Research article

High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families

Gyorgy Szittya[†1], Simon Moxon[†2], Dulce M Santos[3], Runchun Jing[1], Manuel PS Fevereiro[3], Vincent Moulton[2] and Tamas Dalmay*[1]

Address: [1]School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK, [2]School of Computing Science, University of East Anglia, Norwich, NR4 7TJ, UK and [3]Laboratory of Plant Cell Biotechnology, ITQB/IBET – Apt 127, 2781-901 Oeiras, Portugal

Email: Gyorgy Szittya - g.szittya@uea.ac.uk; Simon Moxon - s.moxon@uea.ac.uk; Dulce M Santos - dsantos@itqb.unl.pt; Runchun Jing - r.jing@uea.ac.uk; Manuel PS Fevereiro - psalema@itqb.unl.pt; Vincent Moulton - Vincent.moulton@cmp.uea.ac.uk; Tamas Dalmay* - t.dalmay@uea.ac.uk

* Corresponding author   †Equal contributors

Abstract

Background: High-throughput sequencing technology is capable to identify novel short RNAs in plant species. We used Solexa sequencing to find new microRNAs in one of the model legume species, barrel medic (*Medicago truncatula*).

Results: 3,948,871 reads were obtained from two separate short RNA libraries generated from total RNA extracted from *M. truncatula* leaves, representing 1,563,959 distinct sequences. 2,168,937 reads were mapped to the available *M. truncatula* genome corresponding to 619,175 distinct sequences. 174,504 reads representing 25 conserved miRNA families showed perfect matches to known miRNAs. We also identified 26 novel miRNA candidates that were potentially generated from 32 loci. Nine of these loci produced eight distinct sequences, for which the miRNA* sequences were also sequenced. These sequences were not described in other plant species and accumulation of these eight novel miRNAs was confirmed by Northern blot analysis. Potential target genes were predicted for most conserved and novel miRNAs.

Conclusion: Deep sequencing of short RNAs from *M. truncatula* leaves identified eight new miRNAs indicating that specific miRNAs exist in legume species.

Background

Gene expression is regulated at several layers in plants to ensure optimal temporal and spatial accumulation of proteins. One of the latest discovered regulatory layers involves short RNA (sRNA) molecules 21–24 nucleotides in length that act post-transcriptionally [1]. There are surprisingly many different sRNAs in plant cells indicating an extensive role for these molecules [2]. Plant sRNAs are produced from double stranded RNA (dsRNA) by one of the four Dicer-like proteins (DCL1-4). The different DCL proteins process dsRNAs generated by diverse pathways [1]. MicroRNAs (miRNAs) are produced from partially complementary dsRNA precursor molecules (pre-miRNA) [3]. Pre-miRNAs are originally single stranded RNAs with hairpin structures and recognized by DCL1 [4]. The other large class of plant sRNAs is small interfering RNAs (siRNAs). siRNAs are processed from dsRNAs usually generated by one of the RNA Dependent RNA Polymerases (RDRs). Trans-acting siRNA (ta-siRNAs) precursors are generated by RDR6 [5,6] and heterochromatin siRNA pre-

cursors are made by RDR2 [7]. Another group of siRNAs, the natural antisense siRNAs (nat-siRNAs), are processed from dsRNA produced by overlapping antisense mRNAs [8].

MiRNAs are the best characterized sRNAs in plants [9]. The primary transcript (pri-miRNA) is transcribed by RNA polymerase II and contains an imperfect hairpin structure. DCL1 trims this hairpin structure producing the pre-miRNA and then a second cleavage by DCL1 produces the miRNA/miRNA* duplex [10]. This molecule has a two nucleotide 3' overhang at each side of the duplex and contains a few mismatches [9]. One of the strands of the miRNA/miRNA* duplex is integrated into RISC (RNA induced silencing complex). This strand is called mature miRNA and the partially complementary miRNA* strand gets degraded, although in most cases the miRNA* strand also accumulates at a lower level [9]. RISC finds specific mRNAs because the incorporated mature miRNA can anneal to partially complementary target sites [3]. Target sites show near perfect matches to plant miRNA sequences and initially it was thought that all target mRNAs are cleaved by RISC [3]. Recently it was shown that the translation of plant mRNAs is also suppressed without a cleavage [11].

