Transcriptomic and evolutionary analysis of the mechanisms by which P. argentatum, a rubber producing perennial, responds to drought

Andrew D. L. Nelson

University of Arizona

Grisel Ponciano

USDA-ARS Western Regional Research Center

Colleen McMahan

USDA-ARS Western Regional Research Center

Daniel C. Ilut

Cornell University

N. Ace Pugh

University of Arizona

Diaa Eldin Elshikha

USDA-ARS

Douglas J. Hunsaker

USDA-ARS

Duke Pauli

University of Arizona

Research article

Keywords: Guayule, transcriptome, drought stress, rubber biosynthesis, lncRNAs, comparative genomic

---

Abstract

Background Guayule (Parthenium argentatum Gray) is a drought tolerant, rubber producing perennial shrub native to northern Mexico and the US Southwest. Hevea brasiliensis, currently the world's only source of natural rubber, is grown as a monoculture, leaving it vulnerable to both biotic and abiotic stressors. Isolation of rubber from guayule occurs by mechanical harvesting of the entire plant. It has been reported that environmental conditions leading up to harvest have a profound impact on rubber yield. The link between rubber biosynthesis and drought, a common environmental condition in guayule's native habitat, is currently unclear. Results We took a transcriptomic and comparative genomic approach to determine how drought impacts rubber biosynthesis in guayule. We compared transcriptional profiles of stem tissue, the location of guayule rubber biosynthesis, collected from field-grown plants subjected to water-deficit (drought) and well-watered (control) conditions. Plants subjected to the imposed drought conditions displayed an increase in production of transcripts associated with defense responses and water homeostasis, and a decrease in transcripts associated with rubber biosynthesis. An evolutionary and comparative analysis of stress-response transcripts suggests that more anciently duplicated transcripts shared among the Asteraceae, rather than recently derived duplicates, are contributing to the drought response observed in guayule. In addition, we identified several deeply conserved long non-coding RNAs (lncRNAs) containing microRNA binding motifs. One lncRNA in particular, with origins at the base of Asteraceae, may be regulating the vegetative to reproductive transition observed in water-stressed guayule by acting as a miRNA sponge for miR166. Conclusions These data represent the first genomic analyses of how guayule responds to drought like conditions in agricultural production settings. We identified an inverse relationship between stress-responsive transcripts and those associated with precursor pathways to rubber biosynthesis suggesting a physiological trade-off between maintaining homeostasis and plant productivity. We also identify a number of regulators of abiotic responses, including transcription factors and lncRNAs, that are strong candidates for future projects aimed at modulating rubber biosynthesis under water-limiting conditions common to guayules' native production environment.

Background

Natural rubber is a crucial material with a myriad of uses and applications, making it invaluable to a wide range of industries, and contributing to its economic footprint of ~ 12.7 billion USD (DESA/UNSD). Natural rubber production, which is predominantly sourced from the rubber tree (*Hevea brasiliensis*), is currently threatened posing socioeconomic risks to industries relying on it as raw material (1). Because the species is clonally propagated and is grown as a geographically concentrated monoculture, it is vulnerable to diseases such as South American leaf blight (*Microcyclus ulei*), a fungal pathogen endemic to *Hevea*'s center of origin in the Amazon (2). Due to these growing concerns for the future stability of *Hevea* populations, scientists have continually searched for alternative sources of natural rubber (3,4). One such species, guayule (*Parthenium argentatum* A. Gray), has already been shown to be an attractive source of natural rubber that may be able to help address projected future shortages (3–6).

Guayule grows throughout northern Mexico and much of the American southwest and thus is naturally adapted to arid environments (6,7). Because of this, producers first considered guayule as an alternative source of natural rubber in the early 1900's. Subsequent utilization of guayule as a rubber source has progressed through multiple "boom and bust" phases largely influenced by world markets and import costs surrounding rubber from *H. brasiliensis* (6). Due to over a century of sporadic but intense efforts to harness guayule's rubber producing potential, it is now understood that the crop has practical advantages over *Hevea*; for example, as a hypoallergenic alternative for those that have adverse reactions to latex rubber (8).

Crop scientists are now aware of unique challenges that guayule poses, particularly from a plant breeding perspective. Indeed, genetic improvement of guayule is complicated because the species has two different modes of reproduction and is able to exist as either facultatively apomictic, polyploid individuals or as sporophytic, self-incompatible diploid individuals (9–12). Due to this inherent biological complexity, a modern breeding approach that uses molecular techniques in tandem with traditional phenotypic selection may be the most effective way to increase the rate of genetic gain in the crop, particularly under stress conditions (13).

While the biological purpose for rubber biosynthesis and accumulation in the plant is unclear, its production is believed to be linked to the way the plant responds to abiotic stressors such as drought and temperature (14–17). Abiotic stress has been shown to elicit a dramatic and highly tissue specific reprogramming of the transcriptional profile in many plant systems (18). For instance, reproductive tissue in maize exhibits down regulation of genes associated with cell division and DNA replication during drought exposure, consistent with the observed delay in ear growth (19). In sunflower (Helianthus annuus), a naturally drought tolerant relative of guayule, exposure to drought conditions resulted in elevated levels of genes associated with osmotic adjustment in leaf and vasculature tissue (20), as well as decreases in genes associated with oil metabolism in the seed (21). While guayule exhibits similar drought tolerant characteristics as sunflower, it is unclear if it responds to stress in a similar molecular manner.

One critical, but until recently, overlooked aspect of the stress response in plants lies in the non-coding RNAs (microRNAs and long non-coding RNAs) that help sense and regulate the response to stress (22–24). These non-coding RNAs act at the pre- and post-transcriptional level to modulate expression and activity of other genes necessary for the stress response (25). Due to their highly specific expression patterns and species specificity sequence conservation, these transcript classes are important targets for understanding the unique ways in which plants have evolved to respond to changes in their environment (26–28). As such, all aspects of the transcriptome must be examined to fully understand the link between rubber biosynthesis and drought stress in guayule.

To uncover the molecular mechanisms that facilitate the drought response in guayule, we used a transcriptomic approach to identify differentially expressed transcripts between plants grown under both well-watered and water limited conditions. We used a phylogenetic approach to gain some insight into

whether recent or more ancient gene duplications were contributing to the observed stress response. Finally, we uncovered a number of stress-responsive long non-coding RNAs, several of which harbor conserved miRNA binding motifs, including two miRNAs with known roles in flowering and drought responses. These lncRNAs add an additional layer of regulatory complexity to the guayule drought response. Thus, we present a first glimpse at how guayule responds to drought and offer some molecular targets for plant breeders wishing to study the trade-off between rubber biosynthesis and water conservation.

Results

Examining the impact of drought at the transcriptome-wide level in guayule

Guayule is a drought tolerant species that has likely evolved a number of physiological mechanisms that enable it to mitigate the effects of drought prevalent in its native environment. To gain an understanding of what genes might be involved in guayule's drought response mechanisms, we evaluated the guayule accession AZ-3 grown in plots for 29 months in Maricopa, Arizona having two contrasting irrigation regimes, $I_{100\%}$ and $I_{25\%}$ (Fig. 1a; (29,30). The $I_{100\%}$ (or control treatment) was completely replenished with irrigation water, meeting measured evaporative soil water losses, while the $I_{25\%}$ received only 25% of the irrigation given to $I_{100\%}$. At the time of collection in March of 2015, the 29-month-old $I_{25\%}$ guayule plants were flowering in comparison to those grown at $I_{100\%}$, which were not (Additional file 1: Figure S1). Stem tissue, the predominant location of guayule rubber biosynthesis, was collected from three biological replicates in each irrigation regime for transcriptomic analysis (Fig. 1b).

Given that no guayule genome is currently available for public use, we utilized a previously published *de novo* assembled transcriptome generated from a mixture of 150 and 300 bp reads (13) for read mapping. This transcriptome contains >200,000 transcripts, suggesting the presence of incomplete or redundant (identical) transcripts. The presence of multiple fragments corresponding to the same transcript might confound our attempts to identify genes that are differentially expressed in response to limited water. The Stonebloom and Scheller transcriptome was filtered in two ways (Fig. 2a), collapsing the transcriptome from 219,819 transcripts to 63,672, a figure congruent with expectations. To ensure that filtering had not removed a significant number of actual transcripts, we mapped our RNA-sequencing data to both filtered and unfiltered transcriptomes and compared the number of reads that mapped to both. No differences were observed in mapping rates (~0.5% improvement in mapping to filtered set over unfiltered; Additional file 2: Table S1), suggesting that the filtered transcriptome would be sufficient for differential expression (DE) analyses.

