

Supplementary Information for

Development And Evolution Of Age-Dependent Defenses In Ant-Acacias

Aaron R. Leichty and R. Scott Poethig

R. Scott Poethig

Email: spoethig@sas.upenn.edu

This PDF file includes:

Supplementary Materials and Methods

Figs. S1 to S6

Tables S1 and S2

References for SI reference citations

Other supplementary materials for this manuscript include the following:

Dataset S1

Supplementary Materials and Methods

Confirmation of species identities

The identity of our samples of *V. collinsii* and *V. cornigera* was confirmed by Dr. Daniel Janzen (University of Pennsylvania), based on plant morphology. We also used the sequence of the *MATK* chloroplast gene to perform a phylogenetic analysis of species/accessions used in this study. The coding region of this gene was determined for two samples of each accession using Sanger sequencing of PCR products following previous methods (1), with an additional internal primer used for sequencing a problematic region (trnK_3_seq, Table S2). Sequences were trimmed to the coding region of *MATK* and aligned using MAFFT (2). A maximum likelihood tree was generated using RaxML v8.2 (3) with the GTRGAMMA model of rate heterogeneity using the rapid bootstrapping mode with 1000 searches (Figure S5). *Prosopis glandulosa* was used as an outgroup and trimmed from the final tree using the Ape package (4-5).

Genome sequencing and identification of MIR156/157 and SPL genes

DNA was extracted from leaves of a single *V. collinsii* Belize plant using the DNeasy Plant kit (Qiagen). A short insert library was prepared using the Illumina Truseq PCR-free kit. Briefly, genomic DNA was fragmented using a Biorupter (Diagenode) and size-selected for a mean insert size of 450 bp using 1.5% gels on a Pippin Prep (Sage Science), and concentrated by column purification (Thermo Fisher Scientific GeneJET PCR Purification kit) before beginning the Truseq protocol. The resulting library was sequenced at the University of Pennsylvania Next Generation Sequencing Core, on a single lane of a HiSeq 2500 using 250PE format.

Raw reads were merged using FLASH v1.2.11 with the default settings (6). Both merged and unmerged reads were used to estimate the genome size using the Perl script, `estimate_genome_size.pl` (https://github.com/josephryan/estimate_genome_size.pl). The resulting reads were *de novo* assembled using MaSuRCA (7). Super-reads were assembled using a kmer of 127 bp. The MaSuRCA-generated assembly script was modified to allow reads larger than 200 bp. Genome completeness was determined using BUSCOv3 (8-9).

To identify scaffolds containing *MIR156* and *MIR157*, the *V. collinsii* scaffolds were searched with BLASTn (10) for homology with a database of *MIR156* and *MIR157* hairpin sequences from 6 plant species. Data for the six species—*Arabidopsis thaliana*, *Glycine max*, *Malus domestica*, *Medicago truncatula*, *Populus trichocarpa*, *Vitis vinifera*—were obtained from miRBase v19 (12). The resulting sequences were checked for the existence of a stem-loop structure in a 200-300 bp region centered on the putative miRNA using RNAfold (13).

Scaffolds containing *SQUAMOSA PROTEIN LIKE (SPL)* genes were identified using tBLASTx with a database of all annotated *A. thaliana SPL* genes (10-11). These scaffolds were then annotated with MAKER (14), using SNAP (13) for *ab initio* prediction, and a sampling of mRNA and protein sequences from species within the Fabaceae for homology based prediction. SNAP was run using the precompiled HMM models for *A. thaliana*. A final round of manual annotation was done to merge overlapping contigs producing identical transcripts and contigs corresponding to different haplotypes, to produce a final set of non-redundant *SPL* transcripts. Naming of *SPL* genes was based on the top blast hit between the predicted protein sequence and the *Arabidopsis SPL* family.

qPCR analysis of smRNA and mRNA abundance

Individual alignments of miR156, miR157, miR159, and miR168 sequences downloaded from miRBase v19 (12) were used to design primers specific to each gene family (Table S2). We utilized these primers with the stem-loop RT primers previously described for qPCR of smRNAs in plants (16). The highly conserved microRNAs miR159 and miR168 were used as endogenous controls for smRNA qPCR, which were validated using the first (nodes 1-2) and last (nodes 9-12) time points. Total RNA was extracted using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich). cDNA was synthesized using Invitrogen SuperScript III following the methods of Varkonyi-Gasic et al. (16). For small RNA qPCR, Platinum Taq (Invitrogen) was used with the Roche universal hydrolysis probe #21, and a three step amplification protocol. For mRNA qPCR and semi-quantitative PCR of *SPL* and *MIR156/157* transcripts, cDNA was synthesized using Invitrogen SuperScript III and DNase digestion (Qiagen) following the manufacturer's protocol. For these samples, a SYBR green master mix (Biotools) with a two-step protocol was used with *ACT2* as an endogenous control. Relative measures of abundance were calculated using the $2^{-\Delta\Delta C_t}$ method (17), and for small RNA qPCR miR159 and miR168 threshold values were averaged (18). The miR156 and miR157 primers were tested for selective amplification using synthetic RNA oligos of the four major variants: miR156a-f, miR156g, miR157a-d, and miR157e (Integrated DNA Technologies) (Figure S6A). Rates of relative detection were calculated across three concentrations spanning the experimental amplification thresholds seen in our total RNA samples (Figure S6B) using reactions that were carried out as described above, but with the addition of total RNA from *E. coli* in place of plant total RNA. Analysis of miR156g and miR157e showed their temporal abundance patterns were consistent with their pri-miRNAs (Figure S2A & S6C) but these miRNAs were too lowly abundant to accurately measure across all tissues and time points.

