Dynamically expressed small RNAs, substantially driven by genomic structural variants, contribute to transcriptomic changes during tomato domestication

You Qing[1,2], Yi Zheng[1,2,3,*], Sizolwenkosi Mlotshwa[4,10], Heather N. Smith[5,11], Xin Wang[3], Xuyang Zhai[1,2], Esther van der Knaap[6,7,8], Ying Wang[4,5,*], Zhangjun Fei[3,9,*]

[1]Beijing Key Laboratory for Agricultural Application and New Technique, College of Plant Science and Technology, Beijing University of Agriculture, Beijing, China 102206

[2]Bioinformatics Center, Beijing University of Agriculture, Beijing, China 102206

[3]Boyce Thompson Institute, Cornell University, Ithaca, NY 14853

[4]Department of Molecular Genetics, Ohio State University, Columbus, OH 43210

[5]Department of Biological Sciences, Mississippi State University, Starkville, MS 39759

[6]Center for Applied Genetic Technologies, University of Georgia, Athens, GA, 30602

[7]Institute for Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, 30602

[8]Department of Horticulture, University of Georgia, Athens, GA, 30602

[9]USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, 14853

[10]Present address: Department of Plant Pathology, Ohio State University, Wooster, OH 44691

[11]Present address: Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803

[*]Correspondence: Yi Zheng (yz@moilab.net), Ying Wang (wang@biology.msstate.edu), Zhangjun Fei (zf25@cornell.edu)

Running title: small RNA dynamics during tomato domestication

Summary

Tomato has undergone extensive selections during domestication. Recent progress has shown that genomic structural variants (SVs) have contributed to gene expression dynamics during tomato domestication, resulting in changes of important traits. Here, we performed comprehensive analyses of small RNAs (sRNAs) from nine representative tomato accessions. We demonstrate that SVs substantially contribute to the dynamic expression of the three major classes of plant sRNAs: microRNAs (miRNAs), phased secondary short interfering RNAs (phasiRNAs), and 24-nucleotide heterochromatic siRNAs (hc-siRNAs). Changes in the abundance of phasiRNAs and 24-nucleotide hc-siRNAs likely contribute to the alteration of mRNA gene expression in *cis* during tomato domestication, particularly for genes associated with biotic and abiotic stress tolerance. We also observe that miRNA expression dynamics are associated with imprecise processing, alternative miRNA-miRNA* selections, and SVs. SVs mainly affect the expression of less-conserved miRNAs that do not have established regulatory functions or low abundant members in highly expressed miRNA families. Our data highlight different selection pressures on miRNAs compared to phasiRNAs and 24-nucleotide hc-siRNAs. Our findings provide insights into plant sRNA evolution as well as SV-based gene regulation during crop domestication. Furthermore, our dataset provides a rich resource for mining the sRNA regulatory network in tomato.

Significance statement

The change of epigenetic regulation during crop domestication, particularly in regards to the dynamic expression and function of small RNAs, has rarely been characterized in a systematic manner. Here we report the comprehensive comparative investigation of small RNA expression dynamics during tomato domestication, which reveals differential selection pressures over distinct classes of small RNAs and SVs as a driving force for the dynamic expression of sRNAs in the history of tomato breeding.

Key words: tomato, *Solanum lycopersicum*, *Solanum pimpinellifolium*, domestication, genomic structural variant, miRNA, phasiRNA, 24-nt hc-siRNA

Introduction

Tomato is the world leading fruit crop in terms of total production and market value (http://www.fao.org/faostat). Originally domesticated in Northern Ecuador and Peru, tomato underwent further selections in Central America and Mexico prior to its arrival in Europe in the early 16th century (Blanca *et al.*, 2012; Blanca *et al.*, 2015; Razifard *et al.*, 2020). Along the way, selections had been made for larger fruit, enhanced flavor, and improved resistance to biotic and abiotic stresses (Zhu *et al.*, 2018). These phenotypic changes reflect the alterations in gene sequences and expression.

Genomic structural variants (SVs) are strongly associated with selection pressures over the course of tomato domestication that impact the expression of genes underlying certain agronomic traits (Alonge *et al.*, 2020; Wang *et al.*, 2020). SVs include insertions, deletions, duplications, inversions and translocations, which can be attributed to DNA recombination-based processes, replication-based processes, and transposon activities (Sanseverino *et al.*, 2015; Carvalho and Lupski, 2016). Many SVs serve as the causative genetic variants for diverse crop traits that have been selected during domestication (Alonge *et al.* 2020; Wang *et al.*, 2020). For example, the decrease of fruit lycopene levels is strongly associated with deletions in the promoters of multiple key genes involved in lycopene biogenesis in modern tomato (Wang *et al.*, 2020). However, the molecular basis underlying the link between genome-wide SVs and gene expression often remains elusive.

Small RNA (sRNA)-mediated gene silencing acts as a key mechanism in regulating gene expression in most eukaryotic organisms. In plants, there are three major groups of sRNAs: microRNAs (miRNAs), phased secondary short interfering RNAs (phasiRNAs), and 24-nucleotide heterochromatic siRNAs (hc-siRNAs) (Axtell, 2013a; Borges and Martienssen, 2015). MiRNAs, phasiRNAs and hc-siRNAs are generated by Dicer-like enzymes (DCLs): DCL1, DCL4, and DCL3, respectively (Vazquez, 2006). After production, they are loaded into the RNA-induced silencing complex for function (Baulcombe, 2004). In plants, miRNAs and phasiRNAs with a length of 21 or 22 nucleotides mainly guide cleavage of target mRNAs. By contrast, 24-nucleotide hc-siRNAs, abundantly derived from transposable elements, play a major role in RNA-directed DNA methylation to confer epigenetic regulation over gene expression (McCue and Slotkin, 2012; Axtell, 2013a; Lisch, 2013). It is known that sRNA-based regulation relies on the abundance of

the sRNAs (Bartel, 2009); therefore, sRNA abundance has a significant impact on their regulatory functions.

In plants, miRNAs and phasiRNAs among distinct species are evolutionarily fluid (Chavez Montes *et al.*, 2014; Zheng *et al.*, 2015). Comparative studies on miRNA gene evolution in *Arabidopsis lyrata* and *A. thaliana* that diverged more than 10 million years ago have discovered numerous less conserved miRNA genes exhibiting high divergence in hairpin structures, processing fidelity, and target complementarity (Ma *et al.*, 2010; Cuperus *et al.*, 2011). With the increasing number of analyses on sRNA sequencing (sRNA-Seq) data, more and more less conserved miRNAs and phasiRNA-generating loci (PHASs) have been uncovered in diverse species, from green algae to flowering plants (Chavez Montes *et al.*, 2014; Zheng *et al.*, 2015; Lunardon *et al.*, 2020; Baldrich *et al.*, 2021; Chen *et al.*, 2021; Guo *et al.*, 2022). However, whether and how the expression patterns and functions of sRNAs have been changed in shorter evolutionary times such as during crop domestication is unknown. Tomato evolved from a wild red-fruited progenitor species, *Solanum pimpinellifolium* (SP) approximately 80 thousand years ago into *Solanum lycopersicum* var. *cerasiforme* (SLC) (Razifard *et al.*, 2020; Blanca *et al.*, 2021). Semi-domesticated SLC further evolved into the fully domesticated tomato, *Solanum lycopersicum* var. *lycopersicum* (SLL). The evolution within the red-fruited tomato clade provides a unique system for studying selection and functional divergence of plant sRNAs during crop domestication as a result of its extensive genetic and genomic resources.

Here, we present comprehensive sRNA profiles of cultivated tomatoes (six accessions of SLC and one accession of SLL) and their wild progenitors (two accessions of SP) and the discovery that genomic SVs can substantially influence sRNA expression dynamics. Our analyses show that SVs are an important driving force for the dynamic expression of sRNAs. Moreover, we show that SVs can change the hc-siRNA hotspots in promoters of nearly 100 protein-coding genes, thereby altering their expression. These genes are mostly associated with responses to biotic and abiotic stresses. SVs are also correlated with many rapid birth and death of PHASs that are overwhelmingly related to disease resistance traits. SVs overlapping with miRNA genes can determine the gain or loss of certain less-conserved miRNA genes or affect the expression of miRNAs. Interestingly, the differential expression of miRNAs has a neglectable effect on transcriptomes, in contrast to the changes in mRNA expression correlated with the dynamics of hc-siRNA hotspots. Our findings unravel SV-related differential expression of 24-nucleotide hc-

siRNAs regulating the expression of certain genes associated with responses to biotic and abiotic stresses as well as differential selection pressures over distinct classes of sRNAs during tomato domestication. Our dataset is also valuable to promote other sRNA-related functional studies on tomato development and domestication.

Results

Comprehensive sRNA profiles during tomato domestication

Resulting from domestication, tomato underwent substantial changes in plant morphology, yield, fruit flavor, and adaptation to adverse environments, reflecting certain levels of adjustments in both genomes and transcriptomes. We selected nine accessions, from wild ancestors (SP) to semi-domesticated populations (SLC) to domesticated tomatoes (SLL), for comprehensive transcriptome analyses, representing samples from regions along the domestication and cultivation history (Razifard *et al.*, 2020). These accessions included two SP (BGV006370 collected in Peru and BGV007151 collected in Ecuador), six SLC (BGV005895, BGV007023 and PI 129026 collected in Ecuador; BGV007990 and BGV008189 collected in Peru; BGV008219 collected in Costa Rica), and one SLL (BGV007863 collected in Mexico) (Blanca *et al.*, 2015; Razifard *et al.*, 2020).

For each accession, transcriptome profiles including both sRNA and mRNA profiles were investigated in young leaves, anthesis-stage flowers, and fruits at four different developmental stages (young green, mature green, breaker, and red ripe). The four fruit developmental stages have previously been used to analyze the gene regulatory networks in a single accession of domesticated tomato (Karlova *et al.*, 2013; Zhong *et al.*, 2013). Therefore, our comprehensive dataset would empower detailed dissection on the gene regulatory networks underlying the fruit ripening process in addition to the transcriptome profiles of leaf and flower.