Differentially expressed genes were identified by comparing the $I_{25\%}$ irrigation treatment to the $I_{100\%}$. Of the 63,672 transcripts, 42,711 were expressed (minimum of 0.5 TPM in all replicates) in the control conditions and 43,002 in the samples grown under the limited water. Of these, 251 transcripts were upregulated under the water-limited irrigation regime whereas 393 were downregulated (Fig. 2b and Additional file 3: Table S2; adjusted p-value of 0.01). The transcript most significantly upregulated in the

water-limited treatment, GFTW01080018.1 (Fig. 2c), was expressed 23-fold compared to the control treatment (~9 −fold increase observed with qRT-PCR, Additional file 4: Figure S2). In contrast, the transcript most significantly downregulated, GFTW01080137.1 (Fig. 2d), was reduced more than 200-fold to near imperceptible detection levels, a value confirmed by qRT-PCR (Additional file 4: Figure S2).

To gain an understanding of the cellular mechanisms that are involved in guayule's response to drought, we performed a GO analysis of the significantly up- and down-regulated transcripts. An InterPro ID or shared similarity with an Arabidopsis protein-coding gene allowed us to infer biological processes for 273 of the 393 downregulated, and 163 of the 251 upregulated transcripts (Additional file 4: Table S3). Transcription factors (regulation of transcription) were the most abundant class of both up and down-regulated transcripts (Fig. 3). In agreement with previous data from drought-stressed plants, defense response, trehalose biosynthesis (31), glycosyltransferase activity (32,33), and response to water deficit were among the processes more likely to be upregulated under the water-limited irrigation treatment, whereas isoprenoid/terpenoid biosynthesis, carbohydrate metabolism, and lipid metabolism processes were more likely to be downregulated (Fig. 3).

Next, the most differentially expressed transcripts were assessed. The most significant, highly upregulated transcript, GFTW01080018.1, appears to be orthologous to the Arabidopsis PIP2s (specifically PIP2A, B, and C; Additional file 6: Figure S3), a family of aquaporins important for hydraulic regulation (34). Despite the recovery of numerous PIP2 paralogs in the genomes of *Helianthus annuus* and *Lactuca sativa*, two close relatives of guayule within the Asteraceae (35); Additional file 6: Figure S3), and three paralogs in the guayule transcriptome, only one aquaporin was differentially expressed in response to water deficit ($I_{25\%}$). The most significantly down-regulated transcript, GFTW01080137.1, shares sequence similarity to Arabidopsis Cold Regulated Gene 27 (COR27; AT5G42900). Interestingly, in Arabidopsis, COR27 and another cold regulated gene with little sequence similarity, COR28, are positive regulators of flowering (36). In guayule, putative orthologs for both COR27 and COR28 (GFTW01080137.1 and GFTW01127972.1, respectively) are both significantly repressed under water limited conditions, despite the near uniform flowering that was observed for these plants (Additional file 1: Figure S1). Finally, GFTW01028919.1, the transcript that displayed the greatest decrease in transcription (although not the most significant), at > 900-fold (adjusted p-value < 2E-12; Additional file 7: Figure S4) is a putative ortholog of Arabidopsis Terpene Synthase 3 (AT4G16740) and is one of 12 downregulated guayule transcripts involved in isoprenoid/terpenoid biosynthesis (Fig. 3). In sum, guayule's transcriptomic response to water-limited conditions includes a dramatic increase in aquaporin production and defense response genes, as well as a decrease in terpenoid biosynthesis, carbohydrate metabolism, and oxidation reduction mechanisms.

Examining the evolutionary history of duplicated drought-responsive transcripts

The GO-term analysis revealed that some of the differentially expressed guayule transcripts displayed similarity to the same Arabidopsis gene, suggesting one of three possibilities: 1) an ancient expansion in a stress-responsive gene family, 2) that the transcripts are paralogs that emerged following the cross-

hybridization and polyploidy event that gave rise to AZ-3, or 3) that the transcripts contain the same functional domain but bear no phylogenetic relationship. Specifically, 127 guayule stress-responsive transcripts clustered, in sets of 2 - 4 transcripts each, with 56 Arabidopsis genes. For example, the downregulated guayule terpene synthase ortholog (GFTW01028919.1) groups with AT4G16740 along with two other guayule transcripts (GFTW01072004.1 and GFTW01017460.1). We first determined if the guayule transcripts were indeed the product of a gene duplication by examining codon-guided multiple sequence alignments. Transcripts associated with roughly half (n = 27) of the Arabidopsis gene clusters either did not share a recent evolutionary past (sequence identity < 50%) or there was not enough evidence to support a gene duplication (e.g., guayule gene fragments that did not overlap one another in the alignment). The three guayule transcripts within the terpene synthase cluster with AT4G16740 shared sufficient sequence similarity to proceed forward to phylogenetic analysis, whereas three guayule transcripts that shared similarity with an Arabidopsis mitogen-activated protein kinase (MAPK16, AT5G19010) exhibited little to no similarity outside of the kinase domain and were not considered further.

To determine the timing of the guayule gene duplication events associated with the remaining 29 Arabidopsis gene clusters, we took a comparative and evolutionary approach, searching the genomes of sunflower (*H. annuus*; (35)) and lettuce (*L. sativa*; (37)) for homologs to the stress-responsive guayule transcripts and their putative Arabidopsis orthologs. We then inferred phylogenies for each of these gene families to determine when the observed gene duplication occurred. Two whole genome triplication events are shared between sunflower and guayule, with an additional, species-specific whole genome duplication event occurring in each species (Fig. 4a). Thus, we examined the resulting phylogenies for two patterns that would indicate that the guayule transcripts were the result of an Asteraceae (or earlier) duplication event (Fig. 4b, left; "Asteraceae event"). In this scenario, each of the guayule transcripts would be immediately-sister to a sunflower gene. In the event that the transcript duplication was AZ-3 specific, we would observe the duplicated transcripts first sister to each other and then to a sunflower gene (Fig. 4b, right; "AZ-3 event"). Of the 20 Arabidopsis gene clusters comprised of down-regulated guayule transcripts, 13 contained transcripts where the gene duplication was inferred to be an Asteraceae event (Fig. 4c, purple bar), 7 arose from an AZ-3 event (Fig. 4c, blue bar), and two gene clusters contained both types of duplication events. Of the nine Arabidopsis gene clusters comprised of up-regulated guayule transcripts, three of the paralogs arose from an Asteraceae event, whereas six where AZ-3 specific (Fig. 4c). One example of a AZ-3 event can be seen in the putative guayule orthologs of AT1G01060 (LHY), a transcription factor that regulates flowering and circadian rhythm (Fig. 4d, blue box). These transcripts, all of which were significantly upregulated, fall sister to one another in the phylogeny with strong bootstrap support. In contrast, the terpene synthase gene cluster, contained two guayule transcripts that were each sister to multiple sunflower genes (Fig. 4e, purple box).

Duplication and expression do not necessarily imply that the resulting transcript is capable of encoding for a protein. Pseudogenization or neo-functionalization of a locus (protein-coding gene -> long non-coding RNA) can occur through the disruption of a protein-coding gene's open reading frame (ORF). We examined each of the gene clusters for loss of ORF integrity in at least one (but not all) of the duplicate

guayule transcripts. We found that 6/20 of the down-regulated gene clusters had experienced a pseudogenization event that left them with a single protein-coding gene, whereas 7/9 up-regulated gene clusters were left with a single protein-coding transcript (Fig. 4c, tan bars). Thus, it appears that a number of stress-responsive paralogs with intact ORFs have been retained through multiple speciation events, suggesting they may help guayule mount a response to drought conditions.

A role for long non-coding RNAs in guayule's drought response

The identification of stress-responsive transcripts that are no longer protein-coding raises the possibility of uncovering long non-coding RNAs (lncRNAs) that are also differentially expressed under the water-limited irrigation regime. While not as extensively studied in plants as in vertebrate systems, a number of plant lncRNAs have been reported to differentially expressed in response to abiotic and biotic stress (38–42), where, among many functions, they can act as regulators of transcription, microRNA sponges, and influence alternative splicing (25,43,44). Although not differentially expressed under the imposed irrigation treatments, a homolog of the deeply conserved light responsive lncRNA, *HID1 (45)*, was present in the guayule transcriptome (Fig. 5a). As expected based on prior analyses, the protein interaction domain annotated as SL2 was highly conserved between Asteraceae, Arabidopsis, and rice (Fig. 5a), suggesting a potentially shared role for this lncRNA across flowering plants. In addition, the identification of a guayule *HID1* demonstrates that the Stonebloom and Scheller transcriptome captured polyadenylated lncRNAs as well as protein-coding transcripts.

To identify putative lncRNAs, we focused on the set of differentially expressed transcripts that bore no similarity to any known protein domains (Fig. 5b). We then removed potential transposable elements (TEs) and known housekeeping RNAs (rRNAs and spliceosomal RNAs). To be conservative in our lncRNA identification, we also removed any transcripts that overlapped a protein-coding gene in the *H. annuus* genome, as these guayule transcripts may reflect incompletely assembled protein-coding genes resulting from technical difficulties of *de novo* transcriptome assembly. Following these filters, we recovered 31 putative lncRNAs that were down-regulated and 39 that were up-regulated in response to drought (a complete list can be found in Additional file 8: Table S4).