Most plant miRNA families have been identified by traditional Sanger sequencing method in model species with known genome sequences (Arabidopsis, rice and poplar) and most miRNAs are conserved across plant families [12]. However, some miRNAs are species/family specific and Allen et al. [13] suggested that these "young" miRNAs have evolved recently, in contrary to the conserved miRNAs ("old" miRNAs). Since non-conserved miRNAs are often accumulated at a lower level than conserved miRNAs, traditional small-scale sequencing primarily reveals conserved miRNAs. Establishment of high-throughput technologies has allowed the identification of several non-conserved or lowly expressed miRNAs through deep sequencing, e.g. in Arabidopsis, wheat and tomato [14-17]. Here we describe the deep sequencing of short RNAs extracted from *M. truncatula* leaves and the experimental validation of eight novel miRNAs.

Results
Deep sequencing of **M. truncatula** *short RNAs*
Two separate cDNA libraries of short RNAs were generated from *Medicago truncatula* leaves and both libraries were sequenced by Solexa (Illumina). The first PCR product was quantified by nanodrop and 5 pM was loaded to Solexa, which yielded 5942 clusters and 872,048 sequence reads. The second sample was also quantified on polyacrylamide gel because the capacity of Solexa is higher than this. Since the nanodrop gave an approximately 50 times higher concentration than quantification

on a gel we concluded that the nanodrop overestimates the concentration of the PCR product. To obtain more sequences, we loaded four times of the amount that we would have loaded based only on the nanodrop reading and this yielded 23301 clusters and 3,076,823 sequence reads. It is worth mentioning that loading based on only quantification on gel would have led to overloading the Solexa and would have produced many but unreliable sequences.

The two sets of reads were combined and analysed together (Table 1) using the miRCat pipeline we have developed earlier [18]. The size distribution of sequence reads showed that the 24 nt class was the most abundant group of sRNAs followed by the 21 nt sequences and then the 23, 22 and 20 nt reads (Figure 1). The almost four million reads represented 1,563,959 distinct sequences suggesting that the library is still not saturated. Out of the four million reads 2,168,937 matched to the genome without any mismatches, representing 619,175 sequences.

Conserved miRNAs
First we looked for known miRNAs by comparing our library to known miRNAs from other plant species. 174,504 reads corresponding to 25 conserved miRNA families showed perfect matches to known miRNAs. We analysed the number of reads for conserved miRNAs and miR156, 159 and 166 were represented most frequently in the library (Table 2). Allowing one or two mismatches between sequences in our library and sequences in miR-Base increased the number of conserved miRNA families in *M. truncatula* to 31 (Table 2). Next we predicted target genes and putative targets were identified for 27 out of 31 conserved families (Additional File 1). No targets were found in *M. truncatula* for miR163, miR169, miR394 and miR894. We found homologs of known miRNA target genes for several conserved *M. truncatula* miRNAs, such as SBP for miR156, NAC for miR160, AGO1 for miR168, bZIP for miR165, GRAS for miR171, AP2 for miR172 and low affinity sulphur transporter for miR395. On the other hand we predicted many genes with unknown function and hypothetical genes for miRNA targeting (Additional File 1) and careful analysis of these potential targets will contribute to our understanding of the role of miRNAs in legume plants.

