BMC Research Notes

TECHNICAL NOTE Open Access

MediPlEx - a tool to combine in silico & experimental gene expression profiles of the model legume Medicago truncatula

Kolja Henckel[1*], Helge Küster[2], Leonhard J Stutz[3,4], Alexander Goesmann[3]

Abstract

Background: Expressed Sequence Tags (ESTs) are in general used to gain a first insight into gene activities from a species of interest. Subsequently, and typically based on a combination of EST and genome sequences, microarray-based expression analyses are performed for a variety of conditions. In some cases, a multitude of EST and microarray experiments are conducted for one species, covering different tissues, cell states, and cell types. Under these circumstances, the challenge arises to combine results derived from the different expression profiling strategies, with the goal to uncover novel information on the basis of the integrated datasets.

Findings: Using our new analysis tool, MediPlEx (MEDIcago truncatula multiPLe EXpression analysis), expression data from EST experiments, oligonucleotide microarrays and Affymetrix GeneChips® can be combined and analyzed, leading to a novel approach to integrated transcriptome analysis. We have validated our tool via the identification of a set of well-characterized AM-specific and AM-induced marker genes, identified by MediPlEx on the basis of *in silico* and experimental gene expression profiles from roots colonized with AM fungi.

Conclusions: MediPlEx offers an integrated analysis pipeline for different sets of expression data generated for the model legume *Medicago truncatula*. As expected, *in silico* and experimental gene expression data that cover the same biological condition correlate well. The collection of differentially expressed genes identified via MediPlEx provides a starting point for functional studies in plant mutants. MediPlEx can freely be used at http://www.cebitec.uni-bielefeld.de/mediplex.

Background

Medicago truncatula is a model plant for the functional analysis of legume biology [1]. The ability to interact with beneficial microbial organisms leading to the formation of nitrogen- fixing root nodules [2] and to phosphate-acquiring arbuscular mycorrhizal (AM) roots [3] is one of the main distinctive features of the legume family. AM interactions between the host root and the fungal partner are a particularly interesting field of research, since more than 80% of land plants depend on an efficient AM for the uptake of nutrients, primarily phosphate [4]. By recruiting the basic genetic programme allowing microbial infection during AM [5], legumes such as *Medicago truncatula* evolved the capacity to enter a second beneficial interaction: the nitrogen-fixing symbiosis with the soil bacterium *Sinorhizobium meliloti* [6]. Symbiotic nitrogen fixation allows legume plants such as *Medicago truncatula* to grow on nitrogen-depleted soils and to develop protein-rich seeds, properties exploited in sustainable agriculture. Likewise, apart from direct advantageous effects resulting from an improved plant nutrition, an important indirect benefit of mycorrhization is an enhanced resistance against different abiotic and biotic stress conditions [7].

The great interest in transcriptome studies in *Medicago truncatula* (more than 500 publications in Pubmed [8] by searching for "*Medicago truncatula*" as keywords and the last 5 years as publication time span) is evidenced by the generation and sequencing of more than 70 cDNA libraries, in total yielding more than 250.000 ESTs stored in the DFCI Medicago Gene Index [9].

* Correspondence: khenckel@cebitec.uni-bielefeld.de
[1]Bioinformatics of Signaling Networks, Center for Biotechnology, Bielefeld University, Germany
Full list of author information is available at the end of the article

Parallel to the generation of EST data, thousands of oligonucleotide microarrays were hybridized with targets from different biological conditions [10], using layouts such as Mt16kOLI1 [11] and Mt16kOLI1Plus [12] (Arrayexpress ID: A-MEXP-85/A-MEXP-138). In the last couple of years, Affymetrix Medicago GeneChips® more and more moved into the focus of Medicago transcriptomics, since these more genome-wide tools allow a better comparison of gene expression data from a multitude of conditions [13], leading to more accurate results. Parallel to the development and use of transcriptomics tools, a genome project was conduced for *Medicago truncatula* [14,15].

Different institutes store the various sequence and expression datasets, using them for further analysis, or offering them as downloads. At the J. Craig Venter Institute (TIGR before 2006) EST libraries are clustered and assembled, resulting in species-specific Gene Indices [9] for over 100 species. These GeneIndices, including the *Medicago truncatula* GeneIndex 10.0, are now hosted at the Dana-Farber Cancer Institute (DFCI) [16]. Storing information on how the ESTs were assembled, the GeneIndices allow to relate EST data to the biological conditions used for the generation of cDNA libraries, whilst statistical methods were developed to assess if a gene is differentially expressed under a given condition [17,18]. In contrast to EST data, a range of different databases such as GEO [19,20], Arrayexpress [21], PEPR [22], The Stanford MicroArray Database [23], and PlexDB [24] store microarray and GeneChip® expression data, offering researchers public access to results from transcriptomics experiments. In case of *Medicago truncatula*, the Medicago Gene Expression Atlas [13] has developed into a popular resource for expression profiles relying on Medicago GeneChips®.

To yield novel insights into gene expression, it would be desirable to integrate different kinds of *in silico* and experimental expression data. In case of the model legume *Medicago truncatula*, the TRUNCATULIX data warehouse [25] currently integrates five different sequence databases (MtGI 8.0 [9], MtGI 9.0 [9], *Medicago truncatula* 454 sequencing project [26], *Medicago truncatula* genome project 2.0 [27], Medicago GeneChip® reporter sequences) as well as oligonucleotide microarray and GeneChip® expression experiments from different source databases. The user can quickly scan the complete database for the expression of genes of interest, but downstream analyses of expression data cannot be performed inside the warehouse. This lack of an integrated expression analysis with an easy-to-use interface and a database connection prompted us to create MediPlEx (MEDIcago truncatula multiPLe EXpression analysis). We here report on the design and implementation of this tool and provide a first example for its use to identify genes activated in *Medicago truncatula* AM roots.

