1 **Large-scale comparative small RNA analyses reveal genomic structural variants in driving**

2 **expression dynamics and differential selection pressures on distinct small RNA classes**

3 **during tomato domestication**

4

5 You Qing[1,2], Yi Zheng[1,2,3,*], Sizolwenkosi Mlotshwa[4,10], Heather N. Smith[5,11], Xin Wang[3], Xuyang

6 Zhai[1,2], Esther van der Knaap[6,7,8], Ying Wang[4,5,*], Zhangjun Fei[3,9,*]

7

8 [1]Beijing Key Laboratory for Agricultural Application and New Technique, College of Plant

9 Science and Technology, Beijing University of Agriculture, Beijing, China 102206

10 [2]Bioinformatics Center, Beijing University of Agriculture, Beijing, China 102206

11 [3]Boyce Thompson Institute, Cornell University, Ithaca, NY 14853

12 [4]Department of Molecular Genetics, Ohio State University, Columbus, OH 43210

13 [5]Department of Biological Sciences, Mississippi State University, Starkville, MS 39759

14 [6]Center for Applied Genetic Technologies, University of Georgia, Athens, GA, 30602

15 [7]Institute for Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, 30602

16 [8]Department of Horticulture, University of Georgia, Athens, GA, 30602

17 [9]USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, 14853

18 [10]Present address: Department of Plant Pathology, Ohio State University, Wooster, OH 44691

19 [11]Present address: Department of Biological Sciences, Louisiana State University, Baton Rouge,

20 LA 70803

21 [*]Correspondence: yz@moilab.net, wang@biology.msstate.edu, zf25@cornell.edu

Abstract

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, through 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-nt heterochromatic siRNAs (hc-siRNAs). Changes in the abundance of phasiRNAs and 24-nt hc-siRNAs likely contribute to the alteration of mRNA gene expression during tomato's recent evolution, 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, highlighting different selection pressures on miRNAs compared to phasiRNAs and 24-nt 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.

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 [1-3]. Along the way, selections had been made for larger fruit, enhanced flavor, and improved resistance to biotic and abiotic stresses [4]. These phenotypic changes reflect the alterations in gene sequences and expression.

Recent evidence has demonstrated that genomic structural variants (SVs) are strongly associated with selection pressure over the course of tomato's recent evolution that impact the expression of genes underlying certain agronomic traits [5, 6]. SVs include insertions, deletions, duplications, inversions and translocations, and many of them serve as the causative genetic variants for diverse crop traits that have been selected during domestication [5, 6]. 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 [5]. 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-nt heterochromatic siRNAs (hc-siRNAs) [7, 8]. MiRNAs, phasiRNAs and hc-siRNAs are generated by Dicer-like enzymes (DCLs): DCL1, DCL4, and DCL3, respectively [9]. After production, they are loaded into the RNA-induced silencing complex (RISC) for function [10]. In plants, miRNAs and phasiRNAs with a length of 21 or 22 nt mainly guide cleavage of target mRNAs. By contrast, 24-nt hc-siRNAs play a major role in RNA-directed DNA methylation to confer epigenetic regulation over gene expression [7]. It is known that sRNA-based regulation relies on the abundance of the sRNAs [11]; therefore sRNA abundance has a significant impact on the functions.

In plants, miRNAs and phasiRNAs among distinct species are evolutionarily fluid [12, 13]. 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 that exhibit high divergence in hairpin structures, processing fidelity, and target complementarity [14, 15]. 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

3

70 diverse species from green algae to flowering plants [12, 13, 16, 17]. However, whether and how
71 the expression patterns and functions of sRNAs have been changed in shorter evolutionary times
72 such as during crop domestication is unknown. Tomato evolved from a wild red-fruited progenitor
73 species, *Solanum pimpinellifolium* (SP) approximately 80 thousand years ago into *S. lycopersicum*
74 var. *cerasiforme* (SLC) [3, 18]. Semi-domesticated SLC further evolved into the fully
75 domesticated tomato, *S. lycopersicum* var. *lycopersicum* (SLL). The evolution within the red-
76 fruited tomato clade provides a unique system for studying selection and functional divergence of
77 plant sRNAs in a shorter time scale thanks to its extensive genetic and genomic resources. Analysis
78 of samples from wild and domesticated tomato accessions may reveal novel regulatory details
79 underlying sRNA evolution and function.