To ensure sampling at comparable developmental stages of different tomato accessions, we documented the timing of fruit developments by tracking 10-20 fruits from three to five plants of each accession. As shown in **Figure S1**, the two SP accessions exhibited slightly early ripening, whereas SLL ripened 1 week later. Ripening time in SLC accession varied from as early as SP to as late as SLL. We collected samples at the chosen time points and constructed and sequenced a total of 162 sRNA-Seq libraries (**Table S1**). sRNA data of the three replicates from young green fruits of BGV007023 did not pass the quality check, so only the data from the remaining 159

libraries were used for the subsequent analyses. Principal component analysis showed that biological replicates were tightly clustered and samples at the same developmental stages were clustered together (**Figure S2a**), indicating the high reproducibility and reliability of our data. In addition, sRNA expression profiles in SLC and SLL accessions were more closely related in comparison to sRNA profiles in SP accessions (**Figure S2b**). Analysis on sRNA size distribution showed that 24-nt siRNAs were the most dominant (**Figure S2c**). Interestingly, 24-nt siRNAs were slightly more abundant in leaf and young green fruit samples than in other samples (**Figure S2c**). To improve the sRNA mapping accuracy, we exploited two high-quality reference genomes (genomes of the domesticated Heinz 1706 and an SP accession LA2093) as detailed in a recent study (Wang *et al.*, 2020). It is noteworthy that, as expected, reads from SP samples were mapped to the LA2093 genome with a slightly higher rate, whereas reads from SLC and SLL samples were mapped to the Heinz SL4.0 with a slightly higher rate (**Figure S2d**). RNA-Seq libraries were also generated from the same samples and described in our previous study (Wang *et al.*, 2020; Pereira *et al.*, 2021).

Hc-siRNA hotspots and structural variants

The 24-nt hc-siRNAs play a major role in RNA-directed DNA methylation, a fundamental mechanism in epigenetic regulation (McCue and Slotkin, 2012; Axtell, 2013a; Borges and Martienssen, 2015). We reasoned that comparative analyses on hc-siRNA accumulation patterns may provide insights into the dynamic epigenetic changes underlying trait-related gene expression. When analyzing the global hc-siRNA abundance across the 12 tomato chromosomes, we noted that hc-siRNA abundance displayed a strong genome-wide correlation with the density of SVs between SP LA2093 and SLL Heinz 1706 reported in our previous study (Wang *et al.*, 2020) (**Figure 1a**). We then performed pairwise comparisons to identify differentially expressed (DE) hc-siRNA regions (fold change ≥ 2 and adjusted $P < 0.05$). For each tissue or development stage, we performed pairwise comparisons between SP and SLL accessions (two comparisons), between SP and SLC accessions (12 comparisons, except 10 for the young green fruit stage), and SLC and SLL accessions (six comparisons, except five for young green fruit stage), therefore comprising a total of 117 pairwise comparisons for all six tissues/developmental stages. We found that, for hc-siRNA regions mapped to SVs, DE hc-siRNA regions presented a substantially higher proportion (61.4%) than non-DE hc-siRNA regions (38.6%), whereas, for hc-siRNA regions mapped to gene

bodies or the whole genome, similar proportions were observed for DE and non-DE hc-siRNA regions (52.8 and 47.2% for gene bodies; 52.6 and 47.4% for the whole genome) (**Figure 1b**). This observation infers a model that epigenetic regulation may be substantially influenced by SVs during crop domestication, which could lead to large scale gene expression changes.

To obtain additional evidence in support of this model, we analyzed SV-overlapping hc-siRNA hotspots in promoter regions of protein-coding genes and identified hc-siRNA and protein-coding gene pairs for which the abundances showed negative correlations in our sRNA-Seq and RNA-Seq data in pairwise comparisons. In total, 1,386 protein-coding genes affected by the SV-overlapping hc-siRNA hotspots were identified in at least one comparison. We noted that most of these genes under the control of this epigenetic regulation were expressed in leaf and flower (**Figure 1c**). Gene ontology (GO) term analysis showed that the majority of the protein-coding genes affected by these domestication-associated epigenetic changes were related to pathways in response to biotic and abiotic stresses as well as several developmentally related processes (**Figure 1d and 1e**), implying a selection pressure favoring expression changes in genes related to environmental adaptation including selection for domestication traits. To obtain highly confident negative correlations between differentially accumulated 24-nt hc-siRNAs and the cognate protein-coding genes, we focused on the pairs of hc-siRNAs and the negatively correlated protein-coding genes repeated in at least ten pairwise comparisons, which resulted in the identification of 99 protein-coding genes (**Table S2**). The majority of these 99 genes are involved in plant resistance to pathogens. Therefore, SVs appeared to have played a role in shaping the hc-siRNA hotspots across the tomato genome. This regulation over the dynamics of hc-siRNA hotspots has resulted in differential gene expression during tomato domestication, probably under the selection pressure for plant adaptation to different environments and in agricultural settings.

A notable example is a Bax inhibitor-1 (BI-1) family gene (*Solyc06g084130*). BI-1 proteins are conserved in eukaryotic organisms and associated with cell death during host-pathogen interactions (Chae *et al.*, 2003; Li *et al.*, 2014; Xu *et al.*, 2017). In plants, BI-1 proteins are involved in the autophagy process (Xu *et al.*, 2017) and confer plant resistance to various pathogens (Scotton *et al.*, 2017; Xu *et al.*, 2017; Lu *et al.*, 2018; Hernandez-Lopez *et al.*, 2019). In particular, BI-1 expression is pivotal to the autophagic activity that is critical for *N* gene-mediated resistance to tobacco mosaic virus (Xu *et al.*, 2017), a major viral pathogen of tomato. The repression of *Solyc06g084130* expression was strongly associated with SV-related 24-nt hc-siRNA differential

accumulation from SP to SLC and SLL plants (**Figure 1f**). Similar repression patterns could be found in multiple genes related to responses to biotic stresses (**Table S2**). Therefore, the repression of *Solyc06g084130* and other genes in response to biotic stresses during tomato domestication may affect tomato resistance traits, which is consistent with the observation that cultivated tomatoes are not well adapted to adverse environments as wild progenitors.

Highly dynamic gain/loss of phasiRNA generating loci during domestication

PhasiRNAs are known regulators of gene expression in plants (Fei *et al.*, 2013) and exhibit expression dynamics in response to environmental cues (Zheng *et al.*, 2015). Current models suggest that phasiRNAs serve as negative regulators to modulate the expression of their parental transcripts (Fei *et al.*, 2013; Zheng *et al.*, 2015). Although there is rapid progress in uncovering an enormous amount of phasiRNA-generating loci in various plants (Fei *et al.*, 2013; Zheng *et al.*, 2015) and unraveling their functions in plant development (Jiang *et al.*, 2020; Zhang *et al.*, 2020; Pokhrel *et al.*, 2021), a detailed analysis of phasiRNA dynamics during crop domestication has not been conducted. Using a previously established algorithm (Xia *et al.*, 2013), we analyzed phasiRNA-generating loci across all nine accessions and identified 290 PHASs mapped to the Heinz 1706 genome (SL4.0) and 286 mapped to the LA2093 genome, among which 77 were uniquely mapped to SL4.0 and 73 were uniquely mapped to the LA2093 genome (**Tables S3-S5**). In general, PHASs were mainly mapped to protein-coding genes with diverse functions as shown in the GO term analysis (**Figure S3a**), akin to our previous findings (Zheng *et al.*, 2015).

Interestingly, we found 34 SV-overlapping PHASs mapped to SL4.0 and 39 mapped to the LA2093 genome, among which 12 were uniquely mapped to SL4.0 and 17 were uniquely mapped to the LA2093 genome (**Table S6**). SV-related PHASs were distributed across all chromosomes, except chromosome 4 and 7, and there were numerous SV-related PHASs residing proximal to the terminal region of the long arm of chromosome 11 (**Figure 2a**). Among accessions in different tomato groups, SVs markedly contributed to the changes in PHASs (**Figure S3b**), which were overwhelmingly mapped to disease resistance genes (**Figure S3c**). This observation suggests that selection pressure favors the emergence of phasiRNAs regulating the expression of disease resistance genes in balancing growth and pathogen defense.

One cluster of disease resistance genes resided in a region where an insertion in the Heinz 1706 genome expanded the PHASs on chromosome 12 (**Figure 2b**). All three genes in this inserted

region are involved in tomato resistance to bacterial and oomycete pathogens (Canto-Pastor *et al.*, 2019). We checked the frequency of the insertion in different tomato groups, including SP (n = 51 accessions), SLC (n = 228), heirloom (n = 226), and modern cultivars (n = 52) (Wang *et al.*, 2020). As shown in **Figure 2c**, the frequency of the insertion associated with this gene cluster drastically increased in domesticated tomato to nearly 100%. Productions of phasiRNAs associated with this SV were also highly elevated in the SLL accession (**Figure 2d**).

A large portion of miRNAs exhibit highly dynamic expression patterns

The functions of many miRNAs in tomato growth and fruit development have been well studied in single accession analyses (Moxon *et al.*, 2008; Wang *et al.*, 2011; Karlova *et al.*, 2013). However, detailed analyses on miRNA expression profiles are lacking with respect to inferring the dynamics of miRNA-based gene regulatory network during crop domestication. To this end, we annotated all the miRNAs in our sRNA-Seq dataset based on recently revised criteria (Axtell and Meyers, 2018). We identified 122 miRNA genes mapped to SL4.0 and 126 mapped to the LA2093 genome. Both sets included 72 previously reported tomato miRNAs (Moxon *et al.*, 2008; Tomato Genome Consortium, 2012; Karlova *et al.*, 2013). There were 116 miRNA genes mapped to both SL4.0 and the LA2093 genome, whereas there were six and ten miRNA genes specifically mapped to SL4.0 and the LA2093 genome, respectively (**Table S7**). Based on the mapping results, we summarized the mature miRNAs and miRNA*s as well as the processing variants from the miRNA precursors in **Tables S8-S9**.