Results and Discussion

Software solution

To combine the different kinds of gene expression datasets, we created an analysis tool called MediPlEx. It can be launched via SAMS [28], a Sequence Analysis and Management System, that stores data on Tentative Consensus sequences (TCs = assembled ESTs). Loading the SAMS project for *Medicago truncatula*, users can start a combined expression analysis. To do so, the user first selects the cDNA libraries covering interesting biological conditions. Subsequently, MediPlEx gathers information on the composition of the relevant TCs from SAMS and calculates logarithmic likelihood ratios [17] (c.f. Methods Section), an *in silico* expression measure, for the selected TCs. The microarray experiments to be related to the EST expression data are selected during the next step. Afterwards, MediPlEx fetches the different expression datasets from the TRUNCATULIX data warehouse [25] that stores a variety of expression data being publicly available for *Medicago truncatula*. The results are clustered hierarchically and can subsequently be browsed in an interactive 3D visualization tool implemented in Java [29]. An export option offers the possibility to store the results of the combined expression analysis. A complete list of expression values can be viewed and the result of the hierarchical clustering is shown in a dendrogram. The combined search for gene expression on the basis of EST frequencies and microarray/GeneChip® hybridization data offers the possibility to exploit both *in silico* and experimental expression profiles of various sources to trace novel candidate genes for the biological condition of interest.

Biological findings

We compared the expression based on a selection of different EST-libraries to GeneChip® and microarray analyses performed in the same biological background, in this case the AM symbiosis (Figure 1). To identify AM-specific TCs (and thus AM-specific genes), we used the preselection "arbuscular mycorrhizal root libraries (6) ", consisting of the following selection of EST libraries from the DFCI Medicago GeneIndex:

- MUST contain ESTs (using 'OR' as concatenation):
- #9CR (Medicago truncatula mycorrhized roots 3 weeks)
- #ARB (MTGIM)
- #ARE (MTAMP)
- #GFS (MHAM2)
- 5520 (MtBC)
- T1682 (MHAM)

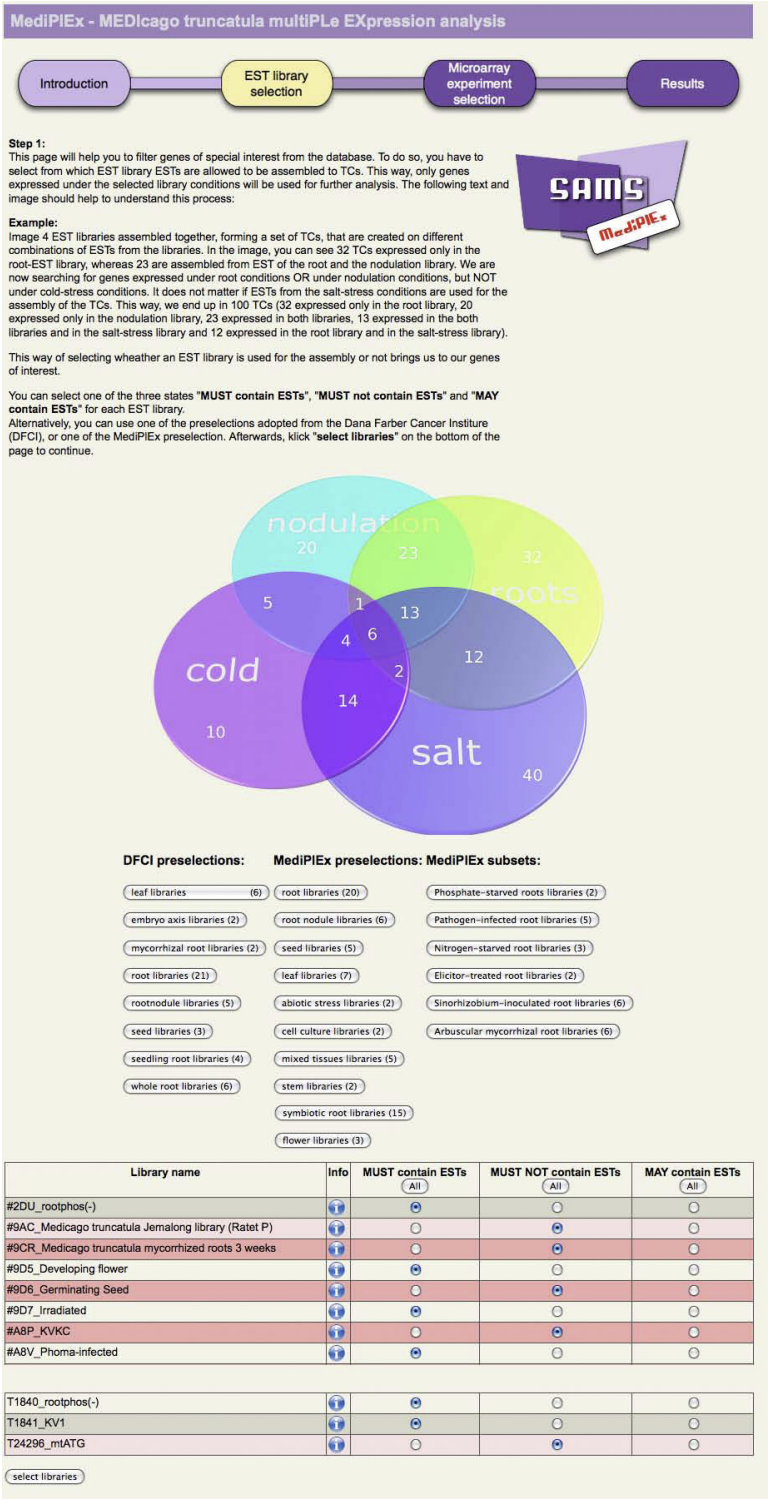

Figure 1 EST library selection. The screenshot shows the MediPlEx web page to select the genes of interest by using the EST library assembly. For each library, the user can select one of the three states: 'MUST contain ESTs', 'MUST NOT contain ESTs', and 'MAY contain ESTs'. 'MUST contain ESTs' means that the resulting TCs have to consist of at least one EST from these libraries. 'MUST NOT contain ESTs' denotes that the resulting TCs are not allowed to have an EST assembled from these libraries at all, and 'MAY contain ESTs' indicates that it does not matter if an EST is assembled from these libraries to a TC or not (see venn diagram). EST library preselection are implemented via buttons, the complete library list is available on the bottom of the page. The EST library list is shortened for this screenshot.