80 Here, we present comprehensive sRNA profiles of cultivated tomatoes and their wild
81 progenitors, and the novel discovery that genomic SVs can substantially influence sRNA
82 expression dynamics. Our findings show that SVs are an important driving force for the dynamic
83 expression of sRNAs. Moreover, we show that SVs can change the hc-siRNA hotspots in
84 promoters of nearly 100 protein-coding genes, thereby altering their expression. These genes are
85 mostly associated with responses to biotic and abiotic stresses. SVs are also correlated with many
86 rapid birth and death of PHASs that are overwhelmingly related to disease resistance traits. SVs
87 overlapping with miRNA genes can determine the gain or loss of certain less-conserved miRNA
88 genes or affect the expression of miRNAs. Interestingly, the differential expression of miRNAs
89 has a neglectable effect on transcriptomes, in contrast to the changes in mRNA expression
90 regulated by the dynamics of hc-siRNA hotspots. Our findings unravel SV-related differential
91 expression of 24-nt hc-siRNAs regulating the expression of certain genes associated with
92 responses to biotic and abiotic stresses as well as differential selection pressures over distinct
93 classes of sRNAs during tomato domestication. Our dataset is also valuable to promote other
94 sRNA-related functional studies on tomato development and domestication.

95

96 **Results**
97 **Comprehensive tomato sRNA profiles highlighting expression dynamics as a consequence of**
98 **domestication**
99 Resulting from domestication, tomatoes underwent substantial changes in plant morphology, yield,
100 fruit flavor and adaptation to adverse environments, reflecting certain levels of adjustments in both

4

101   genomes and transcriptomes. We selected nine accessions spanning from wild ancestors (*Solanum*

102   *pimpinellifolium*), semi-domesticated populations (*S. lycopersicum* var. *cerasiforme*) to

103   domesticated tomatoes (*S. lycopersicum* var. *lycopersicum*) for comprehensive transcriptome

104   analyses. These accessions included two SP accessions (BGV006370 collected in Peru and

105   BGV007151 collected in Ecuador), six SLC accessions (BGV005895, BGV007023 and PI 129026

106   collected in Ecuador; BGV007990 and BGV008189 collected in Peru; BGV008219 collected in

107   Costa Rica), and one SLL accession (BGV007863 collected in Mexico) [1, 3].

108   For each accession, transcriptome profiles including both sRNA and mRNA profiles were

109   investigated in young leaves, anthesis-stage flowers, and fruits at four different developmental

110   stages (young green, mature green, breaker and red ripe). The four fruit developmental stages have

111   previously been used to analyze the gene regulatory networks in a single accession of domesticated

112   tomato [19, 20]. Therefore, our comprehensive dataset would empower detailed dissection on the

113   gene regulatory networks underlying the fruit ripening process in addition to the transcriptome

114   profiles of leaf and flower.

115   To ensure sampling at comparable developmental stages of different tomato accessions,

116   we first documented the timing of fruit developments by tracking 10-20 fruits from 3-5 plants of

117   each accession. As shown in **Fig. S1**, the two SP accessions exhibited slightly early ripening

118   whereas SLL ripened one week later. Ripening time in SLC accession varied from as early as SP

119   to as late as SLL. We collected samples at the chosen time points and constructed a total of 162

120   sRNA-Seq libraries (**Table S1**). The three replicates from young green fruits of BGV007023 did

121   not pass quality check, so we only used the remaining 159 libraries for the subsequent analyses.

122   Principal component analysis (PCA) showed that samples at the same developmental stage were

123   clustered together (**Fig. S2A**). In addition, expression profiles in SLC and SLL accessions were

124   more closely related in comparison to profiles in SP accessions (**Fig. S2B**). Analysis on sRNA size

125   distribution showed that 24-nt siRNAs were the most dominant (**Fig. S2C**). Interestingly, 24-nt

126   siRNAs were slightly more abundant in leaf and young green fruit samples (**Fig. S2C**). To improve

127   the sRNA mapping accuracy, we exploited two high-quality reference genomes (genomes of the

128   domesticated Heinz 1706 and an SP accession LA2093) as detailed in a recent study [5]. It is

129   noteworthy that as expected, reads from SP samples were mapped to the LA2093 genome with a

130   slightly higher rate, while reads from SLC and SLL samples were mapped to the Heinze SL4.0

131 with a slightly higher rate (**Fig. S2D**). RNA-Seq libraries were also generated from the same

132 samples and described in our previous study [5, 21].