For semi-quantitative PCR, single gels were run for each gene individually. All gels were image processed identically using automated white balance and equalize adjustments. In cases where a primer set failed to amplify a predicted transcript, at least one additional primer pair was tested to confirm lack of expression.

To validate the manual annotation of *SPL* genes, we cloned partial *SPL-like* transcripts from *V. collinsii* Belize using a pool of 9-12 mm stipule primordia from node

7. This RNA was used for 3' RACE using the SMARTer RACE 5'/3' kit (Clontech) with a degenerate primer designed from an alignment of the SBP-domain of all 10 miR156-targeted *SPL* genes in *A. thaliana* and blunt end cloning of the resulting multi-product PCR pool (Thermo Scientific CloneJET PCR Cloning Kit). Each *SPL-like* sequence was represented by at least 3 independent colonies. Validated genes are noted in Dataset S1.

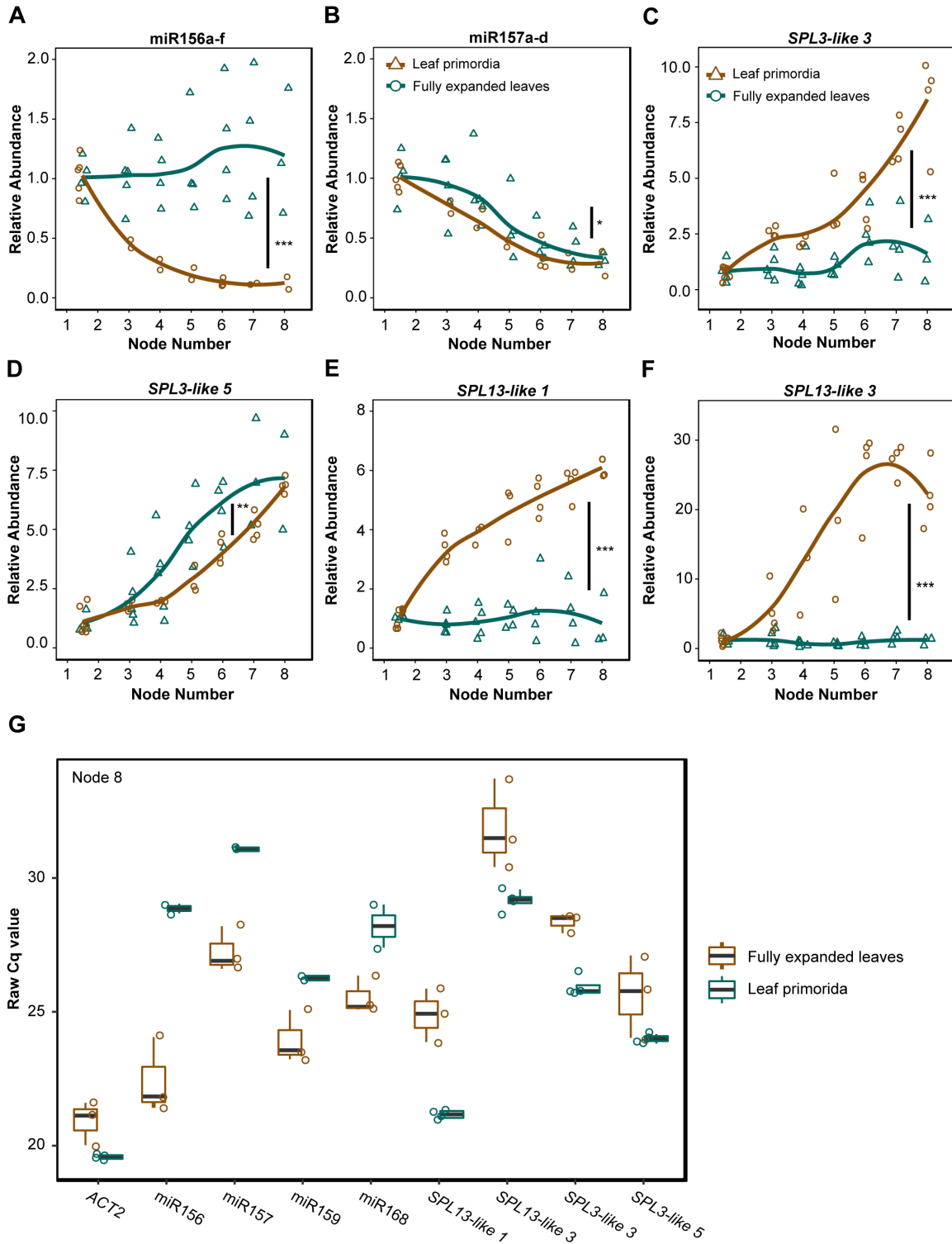


Figure S1. Comparison of miR156/157 and SPL abundance in fully expanded leaves and leaf primordia. (A-F) Relative abundance of miR156a-f (A), miR157a-d (B), *SPL3-*

like 3 (C), *SPL3-like 5 (D)*, *SPL13-like 1 (E)*, *SPL13-like 3 (F)* in fully expanded leaves and leaf primordia. Curves represent conditional means using a Loess smoother. Differences between treatments were tested by ANCOVA. Significance of these tests are indicated: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, ns = $p > 0.05$. **(G)** Raw threshold values from qPCR for fully expanded leaves and leaf primordia for mRNA/miRNA transcripts and their endogenous controls from node 8 samples.