Novel miRNAs
The flanking regions of the 2,168,937 sequence reads matching the genome were subjected to secondary structure analysis. We used the strict criteria suggested by Jones-Rhoades et al. [9] to identify potential miRNA loci. 32 short RNA producing loci corresponding to 26 sequence reads could be folded into step-loop structures (Additional File 2). Nine of these loci produced eight dis-
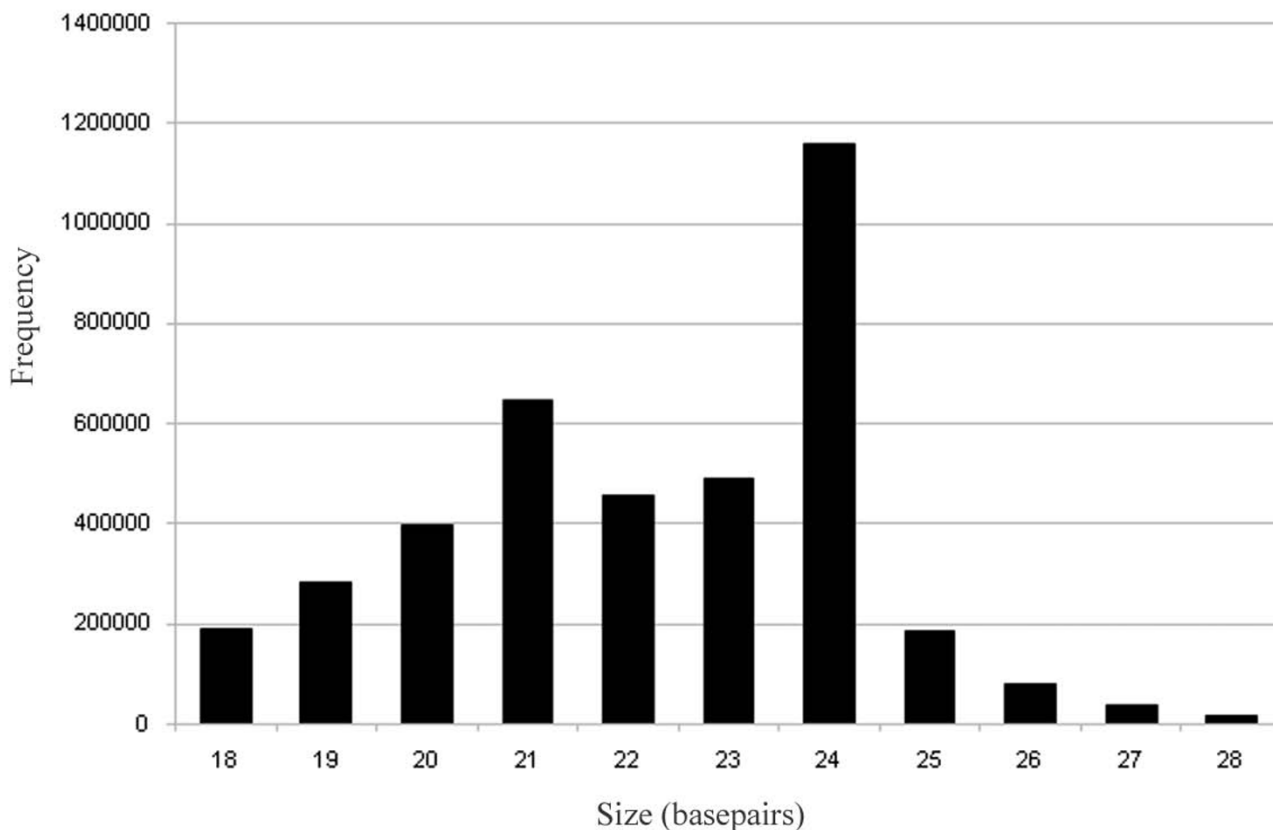
Figure 1
Size distribution of sequenced short RNAs.

tinct sequences, for which we also sequenced the miRNA* sequences (Table 3). These eight sequences were considered as novel legume specific miRNAs based on Rajagopalan et al. [15], who demonstrated that loci producing sequenced mature miRNAs and miRNA* sequences are miRNA loci. The other 23 loci producing 18 sRNAs were considered as putative new miRNA genes since at the moment there is not sufficient evidence to classify them as

miRNA genes. Accumulation of all eight new miRNAs and most miRNA*s were validated by Northern blot analysis (Figure 2).

We also predicted target genes for putative miRNAs (Additional file 3) and the new miRNAs with sequenced miRNA* (Additional file 4). Five out of the eight new miRNAs have potential targets in the available M. trunca-

Table 1: Distribution of short RNAs

	Number of reads	Number of distinct sequences
Total	3948871	1563959
Mapping to the genome without a mismatch	2168937	619175
Known miRNAs (exact match)	174504	25
New miRNA candidates With sequenced miRNA*	3370	8
New miRNA candidates With sequenced miRNA*	1272	18

Number of reads and distinct sequences are shown of different categories of sequences.