133

134 **Hc-siRNA hotspots and structural variants**

135 The 24-nt hc-siRNAs play a major role in RNA-directed DNA methylation, a fundamental

136 mechanism in epigenetic regulation [7, 8]. We reasoned that comparative analyses on hc-siRNA

137 accumulation patterns may provide insights into the dynamic epigenetic changes underlying trait-

138 related gene expression. When analyzing the global hc-siRNA abundance across the 12 tomato

139 chromosomes, we noticed that hc-siRNA abundance displayed a strong genome-wide correlation

140 with the density of SVs (**Fig. 1A**). This observation infers a novel model that epigenetic regulation

141 may be influenced by SVs during crop domestication, which could lead to large scale gene

142 expression changes.

143 To obtain more evidence in support of this model, we analyzed SV-overlapping hc-siRNA

144 hotspots in promoter regions of protein-coding genes, and identified hc-siRNA and protein-coding

145 gene pairs whose abundances showed simultaneous negative correlations in our sRNA-Seq and

146 RNA-Seq data. For each tissue or development stage, we performed pairwise comparisons

147 between SP and SLL accessions (2 comparisons), between SP and SLC accessions (12

148 comparisons except 10 for young green fruit stage), and SLC and SLL accessions (6 comparisons

149 except 5 for young green fruit stage), therefore a total of 117 pairwise comparisons for all six

150 tissues/developmental stages. A total of 1,386 protein-coding genes affected by the SV-

151 overlapping hc-siRNA hotspots were identified in at least one comparison. We noticed that most

152 of these genes under the control of this novel epigenetic regulation were expressed in leaf and

153 flower (**Fig 1B**). Gene ontology (GO) term analysis showed that the majority of the protein-coding

154 genes affected by these domestication-associated epigenetic changes were related to pathways in

155 response to biotic and abiotic stresses as well as development and several developmentally related

156 processes (**Fig. 1C**), implying a selection pressure favoring expression changes in genes related to

157 environmental adaptation including selection for domestication traits. To obtain highly confident

158 negative correlations between differentially accumulated 24-nt hc-siRNAs and the cognate

159 protein-coding genes, we focused on the pairs of hc-siRNAs and the negatively correlated protein-

160 coding genes repeated in at least ten pairwise comparisons, which resulted in the identification of