A close look at the mature miRNAs showed that the majority of known miRNAs were 21-nt in length with "U" as the first nucleotide (**Figure S4**), in line with previous observations (Chavez Montes *et al.*, 2014). Notably, the majority of novel miRNAs were 20-nt in length with "A" as the first nucleotide (**Figure S4**), in contrast to conserved miRNAs with well-established functions that mostly begin with a uracil. To our surprise, we found that a large portion of miRNAs exhibited highly variable expression patterns among accessions. For example, as shown in **Table S10**, the two SP accessions each had over 80 miRNAs that displayed fold changes in expression ≥ 1.5 and adjusted P values less than 0.05 compared to SLC and SLL accessions.

To better describe this dynamic, we plotted miRNA expression profiles across the nine tomato accessions and categorized the patterns into multiple groups. As shown in **Figure 3**, there were eight groups with each having more than six distinct miRNAs. Although the majority of the

miRNAs exhibited dynamic expression profiles in SLC accessions, we found that those in groups 1 and 7 exhibited an overall decrease, whereas those in groups 5, 6, and 8 exhibited an overall increase in expression over the course of tomato domestication and breeding. The presence of eight distinct groups also reflected the fluctuating expression patterns of miRNAs among different tomato accessions. However, we did not observe the corresponding correlated changes in the expression of their predicted targets, indicating that these miRNAs, despite having dynamic expression patterns, may not possess pivotal regulatory functions.

miRNAs and structure variants

Despite the notion that most miRNA gene families had the same number of members mapped to SL4.0 and the LA2093 genome, some underwent deletion or duplication events that changed the number of members in each family. For example, there were two miR10535 genes in the LA2093 reference genome but only one copy in Heinz 1706 SL4.0 (**Figure 4a**). Synteny analysis showed that the loss of one miR10535 gene copy could be attributed to a deletion that had occurred during tomato domestication (**Figure 4a**).

Because a genome deletion caused the loss of one copy of miR10535 gene, we reasoned that SVs might play important roles in determining the presence and the expression levels of miRNA genes. To this end, we identified 25 miRNA genes associated with 32 SVs, including 19 conserved and six novel miRNAs. Twenty-one out of 32 SVs were mapped to promoters of miRNA genes, seven were mapped to miRNA gene bodies, and four were mapped to both promoters and gene bodies of miRNAs (**Table S11**). GO term analysis showed that the computationally predicted target genes (**Table S12**) of these SV-overlapping miRNAs were mainly involved in the development/growth and responses to environmental stimuli (**Figure 4b**). However, most of these miRNAs, including miR10535, do not have empirically confirmed targets according to previous degradome studies in tomato (Karlova *et al.*, 2013; Pan *et al.*, 2017; Zheng *et al.*, 2017a; Zheng *et al.*, 2017c), supporting that these miRNAs may not have established regulatory functions.

SVs in promoter regions may affect miRNA expression. For example, a 17-bp deletion (SV49979) in the promoter of the miR172a/b-2 gene was observed in the genomes of most heirloom and modern tomatoes (**Figure 5a**). It is worth noting that the allele frequency of SV49949 significantly changed during domestication (**Figure 5b**). Correspondingly, miR172a/b-2 had

distinct expression profiles during fruit development among the nine tomato accessions, with the SLL and SLC accessions harboring this deletion generally showing a higher expression than the two SP accessions that did not harbor the deletion (**Figure 5c**). As a result of the low abundance of miR172a/b-2 compared to the total miR172 abundance (**Tables S8 and S9**), this expression change in miR172a/b-2 did not significantly affect the expression of any miR172 targets. This further indicates that miR172a/b-2 do not play any important regulatory roles.

Imprecisely processed miRNAs are differentially selected

Some miRNA precursors tend to generate a population of miRNA-miRNA* pairs as a result of the imprecise processing by DCL1 (Cuperus *et al.*, 2011). The imprecise processing may affect miRNA functions that heavily rely on sequence complementarity between miRNAs and their targets (Rehmsmeier *et al.*, 2004). We found that eight out of 122 SLL miRNA genes and six out of 126 SP miRNA genes generated more than six variants in our dataset, among which six were shared in the SLL and SP reference genomes (**Table S13**). The miR397 gene, a conserved miRNA across different plant lineages, had six miR397-3P and seven miR397-5P variants in tomato (**Figure 6a**). Notably, we observed a shift in the most abundant product of the miR397 precursor: from only miR397-3P that was expressed in SP accessions, transitioning to miR397-5P (the conserved mature miR397 in plants) that became more prevalent in most SLC and SLL accessions (**Figure 6b**). Over-expression of miR397-5P can enhance tomato response to drought stress (Huang *et al.*, 2021), implying that miR397-5P may be beneficial in SLC and SLL accessions in adaptation to adverse environments. Nevertheless, the target of miR397-5P remains unclear despite extensive efforts of degradome analysis (Karlova *et al.*, 2013; Pan *et al.*, 2017; Zheng *et al.*, 2017a; Zheng *et al.*, 2017c), suggesting that it has not yet established an essential regulation.

We also observed a position shift of miRNA:miRNA* in the precursor of miR9472. The miRNA:miRNA* duplex was closer to the terminal loop region in SP and SLC accessions but resided at a more distant region away from the terminal loop in SLL (**Figure S5**). This shift was unlikely to be caused by any change in recognition by DCL1 because the precursor sequences remained the same. Notably, we found seven such examples, as listed in **Table S14**. This observation indicates that DCL1 recognition on miRNA precursors is likely flexible in plants, and the selection pressure plays a role in determining the expression of the final miRNA:miRNA* duplexes.

Discussion

sRNAs are critical regulators of gene expression underlying plant growth and responses to environmental cues (Karlova *et al.*, 2013; Borges and Martienssen, 2015). sRNA abundance is known to directly impact their functions (Bartel, 2009). Comparative studies on sRNA long-time evolution from green algae to flowering plants have been widely reported (Ma *et al.*, 2010; Cuperus *et al.*, 2011; Chavez Montes *et al.*, 2014; Zheng *et al.*, 2015; Lunardon *et al.*, 2020; Baldrich *et al.*, 2021). However, the role of sRNAs in crop domestication (shorter evolutionary time) and whether and how crop domestication affects sRNA dynamics have not been systematically investigated. To gain a better understanding of sRNA dynamic expression during tomato domestication, we generated a comprehensive sRNA dataset using nine representative tomato accessions spanning from the wild SP progenitors, intermediate SLC accessions, and one domesticated accession, also covering samples from leaf, flower, and fruits at four critical developmental stages. Our high-quality dataset fulfills the immediate needs for high-resolution comparative analyses on sRNA expression and inferring their functions in wild, semi-domesticated, and domesticated tomato plants, in addition to establishing a foundation for future exploration of sRNA functions.

Our dataset clearly demonstrates that all three classes of sRNAs (hc-siRNAs, phasiRNAs, and miRNAs) have significant changes in expression during tomato domestication and breeding. Notably, we found that SVs are an important driving force underlying the dynamic expression of these sRNAs. Those SVs, particularly deletions and insertions, probably have a direct impact on the production of all three major types of sRNAs. For example, deletions or insertions result in the differential accumulation of hc-siRNAs in gene promoter regions and the gain or loss of PHASs. SVs also contribute to the birth and death of miRNAs in domesticated tomatoes, as evidenced by the deletion of miR10535 and additional examples listed in **Table S11**. When SVs reside in promoter regions of miRNA genes, they may influence miRNA expression, as evidenced in miR172a/b-2. In addition to SVs, we also note that imprecise processing of miRNA/miRNA* duplexes can lead to the dynamic expression of miRNAs during tomato domestication. Given the rapid changes in the expression of sRNAs and their roles in regulating gene expression, they possibly contribute to some phenotypic changes during crop domestication.

Tomato has a complex history of domestication, selection, and breeding (Blanca *et al.*, 2012; Blanca *et al.*, 2015; Razifard *et al.*, 2020; Blanca *et al.*, 2021). Tomato domestication before cultivation possibly has a selection pressure on plant adaptation to new environments and developmental processes. Domestication and re-domestication processes possibly have posed a selection pressure on tomato flavor and yield. The modern breeding processes possibly have posed a selection pressure on disease resistance. Interestingly, selection pressures appear to have distinct impacts on different classes of sRNAs. The SV-related dynamic expression of 24-nt hc-siRNAs and phasiRNAs predominantly impacts the expression of genes related to stress responses and growth, implying that the selection pressure favors the regulation of those trait-associated genes through hc-siRNA and phasiRNA pathways. By contrast, the conserved miRNAs play a major role in plant development mostly through regulating transcription factors (Ma *et al.*, 2010; Cuperus *et al.*, 2011). Some of those well-established regulations are conserved in most land plants (Huijser and Schmid, 2011; Xia *et al.*, 2017; Millar *et al.*, 2019). Therefore, the selection pressure is unlikely to favor changes in such critical regulations related to plant growth in a relatively short timeframe during domestication. This is consistent with the rare report of conserved miRNAs being differentially regulated during crop domestication, except for miR397 during rice domestication (Swetha *et al.*, 2018). By contrast, many less conserved miRNAs, mostly expressed at low levels and/or having no to very few confirmed targets, exhibit highly dynamic expression during tomato domestication, reflecting little selection pressure on those miRNAs.

Our dataset serves as a foundation for future studies on sRNA function associated with tomato growth and domestication, and beyond. Recent progress has demonstrated that miRNA gene families may exert functions through developmentally-regulated expression of specific members (Lian *et al.*, 2021; O'Maoileidigh *et al.*, 2021). Our dataset can help with mining miRNA family members with developmentally-regulated expression patterns. For example, miR390b, with a U20A substitution at position 20 in comparison to miR390a, was specifically expressed in flowers (**Figure S6**). The flower-specific expression of miR390b tripled the amount of total miR390 in flowers, which markedly promoted production of phasiRNAs from the *TAS3* locus as well as specifically suppressed the expression of *ARF3* and *ARF4* in flowers (**Figure S6**).

The present study represents the pioneering comprehensive comparative investigation of sRNA expression dynamics during crop domestication and provides insights into plant sRNA evolution as well as SV-based gene regulation during crop domestication. Our analyses have

unraveled a strong role of SVs in affecting the expression of different classes of sRNAs, which further contributes to trait changes during tomato domestication. Our study lays a solid foundation for future investigations of the detailed physiological roles of SVs and their associated sRNAs in regulating important domestication traits.

