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

Griesel 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, 2016). 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 (Cornish 2017). 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 (Lieberei, 2007). Due to these growing concerns for the future stability of *Hevea* populations, scientists have continually searched for alternative sources of natural rubber (Mooibroek and Cornish, 2000; van Beilen and Poirier, 2007). 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 (Mooibroek and Cornish, 2000; Ray, 1993; van Beilen and Poirier, 2007(Rasutis et al. 2015)).

Guayule grows throughout northern Mexico and much of the American southwest and thus is naturally adapted to arid environments (Ray, 1993; Thompson, A.E. and Ray, 1989). Because of this, producers first considered guayule as an alternative source of natural rubber in the early 1900's (Ray, 1993; Thompson, A.E. and Ray, 1989). 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* (Huang, 1991; Ray, 1993). 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 (Siler and Cornish, 1994). Crop scientists are also 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 (Bergner, 1946; Estilai and Ray, 1991; Gore et al., 2011; Ilut et al., 2017, 2015; Powers and Rollins, 1945). 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 (Stonebloom and Scheller, 2019).

Rubber biosynthesis in guayule occurs primarily in the stem of the plant, wherein natural rubber accumulates in epithelial cells situated around resin ducts and eventually spreads to cells in the parenchyma (Backhaus and Walsh, 1983). The biological purpose for the production, accumulation, and subsequent storage of rubber in guayule is unclear (Backhaus, 1985). However, rubber biosynthesis and deposition appears to be dependent on abiotic stressors such as drought and temperature, which have been shown to have an effect on several guayule phenotypes including vegetative growth rates and antioxidant metabolism (Allen et al., 1987; Benedict et al., 1947; Benzioni et al., 1989; Downes and Tonnet, 1985; Nakayama and Bucks, 1984; Ramachandra Reddy and Rama Das, 1988; Sundar, D. and Ramachandra Reddy, 2000; Sundar et al., 2004; Veatch-Blohm et al., 2006). In addition, several studies have demonstrated a relationship between these abiotic factors and rubber yields; therefore, much interest lies in understanding the connection between rubber accumulation and environmental conditions (Ramachandra Reddy and Rama Das, 1988; Sundar, D. and Ramachandra Reddy, 2000; Veatch-Blohm et al., 2006 (Stonebloom and Scheller 2019)).

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 to 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.

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 Hunsaker et al., 2019. 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 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 (Fu et al. 2012) 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 (Altschul et al. 1990) 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 (Srivastava et al., n.d.) in CyVerse's Discovery Environment (Merchant et al. 2016). Quantified reads were prepared for differential expression analysis using the tximport

(Soneson, Love, and Robinson 2015) package in R. Differential expression was determined using DESeq2 (Love, Huber, and Anders 2014) with an adjusted p-value of 0.01 as the cutoff for significance.

Functional analysis

GO terms for the differentially expressed transcripts were identified using BLAST2GO (v5.2.5; Götz et al. 2008). 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 (Berardini et al. 2015).

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 (Castillo et al. 2018). Coding sequences were extracted from the top five unique loci in each genome using CoGeBLAST's view FASTA feature. Sequences were aligned using MAFFT (v7.017; Katoh and Standley 2013) in Geneious (v7.0.6; Kearse et al. 2012). 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, (Stamatakis 2014)) 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 (v1.7.5; Nelson et al. 2017), filtering based on length, coding capacity (using CPC; (v2.0; Kang et al. 2017)), 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 (Nelson et al., 2016). 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.

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 $I_{25\%}$ (Figure 1A; Eranki et al., 2017; Hunsaker et al., 2019). 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 (Supplemental Figure 1). Stem tissue, the predominant location of guayule rubber biosynthesis, was collected from three biological replicates in each irrigation regime for transcriptomic analysis. Following mRNA enrichment from these samples, paired-end Illumina RNA-sequencing was performed, generating ~92 million 150 bp reads across six samples, with an average mapping rate of 78% (Figure 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 (Stonebloom and Scheller, 2019) 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. Thus, the Stonebloom and Scheller transcriptome was filtered in two ways: 1) by passing the transcripts through CD-HIT-EST (Fu et al. 2012) to collapse identical transcripts, and 2) further collapsed transcripts that were 100% identical except for a 120-150 nt fragment (corresponding to one misaligned sequencing read; Figure 2A). These two steps collapsed 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; Supplemental Table 1), suggesting that the filtered transcriptome would be sufficient for differential expression (DE) analyses.

Differentially expressed genes were identified using the R package DESeq2 (Love, Huber, and Anders 2014), 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 (Figure 2B and Supplemental Table 2; adjusted p-value of 0.01). The transcript most significantly upregulated in the water-limited treatment, GFTW01080018.1 (Figure 2C), was expressed 23-fold compared to the control treatment. In contrast, the

transcript most significantly downregulated, GFTW01080137.1 (Figure 2D), was reduced more than 200-fold to near imperceptible detection levels.