Table 2: Conserved miRNAs

miR	Exact	Shorter	Longer	1 mismatch	2 mismatches	Total
miR156	37580	5944	1112	14906	11423	**70965**
miR157	0	0	0	0	206	**206**
miR158	14	1	3	3	1	**22**
miR159	32003	30842	5825	1887	2454	**73011**
miR160	89	35	19	7	7	**157**
miR161	2	0	0	0	0	**2**
miR162	49	4	7	9	3	**72**
miR163	6	4	0	0	0	**10**
miR164	11815	794	210	639	213	**13671**
miR165	256	129	49	56	59	**549**
miR166	56402	12792	6349	1276	8253	**85072**
miR167	2729	623	420	114	133	**4019**
miR168	4404	429	173	338	159	**5503**
miR169	83	39	173	13	25	**333**
miR171	26	30	5	21	30	**112**
miR172	195	53	91	39	29	**407**
miR173	6	2	0	0	0	**8**
miR319	27322	944	3414	1075	1076	**33831**
miR390	729	154	32	55	43	**1013**
miR393	20	2	1	2	2	**27**
miR394	0	1	70	0	25	**96**
miR395	0	42	0	0	34	**76**
miR396	453	1279	158	41	466	**2397**
miR397	0	1	0	0	0	**1**
miR398	20	15	3	2	1	**41**
miR399	2	2	0	0	0	**4**
miR400	1	0	0	0	0	**1**
miR403	0	1	0	0	0	**1**

Table 2: Conserved miRNAs *(Continued)*

miR408	293	22	10	35	25	**385**
miR828	0	0	1	1	0	**2**
miR894	5	1	103	2	38	**149**
Total	**174504**	**54185**	**18228**	**20521**	**24705**	**292143**

Number of reads are shown for sequences that are similar to conserved miRNAs identified in other species. Figures are shown separately for sequences with exact matches and length, exact matches but shorter or longer than known miRNAs and for sequences with one or two mismatches to known miRNAs.

tula genome. It is interesting that two of these five new miRNAs potentially target disease resistance genes suggesting a role for miRNAs in the regulation of biotic stress response.

Discussion
Deep sequencing of M. truncatula *short RNAs*
Most conserved plant miRNAs were identified by traditional Sanger sequencing. Due to the relatively small number of sequences generated this method primarily discovered conserved miRNAs. Recent progress in sequencing technologies allows deep sequencing of large libraries of short RNAs [2,7,17,19-23]. The longest reads are obtained by the 454 technology, which currently gives reads of 250–300 base pairs (bp). However, this technology yields much less reads than other techniques (about 400 000 per sample) [17]. The Solexa platform (Illumina) generates shorter reads (up to 35 bp) but yields 1–3 million reads per sample [22]. The other very high-throughput technique, massively parallel sequencing (MPSS), gives more reads than Solexa but the reads are even shorter, only 17 bp [23]. miRNAs are only 21–23 nt sequences therefore even MPSS can identify them, although this technology generates shorter than 21 bp reads. Because of this it is not the ideal choice for miRNA discovery but it is useful for profiling conserved miRNAs. We decided to use the Solexa platform to test whether new miRNAs can be found in the model legume *M. truncatula*.