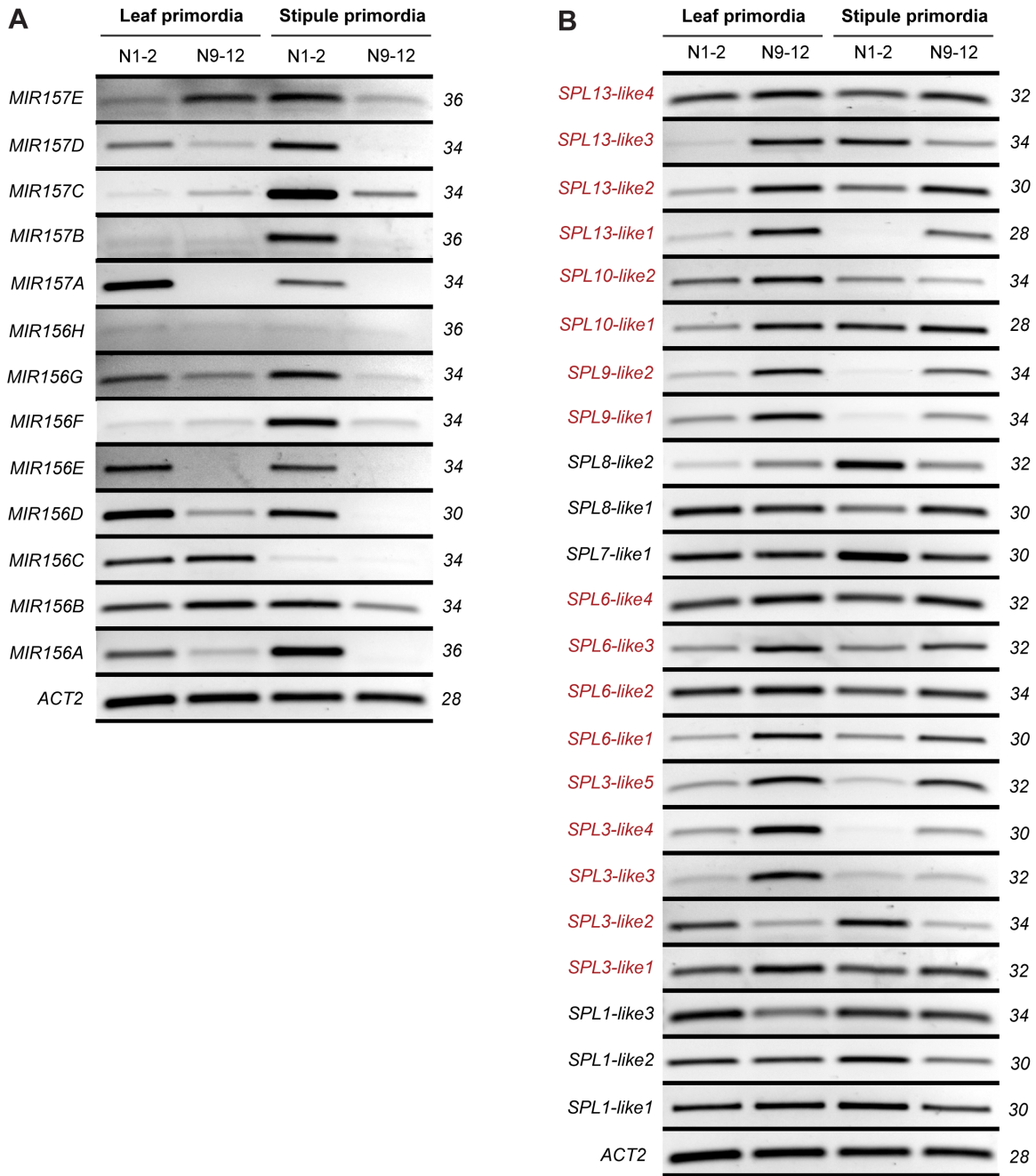


Figure S2. Semi-quantitative PCR of *MIR156/157* and *SPL* gene families in *V. collinsii* Belize. (A) Semi-quantitative PCR of *MIR156/157* genes. Leaf and stipule primordia were sampled at 1-3 mm in size for nodes 1 and 2 (columns 1 and 3) and nodes 9-12 (columns 2 and 4) and pools of 3-5 biological replicates were used for semi-quantitative PCR. Each gene was run on a single gel, allowing for direct comparisons of relative abundance between developmental stage and tissue type. Cycle numbers are on the right. (B) Semi-quantitative PCR of *SPL* genes. Red font represents those genes with

a predicted miR156/157 target site. *ACT2*, the loading control, is identical between *MIR156/157* and *SPL* gels.

