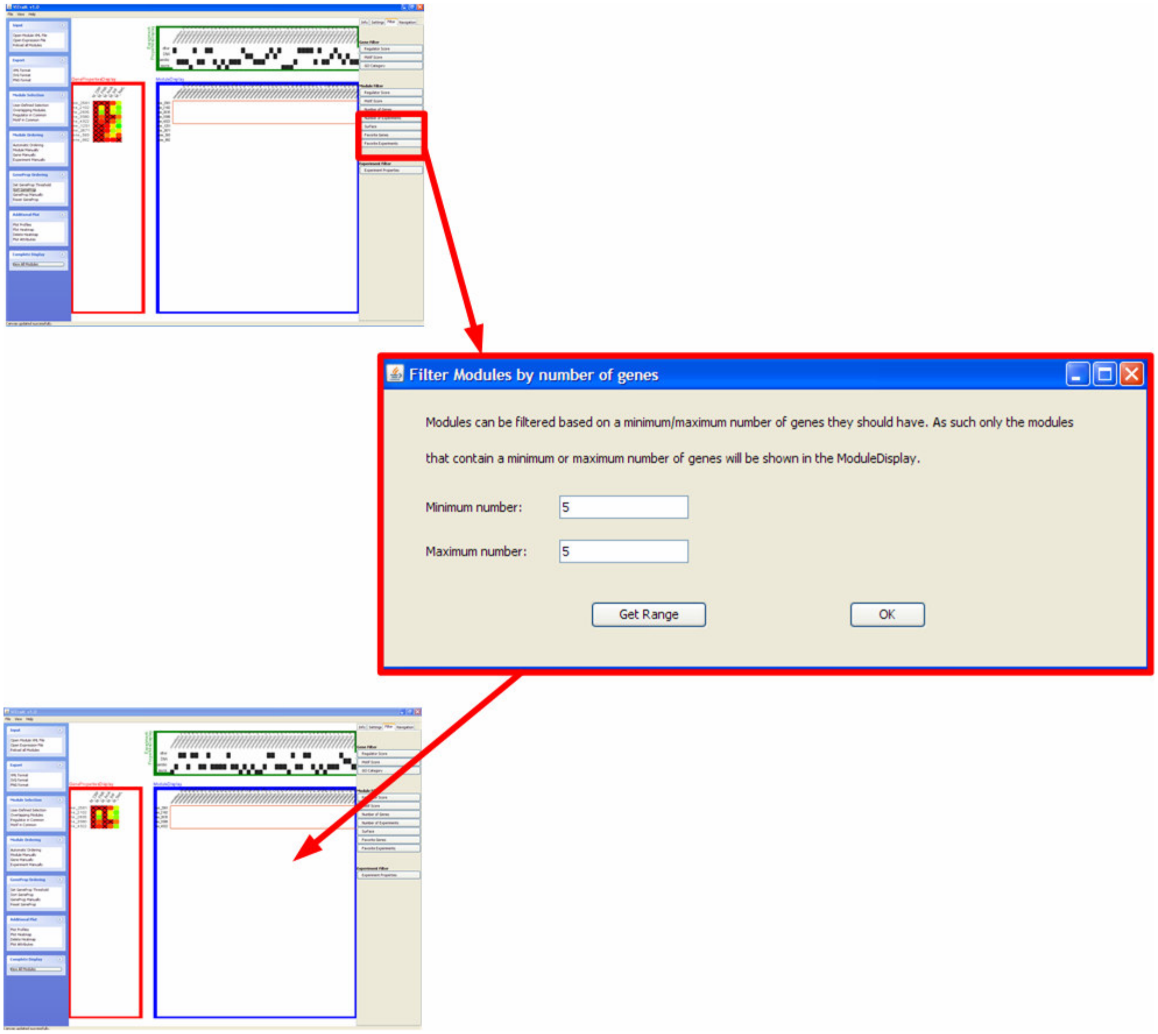


Regulators and motifs in the GenePropertiesDisplay can be ordered based on their scores for the genes in the currently displayed modules.
The regulators or motifs for which the scores for the genes satisfy the gene property threshold the most, will be put first.
Regulators or motifs can also be ordered based on their presence in the modules.

Make one choice please: Sort Motifs based on Motif Score

Figure 6: ViTraM includes several options for ordering the gene properties. In this example, the motifs were ranked according to their score for the genes in the currently displayed modules. Motif M_CRP is the motif with the best scores, followed by motifs M_ArcA and M_FNR.

