The Affymetrix *Medicago* GeneChip® array is applicable for transcript analysis of alfalfa (*Medicago sativa*)

Mesfin Tesfaye, Kevin A. T. Silverstein, Bruna Bucciarelli, Deborah A. Samac and Carroll P. Vance

A Department of Plant Pathology, University of Minnesota, Saint Paul, MN 55108, USA.
B Department of Plant Biology, University of Minnesota, Saint Paul, MN 55108, USA.
C Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108, USA.
D USDA-ARS Plant Science Research Unit, Saint Paul, MN 55108, USA.
E Corresponding author. Email: vance004@umn.edu

This paper originates from a presentation at the Third International Conference on Legume Genomics and Genetics, Brisbane, Queensland, Australia, April 2006.

**Abstract.** The recently released Affymetrix GeneChip® *Medicago* Genome Array contains approximately 52,700 probe sets representing genes in both the model legume *Medicago truncatula* Gaertn. and the closely related crop species *Medicago sativa* L. (alfalfa). We evaluated the utility of the *Medicago* GeneChip® for monitoring genome-wide expression of *M. truncatula* and alfalfa seedlings grown to the first trifoliate leaf stage. We found that approximately 40–54% of the *Medicago* probes were detected in leaf or root samples of alfalfa or *M. truncatula*. Approximately 45–59% of the detected *Medicago* probes were called ‘present’ in all replicate GeneChips of *Medicago* species, indicating a considerable overlap in the number and type of *Medicago* probes detected between root and leaf organs. Nevertheless, gene expression differences between roots and leaf organs accounted for approximately 17% of the total variation, regardless of the *Medicago* species from which the samples were harvested. The result shows that the *Medicago* GeneChip® is applicable for transcript analysis for both alfalfa and *M. truncatula*.

**Keywords:** GeneChip, gene expression, medic, *Medicago*, transcript analysis.

**Introduction**

The plant genus *Medicago* contains approximately 60 species including perennial crops, such as alfalfa (*Medicago sativa*), and the annual species *M. truncatula*. Alfalfa is an allogamous (out-crossing) and an autotetraploid species (2n = 4 × 32) whose cultivars are heterogeneous synthetic populations (Zhu et al. 2005). These traits make alfalfa a genetically complex species, recalcitrant to direct genomics studies. By comparison, *M. truncatula* is an autogamous (self-fertilising), diploid species (2n = 2 × 16) with a small genome approximately 470 Mbp. Recent comparative genetic and physical mapping studies have shown a high degree of conservation of gene content and synteny between alfalfa and *M. truncatula* (Choi et al. 2004; Zhu et al. 2005). Because of the ease of genetic manipulation and small genome size, *M. truncatula* has become a model species for genomic studies of the Fabaceae. In recent years, international collaborations have led to the development of genomic tools for *M. truncatula* including a large collection (over 226,000) of expressed sequence tags (ESTs) released into the public domain (http://www.tigr.org/tdb/mtgi; verified 23 May 2006) and an ongoing *Medicago* genome sequencing effort (VandenBosch and Stacey 2003 and references therein; Young et al. 2005; http://www.medicago.org; verified 23 May 2006). cDNA-based arrays and long oligonucleotide spotted microarrays are also available for the parallel assessment of thousands of genes in a single *Medicago* microarray experiment (Manthey et al. 2004; Hohnjec et al. 2005; Lohar et al. 2006). In addition to spotted arrays, a *Medicago* GeneChip® containing ~10,000 *Medicago* probes was used successfully for gene expression analysis during bacterial symbiosis (Barnett et al. 2004;
In September 2005, a new Medicago GeneChip® was released by Affymetrix for use in whole-genome transcript profiling that includes over 52,000 Medicago probe sets designed from 32,167 M. truncatula ESTs, 18,733 gene predictions from M. truncatula genome sequences and 18,966 cDNAs from alfalfa (Affymetrix). The Affymetrix Medicago GeneChip® also contains 8,305 gene predictions from the genome of the microsymbiont Sinorhizobium meliloti.

Expression profiling when combined with sequence similarity analysis can assist in the identification of candidate genes for particular traits worthy of further investigation and hypothesis building. Microarrays can also be used for the parallel analysis of gene expression in plant species that have limited genomic information available, thereby elucidating commonalities and differences highlighted by the cross-species comparisons (Horvath et al. 2003; Close et al. 2004; Hammond et al. 2006). Using Arabidopsis thaliana L. GeneChips, cross-species microarray experiments have been used successfully to identify genes from related plant species involved in several complex traits (Becher et al. 2004; Weber et al. 2004). Recently, a barley chip microarray analysis applied to wheat–barley chromosome addition lines facilitated a large-scale physical mapping of barley genes (Cho et al. 2006).