We found the quantification of PCR product problematic because the nanodrop overestimated the concentration and using quantitative markers on polyacrylamide gel seemed to underestimate the concentration. The difference between the two approaches was 50 fold. Loading 5
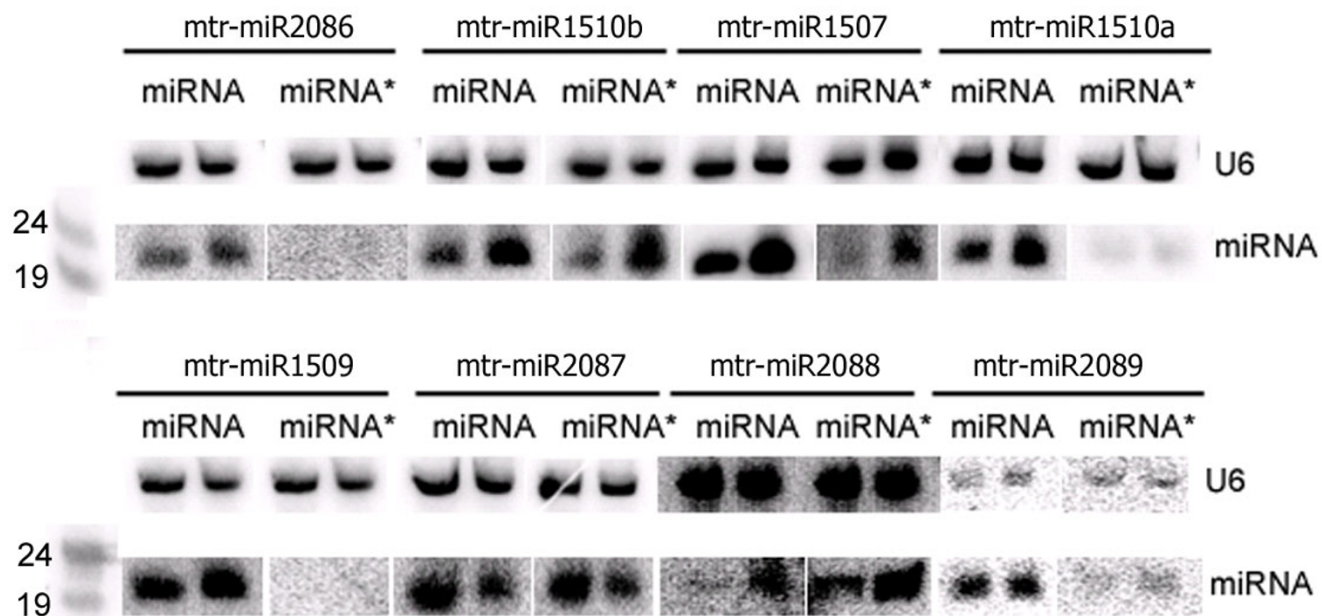

Figure 2
Expression of new miRNAs. Total RNA from stems and leaves of *M. truncatula* was analysed by Northern blot. The numbers correspond to ID numbers in Table 3. Size markers (19 and 24 nt RNA oligonucleotides) are shown on the left. The membranes were also hybridised with a probe specific to U6 [17].

Table 3: Potential new miRNA genes

sRNA	Chr	Start	No reads	Sequence	N. blot	miRNA*
miR2086	MtChr2	22878303	1164	GACATGAATGCAGAACTGGAA	+	Yes
C2	MtChr5	23546032	485	TGAGAATGTTAGATACGGAAC	-	No
miR1510b	MtChr5	17425793	721	ACATGGTCGGTATCCCTGGAA	+	Yes
miR1507	MtChr8	34001142	1170	CCTCGTTCCATACATCATCTAG	+	Yes
C5_1	MtChr4	10481611	414	GGTCCCCGGCAGCGGCGCCA	-	No
C5_2	MtChr4	10570962	414	GGTCCCCGGCAGCGGCGCCA	-	No
C5_3	MtChr4	10515784	414	GGTCCCCGGCAGCGGCGCCA	-	Yes
miR1510a	MtChr5	17429726	170	CGGAGGATTAGGTAAAACAAC	+	No
C7	MtChr2	13077141	50	AAAAGAATACTCATACATAACATT	+	No
C8	MtChr7	2887662	29	TGTAAGGGATGATGCAAATGAGAG	+	No
C9_1	MtChr2	6722889	28	AGGGTGCTTAGATATAGGGACGGA	+	No
C9_2	MtChr8	2492462	28	AGGGTGCTTAGATATAGGGACGGA	+	No
C9_3	MtChr8	10174192	28	AGGGTGCTTAGATATAGGGACGGA	+	No
miR1509	MtChr7	31515912	90	TTAATCTAGGAAAATACGGTG	+	Yes
C12	MtChr2	26140680	19	GAAAGGACTAAAACCAGGAACGGA	-	No
C13	MtChr2	25488088	24	TACGCAGGAGAGATGATGCCG	-	No
C14	MtChr5	28947435	44	CTTGGACTTCAAGAAACAAC	-	No
miR2087	MtChr2	9824368	35	GAAGTAAAGAACCGGCTGCAG	+	Yes
C16	MtChr1	29344468	38	TAACTGCCGAGGCGTGCTGA	-	No
C17	MtChr4	20677608	30	TGATGACGGAAGAAATCCAAA	-	No
C18	MtChr6	2189030	16	AGATGGAGATTAGACGGCTCA	-	No
C19	MtChr5	13285282	16	AGATTTAGGCGGCTTTGAACT	-	No
C20	MtChr6	8477357	22	TGGACTGAAGACAAAGAGAAA	-	No
C21_1	MtChr4	33246760	13	ATTTTTCAGCGGAAGTTAA	-	No
C21_2	MtChr4	27138541	13	ATTTTTCAGCGGAAGTTAA	-	No
miR2088_1	MtChr8	14579972	11	AGGCCTAGATTACATTGGAC	+	Yes
miR2088_2	MtChr8	14675564	11	AGGCCTAGATTACATTGGAC	+	Yes
C23	MtChr5	31612014	12	TGTTGGAACACGGCTTAAGGT	-	No