MAY contain ESTs (IGNORE):

- #IP8 (NOLLY)
- T10174 (kiloclone)
- T11958 (MTUS)
- T12308 (6KUG)

The libraries set to "MAY contain ESTs" were not considered since these mostly represent clone libraries used for microarray construction and thus do not contain information on tissue-specific gene expression. The Medicago GeneChip® datasets selected are derived from the experiment "*Medicago truncatula* AM and phospate-treated roots (Medicago GeneChip log2 expression ratios)", specifically the "*Glomus intraradices* AM roots vs. control roots at 20 miM phosphate" and "*Glomus mosseae* AM roots vs. control roots at 20 miM phosphate" datasets (shown in Figure 2). Following the TC search, 763 TCs fulfilled the specified conditions of an AM-specific EST composition [Additional file 1], and 751 of these were represented by reporters on the Affymetrix Medicago GeneChip® (see Figure 3). Sorting these TCs for the calculated logarithmic likelihood ratio R [17] that provides a measure for differential gene expression under the given conditions, we identified a range of AM marker genes [10,11,30], as was suggested by our search. Remarkably, a TC encoding the mycorrhiza-specific phosphate transporter MtPt4 (TC142142), a key marker gene for an efficient AM symbiosis [31], was identified as the top candidate. In addition, the identification of well-known AM-specific and AM-induced marker genes such as MtBcp1 (TC170722 [11]), MtGlp1 (TC153539 [32]), MtGst1 (TC166174 [33]), MtLec5 TC143161 [34]), MtMYBCC (TC146022 [10]), MtScp1 (TC143816 [35]), MtTi1 (TC152603 [36]) can be regarded as a proof-of-principle for the MediPlEx search strategy.

In general, the different expression values obtained by *in silico* and experimental expression analysis correlated very well. According to the dendrogram of the hierarchical clustering (Figure 4), we subsequently identified four clusters of expression profiles (the created clusters can be found in [Additional file 2]). Alternatively, clusterings with 2-8 cluster were generated, and the sizes of these can be found in Table 1. The generation of the

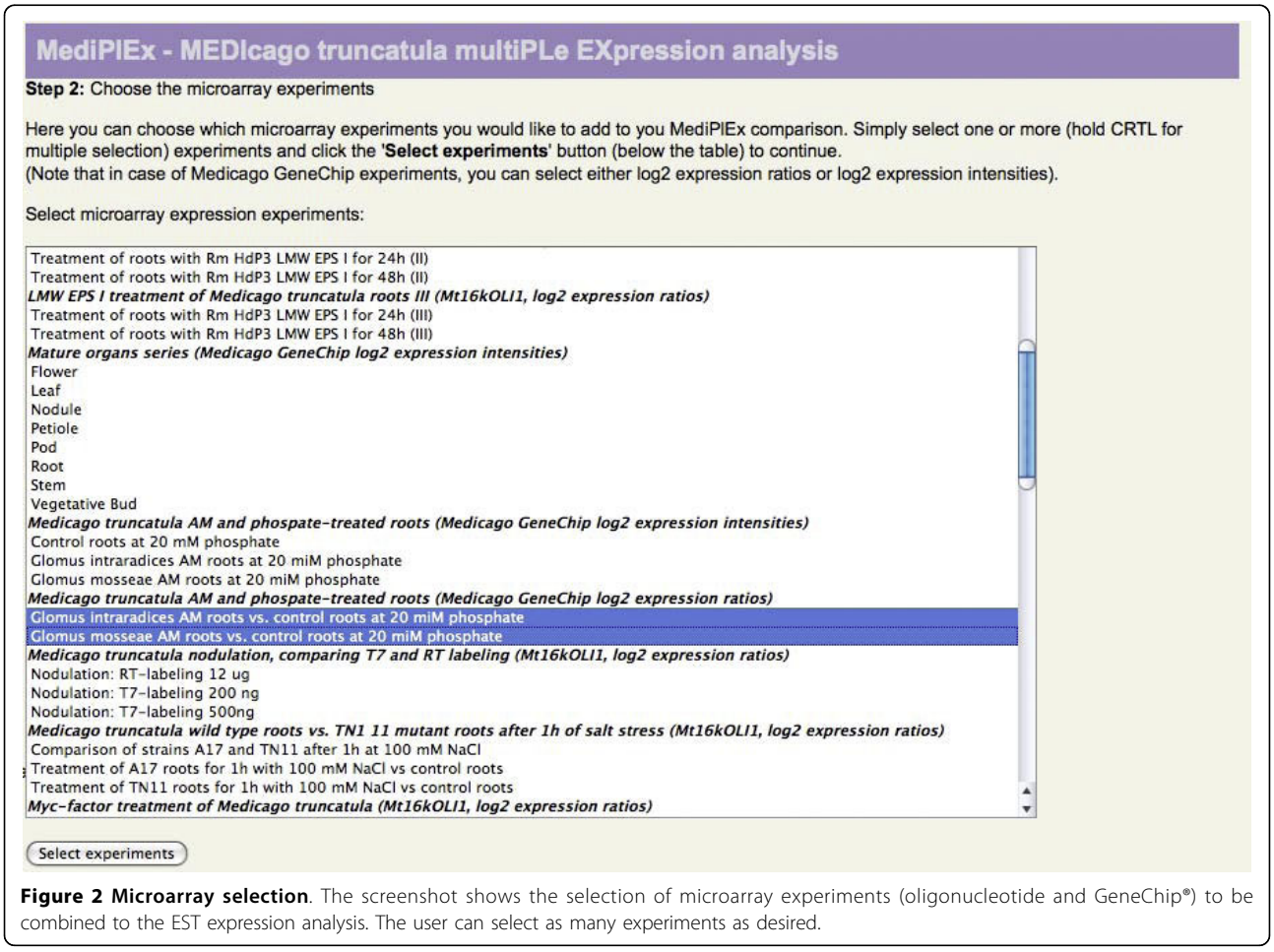

Figure 2 Microarray selection. The screenshot shows the selection of microarray experiments (oligonucleotide and GeneChip®) to be combined to the EST expression analysis. The user can select as many experiments as desired.