We then took an evolutionary approach to identify putative lncRNAs for which we could recover sequence homologs in other species under the premise that conservation implies functionality (26). Of the 70 guayule putative lncRNAs, we identified a sequence homolog for 14 in the sunflower genome (Fig. 5c). We uncovered evidence of conservation for three lncRNAs in the lettuce genome, suggesting that these loci emerged at least ~39 million years ago. Four of the fourteen sunflower conserved lncRNAs were also annotated as lncRNAs in that system, with one also annotated as a lncRNA in lettuce, lending additional confidence in their lncRNA designation (Fig. 5c).

 Next, an attempt to assign a function to these putative lncRNAs beyond "stress-responsive" was made. Our experimental design lacked depth to attempt a "guilt-by-association" analysis, and the absence of a guayule genome precludes the association between a lncRNA and the neighboring protein-coding gene it might regulate. Therefore, we focused on whether the set of guayule lncRNAs might be involved in

sequestering miRNAs away from their intended targets, or in miRNA or phasiRNA, biogenesis. Using psRNAtarget (46), we predicted whether miRNAs might bind to the 14 lncRNAs for which we identified sequence homologs in sunflower. We then scanned the homologous locus in sunflower (and in lettuce) for conservation of the miRNA binding site. Using this approach we identified six lncRNAs with conserved miRNA binding sites (Fig. 5c; Additional file 8: Table S4). One of the guayule lncRNAs conserved and annotated as a lncRNA in both sunflower and lettuce, GFTW01168370.1, harbors a completely conserved binding site for miR166 (Fig. 5d), a microRNA associated with tissue development and whose knockdown in Arabidopsis leads to an enhanced drought response (47). As a miRNA sponge, GFTW01168370.1 would act to recruit miR166 away from its intended target, in short mimicking the knockdown response reported in Arabidopsis. Thus, within the dataset of drought-responsive transcripts, a subset was identified that showed the hallmarks of being lncRNAs. Several of these lncRNAs contain conserved miRNA binding sites, with one in particular likely helping to mediate the guayule drought response.

Discussion

Transcriptome analyses reveal a suite of drought-responsive genes in guayule

As a perennial shrub native to the American Southwest and northern Mexico, guayule is well adapted to long periods of little to no water. Using next-generation sequencing, we examined the molecular mechanisms by which guayule responded to simulated drought conditions via imposed irrigation treatments. By examining stem tissue, the primary location of rubber biosynthesis in guayule, we were also able to consider the impact of drought on this metabolic pathway. We performed our analyses using a published transcriptome for guayule, taking steps to collapse potential isoforms and miss-assembled transcripts. As expected, we identified a number of differentially expressed transcripts involved in signal transduction pathways (e.g., protein phosphorylation), transcriptional regulation, and transmembrane transport. We identified more than 20 up or down-regulated transcripts with similarity to Arabidopsis transcription factors associated with circadian clock regulation. Interestingly, many of these transcripts are annotated as cell-to-cell mobile in Arabidopsis (48), perhaps indicating that our transcriptomic analysis in stem tissue is generating a snapshot of circadian regulation occurring elsewhere in the plant. Regardless, while drought conditions dramatically impact both flowering and the circadian clock in guayule, due to the abundance of transcripts, it is unclear which transcript might be the regulator/sensor that is connecting drought to flowering.

The most upregulated guayule transcript is orthologous to the Arabidopsis aquaporin PIP2 family. Interestingly, despite recent duplications in close relatives, sunflower and lettuce, that are likely shared with guayule, we only observed differential expression for a single aquaporin out of three observed in the transcriptome, suggesting that it is the key regulator of water transport in stem tissue. We also observed twelve transcripts associated with rubber biosynthesis that were down-regulated under water-limited conditions. Although guayule rubber biosynthesis is known to be induced by cold temperatures, little is known about the mechanistic impact drought has on this pathway. However, given the abundance of terpene biosynthesis-associated transcripts and their almost complete down-regulation suggests that

guayule modulates precursors to the rubber biosynthesis pathway when faced with water deficit conditions. This is in agreement with the observation that $I_{100\%}$ plants contained twice the rubber content of those grown at $I_{25\%}$ even though water use efficiency was equivalent (29).

WGD events have added to the complexity of the guayule drought response

Gene duplication, when the resulting duplicate is retained, can result in increased nuance in how plants perceive and respond to abiotic stress (49). The presence of duplicated transcripts in guayule are not surprising, given the multiple reported whole genome duplication (WGD) events leading up to the speciation event of guayule (35). A whole genome triplication event occurred at the base of the Asteraceae and is shared among all family members. More recently, a whole genome duplication has been observed in the formation of the guayule accession used in this analysis, AZ-3. AZ-3 is a complex polyploid formed by the likely hybridization of diploid *P. argentatum* and an unknown Parthenium species. Tetraploid guayule reportedly has increased biomass, rubber yield, and vigor compared to its diploid relatives. Thus, both of these polyploidization events raise the possibility that some of the duplicated genes may be mediating a successful response to drought stress or are contributing to increased vigor in the species.

We searched for evidence of duplication in the stress-responsive transcripts using a parsimony based approach to infer when those duplications occurred. It should be noted that we are not observing all duplicate genes here, only the ones that continue to be stress-responsive following duplication. These transcripts likely retain conservation in their regulatory domains (e.g., promoter elements), but in the absence of a genome, we focused on retention of protein-coding capacity. We were able to infer duplication events for 29 clusters of 68 stress-responsive guayule transcripts, with most (18/29) duplication events shared across Asteraceae. ORFs were retained in a majority of these transcripts (16/29), which, when combined with the shared pattern of differential expression between paralogs and their deep conservation, suggests that these duplicates are functional. However, as most of the observed retained duplicates appear to be shared across Asteraceae, they likely cannot explain the vigor associated with tetraploid guayule.

LncRNAs are helping to mediate the drought response in guayule

Long non-coding RNAs add an additional layer of complexity to plant stress responses through their ability to act as pre- and post-transcriptional regulators of gene expression. Interestingly, we recovered a homolog of *HID1*, a lncRNA that helps mediate shade avoidance in Arabidopsis. Although HID1 is conserved across land plants, this is the first Asterid homolog identified. In agreement with previous reports on *HID1* conservation, guayule *HID1* was conserved in the 5' region believed to be important for protein-binding. Given the role of *HID1* in light signaling it is perhaps not surprising that its expression was not responsive to drought. However, we were able to identify 70 putative lncRNAs that were differentially expressed in response to drought, 14 of which were conserved in the sunflower genome. De novo transcriptome assembly routinely produces fragmented transcripts with disrupted ORFs that would

appear to look like a lncRNA. Thus, we took a more conservative approach than is typically taken when a reference genome is available by filtering out any transcripts that shared sequence similarity with protein-coding genes from related species. Four of the sunflower-conserved lncRNAs were also annotated as lncRNAs in sunflower, lending further support to their classification in guayule. Based on conservation and their stress-responsiveness, we would predict that these lncRNAs are likely functioning to modulate the drought response in guayule.

Functional prediction for lncRNAs is difficult in the absence of genomic context clues or without the ability to apply guilt-by-association strategies through many experimental time points or conditions. Thus, we focused on one functional class of lncRNA, that of miRNA sponge/precursor, as miRNA binding sites are fairly easy to predict computationally. Again, using sequence conservation as a means of boosting predictive confidence, we identified conserved miRNA binding sites in six guayule lncRNAs. One of these putative miRNA sponges in particular harbors a binding site for miR166, a microRNA involved in vegetative growth, floral morphogenesis, and regulating responses to salinity and drought. The lncRNA containing the miR166 binding site is upregulated under drought conditions and therefore could be mediating either the observed floral transition or the drought response.

Conclusions

As a drought tolerant, rubber producing perennial crop, guayule represents a remarkable natural resource for meeting industrial demands for raw products. In the present work, a transcriptomic and comparative evolutionary analysis approach was taken to identify and characterize the molecular response of guayule to drought-like conditions. We found that rubber biosynthesis-associated transcripts were dramatically down-regulated in the plants subjected to water-limited conditions in comparison to the plants in the well-watered control treatment. These results demonstrate that even given guayule's inherent drought tolerance, there is a molecular trade-off occurring between rubber biosynthesis and the plants ability to maintain hydration status and homeostasis. These findings suggest that water and other crop inputs need to be optimized with respect to rubber yield to find an economic balance for potential producers.

Methods

Plant growth and tissue collection

Guayule (*P. argentatum* AZ-3) plants were grown in the field under subsurface drip irrigation at the University of Arizona, Maricopa Agricultural Center in Maricopa, Arizona as described in (29). AZ-3 seed has been deposited and is available from the USDA-ARS National Plant Germplasm System (NPGS; https://www.ars-grin.gov/) in Parlier, CA under the ID PI 599676. On the day of final harvest when plants were 29 months-old (March 2015), 10-15 mm diameter-stem segments from each plant were harvested and immediately frozen in liquid nitrogen and then stored at -80°C until used. Three biological replicates for each treatment were harvested.