Alfalfa is grown as a forage crop on 32 million hectares throughout the world. Non-traditional uses of alfalfa include phytoremediation of contaminated soils (Wang et al. 2005), production of recombinant enzymes and feedstocks for industrial applications (Sarunil et al. 2002; Tesfaye et al. 2005) and for human consumption as a source of flavonoids in health food products, in sprouts and juice. Alfalfa is also being developed as a biomass feedstock for energy production. While alfalfa is an extremely important crop worldwide, its genetic complexity has hindered progress in understanding the genes controlling important agronomic traits. The high degree of DNA sequence identity and similarity analysis can assist in the identification of candidate genes involved in nitrogen assimilation between alfalfa and other leguminous plants, we performed pair-wise comparisons of nucleotide sequence identities of five nitrogen assimilation genes between M. truncatula and other leguminous plants including alfalfa (Table 1). We observed 93% or more DNA sequence identity between protein coding regions of selected homologous genes in alfalfa and M. truncatula (Table 1). Percent nucleotide identity between M. sativa and M. truncatula of five selected genes involved in nitrogen assimilation was 97.5%

### Table 1. Percent nucleotide identity of genes involved in nitrogen assimilation between Medicago truncatula and related legumes

<table>
<thead>
<tr>
<th></th>
<th>Medicago truncatula</th>
<th>Medicago sativa</th>
<th>Pisum sativum</th>
<th>Lotus japonicus</th>
<th>Glycine max</th>
<th>Phaseolus vulgaris</th>
<th>Lupinus albus</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>97.5</td>
<td>97.9</td>
<td>93.4</td>
<td>99.8</td>
<td>98.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPYC</td>
<td>90.1</td>
<td>91.9</td>
<td>92.1</td>
<td>93.0</td>
<td>89.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Our laboratory has cloned and characterised several nitrogen assimilating genes from alfalfa nodules. Using homologous gene sequences in the public domain for M. truncatula and other leguminous plants, we performed results

### Materials and methods

Medicago truncatula Gaertn. and alfalfa (Medicago sativa L.) seedlings were grown to the first trifoliate leaf stage and total RNA was extracted from the first trifoliate leaf and whole root using the RNeasy Plant RNA mini kit (Qiagen, Valencia, CA). Tissues were harvested from three experimental replicates with up to 30 plants in each replicate. Extracted RNA was treated with DNase I enzyme to remove any contaminating DNA. Total RNA (10 μg) was used to produce biotin-labelled cRNA with Affymetrix kits and following suggested procedures for eukaryotic reactions (Affymetrix, Santa Clara, CA). Biotin-labelled cRNA (15 μg) fragmented as suggested by Affymetrix was hybridised to the GeneChip® Medicago Genome Array. The integrity and quality of total RNA, labelled and fragmented biotin-labelled cRNA was verified with the Agilent 2100 Bioanalyzer in RNA 6000 Nano LabChip (Agilent Technologies, Palo Alto, CA). GeneChips were hybridised, washed, stained and scanned by the University of Minnesota Affymetrix core facility.

Probe sets in the Medicago GeneChip® array are composed of 11 pairs of 25-mer oligonucleotides, with each pair consisting of a perfect match (PM) oligonucleotide and a mismatch (MM) control containing a single nucleotide substitution at the thirteenth base position (Affymetrix). The purpose of MM controls is to help distinguish background non-specific hybridisation from true low-level expression of genes (Affymetrix). For each probe set, we calculated both its detection call (‘present’ v ‘absent’) and its ‘expression value’, i.e. a number that reflects the concentration of the probe set’s target RNA species. For detection calls, we used the default MAS 5.0 parameters except changing αc = 0.05 as suggested for 11 probe pairs per probe set by Liu et al. (2002). Microarray data analysis was also conducted using GeneData Expressionist® Pro version 1.0 (GeneData, San Francisco, CA). Signal intensity values were normalised by global median scaling. In all of the data analyses, signals corresponding to Sinorhizobium probes were excluded.