Results of your combines expression analysis

TC Name	Reporter Name	log likelihood ratio ▲	Glomus intraradices AM roots vs. control roots at 20 miM phosphate	Glomus mosseae AM roots vs. control roots at 20 miM phosphate
TC112872	MT009707 / Mtr.43062.1.S1_at	153.1662	10.0172996520996	8.65207958221436
TC131486	Mtr.8434.1.S1_at	135.3479	0.283836007118225	4.6682300567627
TC135802	MT009013 / Mtr.15957.1.S1_at	111.9159	8.98116970062256	8.12963962554932
TC124697	Mtr.40214.1.S1_at	90.2320	8.6655101776123	9.94029998779297
TC128110	MT008641 / Mtr.45893.1.S1_at	72.3016	10.3252000808716	9.25520038604736
TC114740	MT009185 / Mtr.31225.1.S1_at	68.5941	8.15530014038086	7.74411010742188
TC132711	MT008095 / Mtr.7475.1.S1_at	65.0061	9.33572006225586	8.36590003967285
TC128488	Mtr.10657.1.S1_at	51.5611	7.87438011169434	6.6447901725769
TC128939	MT014645 / Mtr.7210.1.S1_at	49.7840	9.31857967376709	8.75129985809326
TC137524	Mtr.37914.1.S1_at	42.1871	-0.0295109990984201	0.0949440002441406
TC123171	MT006798 / Mtr.16454.1.S1_at	39.5035	9.71012020111084	8.40919017791748
TC136093	MT002169 / Mtr.10562.1.S1_at	36.1232	7.80604982376099	7.13514995574951
TC134921	Mtr.10406.1.S1_at	35.4483	9.10970020294189	1.82492995262146
TC113973	MT013816 / Mtr.15653.1.S1_at	32.9652	8.0070104598999	7.64075994491577
TC132245	MT015421 / Mtr.35424.1.S1_at	32.6066	9.55953979492188	0.705334007740021
TC129609	MT002169 / Mtr.10562.1.S1_at	27.5054	7.80604982376099	7.13514995574951
TC124054	MT009704 / Mtr.12500.1.S1_at	24.3734	7.93924999237061	7.79982995986938

Figure 3 Resulttable. The screenshot shows the table listing the *in silico* calculated logarithmic likelihood ratio, as well as the expression datasets of the microarray experiments.

cluster is demonstrated in the dendrogram in Figure 4. The red lines indicate the positions where the cluster-tree is cut, the size of the resulting cluster is denoted at these positions. A 3D visualization can be started after selecting the three experiments for the three axis. Figure 5 shows the 3D visualization and the four cluster obtained in different colors.

Clustering of expression profiles reveals the predominant activation of different GO categories, e.g. ATP binding (5524), protein amino acid phosphorylation (6468), metabolic process (8152), binding (3677) and nucleus (5634) for cluster 1, transport (6810), DNA binding (3677), integral to membrane (16021), ATP binding (5524) and transporter activity (5215) for cluster 3, and intracellular (5622), structural constituent of ribosome (3735), translation (6412), and ribosome (5840) for cluster 4. These functional differentiations could indicate the fine-tuning or coregulation of specific cellular functions during fungal colonization of AM roots. A list of GO categories for the clustered genes can be found in [Additional file 3].

Discussion

Many EST and microarray experiments have been performed throughout the last years, leading to an immense amount of expression datasets. Our newly developed application, MediPlEx, integrates two different gene expression analysis methods (EST- and microarrays/GeneChip® -based transcriptome profiling), and delivers new results that have the potential to yield novel insights into gene expression in the model legume *Medicago truncatula*. Similar to MediPlEx, expression analyses can be performed using Simcluster, a tool developed by Vencio in 2007 [37]. Simcluster can take different expression experiment datasets, which include SAGE [38], MPSS [39], and Digital Northern powered by traditional [40] or, recently developed, EST sequencing-by-synthesis (SBS) technologies [41], and map them to the simplex space [42,43]. This mapping should make the data from different data sources and methods more comparable, as the simplex space does not use absolute values and scales, but relative (relative values to the overall expression for single experiments). Unfortunately, Simcluster is not