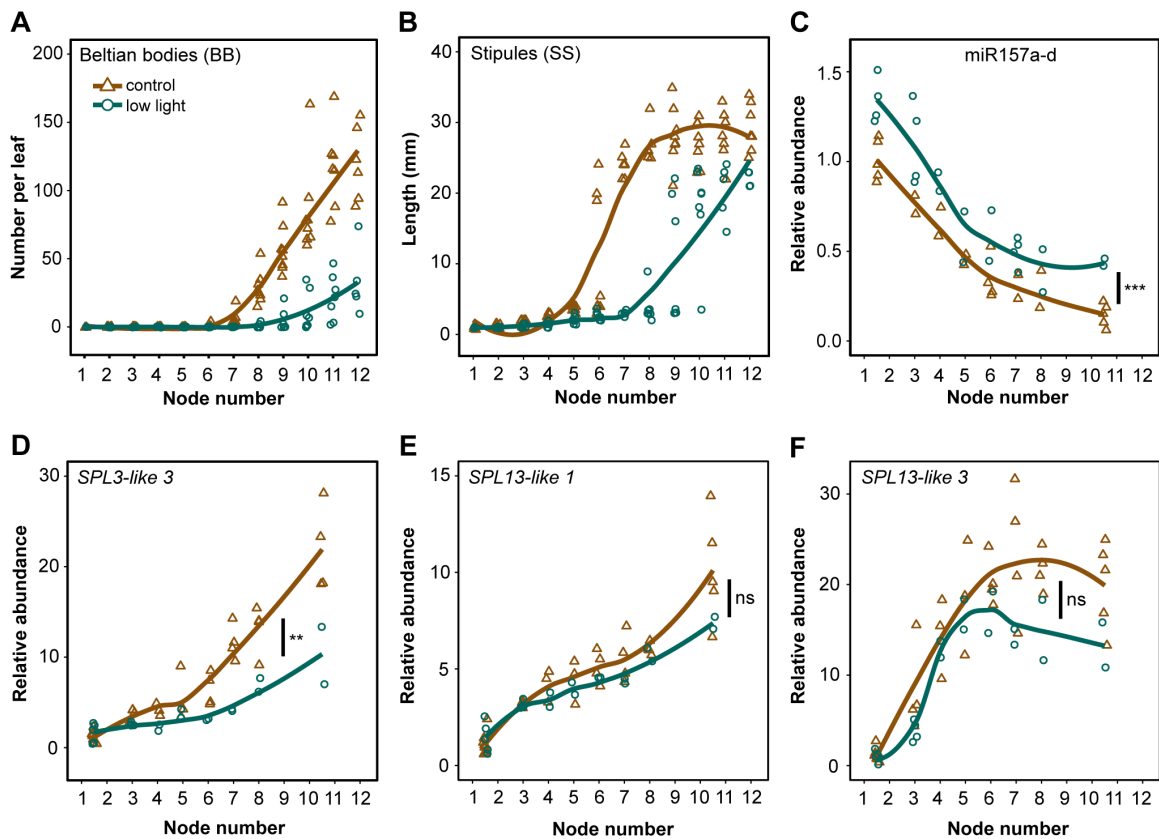


Figure S3. The effect of reduced light intensity on the development of the swollen thorn syndrome in *V. collinsii* Belize. (A-B) Beltian body number and stipule length as a function of nodal position under low light and full light conditions. Curves represent conditional means using a Loess smoother. (C) Relative abundance of miR157a-d in reduced light and control conditions. Plotting as in A. (D-F) Relative abundance of *SPL-like* transcripts in reduced light and control conditions. Plotting as in A. Differences between treatments were tested by ANCOVA. Significance of these tests are indicated: ** = $p < 0.01$, *** = $p < 0.001$, ns = $p > 0.05$.

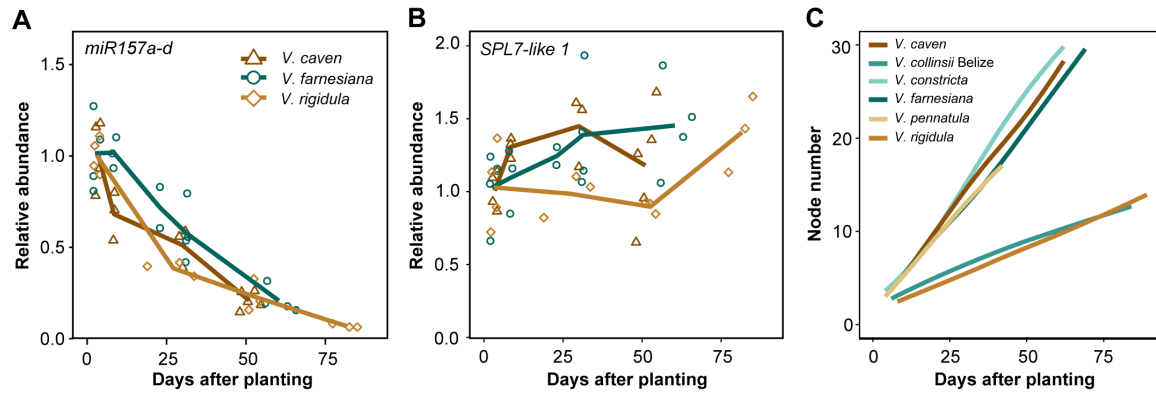


Figure S4. The expression pattern of miR157 and a non-targeted *SPL* gene in non-ant acacias. Relative abundance of miR157a-d (**A**) and *SPL7-like 1* (**B**) as a function of plant age. Lines are plotted through the means of samples grouped by similar age. (**C**) Rates of leaf initiation in ant and non-ant acacias. Curves represent conditional means using a Loess smoother.