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 (Supplemental Table 3). Transcription factors (regulation of transcription) were the most abundant class of both up and down-regulated transcripts (Figure 3). In agreement with previous data from drought-stressed plants, defense response, trehalose biosynthesis (Penna, 2003), glycosyltransferase activity (Li et al., 2015; 2017; Zheng et al., 2016), and response to water deficit were among the processes more likely to be upregulated under the water-limited irrigation treatment, whereas isoprenoid/terpenoid biosynthesis (Munne-Bosch, 2009; Loyola et al., 2011), carbohydrate metabolism, and lipid metabolism (Anh et al., 1985; Gigon et al., 2004) processes were more likely to be downregulated (Figure 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; Supplemental Figure 2), a family of aquaporins important for hydraulic regulation (Maurel et al. 2008). Despite the recovery of numerous PIP2 paralogs in the genomes of *Helianthus annuus* and *Lactuca sativa*, two close relatives of guayule within the Asteraceae (CITE; Supplemental Figure 2)), 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 (Li et al. 2016). 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 (Supplemental Figure 1). 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; Supplemental Figure 3) is a putative ortholog of Arabidopsis Terpene Synthase 3 (AT4G16740) and is one of 12 downregulated guayule transcripts involved in isoprenoid/terpenoid biosynthesis (Figure 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*; (Badouin et al. 2017)) and lettuce (*L. sativa*; (Reyes-Chin-Wo et al. 2017)) 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 (Figure 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 (Figure 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 (Figure 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 (Figure 4C, purple bar), 7 arose from an AZ-3 event (Figure 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 (Figure 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 (Figure 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 (Figure 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 (Figure 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 (Zhao et al. 2018; Yan et al. 2019; Pang et al. 2019; Di et al. 2014; Seo et al. 2017), where, among many functions, they can act as regulators of transcription, microRNA sponges, and influence alternative splicing (Bazin et al. 2018; Bardou et al. 2014; Cho and Paszkowski 2017). Although not differentially expressed under the imposed irrigation treatments, a homolog of the deeply conserved light responsive lncRNA, *HID1 (Wang et al. 2014)*, was present in the guayule transcriptome (Figure 5A). As expected based on prior analyses, the protein interaction domain annotated as SL2 was highly conserved between Asteraceae, Arabidopsis, and rice (Figure 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 (Figure 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 Supplemental Table 4).

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 (Nelson et al. 2016). Of the 70 guayule putative lncRNAs, we identified a sequence homolog for 14 in the sunflower genome (Figure 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 (Figure 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 (Dai, Zhuang, and Zhao 2018), 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 (Figure 5C; Supplemental Table 4). 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 (Figure 5D), a microRNA associated with tissue development and whose knockdown in Arabidopsis leads to an enhanced drought response (Zhang et al. 2018). 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 (Thieme et al. 2015), 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 (Hunsaker et al., 2019).

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 (Panchy, Lehti-Shiu, and Shiu 2016). 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 (Badouin et al. 2017). 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.

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

References

Allen, S.G., Nakayama, F.S., Dierig, D.A., Rasnick, B.A., 1987. Plant Water Relations, Photosynthesis, and Rubber Content of Young Guayule Plants during Water Stress1. Agron. J. 79, 1030. https://doi.org/10.2134/agronj1987.00021962007900060016x

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.

Anh, T. P. T., Borrel-Flood, C., da Silva, J. V., Justin, A. M., Mazliak, P., 1985. Effects of water stress on lipid metabolism in cotton leaves. *Phytochemistry*, *24*(4), 723-727.

Backhaus, R.A., 1985. Rubber Formation In Plants-A Mini-Review. Isr. J. Bot. 34, 283–293. https://doi.org/10.1080/0021213X.1985.10677027

Backhaus, R.A., Walsh, S., 1983. The Ontogeny of Rubber Formation in Guayule, Parthenium argentatum Gray. Bot. Gaz. 144, 391–400. https://doi.org/10.1086/337388

Badouin, Hélène, Jérôme Gouzy, Christopher J. Grassa, Florent Murat, S. Evan Staton, Ludovic Cottret, Christine Lelandais-Brière, et al. 2017. "The Sunflower Genome Provides Insights into Oil Metabolism, Flowering and Asterid Evolution." *Nature* 546 (7656): 148–52.

Bardou, Florian, Federico Ariel, Craig G. Simpson, Natali Romero-Barrios, Philippe Laporte, Sandrine Balzergue, John W. S. Brown, and Martin Crespi. 2014. "Long Noncoding RNA Modulates Alternative Splicing Regulators in Arabidopsis." *Developmental Cell* 30 (2): 166–76.

Bazin, Jérémie, Natali Romero, Richard Rigo, Celine Charon, Thomas Blein, Federico Ariel, and Martin Crespi. 2018. "Nuclear Speckle RNA Binding Proteins Remodel Alternative Splicing and the Non-Coding Arabidopsis Transcriptome to Regulate a Cross-Talk Between Auxin and Immune Responses." *Frontiers in Plant Science* 9 (August): 1209.

Benedict, H.M., McRary, W.L., Slattery, M.C., 1947. Response of Guayule to Alternating Periods of Low and High Moisture Stresses. Bot. Gaz. 108, 535–549. https://doi.org/10.1086/335441

Benzioni, A., Mills, D., Forti, M., 1989. Effect of Irrigation Regimes on the Water Status, Vegetative Growth and Rubber Production of Guayule Plants. Exp. Agric. 25, 189. https://doi.org/10.1017/S0014479700016690

Berardini, Tanya Z., Leonore Reiser, Donghui Li, Yarik Mezheritsky, Robert Muller, Emily Strait, and Eva Huala. 2015. "The Arabidopsis Information Resource: Making and Mining the 'Gold Standard' Annotated Reference Plant Genome." *Genesis* 53 (8): 474–85.

Bergner, A.D., 1946. Polyploidy and aneuploidy in guayule.

Castillo, Andreina I., Andrew D. L. Nelson, Asher K. Haug-Baltzell, and Eric Lyons. 2018. "A Tutorial of Diverse Genome Analysis Tools Found in the CoGe Web-Platform Using Plasmodium Spp. as a Model." *Database*. https://doi.org/10.1093/database/bay030.