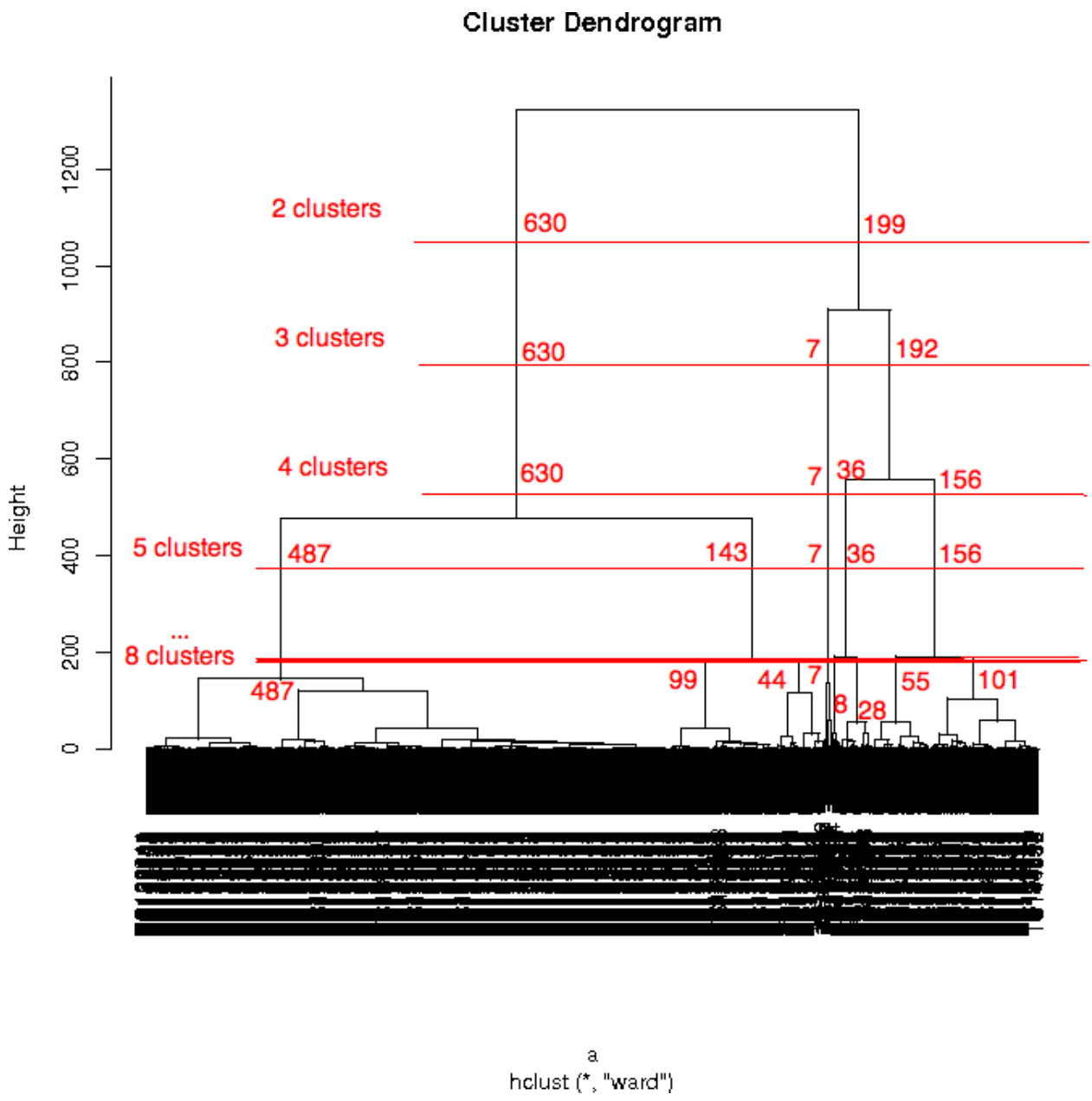

Figure 4 Cluster dendrogram. The cluster dendrogram created by the hierarchical clustering (ward clustering) of the expression profiles. The lines indicate which clusters are created when using the different cluster options. The sizes of the clusters are denoted at the specific branches.

connected to any database, so candidate genes and expression values have to be searched for elsewhere and converted to fit the designated format.

Thus, the database connection for obtaining expression data, combined with an easy-to-use web interface, is a major benefit of the MediPlEx tool. MediPlEx is currently available for two *Medicago truncatula* SAMS projects (SAMS_Medicago_truncatula_DFCI_9 & SAMS_Medicago_truncatula_DFCI_10). Datasets of upcoming expression analysis methods, such as SAGE or MPSS could be integrated as well, that way taking into account also the results of recently developed high-throughput expression profiling strategies.

Conclusions

The newly developed analysis tool MediPlEx offers an approach for combining gene expression values from already performed expression experiments in order to find candidate genes. By relating different experiments, the user can analyze and cluster transcriptomics data, visualize gene expression in 3D and find cluster of genes with correlating expression. Using our method, existing

Table 1 The sizes of the different clusters on the performed clusterings

cluster	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6	cluster7	cluster8
2 cluster	530	221						
3 cluster	530	13	208					
4 cluster	530	13	79	129				
5 cluster	109	421	13	79	129			
6 cluster	109	421	2	11	79	129		
7 cluster	109	421	2	11	79	42	87	
8 cluster	109	421	2	11	21	58	42	87
Total TCs	**751**							

The table shows the number of TCs according to the number of clusters created in the expression experiment.

experimental results can be validated and novel insights into the expression of *Medicago truncatula* genes can be found. The collection of differentially expressed genes identified via MediPlEx provides a starting point for functional studies either in *Medicago truncatula* mutants or via RNA interference approaches.

Methods

MediPlEx extends the Sequence Analysis and Management System (SAMS), developed at Bielefeld University and combines it with the microarray expression datasets stored in TRUNCATULIX. An expression value for EST analyses is calculated my means of the logarithmic likelihood ratio introduced by Stekel *et al.* [17]. The software consists of four parts which are described in the following.

SAMS

SAMS stores the TC sequences (assembled ESTs), the EST composition and sequence data of all TCs of the *Medicago truncatula* GeneIndex 10.0, generated at the J. Craig Venter Institute and annotations of all TCs