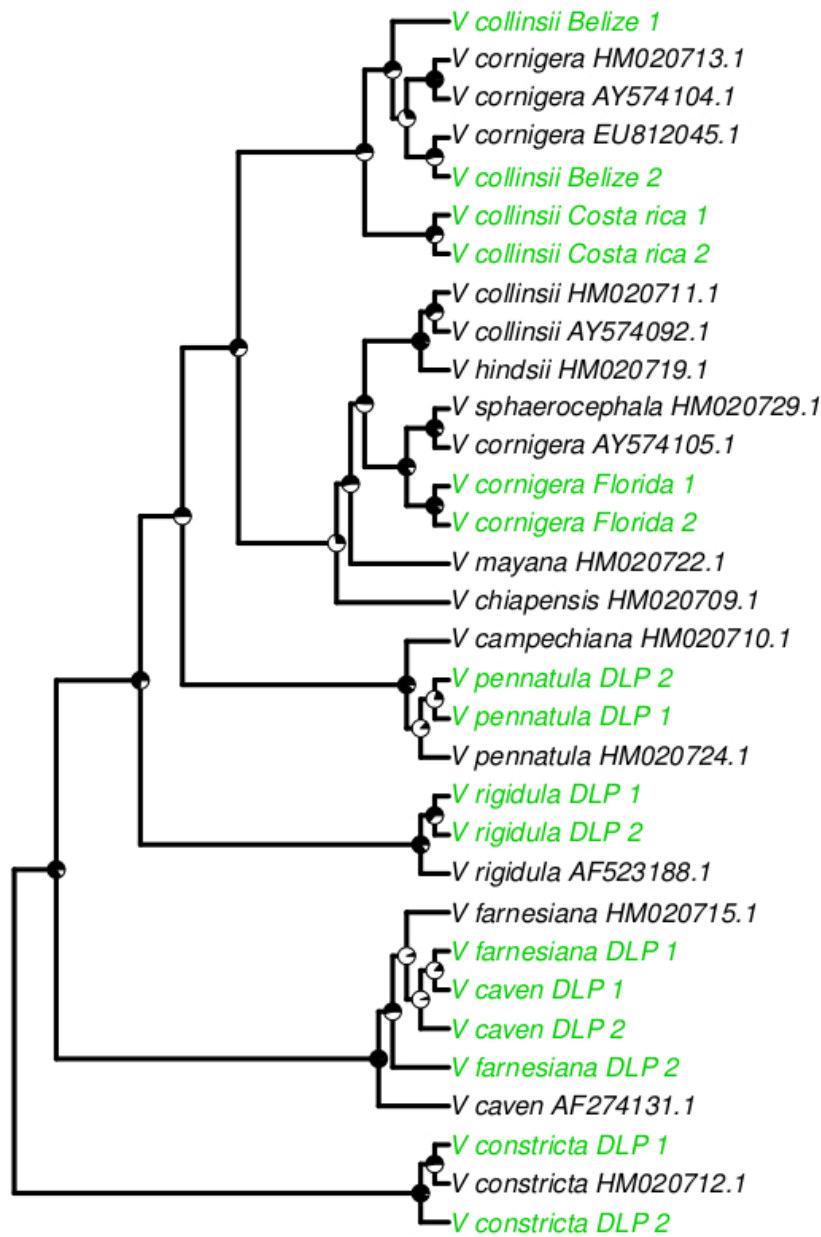


Figure S5. *MATK* gene phylogeny of species used in study. The coding sequence of the chloroplast marker, *MATK* was used to construct a maximum likelihood tree. Sequences from species/accessions in black were obtained from Genbank. Sequences from species in green were from this study. Piecharts represent bootstrap support for nodes. The root was pruned from tree. In general, the number of differences between species was very low.