Experimental Procedures

Plant materials and RNA isolation

Tomato plants were grown in a greenhouse at 25 °C under a 16:8 h light/dark photocycle at Ohio State University (Columbus, OH, USA). For each accession, young leaves, anthesis-stage flowers, and fruits at four different developmental stages (young green, mature green, breaker, and red ripe) were collected with three biological replicates. Total RNA from tomato samples was isolated and fractioned to >200 nt and <200 nt populations using the RNAzol RT reagent (Sigma-Aldrich, St. Louis, MO, USA). sRNA species were further purified using the miRVana miRNA isolation kit (Thermo Fishier Scientific, Grand Island, NY, USA) in accordance with the manufacturer's instructions. mRNA populations were further purified using a Magnetic mRNA isolation kit (NEB, Ipswich, MA, USA).

Library construction and sequencing

sRNA libraries were constructed following the established protocol (Chen *et al.*, 2012). Briefly, 18-30 nt sRNA populations purified on 15% (w/v) polyacrylamide/8 M urea gel were ligated with 3'- and 5'-adapters. sRNA populations with adapters were reverse transcribed, PCR amplified, and then purified from the 8% native PAGE gel. Strand-specific RNA-Seq libraries were constructed using a protocol described previously (Zhong *et al.*, 2011). All the constructed libraries were analyzed and quantified by Bioanalyzer (Agilent, Santa Clara, CA, USA) and sequenced on an Illumina HiSeq 2500 system (Illumina Inc., San Diego, CA, USA).

sRNA sequence processing

sRNA reads were processed to remove adapters using the sRNA cleaning script provided in the VirusDetect package (Zheng *et al.*, 2017b). The trimmed sRNA reads shorter than 15 nt were discarded. The resulting sRNA reads were further cleaned by removing those that perfectly matched to the sequences of tRNAs, snoRNAs, snRNAs (collected from GenBank), or rRNAs

(Quast *et al.*, 2013) using Bowtie (Langmead *et al.*, 2009). Raw counts for each unique sRNAs were derived and normalized into TPM (transcripts per million). Three replicates from young green fruits of BGV007023 were discarded because they did not pass the quality check. The cleaned sRNA reads were mapped to the wild (LA2093) (Wang *et al.*, 2020) and cultivated tomato (Heinz 1706, SL4.0 and ITAG 4.1; https://solgenomics.net) (Hosmani *et al.*, 2019) reference genomes using the ShortStack with the unique-weighing mode, which provides more accurate assignment of multi-mapping reads to genome loci than other methods (Axtell, 2013b; Johnson *et al*., 2016).

A total of 92,523 genomic SVs between SP LA2093 and SLL Heinz 1706 reported in our previous study (Wang *et al.*, 2020) and their genotyping data in 597 tomato accessions [51 SP, 228 SLC, 312 SLL (226 heirloom, 52 modern and 34 other cultivars) and 6 *S. cheesmaniae* and *S. galapagense*], including all nine accessions used in the present study, were used for the subsequent analyses.

Identification and analysis of 24-nt hc-siRNA hotspots

Based on the sRNA alignments, siRNA hotspot regions that were defined by continuously covered sRNAs were identified using ShortStack (Axtell, 2013b). The expression of 24-nt siRNAs was calculated by counting the number of 24-nt siRNA reads mapped to the corresponding regions. Only a region with no less than 10 mapped 24-nt siRNA reads was considered as a 24-nt siRNA hotspot. The 24-nt siRNA expression was normalized to number of reads per kilobase of genome region per million mapped reads (RPKM), based on all mapped reads.

To identify SV-related 24-nt siRNAs, pairwise comparisons were firstly applied between stages or between tissues, and statistical analysis was performed using DESeq2 (Love *et al.*, 2014). Only regions with adjusted $P < 0.05$ and fold change ≥ 2 were considered as significantly changed hotspots. The significantly changed protein-coding genes and corresponding changed 24-nt siRNA hotspots in promoter regions were treated as genes pairs involved in epigenetic regulation. To further exclude potential false positive candidates, we only kept the pairs of 24-nt hotspots and the corresponding protein-coding genes with a negative correlation in expression that occurred in at least 10 samples. Previously reported SVs (Wang *et al.*, 2020) that overlapped with significantly changed 24-nt siRNA hotspots were then identified. The identified SVs were further filtered to

keep those in promoter regions of protein-coding genes that exhibited an opposite expression pattern compared to the changes in abundance of the corresponding 24-nt siRNA hotspots.

Identification of candidate PHAS loci

We used the previously described methods to identify PHAS loci (Zheng *et al.*, 2015) from each of the 159 samples. Following sRNA read alignment, the reference sequences were then scanned with a sliding window of 189 bp (nine 21-nt phase registers). P values and phasing scores for each window were calculated following the methods described previously (De Paoli *et al.*, 2009; Xia *et al.*, 2013). A positive window was considered to contain no less than 10 unique sRNAs, with more than half of unique sRNAs being 21 nt in length and with no less than three 21-nt unique sRNAs falling into the phase registers. Finally, positive windows identified from all samples were combined if they i) shared the same phase registers and ii) fell into the same gene loci. The sequences of PHASs and the flanking regions of 200 bp were retrieved and compared between wild and cultivated tomato reference genomes using BLAST (Camacho *et al.*, 2009). The BLAST results were then processed to categorize candidate PHAS loci into three groups: PHAS loci shared by wild and cultivated tomatoes, specific to wild tomato, and specific to cultivated tomato.

Identification of miRNAs and differential expression analysis

MiRNAs were identified using ShortStack (Axtell, 2013b) from each of the 159 samples, and a series of filtering was applied to obtain high-confidence miRNAs according to the recently revised miRNA annotation criteria (Axtell and Meyers, 2018). Briefly, following alignments, mature miRNAs and corresponding pre-miRNAs were then identified using ShortStack (Axtell, 2013b). The identified miRNAs from these samples were collapsed if they were mapped to the exact locations in the genome. The collapsed miRNAs that existed in at least three samples and expressed at more than 10 TPM in at least one sample were considered as high-confidence miRNAs, which were compared with miRBase (Kozomara *et al.*, 2019) to identify conserved miRNAs, whereas miRNAs that showed no matches in the miRBase were considered as novel miRNAs. We adopted the existing miRNA ID system in miRBase (Kozomara *et al.*, 2019) and PmiREN2.0 (Guo *et al.*, 2022) for our miRNA nomenclature.

Raw counts of the identified miRNAs were processed using DESeq2 (Love *et al.*, 2014) to identify differentially expressed miRNAs among accessions. MiRNAs with adjusted $P < 0.05$ were

considered as differentially expressed. Differentially expressed miRNAs were further clustered into groups according to their expression patterns using DEGreport (Huber *et al.*, 2015). Target genes of differentially expressed miRNAs were predicted using TargetFinder (Fahlgren and Carrington, 2010). GO enrichment analysis was performed on the target genes using GO::TermFinder (Boyle *et al.*, 2004).

RNA-Seq read processing and differential expression

Single-end RNA-Seq reads were processed to remove adapters as well as low-quality bases using Trimmomatic (Bolger *et al.*, 2014) with parameters "ILLUMINACLIP:TruSeq3-SE.fa:2:30:7 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:80". The remaining high-quality reads were subjected to rRNA sequence removal by aligning them to an rRNA database (Quast *et al.*, 2013) using Bowtie (Langmead *et al.*, 2009) allowing up to three mismatches. The cleaned RNA-Seq reads were aligned to the cultivated tomato (Heinz 1706, SL4.0) reference genome using STAR (Dobin *et al.*, 2013) allowing up to two mismatches. Gene expression was measured by counting the number of reads mapped to gene regions (ITAG4.1) and then normalized to the number of reads per kilobase of exon per million mapped reads (RPKM). Differential expression analysis was performed using DESeq2 (Love *et al.*, 2014). To obtain a global comparison among all samples, in particular to identify differentially expressed genes in specific accessions or developmental stages, we followed a previously described linear factorial modeling (Clevenger *et al.*, 2017). We also performed pairwise comparisons to identify differentially expressed genes between stages for different accessions. Genes with adjusted $P < 0.05$ and fold changes ≥ 2 were considered differentially expressed.

Accession numbers

Raw sRNA and RNA-Seq reads have been deposited in the NCBI BioProject database under the accession number PRJNA438371.

Acknowledgements

This work was supported by grants from the US National Science Foundation (IOS-1564366 to EvdK and YW and IOS-1855585 to ZF), the Urban Agriculture Characteristic Platform Construction of Beijing University of Agriculture (5076016177/003 to YZ). This work is dedicated

to the late Prof. Biao Ding (Ohio State University), who participated at the inception stage of this study.

Author contributions

YZ, EvdK, YW and ZF conceived the idea. YZ, YW and ZF supervised the study. SM and YW performed the plant analyses and sampling. YW constructed sRNA-Seq libraries. YQ, YZ, HNS, XW, XZ, YW and ZF analyzed the data. YZ and YW summarized the results. YZ, EvdK, YW and ZF wrote and revised the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Supporting Information

Figure S1. Development timeline of the nine tomato accessions investigated in this study.

Figure S2. sRNA sequencing data.

Figure S3. Characteristics of PHAS loci.

Figure S4. Characteristics of mature miRNAs.

Figure S5. Shifted miRNA-miRNA* pair from the miR9472 precursor.

Figure S6. Expression profiles of the miR390-TAS3-ARF3/ARF4 cascade in the nine tomato accessions.

Table S1. Summary statistics of sRNA sequencing reads.

Table S2. Differentially accumulated hc-siRNAs in promoter regions that regulate the expression of protein-coding genes.

Table S3. List of PHAS loci identified with LA2093 and Heinz 1706 (SL4.0) reference genomes.

Table S4. PhasiRNA regions (LA2093 genome as the reference) and abundances (TPM).

Table S5. PhasiRNA regions (Heinz 1706 genome as the reference) and abundances (TPM).

Table S6. List of SV-related PHAS loci.

Table S7. List of miRNA precursor loci identified with LA2093 and Heinz 1706 (SL4.0) reference genomes.