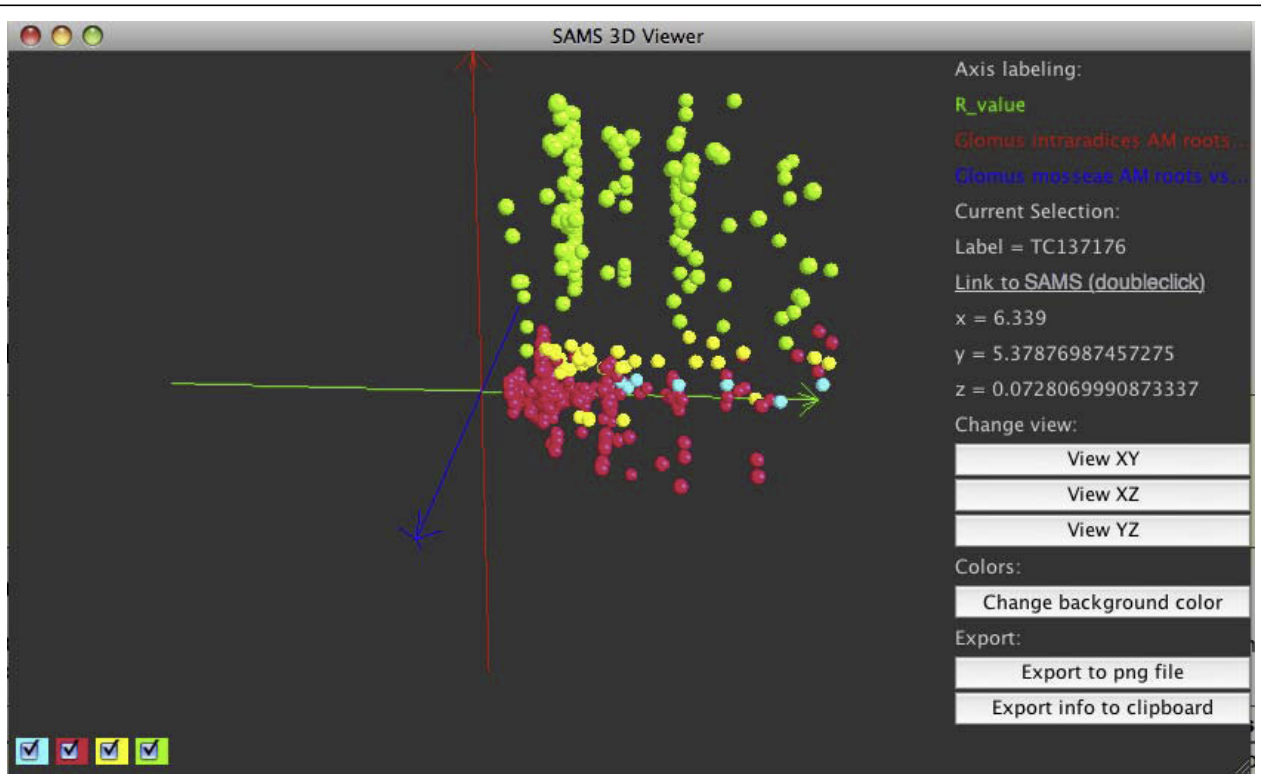

Figure 5 3D Visualization. A screenshot of the interactive 3D application created for the visualization of the MediPlEx results. Genes are painted as spheres in the coordinate system, different colors represent the associated cluster. Clicking on one of the spheres, additional information about the gene (label, expression values, html-link to SAMS) is shown. Each cluster can be activated and deactivated for hiding the corresponding genes in order to gain a better overview. The coordinate system is rotatable and zoomable, a snapshot of the 3D view can be saved by clicking the "Save as png file"-button.

created by an automatic annotation pipeline. The pipeline consists of several bioinformatics tools for gene annotation (BLAST against different sequence databases, Interpro, and HMMER) [44-46]. Using customized BLAST tools, the TCs were mapped to the reporters of the Mt16kOliPlus oligo-microarray chip, as well as to the Affymetrix Medicago GeneChip® reporters.

Logarithmic likelihood ratio

The logarithmic likelihood ratio developed by Stekel *et al.* [17] provides a statistical expression value for each TC for a unique combination of EST libraries. The logarithmic likelihood ratio is calculated as a ratio of ESTs assembled to a TC taking into account the total number of ESTs, the expression ratio of ESTs in each library and the size of the libraries. To calculate this R-value, the available libraries have to be divided into three groups: 'MUST contain ESTs', 'MUST NOT contain ESTs', and 'MAY contain ESTs'. The libraries and ESTs used for the calculation of the R-value are the ones marked as 'MUST contain ESTs' and 'MAY contain ESTs'. All TCs are then scanned for their composition of ESTs from these libraries. According to the ESTs and libraries, an R-value representing the expression is calculated for each TC.

For a more detailed description of the logarithmic likelihood ratio the reader is referred Stekel *et al.* .

TRUNCATULIX

The TRUNCATULIX data warehouse serves as a data source for the microarray expression data used by MediPlEx. It was created in 2008 to allow fast and effective expression search in freely available *Medicago truncatula* sequence and expression data. The data warehouse covers over 100.000 sequences, combined with annotation data and BLAST results. Additionally, the results of over 200 microarray hybridizations are stored in the warehouse. These have been linked to the sequences using a BLAST homology search of the reporter sequences against the gene sequences.

MediPlEx

MediPlEx integrates the results of different gene expression analysis methods to analyze them integratively to find new candidate genes and expression profiles. The user therefore first selects the EST libraries that should be used to filter the sequences. It is possible to select one of three states for each library: 'MUST contain ESTs', 'MUST NOT contain ESTs', and 'MAY contain ESTs'. 'MUST contain ESTs' means that the resulting TCs have to consist of at least one EST from these libraries. 'MUST NOT contain ESTs' denotes that the resulting TCs are not allowed to have an EST assembled from these libraries at all, and 'MAY contain ESTs' indicates that it does not matter if an EST is assembled from these libraries to a TC or not (see Figure 1). Different preselections for the libraries are available, some adopted from the DFCI website, while other are self-created. According to this selection, the TCs are scanned for their composition of ESTs from the libraries. For the TCs that match the query, the logarithmic likelihood is calculated (c.f. previous Section), to compute an expression value for the specific search.

In a second step (see Figure 2), the user selects the microarray experiments he wants to use for an expression analysis. For each of the TCs the results of the BLAST homology search against the two different microarray types Mt16kOliPlus and Affymetrix Medicago GeneChip® are fetched from the SAMS database. These reporters are used to collect the expression datasets of the selected experiments from TRUNCATULIX. The resulting expression values (Mt16kOliPlus arrays: mean of the significance test, Medicago GeneChips®: a1mean) are listed in a table (Figure 3). All fetched expression values, as well as the calculated logarithmic likelihood ratio are used for a hierarchical clustering performed using the statistical analysis software R [47]. The result of the clustering is presented as a dendrogram (Figure 4). The user can then select to create two to eight cluster according to his estimation and the cluster dendrogram. The clustered genes are subsequently visualized in an interactive 3D application (Figure 5). Therefore, the user has to select three of the expression datasets, one for each of the three axis of the coordinate system. The genes are depicted as spheres in the coordinate system, different colors represent the associated cluster. Each cluster can be activated and deactivated for hiding the corresponding genes in order to gain a better overview. The coordinate system is rotatable and zoomable, the genes can be clicked to show the expression values of the experiments. A snapshot of the 3D view can be stored locally by clicking the "Save as png file"-button. The expression information and the clustering results can be exported as csv files, containing the annotation details of the TCs. A link provides direct access to the gene sequence and annotations stored in SAMS.