### Results

Our laboratory has cloned and characterised several nitrogen assimilating genes from alfalfa nodules. Using homologous gene sequences in the public domain for M. truncatula and other leguminous plants, we performed results

### Table 1. Percent nucleotide identity of genes involved in nitrogen assimilation between Medicago truncatula and related legumes

<table>
<thead>
<tr>
<th></th>
<th>Medicago truncatula</th>
<th>Medicago sativa</th>
<th>Pisum sativum</th>
<th>Lotus japonicus</th>
<th>Glycine max</th>
<th>Phaseolus vulgaris</th>
<th>Lupinus albus</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>97.5</td>
<td>97.9</td>
<td>93.4</td>
<td>99.8</td>
<td>98.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPYC</td>
<td>90.1</td>
<td>91.9</td>
<td>92.1</td>
<td>93.0</td>
<td>89.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The application of the Affymetrix Medicago GeneChip® array

One distinct feature of the Medicago GeneChip® over other Medicago microarray platforms is the presence of probes for over 1800 cDNAs from alfalfa. Approximately 75 and 77% of the alfalfa probe sets included in the GeneChip® were detected in M. truncatula first trifoliate leaf and root samples, respectively (Fig. 3), whereas approximately 86 and 85% of the alfalfa probe sets included in the GeneChip® were called ‘present’ in alfalfa first trifoliate and root samples, respectively (Fig. 2). In all, 1091 (61%) alfalfa probe sets were called ‘present’ for leaf and root samples of both Medicago spp.

It is anticipated that transcriptional changes will accompany morphological changes during the course of new organ and tissue development in plants. The MAS 5.0 expression data containing presence and absence calls was used to determine the conservation of detected probe sets or to identify differentially regulated probe sets among first trifoliate and root tissues of Medicago spp. We found that approximately 15,000 probe sets were called ‘present’ in leaf tissues of both alfalfa and M. truncatula (Fig. 3A). In comparison, approximately 17,780 probe sets were called ‘present’ in root tissue of both alfalfa and M. truncatula (Fig. 3B). Most (12,770 probe sets) of the probe sets were called ‘present’ consistently in all replicates of both leaf and root samples (Fig. 3C). Despite the large overlap in gene expression between roots and leaves, unique gene expression patterns were also found in leaves and roots of both species. Of the 612 probes detected only in leaf tissue, 196 probes were unique to M. truncatula while 50 were unique to alfalfa leaf samples (Fig. 3). It was also found that 941 and 73 probe sets
were unique to *Medicago truncatula* and alfalfa roots, respectively (Fig. 3). Of the 50 probes unique to the alfalfa trifoliate leaf, 18 were derived from alfalfa sequences. Similarly, of the 73 unique probes in alfalfa roots, 10 were derived from alfalfa sequences, suggesting that these may be species-specific transcripts.

We utilised principal component analysis to construct a graphical interpretation of gene expression data, shown in Fig. 4. The first two principal components explained approximately 50% of the total gene expression variation. Along the second principal component axis, which explained 17% of the total gene expression variation, gene expression of roots was distinctly separated from first trifoliate leaf regardless of the *Medicago* species from which the samples were harvested.

**Fig. 3.** Comparisons of detected *Medicago* GeneChip® probes between alfalfa and *M. truncatula* (A) leaf, (B) root as well as (C) global comparison between leaf and root. Numbers in parentheses indicate probes unique to an organ. Abbreviations are: TL, *M. truncatula* leaf; TR, *M. truncatula* root; SL, *M. sativa* leaf; SR, *M. sativa* root.

**Fig. 4.** Principal component analysis of gene expression data from first trifoliate and root samples of *Medicago* species. The first and second principal components explained 49% and 17% of the total gene expression variation. The percentages show the variation explained by each principal component.

### Discussion

The Affymetrix GeneChip® *Medicago* genome array is a very recent, largely untested, addition to the genomics tools available for *Medicago* researchers. In this report, we have provided evidence that the GeneChip® is an effective platform for screening the transcriptomes of *M. truncatula* or alfalfa by demonstrating that (a) selected genes between alfalfa and *M. truncatula* showed between 93.4–99.8% nucleotide identity, (b) the total numbers and types of probe sets detected in the *Medicago* GeneChip® were comparable between alfalfa and *M. truncatula*, and (c) considerable gene expression differences were observed between root and leaf tissues, regardless of the *Medicago* species from which the samples were derived. Results also demonstrate that cross-species comparison of gene expression is possible with the *Medicago* GeneChip®.