Table S8. miRNAs abundance (LA2093 genome as the reference).

Table S9. miRNA abundance (Heinz 1706 genome as the reference).

Table S10. Summary of differentially expressed miRNAs identified in pairwise comparisons.

Table S11. List of miRNA genes overlapped with SVs.

Table S12. Expression profiles of predicted target genes of SV-overlapping miRNAs.

Table S13. List of miRNA genes that generate more than six variants.

Table S14. List of miRNA genes that have shifted miRNA:miRNA* pairs.

References

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L. *et al.* (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, **182**, 145–161.e123.

Axtell, M.J. (2013a) Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol*, **64**, 137–159.

Axtell, M.J. (2013b) ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA*, **19**, 740–751.

Axtell, M.J. & Meyers, B.C. (2018) Revisiting criteria for plant microRNA annotation in the era of big data. *Plant Cell*, **30**, 272–284.

Baldrich, P., Bélanger, S., Kong, S., Pokhrel, S., Tamim, S., Teng, C. *et al.* (2021) The evolutionary history of small RNAs in the Solanaceae. *bioRxiv*, 2021.2005.2026.445884.

Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

Baulcombe, D. (2004) RNA silencing in plants. *Nature*, **431**, 356–363.

Blanca, J., Canizares, J., Cordero, L., Pascual, L., Diez, M.J. & Nuez, F. (2012) Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PLoS One*, **7**, e48198.

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Diez, M.J. *et al.* (2015) Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics*, **16**, 257.

Blanca, J., Sanchez-Matarredona, D., Ziarsolo, P., Montero-Pau, J., van der Knaap, E., Díez, M.J. *et al.* (2021) Haplotype analyses reveal novel insights into tomato history and domestication including long-distance migrations and latitudinal adaptations. *bioRxiv*, 2021.2006.2018.448912.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Borges, F. & Martienssen, R.A. (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol*, **16**, 727–741.

Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M. *et al.* (2004) GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, **20**, 3710–3715.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Canto-Pastor, A., Santos, B., Valli, A.A., Summers, W., Schornack, S. & Baulcombe, D.C. (2019) Enhanced resistance to bacterial and oomycete pathogens by short tandem target mimic RNAs in tomato. *Proc Natl Acad Sci U S A*, **116**, 2755–2760.

Carvalho, C.M. & Lupski, J.R. (2016) Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet*, **17**, 224–238.

Chae, H.J., Ke, N., Kim, H.R., Chen, S., Godzik, A., Dickman, M. *et al.* (2003) Evolutionarily conserved cytoprotection provided by Bax inhibitor-1 homologs from animals, plants, and yeast. *Gene*, **323**, 101–113.

Chavez Montes, R.A., de Fatima Rosas-Cardenas, F., De Paoli, E., Accerbi, M., Rymarquis, L.A., Mahalingam, G. *et al.* (2014) Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat Commun*, **5**, 3722.

Chen, C., Li, J., Feng, J., Liu, B., Feng, L., Yu, X. *et al.* (2021) sRNAanno-a database repository of uniformly annotated small RNAs in plants. *Hortic Res*, **8**, 45.

Chen, Y.R., Zheng, Y., Liu, B., Zhong, S., Giovannoni, J. & Fei, Z. (2012) A cost-effective method for Illumina small RNA-Seq library preparation using T4 RNA ligase 1 adenylated adapters. *Plant Methods*, **8**, 41.

Clevenger, J., Chu, Y., Arrais Guimaraes, L., Maia, T., Bertioli, D., Leal-Bertioli, S. *et al.* (2017) Gene expression profiling describes the genetic regulation of *Meloidogyne arenaria* resistance in *Arachis hypogaea* and reveals a candidate gene for resistance. *Sci Rep*, **7**, 1317.

Cuperus, J.T., Fahlgren, N. & Carrington, J.C. (2011) Evolution and functional diversification of MIRNA genes. *Plant Cell*, **23**, 431–442.

De Paoli, E., Dorantes-Acosta, A., Zhai, J., Accerbi, M., Jeong, D.H., Park, S. *et al.* (2009) Distinct extremely abundant siRNAs associated with cosuppression in petunia. *RNA*, **15**, 1965–1970.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Fahlgren, N. & Carrington, J.C. (2010) miRNA target prediction in plants. *Methods Mol Biol*, **592**, 51–57.

Fei, Q., Xia, R. & Meyers, B.C. (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell*, **25**, 2400–2415.

Guo, Z., Kuang, Z., Zhao, Y., Deng, Y., He, H., Wan, M. *et al*. (2022) PmiREN2.0: from data annotation to functional exploration of plant microRNAs. *Nucleic Acids Res*, **50**, D1475–D1482.

Hernandez-Lopez, A., Diaz, M., Rodriguez-Lopez, J., Guillen, G., Sanchez, F. & Diaz-Camino, C. (2019) Uncovering Bax inhibitor-1 dual role in the legume-rhizobia symbiosis in common bean roots. *J Exp Bot*, **70**, 1049–1061.

Hosmani, P.S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L.V., Schijlen, E. *et al.* (2019) An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv*, 767764.

Huang, S., Zhou, J., Gao, L. & Tang, Y. (2021) Plant miR397 and its functions. *Funct Plant Biol*, **48**, 361–370.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S. *et al.* (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*, **12**, 115–121.

Huijser, P. & Schmid, M. (2011) The control of developmental phase transitions in plants. *Development*, **138**, 4117–4129.

Jiang, P., Lian, B., Liu, C., Fu, Z., Shen, Y., Cheng, Z. *et al.* (2020) 21-nt phasiRNAs direct target mRNA cleavage in rice male germ cells. *Nat Commun*, **11**, 5191.

Johnson, N.R., Yeoh, J.M., Coruh, C. & Axtell, M.J. (2016) Improved placement of multi-mapping small RNAs. *G3*, **6,** 2103–2111.

Karlova, R., van Haarst, J.C., Maliepaard, C., van de Geest, H., Bovy, A.G., Lammers, M. *et al.* (2013) Identification of microRNA targets in tomato fruit development using high-throughput sequencing and degradome analysis. *J Exp Bot*, **64**, 1863–1878.

Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res*, **47**, D155–D162.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**, R25.

Li, B., Yadav, R.K., Jeong, G.S., Kim, H.R. & Chae, H.J. (2014) The characteristics of Bax inhibitor-1 and its related diseases. *Curr Mol Med*, **14**, 603–615.

Lian, H., Wang, L., Ma, N., Zhou, C.M., Han, L., Zhang, T.Q. *et al.* (2021) Redundant and specific roles of individual MIR172 genes in plant development. *PLoS Biol*, **19**, e3001044.

Lisch, D. (2013) How important are transposons for plant evolution? *Nat Rev Genet*, **14**, 49–61.

Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.

Lu, P.P., Yu, T.F., Zheng, W.J., Chen, M., Zhou, Y.B., Chen, J. *et al.* (2018) The wheat Bax inhibitor-1 protein interacts with an aquaporin TaPIP1 and enhances disease resistance in Arabidopsis. *Front Plant Sci*, **9**, 20.

Lunardon, A., Johnson, N.R., Hagerott, E., Phifer, T., Polydore, S., Coruh, C. *et al.* (2020) Integrated annotations and analyses of small RNA-producing loci from 47 diverse plants. *Genome Res*, **30**, 497–513.

Ma, Z., Coruh, C. & Axtell, M.J. (2010) Arabidopsis lyrata small RNAs: transient MIRNA and small interfering RNA loci within the Arabidopsis genus. *Plant Cell*, **22**, 1090–1103.

McCue, A.D. & Slotkin, R.K. (2012) Transposable element small RNAs as regulators of gene expression. *Trends Genet*, **28**, 616–623.

Millar, A.A., Lohe, A. & Wong, G. (2019) Biology and function of miR159 in plants. *Plants*, **8**, 255.

Moxon, S., Jing, R., Szittya, G., Schwach, F., Rusholme Pilcher, R.L., Moulton, V. *et al.* (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res*, **18**, 1602–1609.

O'Maoileidigh, D.S., van Driel, A.D., Singh, A., Sang, Q., Le Bec, N., Vincent, C. *et al.* (2021) Systematic analyses of the MIR172 family members of Arabidopsis define their distinct roles in regulation of APETALA2 during floral transition. *PLoS Biol*, **19**, e3001043.

Pan, C., Ye, L., Zheng, Y., Wang, Y., Yang, D., Liu, X. *et al.* (2017) Identification and expression profiling of microRNAs involved in the stigma exsertion under high-temperature stress in tomato. *BMC Genomics*, **18**, 843.

Pereira, L., Sapkota, M., Alonge, M., Zheng, Y., Zhang, Y., Razifard, H. *et al.* (2021) Natural genetic diversity in tomato flavor genes. *Front Plant Sci*, **12**, 642828.

Pokhrel, S., Huang, K. & Meyers, B.C. (2021) Conserved and non-conserved triggers of 24-nucleotide reproductive phasiRNAs in eudicots. *Plant J*, **107**, 1332–1345

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*, **41**, D590–596.

Razifard, H., Ramos, A., Della Valle, A.L., Bodary, C., Goetz, E., Manser, E.J. *et al.* (2020) Genomic evidence for complex domestication history of the cultivated tomato in Latin America. *Mol Biol Evol*, **37**, 1118–1132.

Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.

Sanseverino, W., Henaff, E., Vives, C., Pinosio, S., Burgos-Paz, W., Morgante, M. *et al.* (2015) Transposon insertions, structural variations, and SNPs contribute to the evolution of the melon genome. *Mol Biol Evol*, **32**, 2760–2774.

Scotton, D.C., Azevedo, M.D.S., Sestari, I., Da Silva, J.S., Souza, L.A., Peres, L.E.P. *et al.* (2017) Expression of the *Theobroma cacao* Bax-inhibitor-1 gene in tomato reduces infection by the hemibiotrophic pathogen *Moniliophthora perniciosa*. *Mol Plant Pathol*, **18**, 1101–1112.

Swetha, C., Basu, D., Pachamuthu, K., Tirumalai, V., Nair, A., Prasad, M. *et al.* (2018) Major domestication-related phenotypes in Indica rice are due to loss of miRNA-mediated laccase silencing. *Plant Cell*, **30**, 2649–2662.