Step 7 ViTraM includes a filtering of modules, genes and experiments. The filtering is always performed on the currently displayed modules. It is thus possible to apply several filtering steps sequentially. Modules can for instance be filtered, based on a minimum/ maximum of genes. This means that only modules that have a number of genes in the user-defined range will be shown.



Filter Modules by number of genes

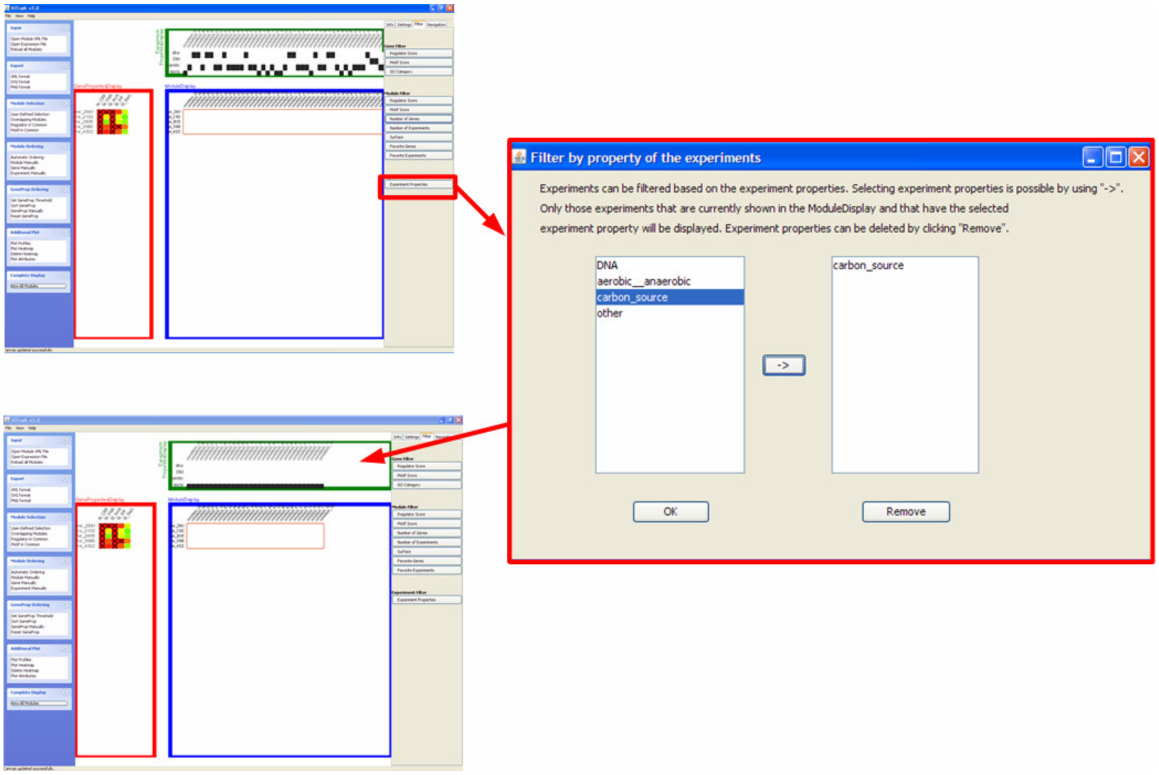
Modules can be filtered based on a minimum/maximum number of genes they should have. As such only the modules that contain a minimum or maximum number of genes will be shown in the ModuleDisplay.

Minimum number: 5

Maximum number: 5

Figure 7: Filtering the modules by the number of genes they contain is possible. When choosing, for instance, modules with exactly 5 genes, only one module of the two satisfies this criterion and only that module is displayed.