Table 3: Potential new miRNA genes *(Continued)*

C24	MtChr3	13221230	12	TAGGTTTGAGAAAATGGGCA	-	No
C25	MtChr8	24335508	11	TCTGTCAGTAGACTCAAT	-	No
C26	MtChr5	28244439	9	AAGGGCATATGTTAAGGAAC	-	No
miR2089	MtChr5	17442776	9	TTACCTATTCCACCAATTCCAT	+	Yes

The first column (sRNA) shows the ID number of the sequences. The second and third columns show the genomic location of the putative miRNAs (Chr: chromosome; Start: starting position on the chromosome). The number of reads are indicated in the fourth column and the sequences themselves are listed in the fifth column. The results of the Northern blot analysis are shown in the sixth column (+: detected; -: not detected). Whether the miRNA* sequence was sequenced or not is indicated in the last column.

pM sample according to the nanodrop result yielded relatively low number of sequences (872,048 reads), well below the capacity of Solexa. However, loading 50 times more would have given poor quality sequences due to overloading. We decided to load four times of the amount suggested by the nanodrop and this approach yielded sufficient number (more than 3 million) and good quality reads. Many short sequences did not map to the genome but it is not known what the explanation is for that. One possibility is that the rate of sequencing error is high and many short reads contains mistakes. The other possibility is that there are polymorphisms between the species we used and the one that was used for the genome sequencing. It is also possible that plants used in these experiments were infected symptomless pathogens and genomes of those organisms contaminated our libraries.

First we searched for conserved miRNAs in the combined, almost four million reads from the two separate sRNA libraries. MiRNAs showing perfect matches to 25 conserved families were found in our combined library, which is more than was identified using bioinformatics approaches [24-26]. Allowing up to two mismatches further increased this number to 31 families. This work experimentally verified all conserved miRNAs predicted by [[24,25] and [26]]. However, none of the predicted new miRNAs [26] were found and the eight new miRNAs we found were not predicted by any of the three studies [[24,25] and [26]] illustrating the benefit of the sequencing approach. Target prediction for conserved miRNAs found several genes showing homology to genes validated in other species. However, several predicted targets are new without known function indicating that more work is required to elucidate the role of miRNAs in legumes.

New miRNAs

Validation of non-conserved miRNAs can be achieved through several ways. One possibility is providing evidence of biogenesis characteristic of miRNAs. This involves the identification of stem-loop structure of the potential pre-miRNAs and sequencing of the miRNA* molecules [9]. In the absence of evidence for biogenesis, functional data can complement the predictions based on

structural analysis. MiRNA mediated cleavage is located in a specific site within the target site and this can be identified by 5'RACE analysis. We have found eight new non-conserved miRNAs supported by biogenesis data since the miRNA* strands were sequenced and detected by Northern blot analysis. We also found 18 other putative miRNAs based on structural data. However, we did not sequence the miRNA* strands for these and we do not have functional proof therefore at the moment these are only putative miRNAs.

The role of miRNAs in development and abiotic stress response is well documented. Several transcription factors involved in development are targeted by miRNAs [27]. Other miRNAs play a role in nutrient assimilation and responses to drought, cold and other abiotic stresses [1,28]. Reports of miRNAs involved in biotic stress response are less common. Navarro et al [29] showed that bacterial flagellin expression induced miR393 and F-box auxin receptor genes were regulated by miR393. Two of the new miRNAs are predicted to target disease resistance genes suggesting a more extensive role for miRNAs in biotic stress response. Targets of the other new miRNAs also include beta-glucan-binding protein, peptidyl-prolyl cis-trans isomerise and a hypothetical protein. The biological importance of the potential regulation of these genes by miRNAs needs further investigation.