Cho, Jungnam, and Jerzy Paszkowski. 2017. "Regulation of Rice Root Development by a Retrotransposon Acting as a microRNA Sponge." *eLife* 6 (August). https://doi.org/10.7554/eLife.30038.

Cornish, Katrina. 2017. "Alternative Natural Rubber Crops: Why Should We Care?" *Technology & Innovation*. https://doi.org/10.21300/18.4.2017.245.

Dai, Xinbin, Zhaohong Zhuang, and Patrick Xuechun Zhao. 2018. "psRNATarget: A Plant Small RNA Target Analysis Server (2017 Release)." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky316.

Di, Chao, Jiapei Yuan, Yue Wu, Jingrui Li, Huixin Lin, Long Hu, Ting Zhang, et al. 2014. "Characterization of Stress-Responsive lncRNAs in Arabidopsis Thaliana by Integrating Expression, Epigenetic and Structural Features." *The Plant Journal: For Cell and Molecular Biology* 80 (5): 848–61.

Downes, R., Tonnet, M., 1985. Effect of environmental conditions on growth and rubber production of guayule (Parthenium argentatum). Aust. J. Agric. Res. 36, 285. https://doi.org/10.1071/AR9850285

Estilai, A., Ray, D.T., 1991. Genetics, cytogenetics, and breeding of guayule, Guayule Natural Rubber. A Technical Publication with Emphasis on Recent Findings.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. "CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data." *Bioinformatics* 28 (23): 3150–52.

Gigon, A., Matos, A. R., Laffray, D., Zuily-Fodil, Y., Pham-Thi, A. T., 2004. Effect of drought stress on lipid metabolism in the leaves of Arabidopsis thaliana (ecotype Columbia). *Annals of botany*, *94*(3), 345-351.

Gore, M.A., Coyle, G., Friebe, B., Coffelt, T.A., Salvucci, M.E., 2011. Complex Ploidy Level Variation in Guayule Breeding Programs. Crop Sci. 51, 210. https://doi.org/10.2135/cropsci2010.05.0283

Götz, Stefan, Juan Miguel García-Gómez, Javier Terol, Tim D. Williams, Shivashankar H. Nagaraj, María José Nueda, Montserrat Robles, Manuel Talón, Joaquín Dopazo, and Ana Conesa. 2008. "High-Throughput Functional Annotation and Data Mining with the Blast2GO Suite." *Nucleic Acids Research* 36 (10): 3420–35.

Huang, H.T., 1991. Guayule natural rubber. University of Arizona, Tucson.

Hunsaker, D. J., D. M. Elshikha, and K. F. Bronson. 2019. "High Guayule Rubber Production with Subsurface Drip Irrigation in the US Desert Southwest." *Agricultural Water Management*. 220, 1-12. https://doi.org/10.1016/j.agwat.2019.04.016.

Ilut, D.C., Sanchez, P.L., Coffelt, T.A., Dyer, J.M., Jenks, M.A., Gore, M.A., 2017. A century of guayule: Comprehensive genetic characterization of the US national guayule (Parthenium argentatum A. Gray) germplasm collection. Ind. Crops Prod. 109, 300–309. https://doi.org/10.1016/J.INDCROP.2017.08.029

Ilut, D.C., Sanchez, P.L., Costich, D.E., Friebe, B., Coffelt, T.A., Dyer, J.M., Jenks, M.A., Gore, M.A., 2015. Genomic diversity and phylogenetic relationships in the genus Parthenium (Asteraceae). Ind. Crops Prod. 76, 920–929. https://doi.org/10.1016/J.INDCROP.2015.07.035

Kang, Yu-Jian, De-Chang Yang, Lei Kong, Mei Hou, Yu-Qi Meng, Liping Wei, and Ge Gao. 2017. "CPC2: A Fast and Accurate Coding Potential Calculator Based on Sequence Intrinsic Features." *Nucleic Acids Research* 45 (W1): W12–16.

Katoh, Kazutaka, and Daron M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30 (4): 772–80.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. 2012. "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." *Bioinformatics* 28 (12): 1647–49.

Li, Y. J., Wang, B., Dong, R. R., Hou, B. K., 2015. AtUGT76C2, an Arabidopsis cytokinin glycosyltransferase is involved in drought stress adaptation. *Plant Science*, *236*, 157-167.

Li, P., Li, Y. J., Zhang, F. J., Zhang, G. Z., Jiang, X. Y., Yu, H. M., Hou, B. K., 2017. The Arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. *The Plant Journal*, *89*(1), 85-103.

Li, Xu, Dingbang Ma, Sheen X. Lu, Xinyi Hu, Rongfeng Huang, Tong Liang, Tongda Xu, Elaine M. Tobin, and Hongtao Liu. 2016. "Blue Light- and Low Temperature-Regulated COR27 and COR28 Play Roles in the Arabidopsis Circadian Clock." *The Plant Cell* 28 (11): 2755–69.

Lieberei, R., 2007. South American Leaf Blight of the Rubber Tree (Hevea spp.): New Steps in Plant Domestication using Physiological Features and Molecular Markers. Ann. Bot. 100, 1125–1142. https://doi.org/10.1093/aob/mcm133

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.

Loyola, J., Verdugo, I., González, E., Casaretto, J. A., Ruiz-Lara, S., 2012. Plastidic isoprenoid biosynthesis in tomato: physiological and molecular analysis in genotypes resistant and sensitive to drought stress. Plant Biology, 14(1), 149-156.