Availability and requirements

Project name: MediPlEx

Project home page: http://www.cebitec.uni-bielefeld.de/mediplex

Operating system(s): Platform independent

Programming language: Perl, R

Additional material

Additional file 1: File containing the 751 genes and expression values. The .tsv file contains a table, separated using tab-stop as delimiter. The table stores the results of the expression analysis: The name of the found genes, the matching reporters of the two different layouts, the logarithmic likelihood ratio, as well as the expression values of the microarray expression experiments and the annotation of the genes.

Additional file 2: File containing the results of the clustering of the found genes. The .tsv file contains a table, separated using tab-stop as delimiter. The table stores the four clusters created. Each cluster contains the gene names and the annotation of the genes.

Additional file 3: File containing the GO categories of the 4 created clusters. The .xls file stores the GO categories of the genes in the four created clusters.

Acknowledgements
KH thanks the International NRW Graduate School in Bioinformatics and Genome Research for funding the project.

Author details
[1]Bioinformatics of Signaling Networks, Center for Biotechnology, Bielefeld University, Germany. [2]Unit IV - Plant Genomics, Institute for Plant Genetics, Leibniz Universität Hannover, Germany. [3]Computational Genomics, Center for Biotechnology, Bielefeld University, Germany. [4]Technical Faculty, Bielefeld University, Germany.

Authors' contributions
KH initiated the project, implemented the backend and the frontend, computed the annotations for the sequence data, and is the main author of the manuscript. HK coordinated most of the Mt16kOliPlus microarray experiments and helped with the biological interpretation of the results. LJS implemented the 3D viewer. AG supervised the project. All authors revised and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 22 September 2010 Accepted: 19 October 2010
Published: 19 October 2010

References
1. Barker D, Bianchi S, Blondon F, Dattée Y, Duc G, Essad S, Flament P, Gallusci P, Génier G, Guy P, Muel X, Tourneur J, Dénarié J, Huguet T: **Medicago truncatula, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis.** *Plant Molecular Biology Reporter* 1990, **8**:40-49.
2. Brewin NJ: **Development of the legume root nodule.** *Annu Rev Cell Biol* 1991, **7**:191-226.
3. Harrison MJ: **Molecular and cellular aspects of the arbuscular mycorrhizal symbiosis.** *Annu Rev Plant Physiol Plant Mol Biol* 1999, **50**:361-389.
4. Schüssler A, Schwarzott D, C W: **A new fungal phylum, the Glomeromycota: phylogeny and evolution.** *Mycol Res* 2001, **105(12)**:1413-1421.
5. Parniske M: **Arbuscular mycorrhiza: the mother of plant root endosymbioses.** *Nat Rev Microbiol* 2008, **6(10)**:763-75.
6. Oldroyd GED, Harrison MJ, Udvardi M: **Peace talks and trade deals. Keys to long-term harmony in legume-microbe symbioses.** *Plant Physiol* 2005, **137(4)**:1205-10.
7. Smith S, Read D: *Mycorrhizal Symbiosis* Academic Press, Second 1997, **2.**
8. **Pubmed.** [http://www.ncbi.nlm.nih.gov/pubmed/].
9. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J: **The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species.** *Nucleic Acids Res* 2001, **29**:159-164.
10. Küster H, Becker A, Firnhaber C, Hohnjec N, Manthey K, Perlick A, Bekel T, Dondrup M, Henckel K, Goesmann A, Meyer F, Wipf D, Requena N, Hildebrandt U, Hampp R, Nehls U, Krajinski F, Franken P, Pühler A: **Development of bioinformatic tools to support EST-sequencing, in silico- and microarray-based transcriptome profiling in mycorrhizal symbioses.** *Phytochemistry* 2007, **68**:19-32.
11. Hohnjec N, Vieweg M, Pühler A, Becker A, Küster H: **Overlaps in the transcriptional profiles of Medicago truncatula roots inoculated with two different Glomus fungi provide insights into the genetic program activated during arbuscular mycorrhiza.** *Plant Physiol* 2005, **137**:1283-1301.
12. Thompson R, Ratet P, Küster H: **Identification of gene functions by applying TILLING and insertional mutagenesis strategies on microarray-based expression data.** *Grain Legumes* 2005, **41**:20-22.
13. Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T, Moreau S, Niebel A, Frickey T, Weiller G, He J, Dai X, Zhao PX, Tang Y, Udvardi MK: **A gene expression atlas of the model legume Medicago truncatula.** *Plant J* 2008, **55(3)**:504-13.
14. Cannon SB, May GD, Jackson SA: **Three sequenced legume genomes and many crop species: rich opportunities for translational genomics.** *Plant Physiol* 2009, **151(3)**:970-7.
15. Young ND, Udvardi M: **Translating Medicago truncatula genomics to crop legumes.** *Curr Opin Plant Biol* 2009, **12(2)**:193-201.
16. **Dana-Farber Cancer Institute.** [http://www.dana-farber.org].
17. Stekel DJ, Git Y, Falciani F: **The comparison of gene expression from multiple cDNA libraries.** *Genome Res* 2000, **10(12)**:2055-2061.
18. Journet EP, van Tuinen D, Gouzy J, Crespeau H, Carreau V, Farmer MJ, Niebel A, Schiex T, Jaillon O, Chatagnier O, Godiard L, Micheli F, Kahn D, Gianinazzi-Pearson V, Gamas P: **Exploring root symbiotic programs in the model legume Medicago truncatula using EST analysis.** *Nucleic Acids Res* 2002, **30(24)**:5579-92.
19. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207-10.
20. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R: **NCBI GEO: mining tens of millions of expression profiles-database and tools update.** *Nucleic Acids Res* 2007, , **35 Database:** D760-5.
21. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R, Rayner T, Sharma A, William E, Sarkans U, Brazma A: **ArrayExpress-a public database of microarray experiments and gene expression profiles.** *Nucleic Acids Res* 2007, **35**:747-50.
22. Chen J, Zhao P, Massaro D, Clerch L, Almon R, DuBois D, Jusko W, Hoffman E: **The PEPR GeneChip data warehouse, and implementation of a dynamic time series query tool (SGQT) with graphical interface.** *Nucleic Acids Res* 2004, **1(32)**:578-81.
23. Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese J, Dwight S, Kaloper M, Weng S, Jin H, Ball C, Eisen M, Spellman P, Brown P, Botstein D, Cherry J: **The Stanford Microarray Database.** *Nucleic Acids Res* 2001, **1, 29(1)**:152-5.
24. Wise RP, Caldo RA, Hong L, Shen L, Cannon E, Dickerson JA: **BarleyBase/PLEXdb.** *Methods Mol Biol* 2007, **406**:347-363.
25. Henckel K, Runte K, Bekel T, Dondrup M, Jakobi T, Küster H, Goesmann A: **TRUNCATULIX - a data warehouse for the legume community.** *BMC Plant Biol* 2009, **9**:19.
26. Cheung F, Haas B, Goldberg S, May G, Xiao Y, Town C: **Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology.** *BMC Genomics* 2006, **7(272)**.
27. Town C: **Annotating the genome of Medicago truncatula.** *Current Opinion in Plant Biology* 2006, **9(2)**:122-127.
28. Bekel T, Henckel K, Küster H, Meyer F, Mittard Runte V, Neuweger H, Paarmann D, Rupp O, Zakrzewski M, Pühler A, Stoye J, Goesmann A: **The Sequence Analysis and Management System - SAMS-2.0: data management and sequence analysis adapted to changing requirements from traditional sanger sequencing to ultrafast sequencing technologies.** *J Biotechnol* 2009, **140(1-2)**:3-12.
29. **Java.** [http://java.sun.com/].
30. Hohnjec N, Henckel K, Bekel T, Gouzy J, Dondrup M, Goesmann A, Küster H: **Transcriptional snapshots provide insights into the molecular basis of**