RNA extraction and Illumina library preparation

Approximately 2g of stem tissue was used for total RNA extracted following the Laudencia et al. 2007 (50) protocol with the following modifications: (i) acid phenol:chloroform MB grade (Ambion, USA) was used for the phenol:chloroform extraction step; (ii) the precipitated RNA was further cleaned with Qiagen RNeasy Plant Mini Kit (Qiagen, USA); and (iii) the cleaned RNA was treated with DNA-*free*[TM] kit (Ambion, USA). PolyA-RNA was prepared employing Qiagen RNeasy/QIAshredder protocols (Qiagen, USA). RNA-sequencing libraries were prepared using the KAPA stranded mRNA-seq kit for Illumina (KK8420) according to the manufacturer's protocol (KR0960 - v3.15). RNA-sequencing was performed on the Illumina HiSeq2000 with 150 bp paired-end reads. A total of 98,430,986 reads was generated for the six samples.

Transcriptomic analysis

A condensed version of the Stonebloom and Scheller transcriptome was generated by initially filtered using CD-HIT-EST v.4.6.8 (51) with a global sequence identity of 1 (100%). To identify potentially identical transcripts that contained a single mis-aligned read, 150 nts were removed from either the 5' or 3' end of the transcript, and if the resulting transcript was greater than 150 nts, was used as query in a BLASTn (52) against all other transcripts. Hits against self were removed, and then all other hits with 100% coverage of one of the sequences, as well as 100% identity, were collapsed into one transcript, with the longest transcript being retained. Read mapping and quantification was performed using Salmon v0.81 (53) in CyVerse's Discovery Environment (54). Quantified reads were prepared for differential expression analysis using the tximport (55) package in R. Differential expression was determined using DESeq2 (56) with an adjusted p-value of 0.01 as the cutoff for significance.

Quantitative PCR (qPCR) analysis

Stem bark tissue was the source of RNA for qPCR analysis. For each of three biological replicates (of both water-deficit and well-watered plants), total RNA was extracted with RNAqueous™ kit (Invitrogen, USA) and traces of DNA removed with DNA-free[TM] kit (Life Technologies, USA). Two micrograms of total RNA were the template for oligo $(dT)_{20}$-generated cDNA with SuperScriptIII First-Strand Synthesis System for qPCR (Life Technologies, USA) following manufacturer instructions. The qPCR reactions were carried out using Applied Biosystems 7500 Fast Real Time PCR System and SYBR Green chemistry (Life Technologies, USA) in 20µl volume reactions containing 400 ng of template cDNA, 900nM of each forward and reverse primer, 10µl of Fast SYBR® Green Master Mix, and water as needed. The following combinations of forward/reverse primers were used: for GFTW01080018.1, 5'-TGCCGTATTCATGGTTCACTTG -3' / 5'- GGGCCGGGTTGATTCC -3'; for GFTW01080137.1, 5'-TTTGTGGAGCAGGAGGAGAG-3' / 5'-GCCAGATGAAACTGTATCAGAGC-3'; for GFTW01028919.1, 5'-ACTTCAAAGGTCGTTCCAAGAC-3' / 5'-TGCCTCGCATTTTTCTCCAG-3'; and for *Pa18S*, 5'-TACTATGGTGGTGACGGGTG -3'/5'- ATTGTCACTACCTCCCCGTG -3'. Thermocycler temperature regime was 95°C for 20s, followed by 40 cycles of 95°C for 3s and 60°C for 30s. Data were analyzed using the

7500 Fast System Detection Software (Life Technologies, USA) with manually set threshold. Expression of each target gene was calculated with the Livak and Schmittgen 2001 method (57), normalized to expression of the endogenous reference gene *Pa18S*, and then to its expression in a calibrator (well-watered control plant). Four technical replicates qPCR reactions were run for each target gene, and the whole experiment was performed twice using the same RNA but freshly synthesized cDNA.

Functional analysis

GO terms for the differentially expressed transcripts were identified using BLAST2GO (58). First, BLASTx was performed against a database of Arabidopsis protein-coding genes with an e-value of 1E-3 and word size of 3. Protein domains were identified using InterProScan with default parameters. For guayule transcripts sharing similarity with an Arabidopsis protein-coding gene as determined by BLASTx analysis, but for which no functional annotation was obtained through BLAST2GO, we extracted Biological Processes directly from TAIR (59).

Duplication event timing and phylogenetic analysis

To determine timing of duplication, gene families were first generated by identifying sequences in the H. annuus (CoGe ID 37147) and L. sativa (CoGe ID 37106) genomes that shared sequence similarity with both the Arabidopsis and guayule sequences using CoGe BLAST with default parameters and an E-value of 1E-10 (60). Coding sequences were extracted from the top five unique loci in each genome using CoGeBLAST's view FASTA feature. Sequences were aligned using MAFFT (61) in Geneious (62). The 5' and 3' UTRs of guayule transcripts were trimmed based on the multiple sequence alignment so that all sequences started with an "ATG" and ended with a stop codon. These alignments were then used to infer phylogenetic relationships with RAxML (v7.2.8, (63)) with the GTR GAMMA substitution model and 100 bootstraps. Trees with poor support (< 70) specifically at the guayule-sunflower node were realigned with fewer sequences or different MAFFT parameters until the support increased above 70. The sister branch to the query guayule sequence, whether it was sunflower or a guayule paralog, was used to infer timing of the duplication event based on the known organismal phylogeny.

LncRNA identification, conservation, and functional assessment

Differentially expressed lncRNAs were identified by first filtering out differentially expressed transcripts that displayed any similarity with known proteins or annotated protein domains (BLASTx, 1E-3). Transcripts were then filtered using Evolinc (64), filtering based on length, coding capacity (using CPC; (65)), and similarity to annotated proteins from the same species (using the set of differentially expressed transcripts predicted to be protein-coding). Sequence homologs for these lncRNAs were identified in the sunflower and lettuce genomes using CoGe BLAST, with an E-value of 1E-20 (26). Guayule transcripts for which a sequence homolog in another species corresponded to an annotated protein-coding gene were removed. These cleared transcripts were then scanned for miRNA motifs using psRNATarget (2017 update). Putative miRNA motifs were examined for conservation using multiple sequence alignments generated by MAFFT and visualized in Geneious.

Abbreviations

lncRNA = long non-coding RNA; USD = United States Dollars; miRNA = microRNA; bp = base pairs; mRNA = messenger RNA; nt = nucleotide; DE = differential expression; TPM = transcript per kilobase million; GO = gene ontology; PIP = Plasma membrane intrinsic protein; LHY = Late elongated hypocotyl; HID1; Hidden Treasure 1; rRNA = ribosomal RNA; phasiRNA = phased, secondary, small interfering RNAs.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

RNA-seq data have been uploaded to NCBI's SRA under the BioProject ID PRJNA400611.

Competing interests

None declared

Funding

Field trial and RNA sequencing was supported by the USDA Biomass Research and Development Initiative (BRDI) grant USDA-NIFA 201210006-19391 OH. Analyses were funded by USDA-NIFA Coordinated Agricultural Program, Sustainable Bioeconomy for Arid Regions (SBAR) 2017-68005-26867 and National Science Foundation grant IOS-1758532 to ADLN. DP was supported by the University of Arizona start-up funds and SBAR.

Author contributions

ADLN - led and oversaw transcriptomic analyses and prepared the manuscript; GP - sample collection, library preparation, and processing of RNAseq data; CM - conceived and designed experiments and collected samples; DCI - analysis of RNAseq data; NAP - bioinformatic support and preparation of manuscript; EDS and DJH - experimental design and management of the field trial and sample collection; and DP - overall project management, analysis of data, and preparation of manuscript. All authors contributed to revising and editing the manuscript.

Acknowledgements

The authors would like to thank Greg Leake (USDA-ARS, retired) for field harvest assistance; Naxin Hou (University of California, Davis) for laboratory assistance; and anonymous reviewers for helpful suggestions on the manuscript. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity employer.

References

1. Cornish K. Alternative Natural Rubber Crops: Why Should We Care? [Internet]. Vol. 18, Technology & Innovation. 2017. p. 244–55. Available from: http://dx.doi.org/10.21300/18.4.2017.245

2. Lieberei R. South American leaf blight of the rubber tree (Hevea spp.): new steps in plant domestication using physiological features and molecular markers. Ann Bot. 2007 Dec;100(6):1125–42.

3. Mooibroek H, Cornish K. Alternative sources of natural rubber. Appl Microbiol Biotechnol. 2000 Apr;53(4):355–65.

4. van Beilen JB, Poirier Y. Guayule and Russian dandelion as alternative sources of natural rubber. Crit Rev Biotechnol. 2007 Oct;27(4):217–31.