161 99 protein-coding genes (**Table S2**). The majority of these 99 genes are involved in plant resistance
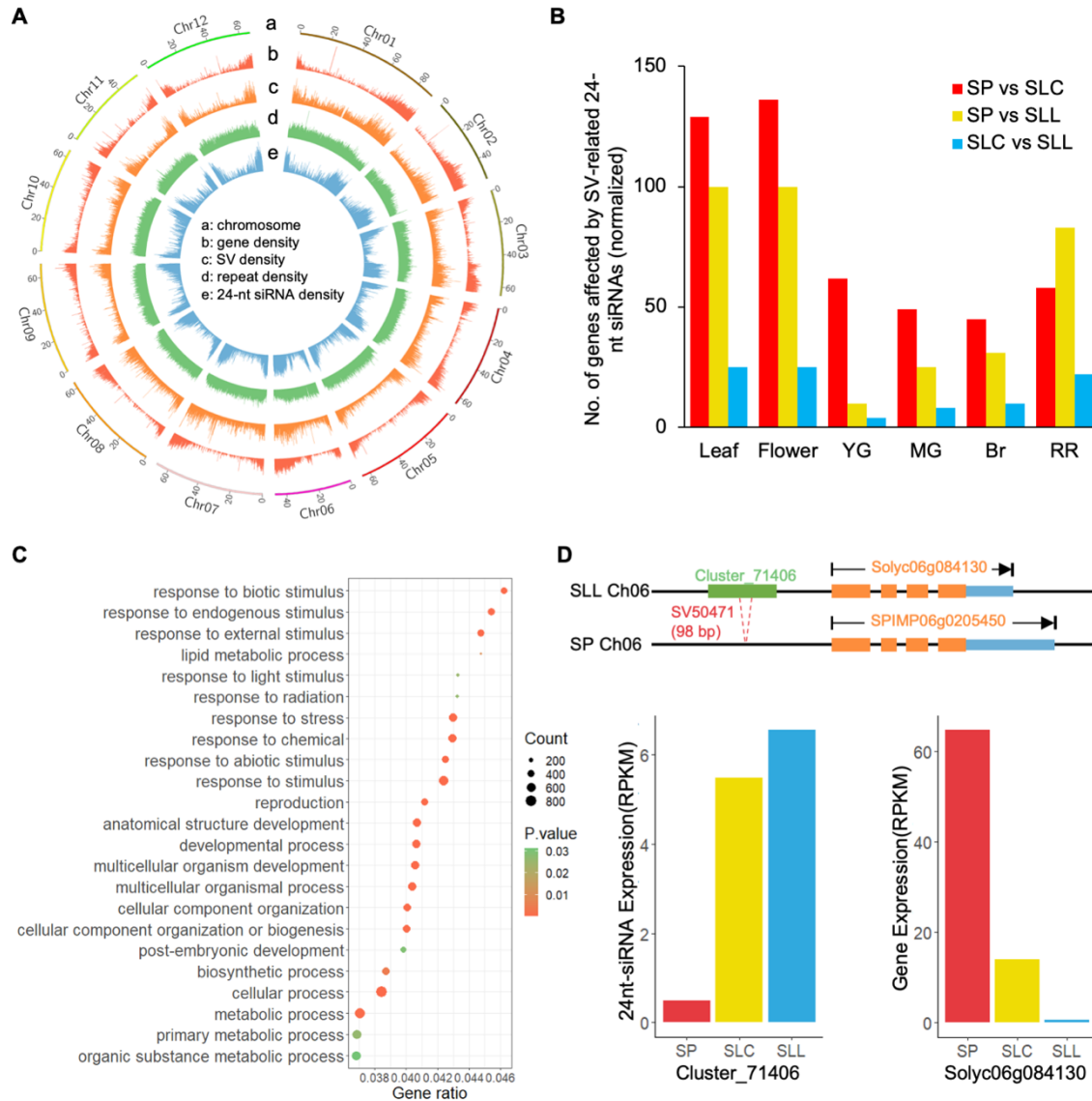
Figure 1. Functional analysis of SV-related 24-nt hc-siRNA regions among wild and cultivated tomatoes. **A.** Circos plot of the densities of 24-nt hc-siRNAs, genes, repeat sequences, and SVs across the tomato genome. **B.** Differentially expressed genes showing a negative correlation with the corresponding 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. **C.** GO term enrichment analysis of differentially expressed genes in (**B**). **D.** 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*).

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.

7

176   A notable example is a BAX inhibitor-1 (BI-1) family gene (*Solyc06g084130*). BI-1
177   proteins are conserved in eukaryotic organisms and associated with cell death during host-
178   pathogen interactions [22-24]. In plants, BI-1 proteins regulate the autophagy process [23] and
179   confer plant resistance to various pathogens [23, 25-27]. In particular, BI-1 expression regulates
180   autophagic activity that is critical for *N* gene-mediated resistance to tobacco mosaic virus [23], a
181   major viral pathogen of tomato. The repression of *Solyc06g084130* expression was strongly
182   associated with SV-related 24-nt hc-siRNA differential accumulation from SP to SLC and SLL
183   plants (**Fig. 1D**). Similar repression patterns could be found in multiple genes related to responses
184   to biotic stresses (**Table S2**). Therefore, the repression of *Solyc06g084130* and other genes in
185   response to biotic stresses during tomato domestication may affect tomato resistance traits, which
186   is consistent with the observation that cultivated tomatoes are not well adapted to adverse
187   environments as wild tomatoes.