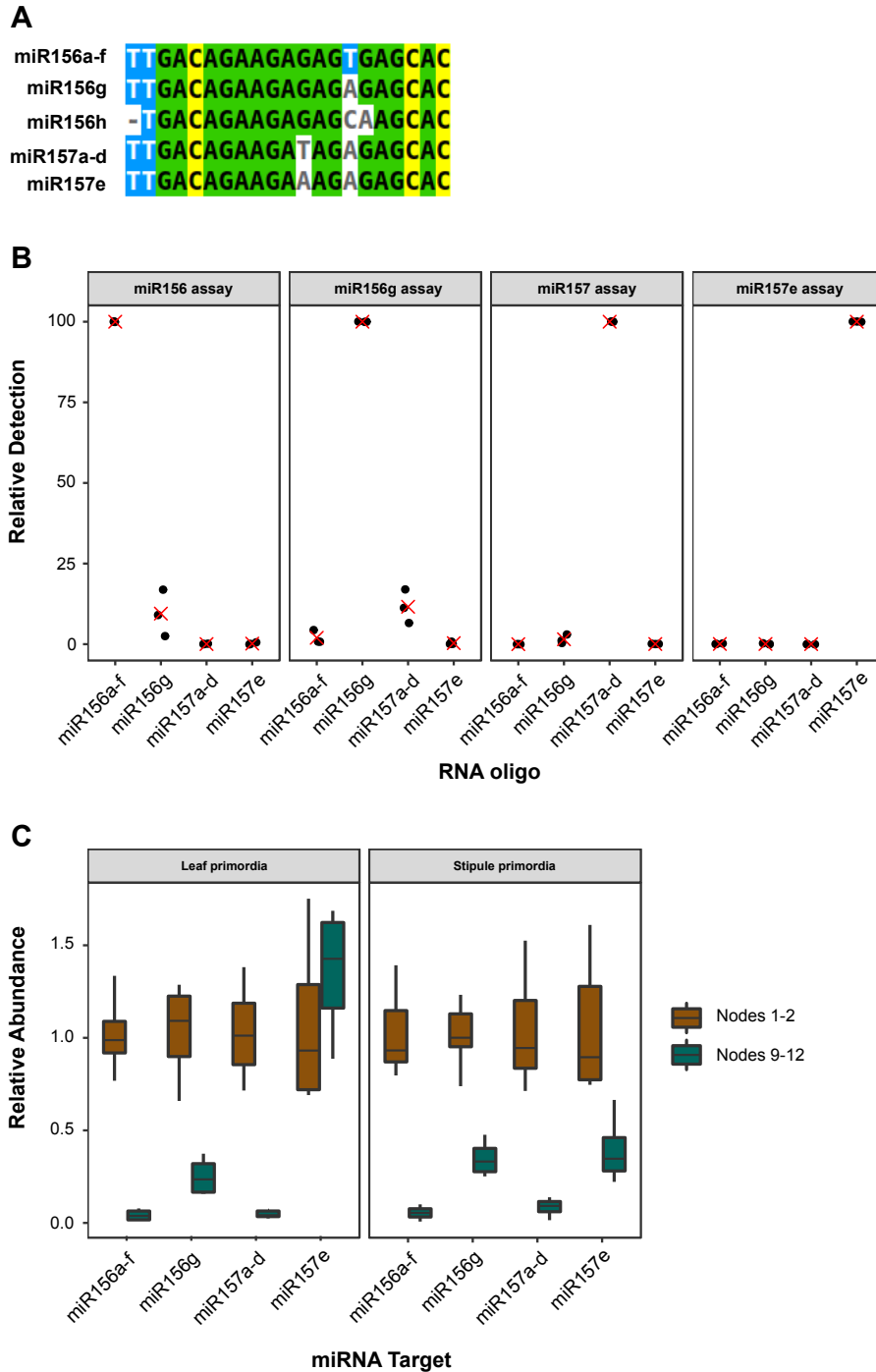


Figure S6. qPCR of miR156/157 variants. (A) Alignment of predicted miR156/miR157 variants identified from genome sequencing. (B) Relative detection assays for four different sets of qPCR primers intended to selectively amplify miRNA products from the following genes: *MIR156a-f*, *MIR156g*, *MIR157a-d*, *MI157e*. Points represent averages from 3 technical replicates for each of three different concentrations of synthetic RNA

oligos. The red “X” represents the average across concentrations. (C) qPCR of the four major miR156/157 variants in the first and last nodes sampled in both leaf and stipule primordia. Boxes bound 1st and 3rd quartile, center line marks the median.

Table S1. Genome assembly statistics for *V. collinsii* Belize (Vcoll-BR v1).

Kmer estimated genome size	518,254,840 bp
Assembly length (percent of total)	461,064,704 (89.0%)
Number of contigs (scaffolds)	122,266 (122,260)
NG50	6,528
Longest contig/scaffold	301,273
complete BUSCOs	86.9%
missing BUSCOs	7.6%

Table S2. Primers used in this study.		
ID	Use	Sequence (5' to 3')
Phylogenetics		
trnK685F	MATK sequencing; From Wojciechowski et al., 2004	GTATCGCACTATGTATCATTGA
trnK2R*	MATK sequencing; From Wojciechowski et al., 2004	CCCGGAACTAGTCGGATGG
trnK_3'_seq	MATK sequencing	TACTACATGAGCATTTCCTAATCCAT
smRNA-qPCR		
SPL_degenerate_motif_F1	3'RACE	MGBTTYTGYCARCARTGYAGYAGGTTYCA
miR156_Forward	miR156 qPCR	GCGGCGTTGACAGAAGAGAGT
miR156g_Forward	miR156g qPCR	GCGGCGTTGACAGAAGAGAGA
miR157_Forward	miR157 qPCR	GCGGCGTTGACAGAAGATAGA
miR157e_Forward	miR157e qPCR	GCGGCGTTGACAGAAGAAAGA
miR159_Forward	miR159 qPCR	GCGGCGTTTGGATTGAAGGGA
miR168_Forward3	miR168 qPCR	AAGGCGTCGCTTGGTG
miR156/7_RT_primer	miR156/7 reverse transcription	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCG CACCAGAGCCAACGTGCTC
miR159_RT_primer	miR159 reverse transcription	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCG CACCAGAGCCAAGTAGAGC
miR168_RT_primer	miR168 reverse transcription	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCG CACCAGAGCCAACCTCCCG