5. Rasutis D, Soratana K, McMahan C, Landis AE. A sustainability review of domestic rubber from the guayule plant [Internet]. Vol. 70, Industrial Crops and Products. 2015. p. 383–94. Available from: http://dx.doi.org/10.1016/j.indcrop.2015.03.042

6. Ray DT. Guayule: A source of natural rubber. New crops [Internet]. 1993;33843. Available from: http://hort.purdue.edu/newcrop/proceedings1993/V2-338.html

7. Thompson AE, Ray DT. Breeding guayule. Plant Breed Rev. 1989;6:93–165.

8. Siler DJ, Cornish K. Hypoallergenicity of guayule rubber particle proteins compared to Hevea latex proteins. Ind Crops Prod. 1994 Sep 1;2(4):307–13.

9. Estilai A, Ray DT. Genetics, cytogenetics, and breeding of guayule. Guayule Natural Rubber A Technical Publication with Emphasis on Recent Findings. 1991;47–92.

10. Gore MA, Coyle G, Friebe B, Coffelt TA, Salvucci ME. Complex Ploidy Level Variation in Guayule Breeding Programs. Crop Sci. 2011;51:210–6.

11. Ilut DC, Sanchez PL, Costich DE, Friebe B, Coffelt TA, Dyer JM, et al. Genomic diversity and phylogenetic relationships in the genus Parthenium (Asteraceae). Ind Crops Prod. 2015 Dec 15;76:920–9.

12. Ilut DC, Sanchez PL, Coffelt TA, Dyer JM, Jenks MA, Gore MA. A century of guayule: Comprehensive genetic characterization of the US national guayule (Parthenium argentatum A. Gray) germplasm collection. Ind Crops Prod. 2017 Dec 15;109:300–9.

13. Stonebloom SH, Scheller HV. Transcriptome analysis of rubber biosynthesis in guayule (Parthenium argentatum gray). BMC Plant Biol. 2019 Feb 12;19(1):71.

14. Allen SG, Nakayama FS, Dierig DA, Rasnick BA. Plant Water Relations, Photosynthesis, and Rubber Content of Young Guayule Plants during Water Stress 1. Agron J. 1987;79(6):1030–5.

15. Reddy AR, Ramachandra Reddy A, Rama Das VS. Enhanced Rubber Accumulation and Rubber Transferase Activity in Guayule Under Water Stress [Internet]. Vol. 133, Journal of Plant Physiology. 1988. p. 152–5. Available from: http://dx.doi.org/10.1016/s0176-1617(88)80130-5

16. Sundar D, Chaitanya KV, Jutur PP, Ramachandra Reddy A. Low temperature-induced changes in antioxidative metabolism in rubber-producing shrub, guayule (Parthenium argentatum Gray). Plant Growth Regul. 2004 Oct 1;44(2):175–81.

17. Veatch-Blohm ME, Ray DT, McCloskey WB. Water-Stress-Induced Changes in Resin and Rubber Concentration and Distribution in Greenhouse-Grown Guayule. Agron J. 2006;98:766–73.

18. Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K. Effects of abiotic stress on plants: a systems biology perspective [Internet]. Vol. 11, BMC Plant Biology. 2011. p. 163. Available from: http://dx.doi.org/10.1186/1471-2229-11-163

19. Danilevskaya ON, Yu G, Meng X, Xu J, Stephenson E, Estrada S, et al. Developmental and transcriptional responses of maize to drought stress under field conditions. Plant Direct. 2019 May 7;3(5):e00129.

20. Moschen S, Di Rienzo JA, Higgins J, Tohge T, Watanabe M, González S, et al. Integration of transcriptomic and metabolic data reveals hub transcription factors involved in drought stress response in sunflower (Helianthus annuus L.). Plant Mol Biol. 2017 Jul;94(4-5):549–64.

21. Hussain M, Farooq S, Hasan W, Ul-Allah S, Tanveer M, Farooq M, et al. Drought stress in sunflower: Physiological effects and its management through breeding and agronomic alternatives. Agric Water Manage. 2018 Mar;201:152–66.

22. Zhao J, He Q, Chen G, Wang L, Jin B. Regulation of Non-coding RNAs in Heat Stress Responses of Plants. Front Plant Sci. 2016 Aug 18;7:1213.

23. Kindgren P, Ard R, Ivanov M, Marquardt S. Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation. Nat Commun. 2018 Nov 1;9(1):4561.

24. Cagirici HB, Alptekin B, Budak H. RNA Sequencing and Co-expressed Long Non-coding RNA in Modern and Wild Wheats. Sci Rep. 2017 Sep 6;7(1):10670.

25. Bardou F, Ariel F, Simpson CG, Romero-Barrios N, Laporte P, Balzergue S, et al. Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. Dev Cell. 2014 Jul 28;30(2):166–76.

26. Nelson ADL, Forsythe ES, Devisetty UK, Clausen DS, Haug-Batzell AK, Meldrum AMR, et al. A Genomic Analysis of Factors Driving lincRNA Diversification: Lessons from Plants. G3 . 2016 Sep 8;6(9):2881–91.

27. Gálvez S, Mérida-García R, Camino C, Borrill P, Abrouk M, Ramírez-González RH, et al. Hotspots in the genomic architecture of field drought responses in wheat as breeding targets. Funct Integr Genomics. 2019 Mar;19(2):295–309.

28. Budak H, Hussain B, Khan Z, Ozturk NZ, Ullah N. From Genetics to Functional Genomics: Improvement in Drought Signaling and Tolerance in Wheat. Front Plant Sci. 2015 Nov 19;6:1012.

29. Hunsaker DJ, Elshikha DM, Bronson KF. High guayule rubber production with subsurface drip irrigation in the US desert Southwest. Agric Water Manage. 2019 Jul 1;220:1–12.

30. Eranki PL, El-Shikha D, Hunsaker DJ, Bronson KF, Landis AE. A comparative life cycle assessment of flood and drip irrigation for guayule rubber production using experimental field data [Internet]. Vol. 99, Industrial Crops and Products. 2017. p. 97–108. Available from: http://dx.doi.org/10.1016/j.indcrop.2017.01.020

31. Penna S. Building stress tolerance through over-producing trehalose in transgenic plants. Trends Plant Sci. 2003 Aug;8(8):355–7.

32. Li Y-J, Wang B, Dong R-R, Hou B-K. AtUGT76C2, an Arabidopsis cytokinin glycosyltransferase is involved in drought stress adaptation. Plant Sci. 2015 Jul;236:157–67.

33. Zheng Y, Liao C, Zhao S, Wang C, Guo Y. The Glycosyltransferase QUA1 Regulates Chloroplast-Associated Calcium Signaling During Salt and Drought Stress in Arabidopsis. Plant Cell Physiol. 2017 Feb 1;58(2):329–41.

34. Maurel C, Verdoucq L, Luu D-T, Santoni V. Plant aquaporins: membrane channels with multiple integrated functions. Annu Rev Plant Biol. 2008;59:595–624.

35. Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature. 2017 Jun 1;546(7656):148–52.

36. Li X, Ma D, Lu SX, Hu X, Huang R, Liang T, et al. Blue Light- and Low Temperature-Regulated COR27 and COR28 Play Roles in the Arabidopsis Circadian Clock. Plant Cell. 2016 Nov;28(11):2755–69.

37. Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikit S, Song C, et al. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. Nat Commun. 2017 Apr 12;8:14953.

38. Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nat Commun. 2018 Nov 29;9(1):5056.

39. Yan Q, Wu F, Yan Z, Li J, Ma T, Zhang Y, et al. Differential co-expression networks of long non-coding RNAs and mRNAs in Cleistogenes songorica under water stress and during recovery. BMC Plant Biol. 2019 Jan 11;19(1):23.

40. Pang J, Zhang X, Ma X, Zhao J. Spatio-Temporal Transcriptional Dynamics of Maize Long Non-Coding RNAs Responsive to Drought Stress. Genes [Internet]. 2019 Feb 13;10(2). Available from: http://dx.doi.org/10.3390/genes10020138

41. Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, et al. Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. Plant J. 2014 Dec;80(5):848–61.

42. Seo JS, Sun H-X, Park BS, Huang C-H, Yeh S-D, Jung C, et al. ELF18-INDUCED LONG-NONCODING RNA Associates with Mediator to Enhance Expression of Innate Immune Response Genes in Arabidopsis. Plant Cell. 2017 May;29(5):1024–38.

43. Bazin J, Romero N, Rigo R, Charon C, Blein T, Ariel F, et al. Nuclear Speckle RNA Binding Proteins Remodel Alternative Splicing and the Non-coding Arabidopsis Transcriptome to Regulate a Cross-Talk Between Auxin and Immune Responses. Front Plant Sci. 2018 Aug 21;9:1209.

44. Cho J, Paszkowski J. Regulation of rice root development by a retrotransposon acting as a microRNA sponge. Elife [Internet]. 2017 Aug 26;6. Available from: http://dx.doi.org/10.7554/eLife.30038

45. Wang Y, Fan X, Lin F, He G, Terzaghi W, Zhu D, et al. Arabidopsis noncoding RNA mediates control of photomorphogenesis by red light. Proc Natl Acad Sci U S A. 2014 Jul 15;111(28):10359–64.

46. Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release) [Internet]. Vol. 46, Nucleic Acids Research. 2018. p. W49–54. Available from: http://dx.doi.org/10.1093/nar/gky316

47. Zhang J, Zhang H, Srivastava AK, Pan Y, Bai J, Fang J, et al. Knockdown of Rice MicroRNA166 Confers Drought Resistance by Causing Leaf Rolling and Altering Stem Xylem Development. Plant Physiol. 2018 Mar;176(3):2082–94.

48. Thieme CJ, Rojas-Triana M, Stecyk E, Schudoma C, Zhang W, Yang L, et al. Endogenous Arabidopsis messenger RNAs transported to distant tissues. Nat Plants. 2015 Mar 23;1(4):15025.

49. Panchy N, Lehti-Shiu M, Shiu S-H. Evolution of Gene Duplication in Plants. Plant Physiol. 2016 Aug;171(4):2294–316.

50. Laudencia-Chingcuanco DL, Stamova BS, You FM, Lazo GR, Beckles DM, Anderson OD. Transcriptional profiling of wheat caryopsis development using cDNA microarrays [Internet]. Vol. 63, Plant Molecular Biology. 2007. p. 651–68. Available from: http://dx.doi.org/10.1007/s11103-006-9114-y

51. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012 Dec 1;28(23):3150–2.

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.

53. Srivastava A, Malik L, Sarkar H, Zakeri M, Soneson C, Love MI, et al. Alignment and mapping methodology influence transcript abundance estimation [Internet]. Available from: http://dx.doi.org/10.1101/657874

54. Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, et al. The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. PLoS Biol. 2016 Jan;14(1):e1002342.

55. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015 Dec 30;4:1521.

56. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

57. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001 Dec;25(4):402–8.

58. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008 Jun;36(10):3420–35.

59. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis. 2015 Aug;53(8):474–85.

60. Castillo AI, Nelson ADL, Haug-Baltzell AK, Lyons E. A tutorial of diverse genome analysis tools found in the CoGe web-platform using Plasmodium spp. as a model [Internet]. Vol. 2018, Database. 2018. Available from: http://dx.doi.org/10.1093/database/bay030

61. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013 Apr;30(4):772–80.

62. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012 Jun 15;28(12):1647–9.

63. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May 1;30(9):1312–3.

64. Nelson ADL, Devisetty UK, Palos K, Haug-Baltzell AK, Lyons E, Beilstein MA. Evolinc: A Tool for the Identification and Evolutionary Comparison of Long Intergenic Non-coding RNAs. Front Genet. 2017 May 9;8:52.

65. Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. 2017 Jul 3;45(W1):W12–6.

Additional Files

Additional file 1: Figure S1. Representative image of water-restricted guayule flowering relative to non-flowering control plants of the same age.

Additional file 2: Table S1. Mapping rate comparison between filtered and unfiltered guayule transcriptome.

Additional file 3: Table S2. Complete results of differential expression analysis.

Additional file 4: Figure S2. qRT-PCR validation of the most differentially expressed guayule drought responsive genes.

Additional file 5: Table S3. Complete list of GO results for differentially expressed transcripts.

Additional file 6: Figure S3. Gene tree of the guayule aquaporin PIP2A ortholog.

Additional file 7: Figure S4. Gene tree of the guayule gene GFTW010289191.1, a putative ortholog of Arabidopsis AT4G16740, a gene involved in terpene biosynthesis.

Additional file 8: Table S4. List of putative guayule lncRNAs organized by conservation within the Asteraceae and presence of miRNA binding motifs.