Maurel, Christophe, Lionel Verdoucq, Doan-Trung Luu, and Véronique Santoni. 2008. "Plant Aquaporins: Membrane Channels with Multiple Integrated Functions." *Annual Review of Plant Biology* 59: 595–624.

Merchant, Nirav, Eric Lyons, Stephen Goff, Matthew Vaughn, Doreen Ware, David Micklos, and Parker Antin. 2016. "The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences." *PLoS Biology* 14 (1): e1002342.

Mooibroek, H., Cornish, K., 2000. Alternative sources of natural rubber. Appl. Microbiol. Biotechnol. 53, 355–365. https://doi.org/10.1007/s002530051627

Munné-Bosch, S., Falara, V., Pateraki, I., López-Carbonell, M., Cela, J., & Kanellis, A. K., 2009. Physiological and molecular responses of the isoprenoid biosynthetic pathway in a drought-resistant Mediterranean shrub, Cistus creticus exposed to water deficit. Journal of Plant Physiology, 166(2), 136-145.

Nakayama, F.S., Bucks, D.A., 1984. Crop Water Stress Index, Soil Water, and Rubber Yield Relations for the Guayule Plant1. Agron. J. 76, 791. https://doi.org/10.2134/agronj1984.00021962007600050019x

Nelson, Andrew D. L., Upendra K. Devisetty, Kyle Palos, Asher K. Haug-Baltzell, Eric Lyons, and Mark A. Beilstein. 2017. "Evolinc: A Tool for the Identification and Evolutionary Comparison of Long Intergenic Non-Coding RNAs." *Frontiers in Genetics* 8 (May): 52.

Nelson, Andrew D. L., Evan S. Forsythe, Upendra K. Devisetty, David S. Clausen, Asher K. Haug-Batzell, Ari M. R. Meldrum, Michael R. Frank, Eric Lyons, and Mark A. Beilstein. 2016. "A Genomic Analysis of Factors Driving lincRNA Diversification: Lessons from Plants." *G3* 6 (9): 2881–91.

Panchy, Nicholas, Melissa Lehti-Shiu, and Shin-Han Shiu. 2016. "Evolution of Gene Duplication in Plants." *Plant Physiology* 171 (4): 2294–2316.

Pang, Junling, Xia Zhang, Xuhui Ma, and Jun Zhao. 2019. "Spatio-Temporal Transcriptional Dynamics of Maize Long Non-Coding RNAs Responsive to Drought Stress." *Genes* 10 (2). https://doi.org/10.3390/genes10020138.

Penna, S., 2003. Building stress tolerance through over-producing trehalose in transgenic plants. *Trends in Plant Science*, *8*(8), 355-357.

Powers, L., Rollins, R.C., 1945. Reproduction and pollination studies on guayule Parthenium argentatum Gray and P. incanum H.B.K. Agron. J. 96–112.

Ramachandra Reddy, A., Rama Das, V.S., 1988. Enhanced Rubber Accumulation and Rubber Transferase Activity in Guayule Under Water Stress. J. Plant Physiol. 133, 152–155. https://doi.org/10.1016/S0176-1617(88)80130-5

Rasutis, Daina, Kullapa Soratana, Colleen McMahan, and Amy E. Landis. 2015. "A Sustainability Review of Domestic Rubber from the Guayule Plant." *Industrial Crops and Products*. https://doi.org/10.1016/j.indcrop.2015.03.042.

Ray, D.T., 1993. Guayule: A Source of Natural Rubber. Wiley, New York.

Reyes-Chin-Wo, Sebastian, Zhiwen Wang, Xinhua Yang, Alexander Kozik, Siwaret Arikit, Chi Song, Liangfeng Xia, et al. 2017. "Genome Assembly with in Vitro Proximity Ligation Data and Whole-Genome Triplication in Lettuce." *Nature Communications* 8 (April): 14953.

Seo, Jun Sung, Hai-Xi Sun, Bong Soo Park, Chung-Hao Huang, Shyi-Dong Yeh, Choonkyun Jung, and Nam-Hai Chua. 2017. "ELF18-INDUCED LONG-NONCODING RNA Associates with Mediator to Enhance Expression of Innate Immune Response Genes in Arabidopsis." *The Plant Cell* 29 (5): 1024–38.

Siler, D.J., Cornish, K., 1994. Hypoallergenicity of guayule rubber particle proteins compared to Hevea latex proteins. Ind. Crops Prod. 2, 307–313. https://doi.org/10.1016/0926-6690(94)90122-8

Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. 2015. "Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences." *F1000Research* 4 (December): 1521.

Srivastava, Avi, Laraib Malik, Hirak Sarkar, Mohsen Zakeri, Charlotte Soneson, Michael I. Love, Carl Kingsford, and Rob Patro. n.d. "Alignment and Mapping Methodology Influence Transcript Abundance Estimation." https://doi.org/10.1101/657874.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13.

Stonebloom, S.H., Scheller, H.V., 2019. Transcriptome analysis of rubber biosynthesis in guayule (Parthenium argentatum gray). BMC Plant Biol. 19, 71. https://doi.org/10.1186/s12870-019-1669-2

Sundar, D. and Ramachandra Reddy, A., 2000. Low night temperature-induced changes in photosynthesis and rubber accumulation in guayule (Parthenium argentatum Gray). Photosynthetica 38, 421–427.

Sundar, D., Chaitanya, K.V., Jutur, P.P., Ramachandra Reddy, A., 2004. Low temperature-induced changes in antioxidative metabolism in rubber-producing shrub, guayule (Parthenium argentatum Gray). Plant Growth Regul. 44, 175–181. https://doi.org/10.1007/s10725-004-4030-0

Thieme, Christoph J., Monica Rojas-Triana, Ewelina Stecyk, Christian Schudoma, Wenna Zhang, Lei Yang, Miguel Miñambres, et al. 2015. "Endogenous Arabidopsis Messenger RNAs Transported to Distant Tissues." *Nature Plants* 1 (4): 15025.