188

189   **Highly dynamic gain/loss of phasiRNA generating loci during domestication**
190   PhasiRNAs are known regulators of gene expression in plants [28] and exhibit expression
191   dynamics in response to environmental cues [13]. Current models suggest that phasiRNAs serve
192   as negative regulators to modulate the expression of their parental transcripts [13, 28]. Although
193   there is rapid progress in uncovering an enormous amount of phasiRNA-generating loci in various
194   plants [13, 28] and unraveling their functions in plant development [29-31], a detailed analysis of
195   phasiRNA dynamics during crop domestication has not been conducted. Using a previously
196   established algorithm [32], we analyzed phasiRNA-generating loci across all nine accessions and
197   identified 290 PHASs mapped to the Heinz 1706 genome (SL4.0) and 286 mapped to the LA2093
198   genome, among which 77 were uniquely mapped to SL4.0 and 73 uniquely mapped to the LA2093
199   genome (**Tables S3-5**). In general, PHASs were mainly mapped to protein-coding genes with
200   diverse functions as shown in the GO term analysis (**Fig. S3A**), akin to our previous findings [13].
201   Interestingly, we found 34 SV-overlapping PHASs mapped to SL4.0 and 39 mapped to the
202   LA2093 genome, among which 12 were uniquely mapped to SL4.0 and 17 uniquely mapped to
203   the LA2093 genome (**Table S6**). SV-related PHASs were distributed across all chromosomes
204   except chromosome 4 and 7, and there were numerous SV-related PHASs residing proximal to
205   the terminal region of the long arm of chromosome 11 (**Fig. 2A**). SVs markedly contributed to the
206   changes in PHASs among accessions in different tomato groups (**Fig. S3B**), and these PHASs
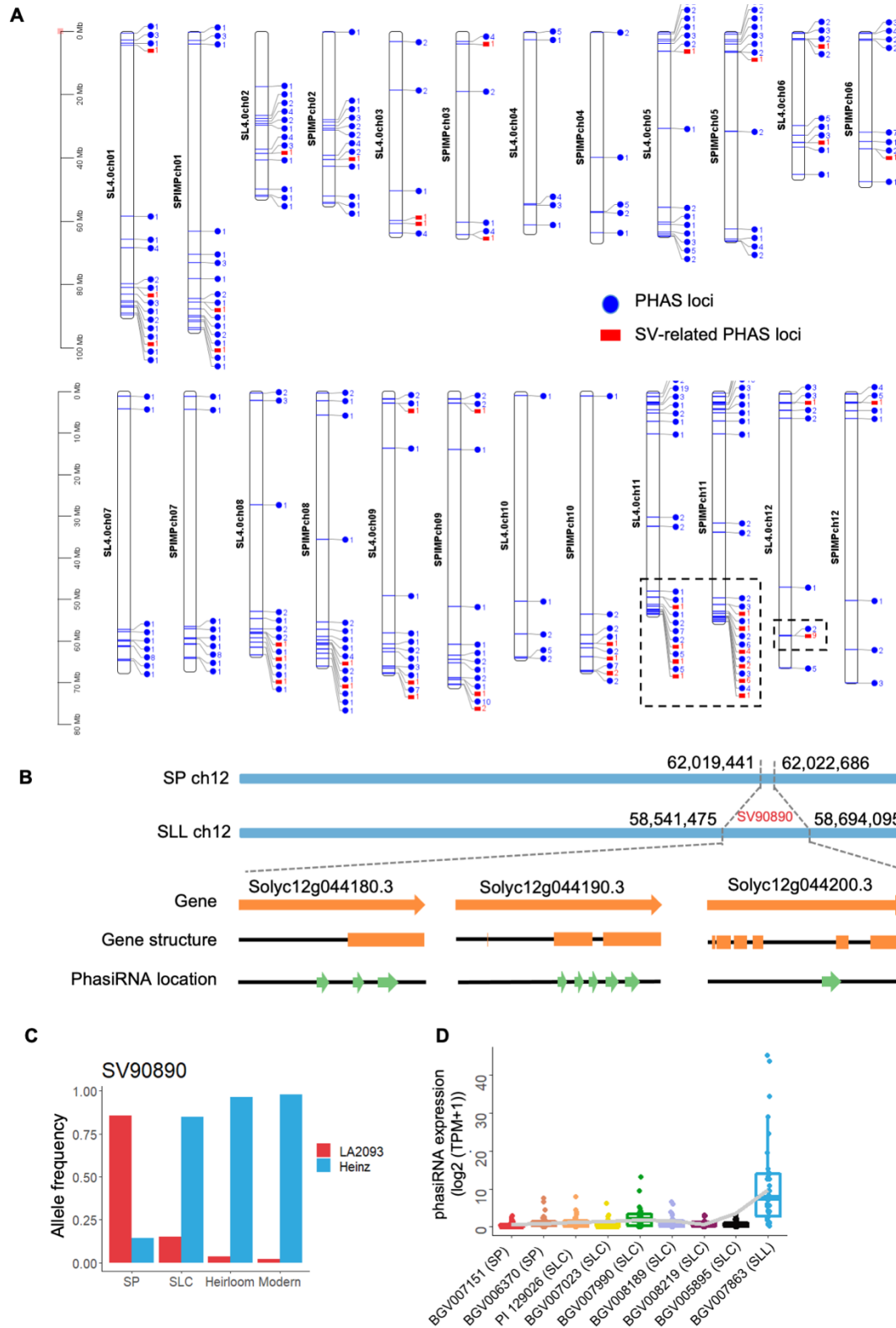
Figure 2. SV-related PHASs. **A.** Distribution of PHAS loci and SV-related PHAD loci across the 12 tomato chromosomes. The number depicts the distinct PHASs at each locus. **B.** SV90890 causes gain/loss of PHASs and protein-coding genes. **C.** Allele frequency of SV90890 in different tomato groups. **D.**