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.

Vazquez, F. (2006) Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci*, **11**, 460–468.

Wang, X., Gao, L., Jiao, C., Stravoravdis, S., Hosmani, P.S., Saha, S. *et al.* (2020) Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat Commun*, **11**, 5817.

Wang, Y., Itaya, A., Zhong, X., Wu, Y., Zhang, J., van der Knaap, E. *et al.* (2011) Function and evolution of a MicroRNA that regulates a Ca2+-ATPase and triggers the formation of phased small interfering RNAs in tomato reproductive growth. *Plant Cell*, **23**, 3185–3203.

Xia, R., Meyers, B.C., Liu, Z., Beers, E.P., Ye, S. & Liu, Z. (2013) MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. *Plant Cell*, **25**, 1555–1572.

Xia, R., Xu, J. & Meyers, B.C. (2017) The emergence, evolution, and diversification of the miR390-TAS3-ARF pathway in land plants. *Plant Cell*, **29**, 1232–1247.

Xu, G., Wang, S., Han, S., Xie, K., Wang, Y., Li, J. *et al.* (2017) Plant Bax Inhibitor-1 interacts with ATG6 to regulate autophagy and programmed cell death. *Autophagy*, **13**, 1161–1175.

Zhang, Y.C., Lei, M.Q., Zhou, Y.F., Yang, Y.W., Lian, J.P., Yu, Y. *et al.* (2020) Reproductive phasiRNAs regulate reprogramming of gene expression and meiotic progression in rice. *Nat Commun*, **11**, 6031.

Zheng, Y., Ding, B., Fei, Z. & Wang, Y. (2017a) Comprehensive transcriptome analyses reveal tomato plant responses to tobacco rattle virus-based gene silencing vectors. *Sci Rep*, **7**, 9771.

Zheng, Y., Gao, S., Padmanabhan, C., Li, R., Galvez, M., Gutierrez, D. *et al.* (2017b) VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology*, **500**, 130–138.

Zheng, Y., Wang, Y., Ding, B. & Fei, Z. (2017c) Comprehensive transcriptome analyses reveal that potato spindle tuber viroid triggers genome-wide changes in alternative splicing, inducible trans-acting activity of phasiRNAs and immune responses. *J Virol*. **91**, e00247-17

Zheng, Y., Wang, Y., Wu, J., Ding, B. & Fei, Z. (2015) A dynamic evolutionary and functional landscape of plant phased small interfering RNAs. *BMC Biol*, **13**, 32.

Zhong, S., Fei, Z., Chen, Y.R., Zheng, Y., Huang, M., Vrebalov, J. *et al.* (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat Biotechnol*, **31**, 154–159.

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y. *et al.* (2011) High-throughput Illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc*, **2011**, 940–949.

Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C. *et al.* (2018) Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell*, **172**, 249–261.e212.

Figure legend

Figure 1. Functional analysis of SV-related 24-nt hc-siRNA regions among wild and cultivated tomatoes. (a) Circos plot of the densities of 24-nt hc-siRNAs, genes, repeat sequences and SVs across the tomato genome. (b) Percentage of differentially expressed (DER) and unchanged (non-DER) 24-nt siRNA regions mapped to SVs, gene bodies and the whole genome. (c) Differentially expressed genes showing a negative correlation with the corresponding SV-related 24-nt hc-siRNA clusters in their promoters in at least one pairwise comparison. YG, young green fruit; MG, mature green fruit; Br, fruit at the breaker stage; RR, red ripe fruit. (d,e) GO term enrichment analysis of differentially expressed genes in (c) using all tomato genes (d) or SV-related genes (e) as the background. (f) SV-related 24-nt hc-siRNA (Cluster_71406) enrichment leading to the repression of *Solyc06g084130* expression in SLL (*Solanum lycopersicum* var. *lycopersicum*) and SLC (*S. lycopersicum* var. *cerasiforme*) compared to SP (*S. pimpinellifolium*). RPKM, reads per kilobase of genome region (for siRNAs) or exon model (for genes) per million mapped reads.

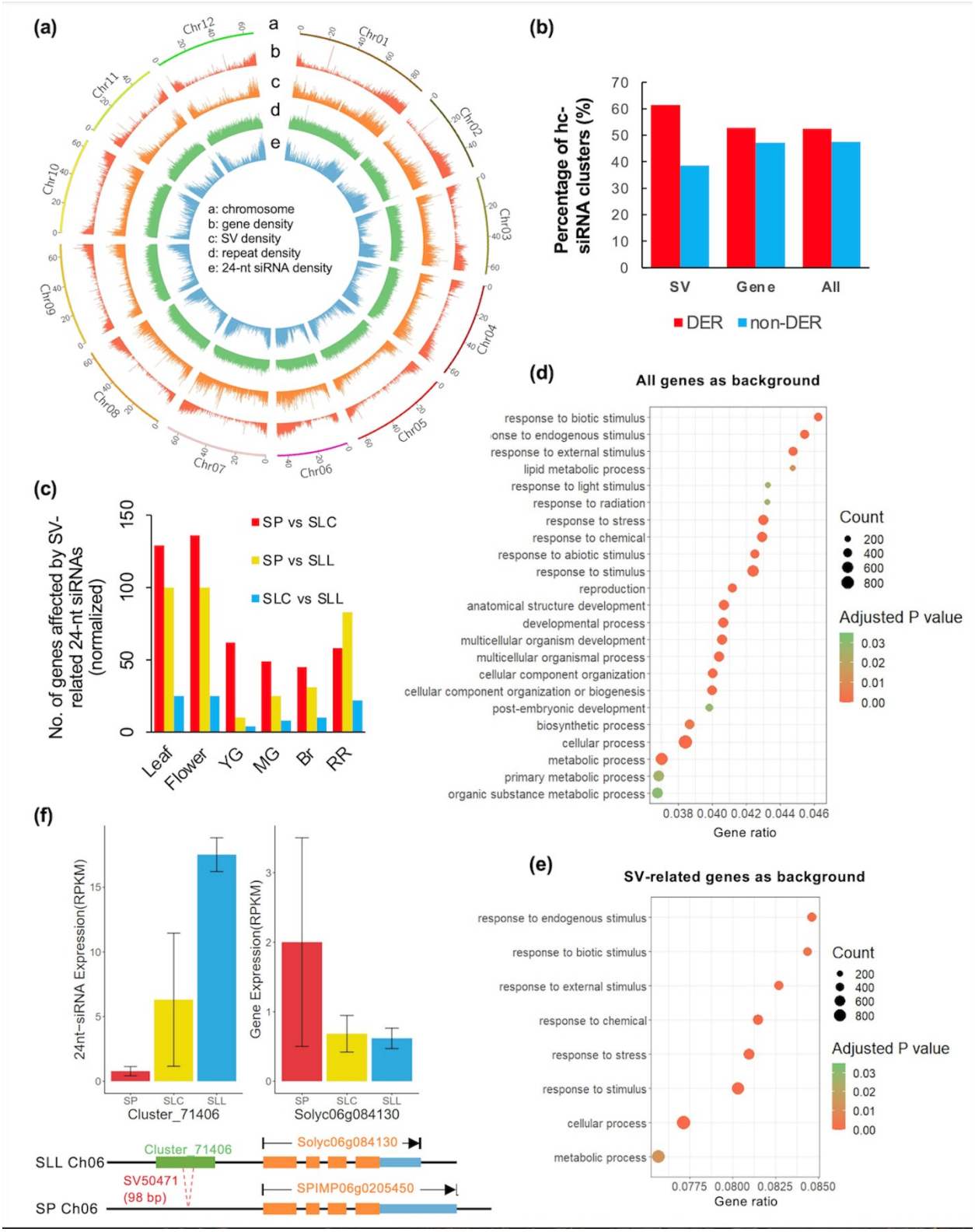
Figure 2. SV-related PHASs. (a) Distribution of PHAS loci and SV-related PHAS loci across the 12 tomato chromosomes. The number depicts the distinct PHASs at each locus. (b) SV90890 causes gain/loss of PHASs and protein-coding genes. (c) Allele frequency of SV90890 in different tomato groups. (d) Abundance of phasiRNAs in the genome region of SLL Heinz 1706 containing SV90890 in the nine tomato accessions. For each box plot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median. The whisker represents 1.5× interquartile range of the lower or upper quartile. TPM, transcripts per million.

Figure 3. Distinct expression patterns of miRNAs. MiRNA abundances were normalized to Z-scores where expression values in transcripts per million (TPM) were centered to the mean and scaled to the standard deviation by each miRNA. For each box plot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median. The whisker represents 1.5× interquartile range of the lower or upper quartile.

Figure 4. MiRNA genes overlapped with SVs. (a) Synteny diagram of miR10535 gene(s) in SP and SLL reference genomes. (b) Enriched GO terms of predicted target genes of SV-related miRNAs.

Figure 5. SV affects miRNA expression. (a) Diagram showing SV49949 (a 17-bp indel) in the promoter of miR172a/b-2 gene. (b) Allele frequency of SV49949 in different tomato groups. (c) Expression profiles of miR172a/b-2 in different tomato accessions. YG, young green fruits. MG, mature green fruits. Br, fruits at the breaker stage. RR, red ripe fruits. TPM, transcripts per million.

Figure 6. Changes in miR397 expression during tomato domestication. (a) Diagram showing the six 5P and seven 3P processing variants from the miR397 precursor. (b) Abundance of each miR397 variant in different tomato accessions. TPM, transcripts per million.