Thompson, A.E. and Ray, D.T., 1989. "Breeding Guayule," in: Plant Breeding Reviews. pp. 93–165.

van Beilen, J.B., Poirier, Y., 2007. Guayule and Russian Dandelion as Alternative Sources of Natural Rubber. Crit. Rev. Biotechnol. 27, 217–231. https://doi.org/10.1080/07388550701775927

Veatch-Blohm, M.E., Ray, D.T., McCloskey, W.B., 2006. Water-Stress-Induced Changes in Resin and Rubber Concentration and Distribution in Greenhouse-Grown Guayule. Agron. J. 98, 766. https://doi.org/10.2134/agronj2005.0203

Wang, Yuqiu, Xiuduo Fan, Fang Lin, Guangming He, William Terzaghi, Danmeng Zhu, and Xing Wang Deng. 2014. "Arabidopsis Noncoding RNA Mediates Control of Photomorphogenesis by Red Light." *Proceedings of the National Academy of Sciences of the United States of America* 111 (28): 10359–64.

Yan, Qi, Fan Wu, Zhuanzhuan Yan, Jie Li, Tiantian Ma, Yufei Zhang, Yufeng Zhao, Yanrong Wang, and Jiyu Zhang. 2019. "Differential Co-Expression Networks of Long Non-Coding RNAs and mRNAs in Cleistogenes Songorica under Water Stress and during Recovery." *BMC Plant Biology* 19 (1): 23.

Zhang, Jinshan, Hui Zhang, Ashish Kumar Srivastava, Yujie Pan, Jinjuan Bai, Jingjing Fang, Huazhong Shi, and Jian-Kang Zhu. 2018. "Knockdown of Rice MicroRNA166 Confers Drought Resistance by Causing Leaf Rolling and Altering Stem Xylem Development." *Plant Physiology* 176 (3): 2082–94.

Zhao, Xinyue, Jingrui Li, Bi Lian, Hanqing Gu, Yan Li, and Yijun Qi. 2018. "Global Identification of Arabidopsis lncRNAs Reveals the Regulation of MAF4 by a Natural Antisense RNA." *Nature Communications* 9 (1): 5056.

Zheng, Y., Liao, C., Zhao, S., Wang, C., & Guo, Y., 2016. The glycosyltransferase QUA1 regulates chloroplast-associated calcium signaling during salt and drought stress in Arabidopsis. *Plant and Cell Physiology*, *58*(2), 329-341.