URP	universal reverse for smRNA qPCR	GTGCAGGGTCCGAGGT
miR156_RNA	Relative detection assay	rUrUrGrArCrArGrArArGrArGrArGrTrGrArGrCrArC
miR156g_RNA	Relative detection assay	rUrUrGrArCrArGrArArGrArGrArGrArGrArGrCrArC
miR157_RNA	Relative detection assay	rUrUrGrArCrArGrArArGrArTrArGrArGrArGrCrArC
miR157e_RNA	Relative detection assay	rUrUrGrArCrArGrArArGrArArArGrArGrArGrCrArC
mRNA-qPCR		
legume_ACT2_F	ACT2-LIKE qPCR	TGGCTCCACCAGAGAGAAAGTA
legume_ACT2_R	ACT2-LIKE qPCR	GAAGGTGCTGAGGGATGCAAG
VcoSPL1-like1_F1	semi-qPCR	TTGTAGGGCCGATCTTGGTAATGC
VcoSPL1-like1_R1	semi-qPCR	GGATGTGTCTTCCTCCTCCTCCTG
VcoSPL1-like2_F1	semi-qPCR	TTGAATGAGGGAGCAGGCGGCC
VcoSPL1-like2_R1	semi-qPCR	CCAGCCAATCTAGTTTTCTTTCCGTTC
VcoSPL1-like3_F1	semi-qPCR	CACCGCCGACACAAAGTTTGTGAGCT
VcoSPL1-like3_R1	semi-qPCR	ATTGGTTAGCAACGGAGGCTTTAGAAG
VcoSPL3-like1_F1	semi-qPCR	AGGGTTTGCCGGAGGAGGAAGTAG
VcoSPL3-like1_R1	semi-qPCR	CATCACTGAGCTCCGCCGTGCAG
VcoSPL3-like2_F2	semi-qPCR	GAGAGGTGTGGAGCTAACTTGGC
VcoSPL3-like2_R2	semi-qPCR	TTAGCCCTGCAACGATCACAACAG
VcoSPL3-like3_F1	semi-qPCR and qPCR	ATGGCGGTGTTGCAGAAGAGCAG
VcoSPL3-like3_R1	semi-qPCR and qPCR	GACCCTCTTCTTCATTTCATGA

VcoSPL3-like4_F1	semi-qPCR	TGGCTCTGGAGAAGAACATGGCC
VcoSPL3-like4_R1	semi-qPCR	CTTCCAGCTGGTTTGGGGCATCTC
VcoSPL3-like5_F1	semi-qPCR and qPCR	GATGGGAAGAAGAAGATGAGAGATCGG
VcoSPL3-like5_R1	semi-qPCR and qPCR	GTGCCATTGAGGAGGAAGAAGAAGAC
VcoSPL6-like1_F1	semi-qPCR	CAGCCATTGTTTTTAGGTTCTCTAGTTGAATC
VcoSPL6-like1_R1	semi-qPCR	GACCGAGCTCTCCTTGCAATTTTATC
VcoSPL6-like2_F1	semi-qPCR	AAAGCGTTGAGTTTGTGGACTTAGGG
VcoSPL6-like2_R1	semi-qPCR	AGTTTGATGCTTCCAGTGGTTGGC
VcoSPL6-like3_F1	semi-qPCR	AGGACGAGTTGGTGATCAAAGAGATGC
VcoSPL6-like3_R1	semi-qPCR	TTTCTACGACGCTCATTATGGCCTG
VcoSPL6-like4_F1	semi-qPCR	GGATGTTCCAAGGGCCAGGTTTTG
VcoSPL6-like4_R1	semi-qPCR	GCTCAACTCTCTGAAGCTGTGAC
VcoSPL7-like_F1	semi-qPCR and qPCR	CGAAGCTGTAGAAGAAAATTAGAGCGC
VcoSPL7-like_R1	semi-qPCR and qPCR	GGCAGTTTCTACGTCGTTTATCTCAC
VcoSPL8-like1_F1	semi-qPCR	ATGGCCTCGTCGAACTCGCCTC
VcoSPL8-like1_R1	semi-qPCR	GTGGTAGTGCTTGGCTTGTGATAG
VcoSPL8-like2_F1	semi-qPCR	GGGCTCTACCATGTGCGTTGAACAAC
VcoSPL8-like2_R1	semi-qPCR	GATGATAGTGCTTCGCGTGGGAG
VcoSPL9-like1_F1	semi-qPCR	CTATTCAGGCATAAAGTCTGTGGC
VcoSPL9-like1_R1	semi-qPCR	AGACCAGCAACAATGACTTTAGGGC

VcoSPL9-like2_F1	semi-qPCR	TGTCAGGTGGAGGGCTGCAAAGTG
VcoSPL9-like2_R1	semi-qPCR	CTCTTGACCTGCGACAACGACCCT
VcoSPL10-like1_F1	semi-qPCR	AAACGACTGTACTTTGAAGATGTCTGTG
VcoSPL10-like1_R1	semi-qPCR	GATAGGTCAAGTTTACAGCCTTCAACC
VcoSPL10-like2_F1	semi-qPCR	TTGTACGGAGGAGGCGGTGTGAAG
VcoSPL10-like2_R1	semi-qPCR	CAACCCTAACACCACAACCTTCGGT
VcoSPL13-like1_F1	semi-qPCR and qPCR	GGAGAACAACCAGAGGAACAATGACC
VcoSPL13-like1_R1	semi-qPCR and qPCR	TCGACGTGAAGTGTAGCTGCCGG
VcoSPL13-like2_F1	semi-qPCR	TGGGGTCTTCTGAAAGCGCCAGAA
VcoSPL13-like2_R1	semi-qPCR	CAGAGCTTCAACACATGGCTTCTCAAG
VcoSPL13-like3_F2	semi-qPCR and qPCR	TTGGTGGATGGATGCAGCTTTGAC
VcoSPL13-like3_R2	semi-qPCR and qPCR	AGGTTTCCTTCTGCGCTTATTATGG
VcoSPL13-like4_F1	semi-qPCR	ATCGGCGACATAGAGTGTGTGAGAAG
VcoSPL13-like4_R1	semi-qPCR	CCCCTACCAACACAATGGGCGTC
VcoMIR156a_F2	semi-qPCR	GATGAATCGAGATCAGTTTTAGCTGC
VcoMIR156a_R2	semi-qPCR	CTTCCCTTACCCTTCAAAGTCTCCATT
VcoMIR156b_F2	semi-qPCR	CTGCCTTCTTGTGGTTGATTTTTCTTC
VcoMIR156b_R1	semi-qPCR	GCACCGCGTATGTATGCCTCTC
VcoMIR156c_F1	semi-qPCR	CTCGACTCCTATAGCTTCTTTCATATG
VcoMIR156c_R1	semi-qPCR	GAACCAGAAATAGCAGTGAGAGACTC