Figures
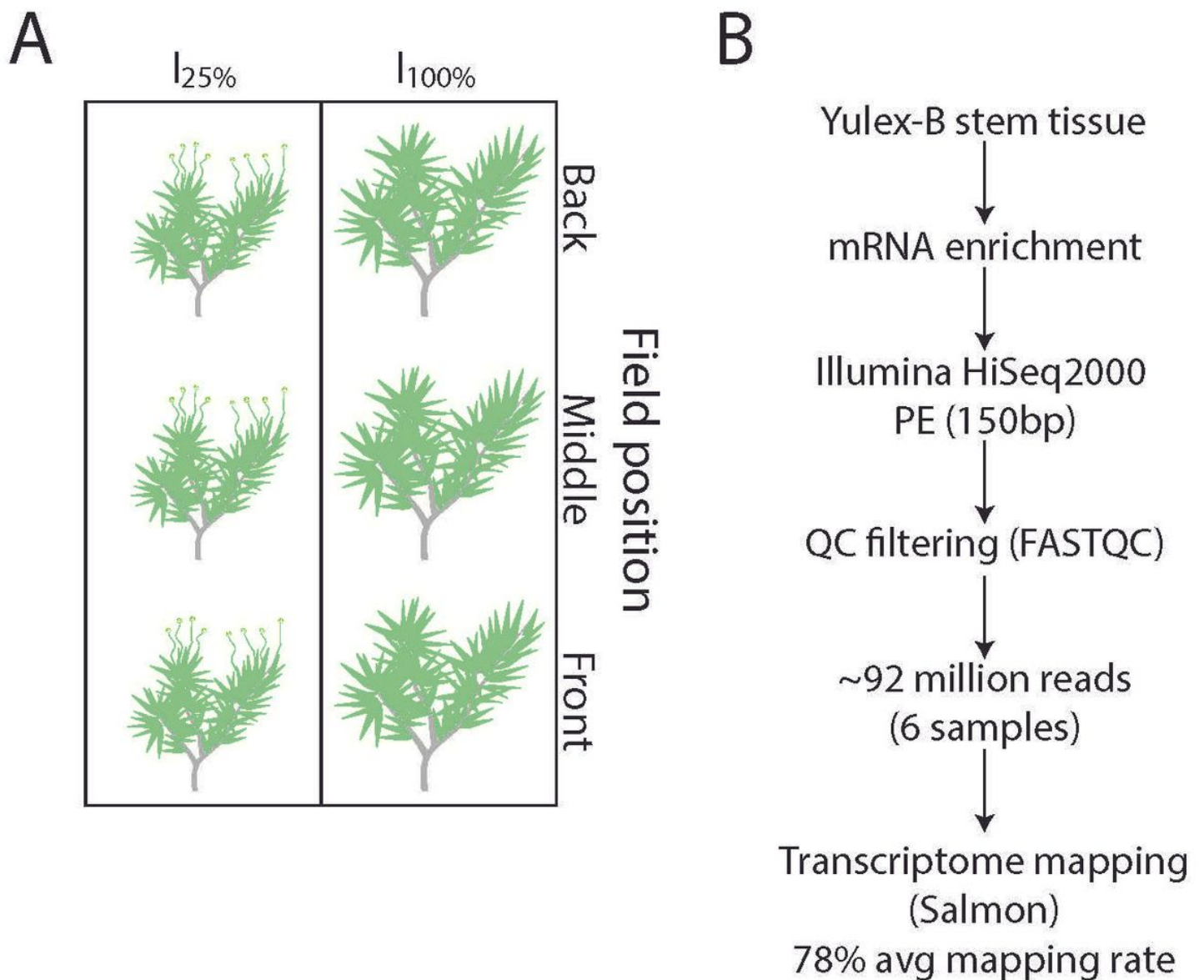
Figure 1

Irrigation and collection scheme for water-limited guayule. A) Schematic representation of irrigation and collection conditions of field grown guayule. Note that guayule grown under water-deficit conditions (25% of control, with control receiving sufficient irrigation to meet measured evaporative soil water losses) were flowering whereas control plants were not. B) Experimental design for transcriptomic profiling.
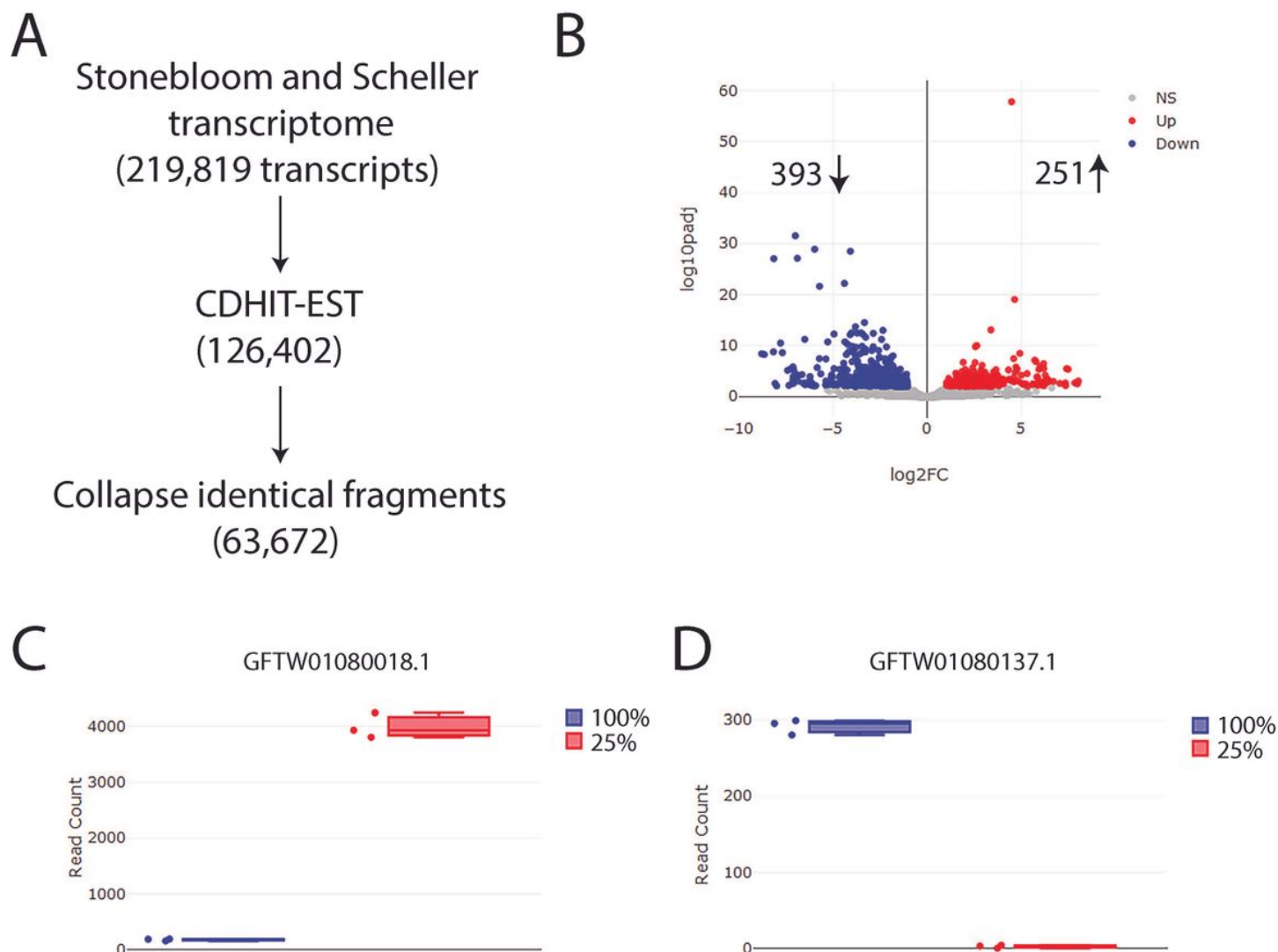
Figure 2

Transcriptomic comparison of plants grown under water-deficit conditions relative to control using a transcriptome-guided approach. A) Schematic describing the approach taken to filter the Stonebloom and Scheller (2019) de novo assembled transcriptome. B) Volcano plot representation of the transcripts differentially expressed under drought relative to control conditions. Log2 fold change (x-axis) is plotted relative to log10 adjusted p-value (y-axis). Transcripts upregulated under drought conditions and with an adjusted p-value < 0.01 are shown in red, whereas those downregulated are shown in blue. C) Box and whiskers expression profile, as denoted by the number of reads mapped to the transcript (read count, y-axis), for the transcript most upregulated under water-deficit conditions (red bar, I25%). The three dots next to each bar represent the three biological replicates for each condition. D) A similar expression profile for the transcript most down-regulated by water-deficit conditions.
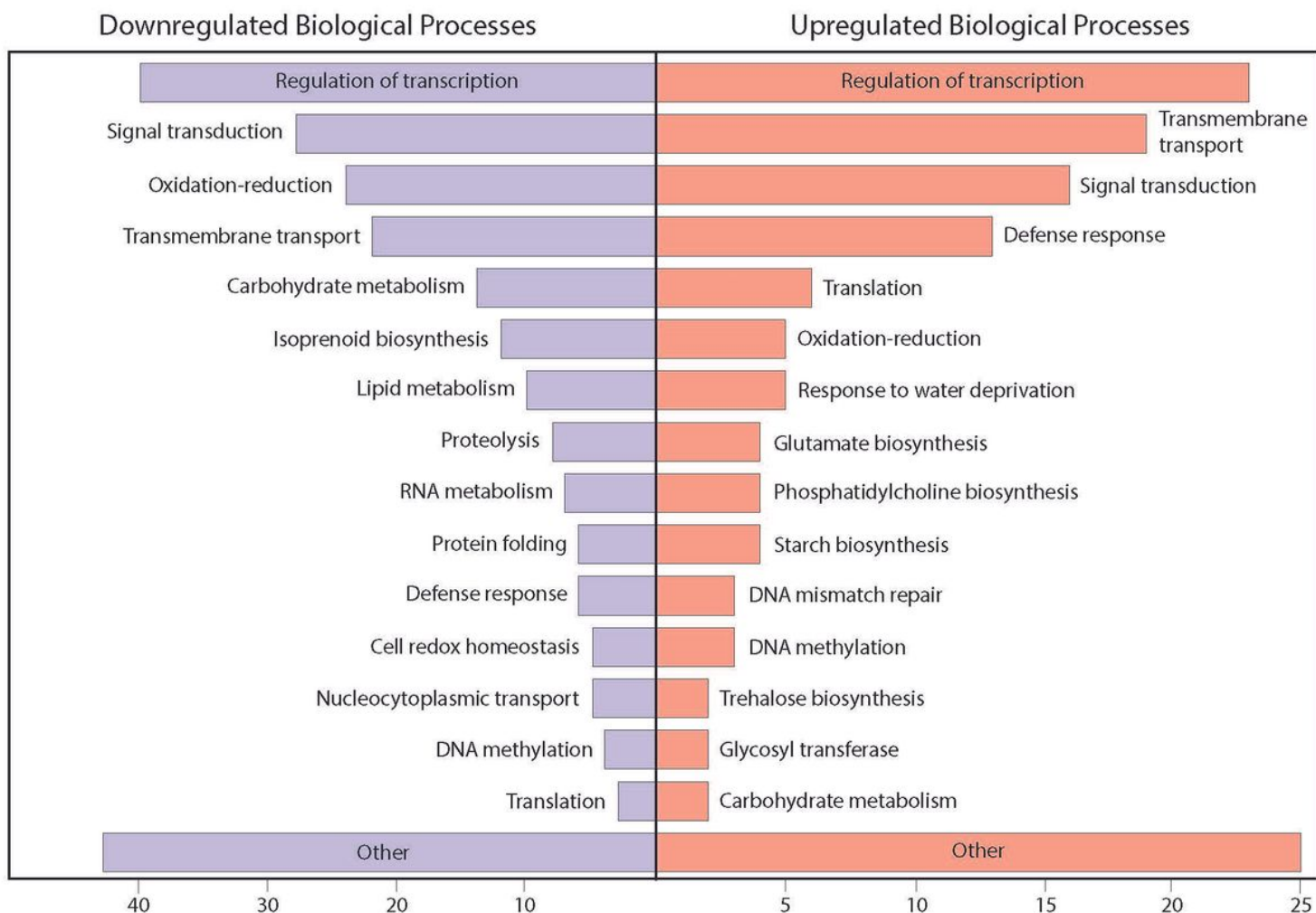
Figure 3