Figures
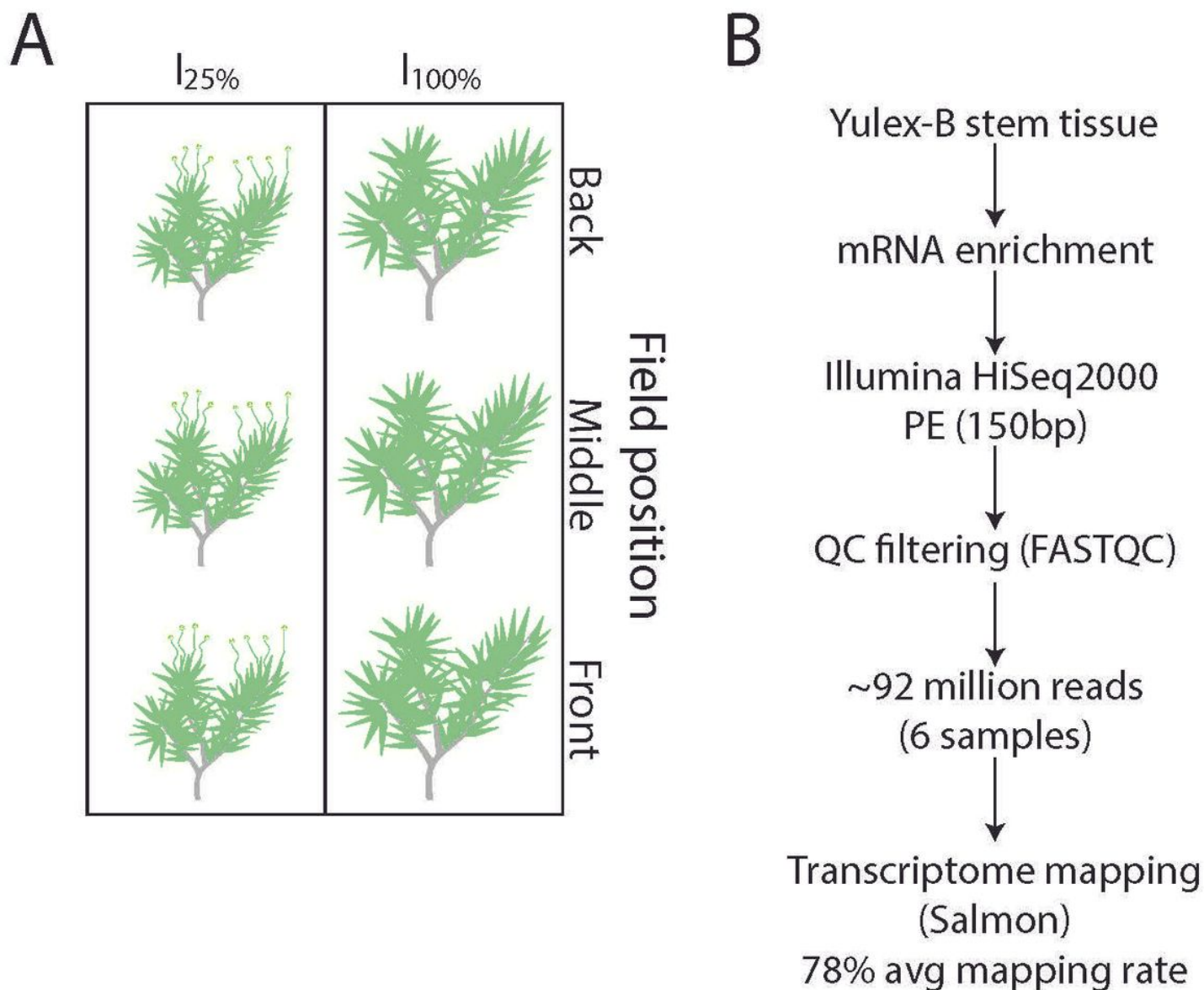
Figure 1

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.
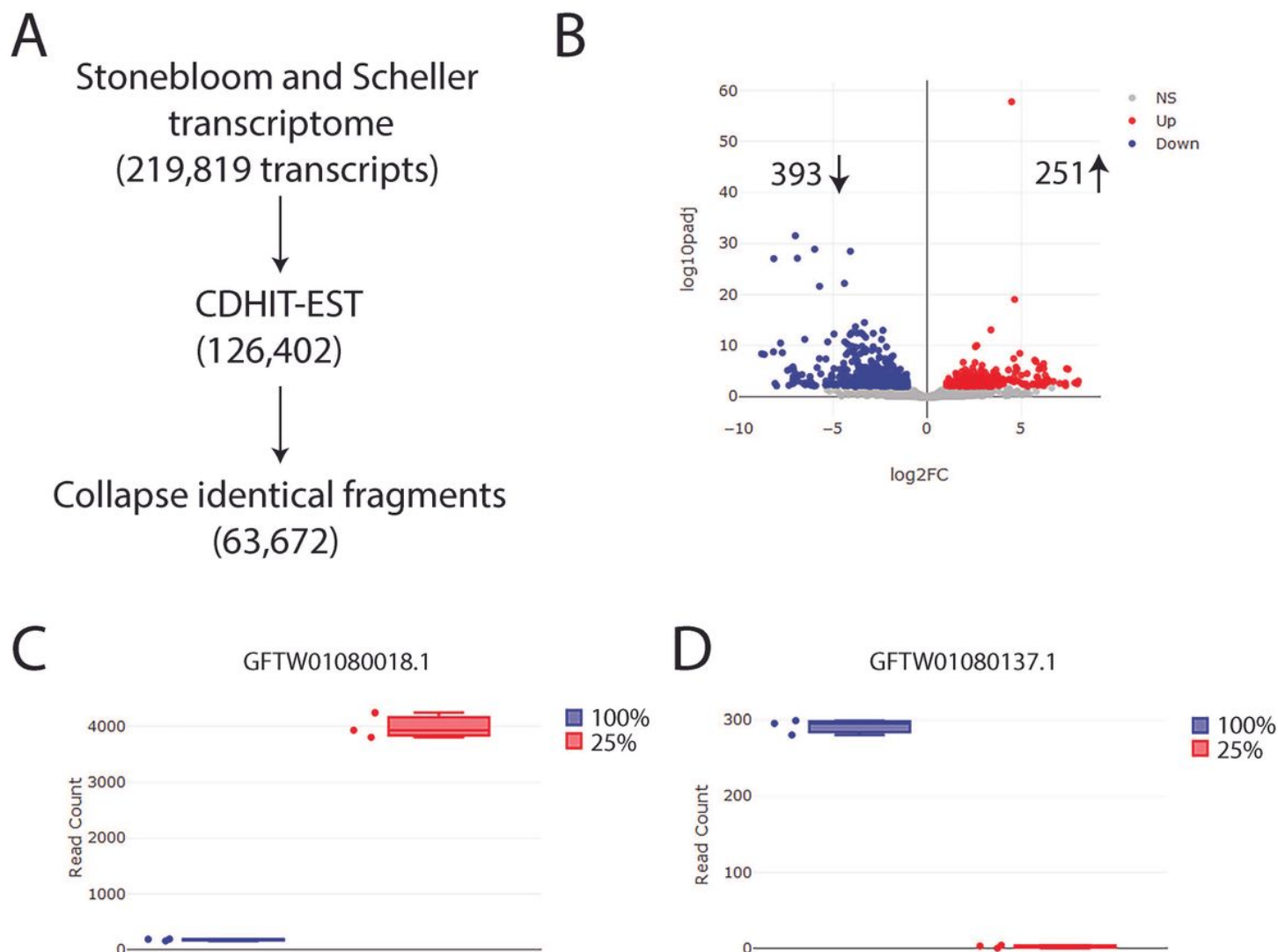
Figure 2