764

765 Figure 1

766

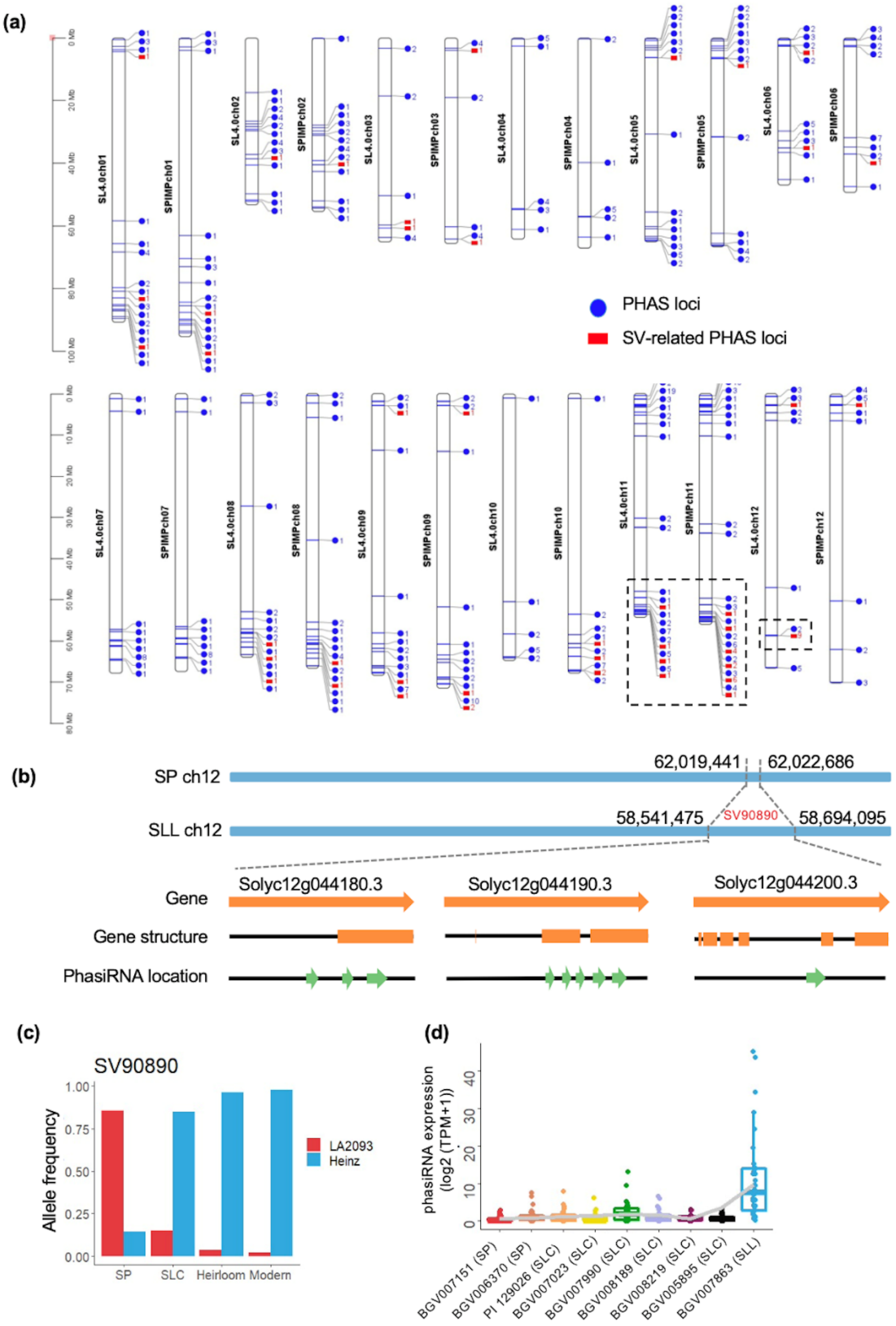
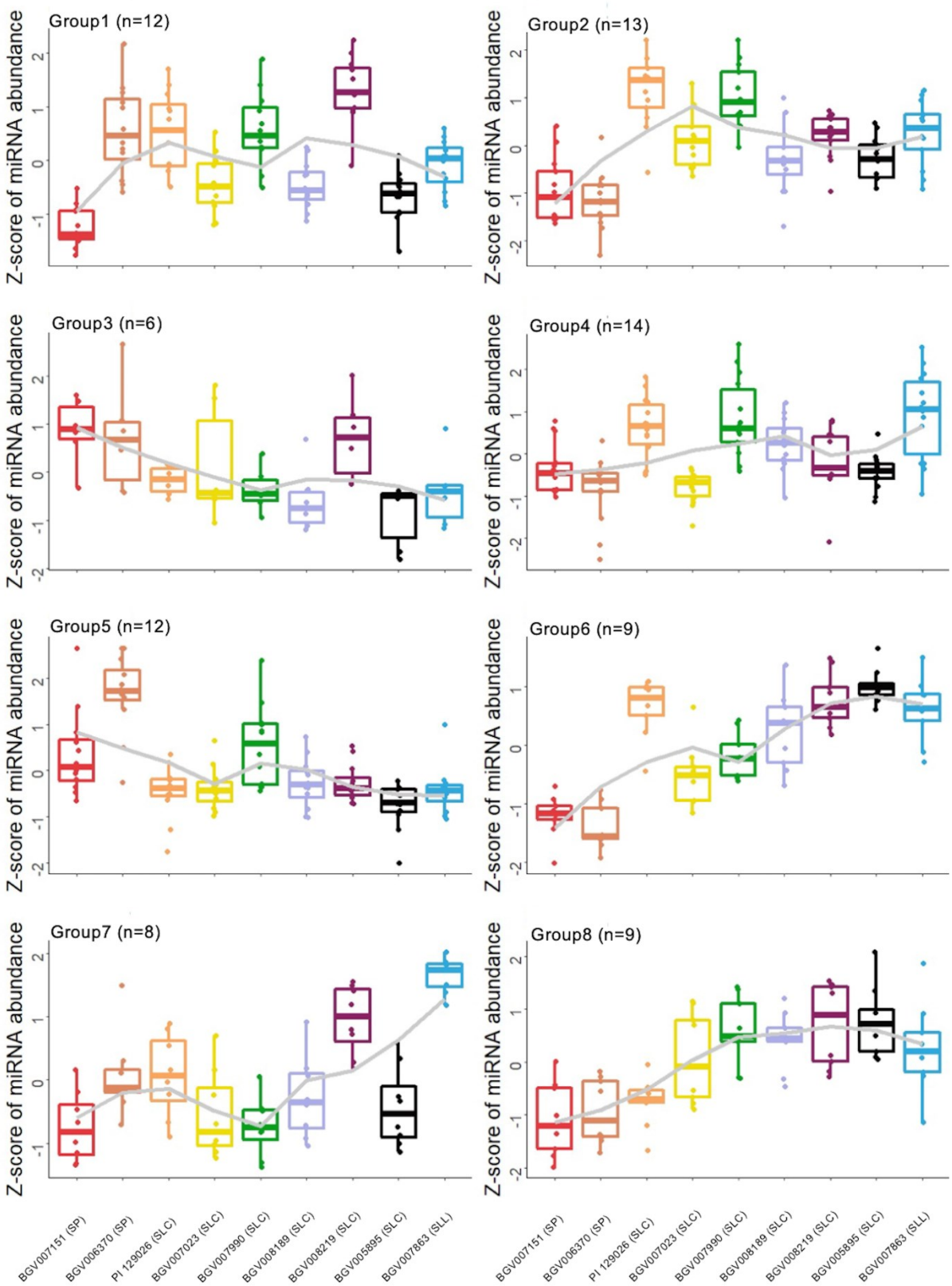
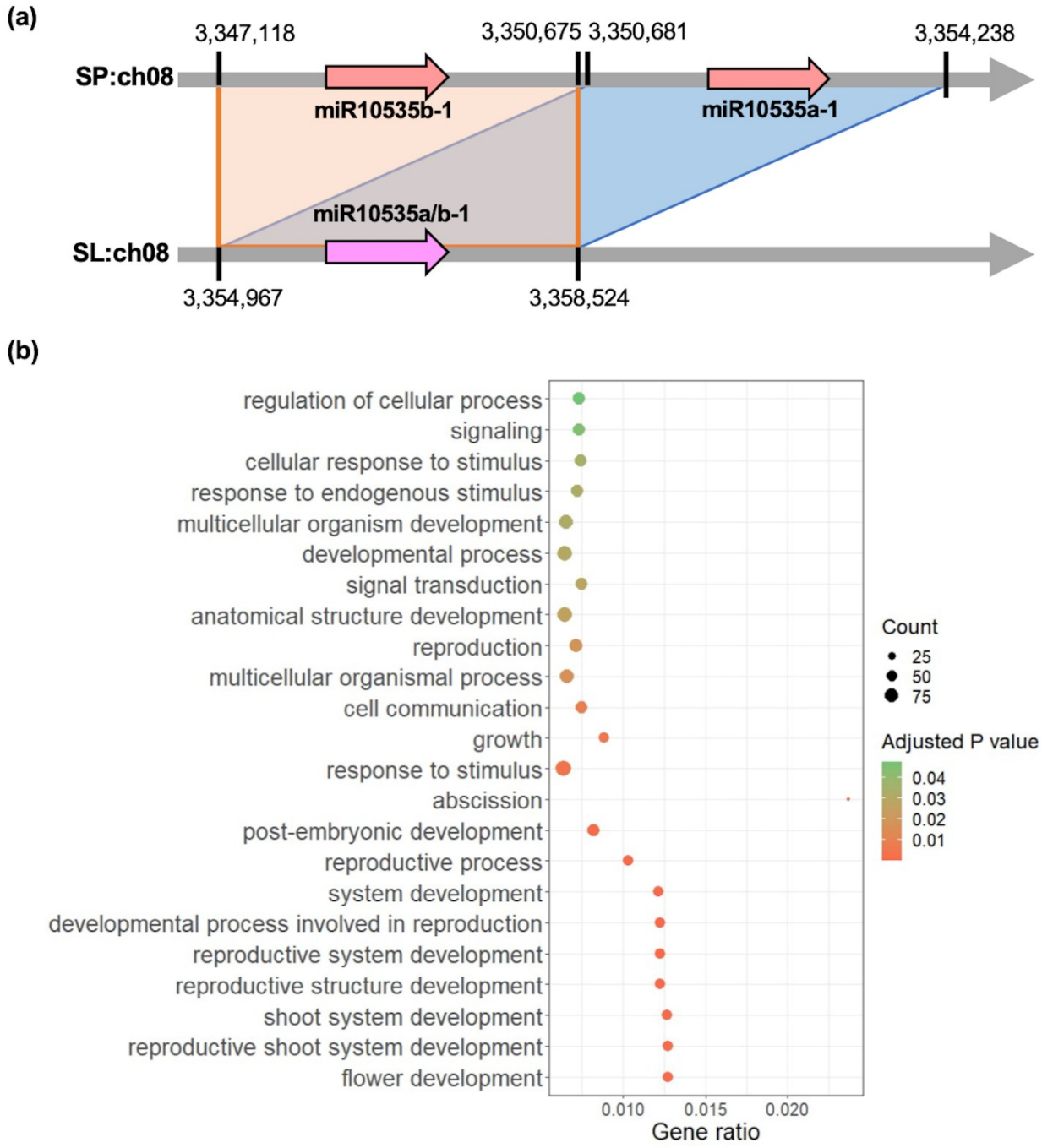