The GeneChip® provides sensitive and reproducible (R²≥0.92) detection of the transcriptomes of both *Medicago truncatula* and alfalfa. Depending on the tissue surveyed, we detected 46–54% of the more than 52,000 plant probe sets for *M. truncatula*, and 41–44% in alfalfa. These percentages are comparable to, or higher than, those reported previously by other workers for other cross-species microarray hybridisations (Becher et al. 2004; Close et al. 2004; Weber et al. 2004). The relatively high numbers of ‘present’ calls (73–88% of the 1896 alfalfa probes) for leaf and root samples of both *Medicago* species are consistent with observations of very high nucleotide similarities between the coding regions of the *M. truncatula* and alfalfa genes. Close and associates (2004) examined gene expression of first leaf stage green seedlings of barley, wheat, oat, rice and sorghum with the 22K Barley1 GeneChip. On average, 45% (barley), 25% (wheat), 12% (oat), 9% (rice), 8% (sorghum) and 6% (maize) of the 22K Barley1 GeneChip probe array showed ‘present’ calls (Close et al. 2004). The percent of detected probe sets was increased to 58% when pathogen-challenged barley leaf tissue was used (Close et al. 2004). In another study, an *Arabidopsis* cDNA microarray was used in a broad cross-species comparison with wild oat (*Avena fatua* L.), poplar (*Populus deltoids* Bartram ex Marshall), and leafy spurge (*Euphorbia esula* L.) to identify genes involved in shoot growth (Horvath et al. 2003). Over 23, 34 and 47% of the 11,522 *Arabidopsis* probes produced signals that were greater than the threshold when hybridised with wild oat, poplar and leafy spurge cRNA, respectively, (Horvath et al. 2003).

Despite a much closer agreement in the number and types of detected probes between *M. truncatula* and alfalfa, the number of expressed genes in roots was higher than in leaves. Similar observations were made in *Arabidopsis* by Schmid et al. (2005). Gene expression profiles of roots and leaves in *Arabidopsis* were clearly distinct, with gene expression in root tissues showing generally higher transcript
abundance than in leaves (Schmid et al. 2005). Consistent with our observation of putative organ-specific or species-specific probe sets in alfalfa or M. truncatula, Schmid et al. (2005) also reported specific marker genes for the major organs of Arabidopsis. Additional data analysis or evaluation by a more sensitive method such as quantitative PCR (qPCR) analysis is needed to confirm putative marker probe sets in Medicago species.

Cross-species microarray transcript profiling with the A. thaliana GeneChip was used successfully to identify candidate genes involved in zinc tolerance in Arabidopsis halleri (L.) O’Kane & Al-Shebaz, a naturally selected candidate genes involved in zinc tolerance in Arabidopsis halleri. The nucleotide level within coding regions. A. thaliana Medicago was used successfully to identify a nucleotide level within coding regions. A. thaliana Medicago analysis is needed to confirm putative marker probes sets in Medicago species. GeneChip was used successfully to identify a nucleotide level within coding regions. A. thaliana Medicago analysis is needed to confirm putative marker probes sets in Medicago species. Analysis is needed to confirm putative marker probes sets in Medicago species.

Becher M, Talke IN, Krall L, Krämer U (2004) Cross-species microarray transcript profiling with the A. thaliana GeneChip was active in untreated plant roots of A. thaliana and M. halleri, respectively. Following zinc treatment, 63.9 and 36.8% of the total probes on the chip were shown to be active in leaves of A. thaliana and A. halleri, respectively (Becher et al. 2004). Candidate gene family members identified by transcript profiling were subsequently shown to confer enhanced heavy metal tolerance to transformed yeast (Becher et al. 2004; Weber et al. 2004). Our results also suggest that genome-wide transcript profiling with the GeneChip® Medicago genome array will aid in the identification of biologically meaningful gene expression patterns in Medicago spp. and will facilitate gene functional discovery in the cultivated crop, M. sativa.

It is estimated that only 10% of the gene annotation in Arabidopsis genome is based on experimental data on gene or protein function (Brown et al. 2005). The majority of genes (62±2%) are annotated based on conserved domains or assigned to a gene family based on homology that clusters the genes into broad categories. Currently, the Medicago GeneChip® lacks functional annotation. Future Medicago genomics efforts should include such an undertaking to enhance the usability of gene expression data.

References


Manuscript received 29 March 2006, accepted 28 April 2006

http://www.publish.csiro.au/journals/fpb