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.
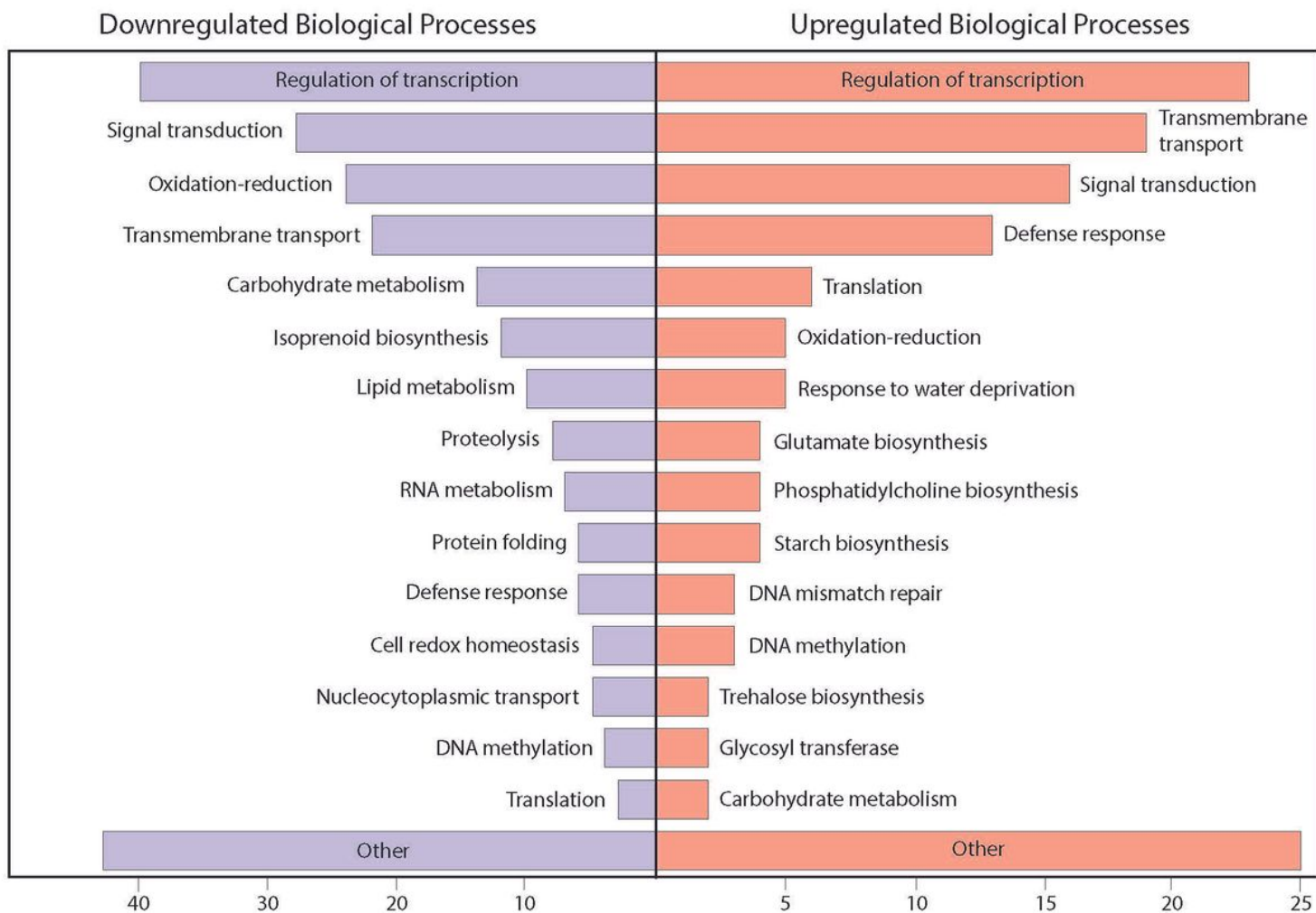
Figure 3

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.
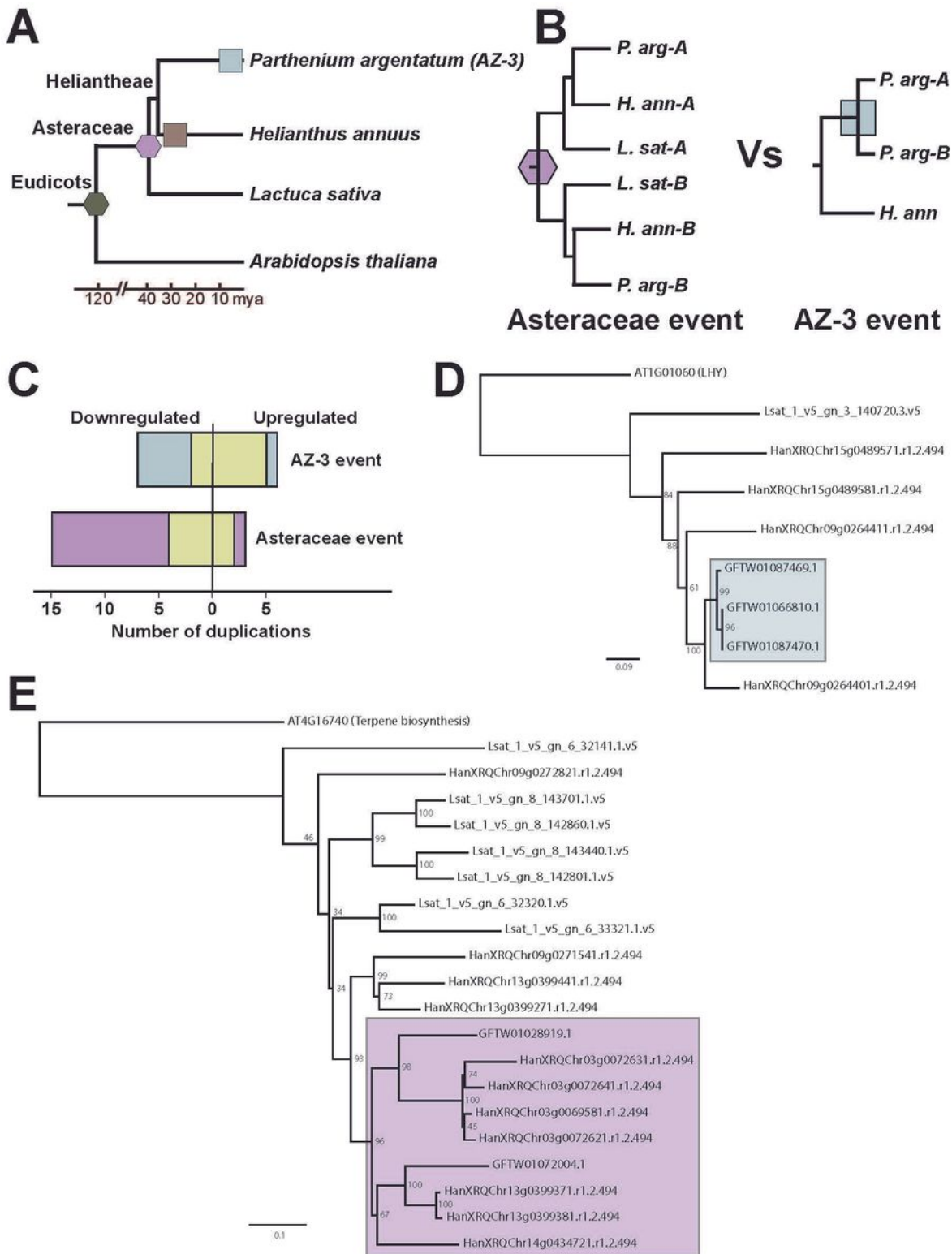
Figure 4