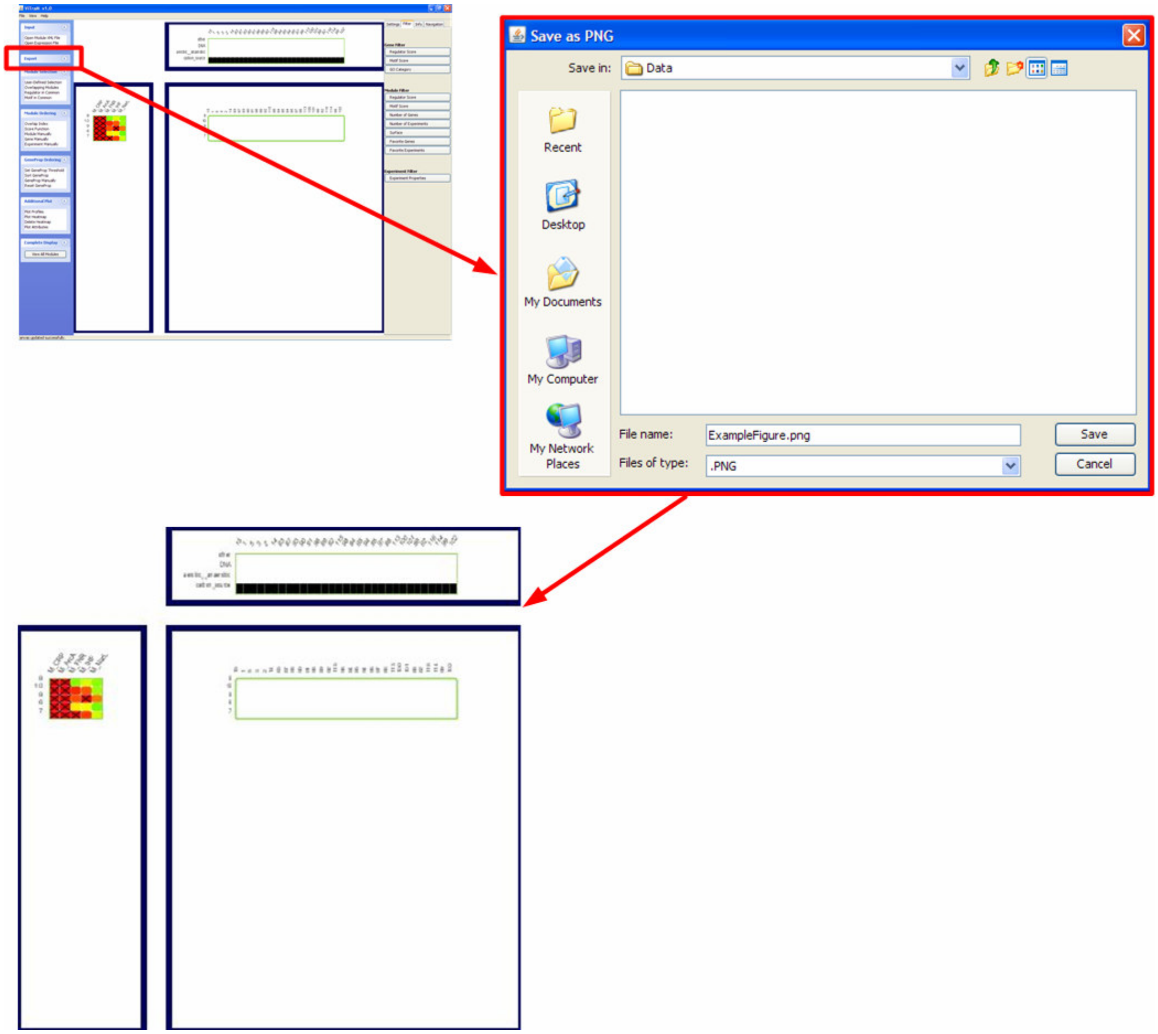
Step 8 The ExperimentPropertiesDisplay shows that many experiments in which the genes of this module were co-expressed belong to either the carbon_source category or the anaerobic_aerobic category. Filtering the experiments by carbon_source reduces the number of experiments.



Experiments can be filtered based on the experiment properties. Selecting experiment properties is possible by using "->". Only those experiments that are currently shown in the ModuleDisplay and that have the selected experiment property will be displayed. Experiment properties can be deleted by clicking "Remove".

Figure 8: Several filtering procedures are included in ViTraM. Filtering the experiments by their properties is one of the options. If., for example, a filtering is done based on the property carbon_source, only those experiments that belong to this conditional class will be shown. This results in a reduced number of experiments that are displayed.

Step 9 The selected modules can be exported to a module XML file. This file can later on be imported again in ViTraM. The layout of the modules in the ModuleDisplay can also be exported as an SVG or PNG figure.



Save as PNG

Save in: Data

File name: ExampleFigure.png

Files of type: .PNG

Figure 9: Saving the current display as a figure. The current display can be saved as either PNG or SVG format.

REFERENCES

1. Lemmens K, De Bie T, Dhollander T, De Keersmaecker S, Thijs I, Schoofs G, De Weerdt A, De Moor B, Vanderleyden J,Collado-Vides J, Engelen K, Marchal K**. Condition-dependent combinatorial regulation in Escherichia coli**. submitted 2009.