Functional analysis of differentially expressed transcripts. Biological processes inferred from gene ontological (GO) terms associated with either InterPro IDs or Arabidopsis orthologs were grouped into major categories. Note difference in scale of x-axis between down and up-regulated GO-terms.
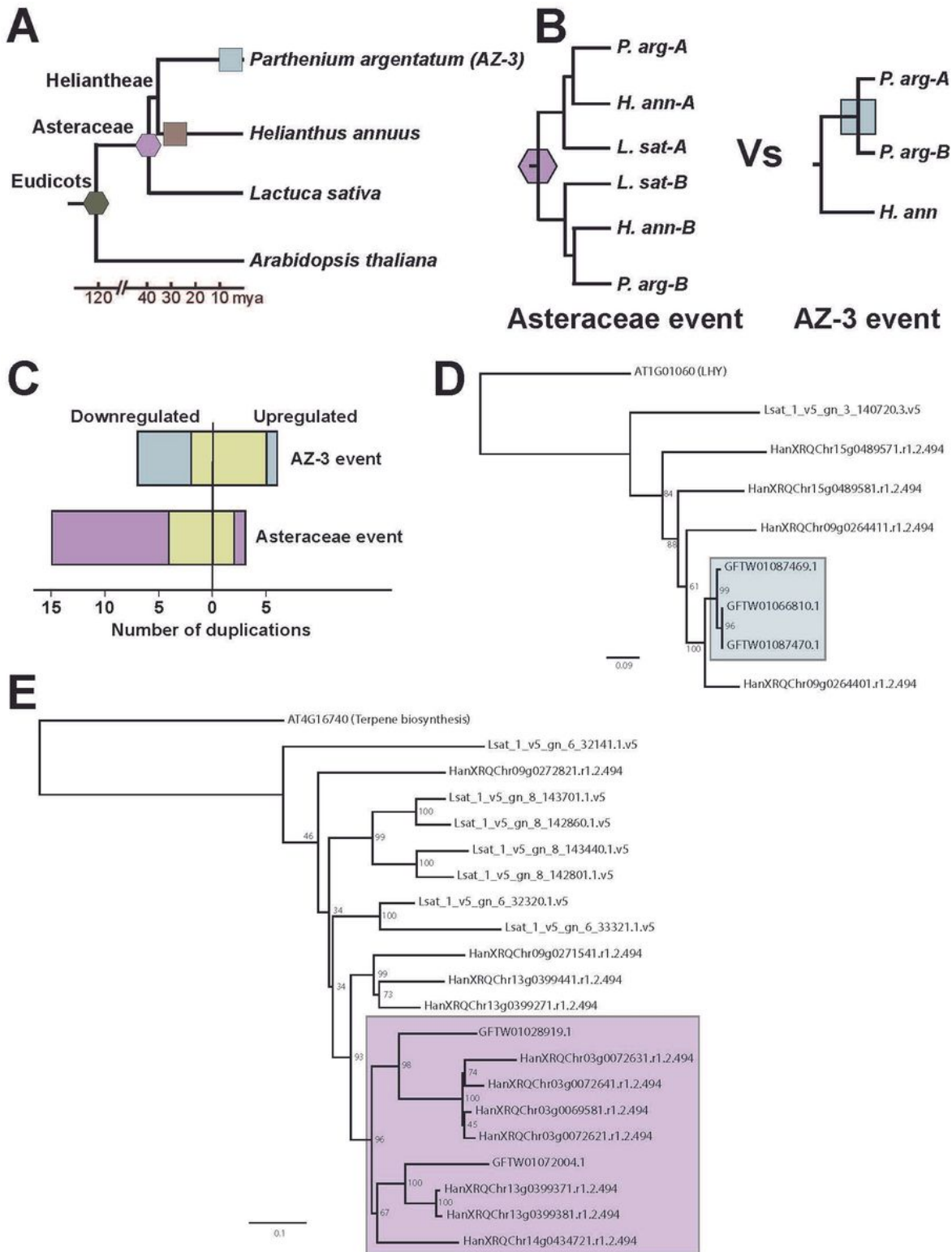
Figure 4

Phylogenetic inference of the timing of duplication for differentially expressed guayule transcripts. A) Chronogram of the four species used to build gene trees for this analysis. Placement of markers representing whole genome triplication (hexagons) and duplication (squares) indicates whether an event occurred in a common ancestor and is therefore shared (e.g. purple hexagon indicates a genome triplication event in the last common ancestor to all Asteraceae), or if it is species-specific (e.g., the light

blue square for guayule represents the duplication event in the accession examined in this study, AZ-3). B) The two phylogenetic models used to infer timing of the duplicated transcripts observed in guayule. Left, duplicated guayule transcripts, represented by P. arg-A and -B are sister to H. annuus paralogs, represented by H. ann-A and H. ann-B and thus likely originated from the whole genome triplication event at the base of the Asteraceae (purple hexagon). Right, guayule paralogs are sister to one another in the gene tree and then to a H. annuus ortholog, suggesting a guayule specific duplication event (light blue square). C) Bar plot indicating the number of differentially expressed guayule transcripts associated with each duplication event. Duplication events inferred to have arisen specifically in guayule (AZ-3) are shown in light blue, whereas those likely originating from the ancient Asteraceae hexaploidy event are shown in purple, using the same color scheme from A and B. Pseudogenization of one of the guayule paralogs is indicated by the tan bar. D) Gene tree representing an AZ-3 specific duplication event (blue box). E) Gene tree representing an Asteraceae event (purple box). In D and E, gene trees were rooted using the Arabidopsis ortholog.
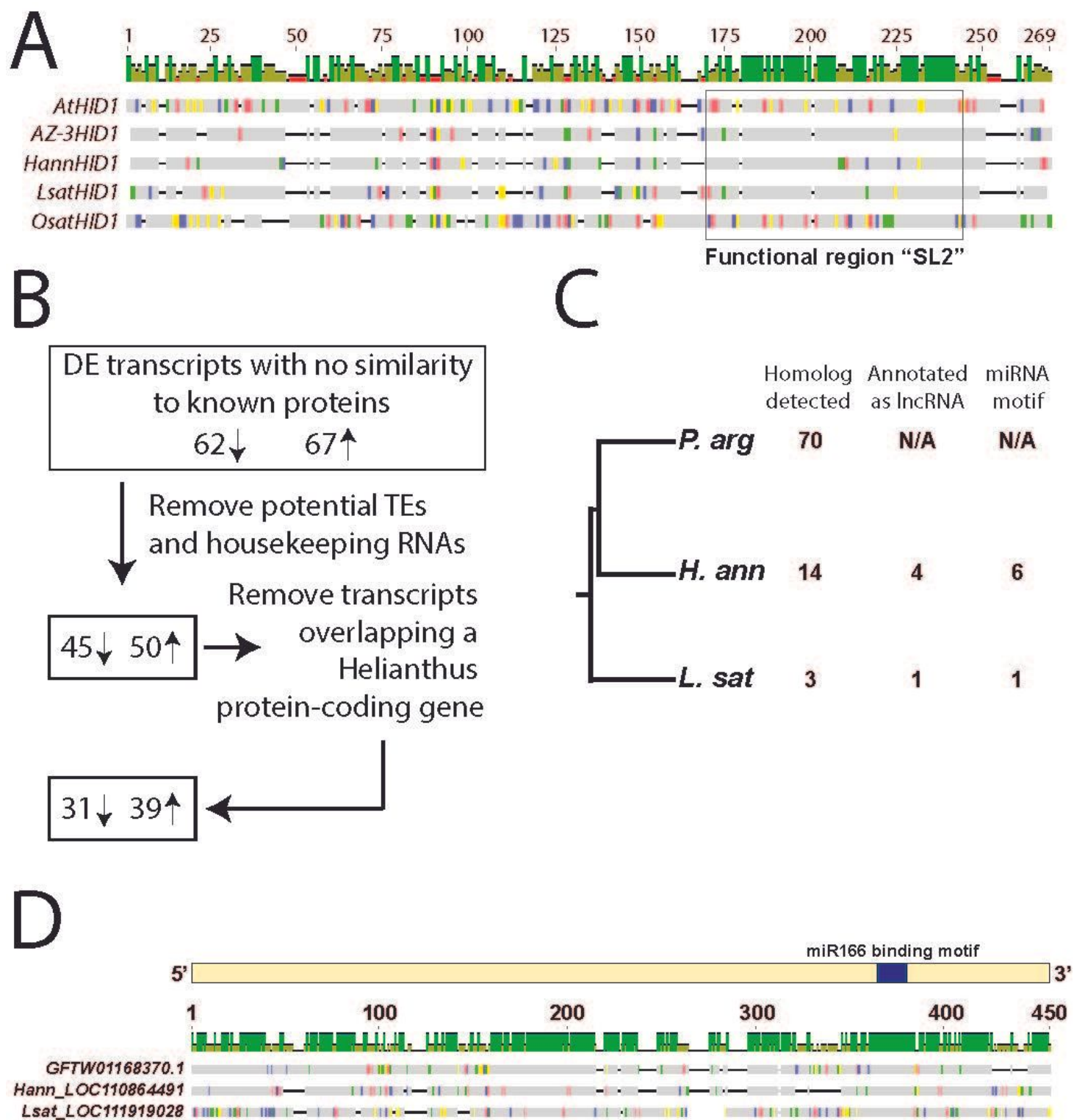
Figure 5

Identification and function inference of guayule stress-responsive lncRNAs. A) Graphical representation of a multiple sequence alignment (MSA) of guayule HID1, along with sequence homologs from Arabidopsis (AtHID1), sunflower (HannHID1), lettuce (LsatHID1), and rice (OsatHID1). 100% sequence identity between all sequences in the MSA are represented by green in the coverage bar across the top. B) Experimental design to identify putative guayule lncRNAs. "Known proteins" refers to proteins with

annotated domains or that are found in the InterPro database. TEs = transposable elements. C) Phylogenetic representation of the number of sequence homologs identified for the guayule lncRNAs. Number of lncRNA sequence homologs annotated as a lncRNA in either H. annuus or L. sativa is shown. Number of conserved guayule lncRNAs for which a miRNA binding motif is conserved is also indicated. D) Graphical representation of an MSA of the putative miRNA sponge, GFTW01168370.1, with the 100% conserved miRNA binding site shown by the blue box along the top of the alignment. The corresponding lncRNA IDs for sunflower and lettuce are shown in this alignment.