VcoMIR156d_F1	semi-qPCR	CACTCATGATGTTGCTTCTTTCTTGGG
VcoMIR156d_R1	semi-qPCR	AAAAGGGAACGAAGCAGAGCCAGA
VcoMIR156e_F2	semi-qPCR	TCCGATTATTGCAGATTGCAGAGTAG
VcoMIR156e_R2	semi-qPCR	CTGTATAAGTAGTAGCAGAAAGTAAAAGG
VcoMIR156f_F1	semi-qPCR	ACGTGAACAAACAAAGTAGAGATTAATGAC
VcoMIR156f_R1	semi-qPCR	CACCCAAGTCAAATGTATGAACG
VcoMIR156g_F1	semi-qPCR	AGAAGAAGAAGAAGATAAGTTATGTGTTG
VcoMIR156g_R1	semi-qPCR	CACAGCTTTTTAGTCATTCCTGGG
VcoMIR156h_F2	semi-qPCR	TATTGTGTGAYCGTGTGTGTGTAGAG
VcoMIR156h_R2	semi-qPCR	CACCACCCTCCCTCTTTTCCTTC
VcoMIR157a_F1	semi-qPCR	GATAAAGAGGAATGGCTATGGCGATG
VcoMIR157a_R1	semi-qPCR	GATGCAATATTGATTTCTCGGTTTCAG
VcoMIR157b_F2	semi-qPCR	CCTTTTTTGCTAATTGTTAAGAGGAGCTG
VcoMIR157b_R2	semi-qPCR	CATCACCAAATTAACACGCCTCCCTT
VcoMIR157c_F2	semi-qPCR	GAAAGAGATACATCAGGCAGGAGGC
VcoMIR157c_R2	semi-qPCR	AATGAGGAGATAGGGACTGGAGGC
VcoMIR157d_F1	semi-qPCR	GAGGGAGATCGAGGAGGTTTCAG
VcoMIR157d_R1	semi-qPCR	AGATGATGATTTGCATGAACTCACTGC
VcoMIR157e_F2	semi-qPCR	CAGGTTTGATTGTTCCGTGATGCAC
VcoMIR157e_R2	semi-qPCR	TGCCGCTTCTTCTTCTTCAGC

Dataset S1

File contains contig identities from Vcoll-BR genome and associated annotation of MIR156, MIR157 and SPL transcripts and hairpin sequences.

REFERENCES

1. M.F. Wojciechowski, M. Lavin, M.J. Sanderson, A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gen resolves many well-supported subclades within the family. *American Journal of Botany* **91**, 1846-1862 (2004).
2. K. Katoh, D.M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
3. A. Stamatakis, RaxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
4. E. Paradis, J. Claude, K. Strimmer APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
5. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> (2015).
6. T. Magoc, S.L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963 (2011).
7. A.V. Zimin, G. Marcais, D. Puiu, M. Roberts, S.L. Salzber, J.A. Yorke, The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677 (2013).
8. R.M. Waterhouse, M. Seppey, F.A. Simao, M. Manni, P. Ioannidis, G. Klioutchnikov, E.V. Kriventseva, E.M. Zdobnov, BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35**, 543–548 (2017).
9. F.A. Simao, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

10. S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410 (1990).
11. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
12. S. Griffiths-Jones, R.J. Grocock, S. van Dogen, A. Bateman, A.J. Enright, MiRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* **34**, D140-144 (2006).
13. R. Lorenz, S.H. Bernhart, C. Hoener zu Siederdisen, H. Tafer, C. Flamm, P.F. Stadler, R.F. Hofacker, ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 26 (2011).
14. B.L. Cantarel, I. Korf, S.M.C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A.S. Alvarado, M. Yandell, MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**, 188-196 (2008).
15. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
16. E. Varkonyi-Gasic, R. Wu, M. Wood, E.F. Walton, R.P. Hellens, Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* **3**, 12 (2007).
17. K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* **25**, 402-408 (2001).
18. G. Riedel, U. Rudrich, N. Fekete-Drimusz, M.P. Manns, F.W.R. Vondran, M. Bock, An extended ΔC_T -method facilitation normalisation with multiple reference genes suited for quantitative RT-PCR analyses of human hepatocyte-like cells. *PLOS ONE* **9**, e93031 (2014).