arbuscular mycorrhiza in the model legume Medicago truncatula. *Functional Plant Biology* 2006, **33(8)**:737-748.
31. Javot H, Penmetsa RV, Terzaghi N, Cook DR, Harrison MJ: **A Medicago truncatula phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis.** *Proc Natl Acad Sci USA* 2007, **104(5)**:1720-5.
32. Doll J, Hause B, Demchenko K, Pawlowski K, Krajinski F: **A member of the germin-like protein family is a highly conserved mycorrhiza-specific induced gene.** *Plant Cell Physiol* 2003, **44(11)**:1208-14.
33. Wulf A, Manthey K, Doll J, Perlick AM, Linke B, Bekel T, Meyer F, Franken P, Küster H, Krajinski F: **Transcriptional changes in response to arbuscular mycorrhiza development in the model plant Medicago truncatula.** *Mol Plant Microbe Interact* 2003, **16(4)**:306-14.
34. Frenzel A, Manthey K, Perlick AM, Meyer F, Pühler A, Kuster H, Krajinski F: **Combined transcriptome profiling reveals a novel family of arbuscular mycorrhizal-specific Medicago truncatula lectin genes.** *Mol Plant Microbe Interact* 2005, **18(8)**:771-82.
35. Liu J, Blaylock LA, Endre G, Cho J, Town CD, VandenBosch KA, Harrison MJ: **Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis.** *Plant Cell* 2003, **15(9)**:2106-23.
36. Grunwald U, Nyamsuren O, Tamasloukht M, Lapopin L, Becker A, Mann P, Gianinazzi-Pearson V, Krajinski F, Franken P: **Identification of mycorrhiza-regulated genes with arbuscule development-related expression profile.** *Plant Mol Biol* 2004, **55(4)**:553-66.
37. Vencio RZN, Varuzza L, de B Pereira CA, Brentani H, Shmulevich I: **Simcluster: clustering enumeration gene expression data on the simplex space.** *BMC Bioinformatics* 2007, **8**:246.
38. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270(5235)**:484-487.
39. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays.** *Nat Biotechnol* 2000, **18(6)**:630-634.
40. Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, Matsubara K: **Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression.** *Nat Genet* 1992, **2(3)**:173-179.
41. Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, Delaney A, Griffith M, Hickenbotham M, Magrini V, Mardis ER, Sadar MD, Siddiqui AS, Marra MA, Jones SJM: **Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach.** *BMC Genomics* 2006, **7**:246.
42. Aitchison J: *The Statistical Annalysis of Compositional Data. Monographs on Statistics and Applied Probability* London: Chapman and Hall 1986.
43. Aitchison J: **Simplicial inference. Algebraic methods in statistics and probability. AMS special session on algebraic methods in statistics.** *American Mathematical Society. Contemp. Math* 2001, **287**:1-22.
44. Altschul S, Gish W, Miller W, Myers E, Lipman D: **Basic Local Alignment Search Tool.** *J Mol Biol* 1990, **215**:402-410.
45. Mulder N, Apweiler R: **InterPro and InterProScan: tools for protein sequence classification and comparison.** *Methods Mol Biol* 2007, **396**:59-70.
46. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14(9)**:755-63.
47. R Development Core Team: *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria 2008 [http://www.R-project.org], [ISBN 3-900051-07-0.

doi:10.1186/1756-0500-3-262
Cite this article as: Henckel *et al.*: MediPlEx - a tool to combine in silico & experimental gene expression profiles of the model legume Medicago truncatula. *BMC Research Notes* 2010 **3**:262.