(a)
PHAS loci
SV-related PHAS loci
(b)
SP ch12
SLL ch12
62,019,441
62,022,686
58,541,475
SV90890
58,694,095
Gene
Gene structure
PhasiRNA location
Solyc12g044180.3
Solyc12g044190.3
Solyc12g044200.3
(c)
SV90890
Allele frequency
LA2093
Heinz
SP
SLC
Heirloom
Modern
(d)
phasiRNA expression
(log2 (TPM+1))
BGV007151 (SP)
BGV006370 (SP)
PI 129026 (SLC)
BGV007023 (SLC)
BGV007990 (SLC)
BGV008189 (SLC)
BGV008219 (SLC)
BGV005895 (SLC)
BGV007863 (SLL)

Figure 2



769

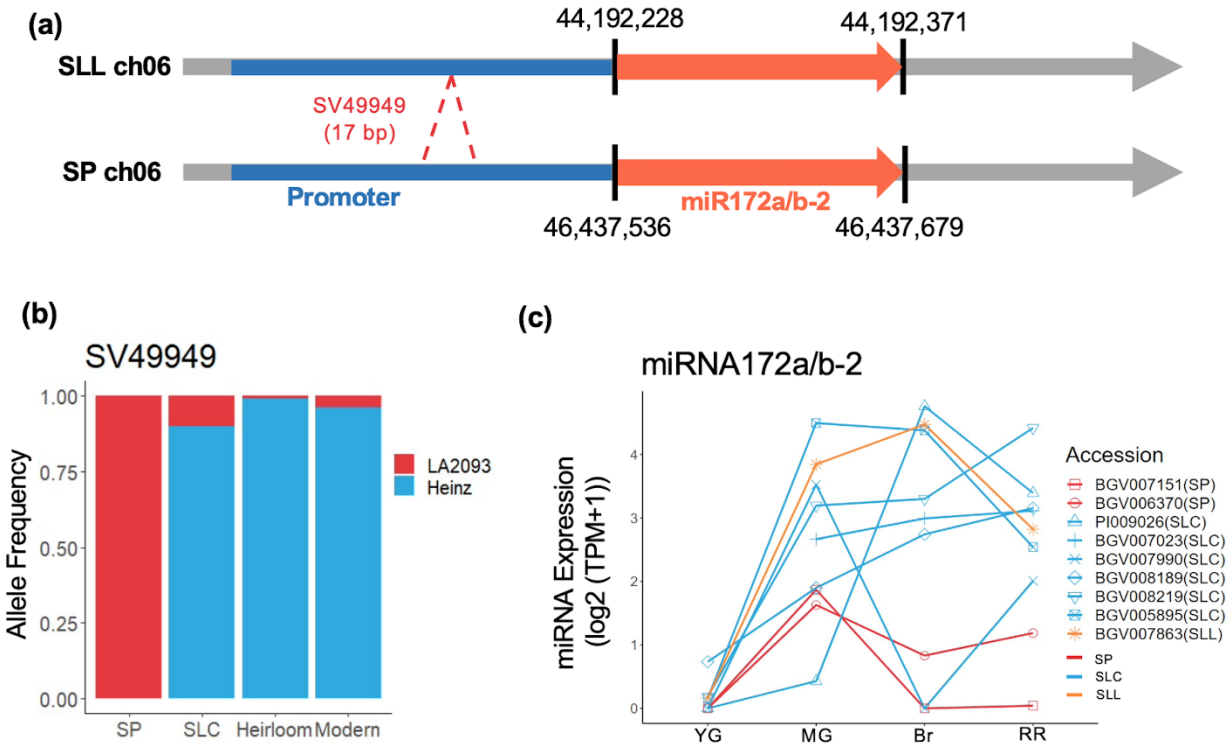
770 Figure 3



771

772 Figure 4

773



774

775 Figure 5

776

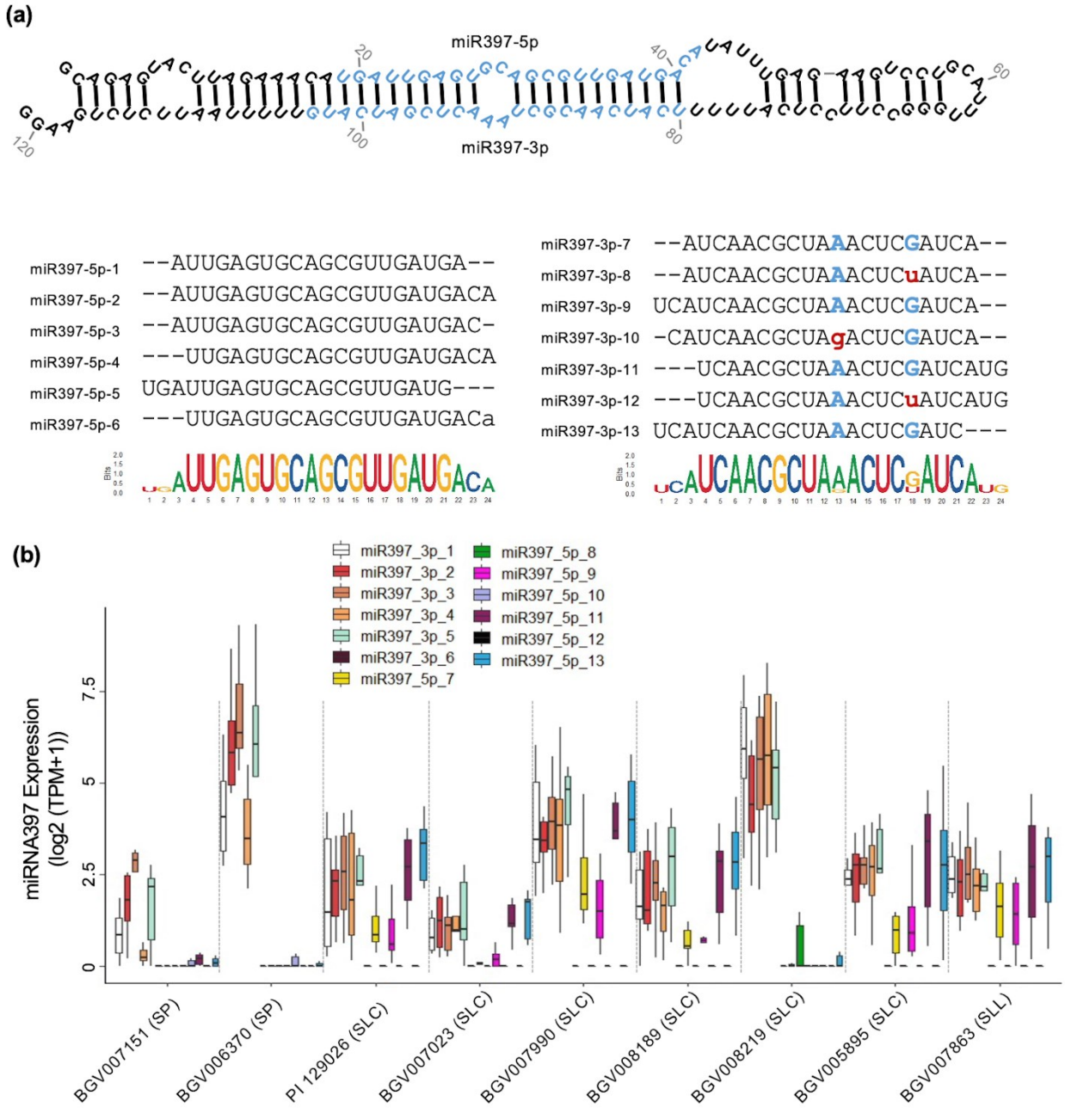

(a)
miR397-5p
miR397-3p
miR397-5p-1 --AUUGAGUGCAGCGUUGAUGA--
miR397-5p-2 --AUUGAGUGCAGCGUUGAUGACA
miR397-5p-3 --AUUGAGUGCAGCGUUGAUGAC-
miR397-5p-4 ---UUGAGUGCAGCGUUGAUGACA
miR397-5p-5 UGAUUGAGUGCAGCGUUGAUG---
miR397-5p-6 ---UUGAGUGCAGCGUUGAUGACa
miR397-3p-7 --AUCAACGCUAAACUCGAUCA--
miR397-3p-8 --AUCAACGCUAAACUCuAUCA--
miR397-3p-9 UCAUCAACGCUAAACUCGAUCA--
miR397-3p-10 -CAUCAACGCUAgACUCGAUCA--
miR397-3p-11 ---UCAACGCUAAACUCGAUCAUG
miR397-3p-12 ---UCAACGCUAAACUCuAUCAUG
miR397-3p-13 UCAUCAACGCUAAACUCGAUC---
Bits
(b)
miR397_3p_1
miR397_3p_2
miR397_3p_3
miR397_3p_4
miR397_3p_5
miR397_3p_6
miR397_5p_7
miR397_5p_8
miR397_5p_9
miR397_5p_10
miR397_5p_11
miR397_5p_12
miR397_5p_13
miRNA397 Expression (log2 (TPM+1))
0
2.5
5
7.5
BGV007151 (SP)
BGV006370 (SP)
PI 129026 (SLC)
BGV007023 (SLC)
BGV007990 (SLC)
BGV008189 (SLC)
BGV008219 (SLC)
BGV005895 (SLC)
BGV007863 (SLL)

Figure 6