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.
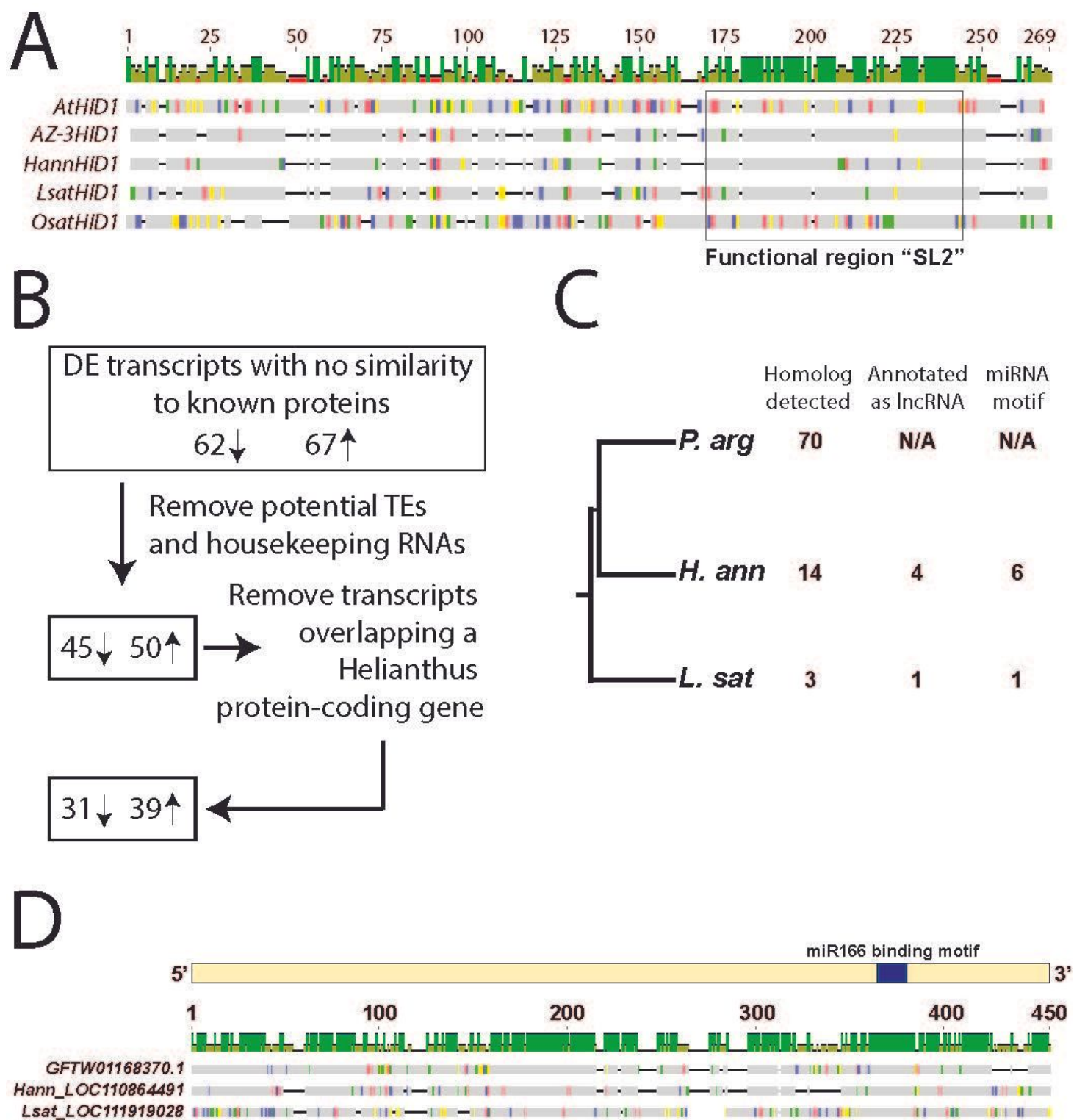
Figure 5

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.