Conclusion

High-throughput sequencing analysis of sRNAs from *M. truncatula* leaves identified eight new miRNAs, which were not found in other species. Sequencing sRNAs from specific tissues of legume plants by deep sequencing is expected to reveal more new miRNAs.

Methods

Cloning of M. truncatula sRNAs and Northern blot analysis

RNA was extracted using the miRVana kit (Ambion) from the aerial part (including stems and leaves) of nine week old *Medicago truncatula* Gaertn. cv. Jemalong seedlings. Plants were grown in pots with soil in a growth chamber having a photoperiod of 16/8 h day/night, a thermope-

riod of 25/18°C day/night and humidity of 40%. Small RNA fraction between 19–24 nt was isolated from 15% denaturing polyacrylamide gel and 15 μg was ligated to adaptors without de-phosphorylating and re-phosphorylating [30]. The short RNAs were converted to DNA by RT-PCR and the DNA was sequenced on a Solexa machine (Illumina). 15 μg of RNA extracted from *M. truncatula* leaves was analysed by Northern blot as described by Pall et al. [31].

Sequence analysis

The *Medicago truncatula* genome sequence (version 2.0) was downloaded from the Medicago Sequencing Resource website http://www.medicago.org/genome/downloads/Mt2/ and sequence reads were mapped to the genome using PatMaN [32]. miRNA candidates were generated by miRCat (http://srna-tools.cmp.uea.ac.uk/mircat/; [33]) using default parameters. DFCI Medicago Gene Index v 9.0 was used for the target predictions: ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Medicago_truncatula/MTGI.release_9.zip. Target searches were performed on the srna-tools website: http://srna-tools.cmp.uea.ac.uk/targets/[33].

The sequences can be found in GEO under GPL7704 platform (High-throughput sequencing of small RNAs in *Medicago truncatula*), series GSE13761, samples GSM346592 and GSM346593.

Authors' contributions

SG cloned the short RNAs and carried out most of the Northern blot analysis; SM analysed the sequences and was supervised by VM; DMS grew the plants and prepared the RNA; RJ carried out some of the Northern blot analysis; MPSF and TD designed and coordinated the study; TD analysed the results and wrote the manuscript.

Additional material

Additional file 1
Predicted targets of conserved M. truncatula *miRNAs. The data provided describes the predicted target genes of conserved miRNAs found in* M. truncatula.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-593-S1.xls]

Additional file 2
Predicted secondary structures of new and putative M. truncatula. *The data shows the predicted secondary structures of validated and putative new* M. truncatula *miRNAs.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-593-S2.pdf]

Additional file 3
Predicted targets of putative miRNAs. This table provides a list of predicted target genes of potential but not validated M. truncatula *miRNAs.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-593-S3.xls]

Additional file 4
Predicted targets of new validated M. truncatula *miRNAs. This table provides a list of predicted target genes of new* M. truncatula *miRNAs with sequenced miRNA*.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-593-S4.doc]

References
1. Phillips J, Dalmay T, Bartels D: **The role of small RNAs in abiotic stress.** *FEBS Letters* 2007, **581(19):**3592-7.
2. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ: **Elucidation of the small RNA component of the transcriptome.** *Science* 2005, **309:**1567-1569.
3. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function.** *Cell* 2004, **116:**281-297.
4. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP: **MicroRNAs in plants.** *Genes Dev* 2002, **16:**1616-1626.
5. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS: **SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNA in Arabidopsis.** *Genes Dev* 2004, **18:**2368-2379.
6. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, Hilbert JL, Bartel DP, Crete P: **Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs.** *Mol Cell* 2004, **16:**69-79.
7. Lu C, Kulkarni K, Souret FF, MuthuValliappan R, Tej SS, Poethig RS, Henderson IR, Jacobsen SE, Wang W, Green PJ, Meyers BC: **MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant.** *Genome Research* 2006, **16:**1276-1288.
8. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK: **Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis.** *Cell* 2005, **123:**1279-1291.
9. Jones-Rhoades MW, Bartel DP, Bartel B: **MicroRNAS and their regulatory roles in plants.** *Annu Rev Plant Biol* 2006, **57:**19-53.
10. Kurihara Y, Watanabe Y: **Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions.** *Proc Natl Acad Sci USA* 2004, **101:**12753-12758.
11. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O: **Widespread translational inhibition by plant miRNAs and siRNAs.** *Science* 2008, **320(5880):**1185-90.
12. Axtell MJ, Bartel DP: **Antiquity of microRNAs and their targets in land plants.** *Plant Cell* 2005, **17:**1658-1673.
13. Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC: **Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana.** *Nat Genet* 2004, **36:**1282-1290.
14. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC: **High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes.** *PLoS ONE* 2007, **2(2):**e219.

15. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP: **A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana.** *Genes Dev* 2006, **20:**3407-25.
16. Yao Y, Guo G, Ni Z, Sunkar R, Du J, Zhu JK, Sun Q: **Cloning and characterization of microRNAs from wheat (Triticum aestivum L.).** *Genome Biology* 2007, **8(6):**R96.
17. Moxon S, Jing R, Szittya G, Schwach F, Rusholme Pilcher RL, Moulton V, Dalmay T: **Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening.** *Genome Research* 2008 in press.
18. Moxon S, Schwach F, MacLean D, Dalmay T, Studholme DJ, Moulton V: **A toolkit for analysing large-scale plant small RNA datasets.** *Bioinformatics* 2008 in press.
19. Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE: **Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning.** *Nature Genetics* 2006, **38:**721-725.
20. Berezikov E, Thuemmler F, van Laake L, Kondova I, Bontrop R, Cuppen E, Plasterk R: **Diversity of microRNAs in human and chimpanzee brain.** *Nature Genetics* 2006, **38:**1375-1377.
21. Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP: **Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans.** *Cell* 2006, **127:**1193-1207.
22. Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML: **A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach.** *Genome Research* 2008, **18(6):**957-64.
23. Lu C, Kulkarni K, Souret FF, MuthuValliappan R, Tej SS, Poethig RS, Henderson IR, Jacobsen SE, Wang W, Green PJ: **MicroRNAs and other small RNAs enriched in the Arabidopsis RNAdependent RNA polymerase-2 mutant.** *Genome Research* 2006, **16(10):**1276-1288.
24. Sunkar R, Jagadeeswaran G: **In silico identification of conserved microRNAs in large number of diverse plant species.** *BMC Plant Biology* 2008, **8:**37-50.
25. Zhou ZS, Huang SQ, Yang ZM: **Bioinformatic identification and expression analysis of new microRNAs from Medicago truncatula.** *Biochemical and Biophysical Research Communications* 2008, **374:**538-542.
26. Wen J, Frickey T, Weiller GF: **Computational prediction of candidate miRNAs and their targets from Medicago truncatula non-protein-coding transcripts.** *In Silico Biol* 2008, **8(3-4):**291-306.
27. Kidner CA, Martienssen RA: **The developmental role of microRNA in plants.** *Curr Opin Plant Biol* 2005, **8(1):**38-44.
28. Sunkar R, Zhu JK: **Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis.** *Plant Cell* 2004, **16(8):**2001-19.
29. Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JD: **A plant miRNA contributes to antibacterial resistance by repressing auxin signaling.** *Science* 2006, **312(5772):**436-9.
30. Rathjen T, Pais H, Sweetman D, Moulton V, Munsterberg A, Dalmay T: **High throughput sequencing of microRNAs in chicken somites.** *BMC Genomics* 2008 in press.
31. Pall GS, Codony-Servat C, Byrne J, Ritchie L, Hamilton A: **Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot.** *Nucleic Acids Research* 2007, **35(8):**e60.
32. Prufer K, Stenzel U, Dannemann M, Green REE, Lachmann M, Kelso J: **PatMaN: rapid alignment of short sequences to large databases.** *Bioinformatics* 2008, **24(13):**1530-1.
33. Moxon S, Schwach F, Maclean D, Dalmay T, Studholme DJ, Moulton V: **A toolkit for analysing large-scale plant small RNA datasets.** *Bioinformatics* 2008, **24(19):**2252-3.