9

211 Abundance of phasiRNAs in the genome region of SLL Heinz 1706 containing SV90890 in the nine tomato
212 accessions. For each box plot, the lower and upper bounds of the box indicate the first and third quartiles,
213 respectively, and the center line indicates the median. The whisker represents 1.5× interquartile range of
214 the lower or upper quartile.
215

216 were overwhelmingly mapped to disease resistance genes (**Fig. S3C**). This observation hints the

217 possibility that selection pressure favors the emergence of phasiRNAs regulating the expression

218 of disease resistance genes in balancing growth and pathogen defense.

219   One cluster of disease resistance genes resided in a region where an insertion in the Heinz

220 1706 genome expanded the PHASs on chromosome 12 (**Fig. 2B**). All three genes in this inserted

221 region are involved in tomato resistance to bacterial and oomycete pathogens [33]. As shown in

222 **Fig. 2C**, the frequency of the insertion associated with this gene cluster drastically increased in

223 domesticated tomato to nearly 100%. Productions of phasiRNAs associated with this SV were also

224 highly elevated in the SLL accession (**Fig. 2D**).

225

226 **A large portion of miRNAs exhibit highly dynamic expression patterns**

227 The functions of many miRNAs in tomato growth and fruit development have been well studied

228 in single accession analyses [20, 34, 35]. However, detailed analyses on miRNA expression

229 profiles are lacking to infer the dynamics of miRNA-based gene regulatory network during crop

230 domestication. To this end, we annotated all the miRNAs in our sRNA-Seq dataset based on

231 recently revised criteria [36]. We identified 122 miRNA genes mapped to SL4.0 and 126 mapped

232 to the LA2093 genome. Both sets included 72 previously reported tomato miRNAs [20, 37, 38].

233 There were 116 miRNA genes mapped to both SL4.0 and the LA2093 genome, while there were

234 six and ten miRNA genes specifically mapped to SL4.0 and the LA2093 genome, respectively

235 (**Table S7**). Based on the mapping results, we summarized the mature miRNAs and miRNA*s as

236 well as the processing variants from the miRNA precursors in **Table S8-9**.

237   A close look at the mature miRNAs showed that the majority of known miRNAs were 21-

238 nt in length with "U" as the first nucleotide (**Fig. S4**), in line with previous observations [12].

239 Notably, the majority of novel miRNAs were 20-nt in length with "A" as the first nucleotide (**Fig.**

240 **S4**), which implied that those miRNAs could lack regulatory functions. To our surprise, we found

241 that a large portion of miRNAs exhibited highly variable expression patterns among accessions.

242 For example, as shown in **Table S10**, the two SP accessions each has ~120 miRNAs that displayed

10

243   1.5-fold changes in expression with an adjusted P value below 0.05 compared with SLC and SLL

244   accessions.

245   To better describe this dynamic, we plotted miRNA expression profiles across the nine

246   tomato accessions and categorized the patterns into multiple groups. As shown in **Fig. 3**, there

247   were eight groups with each having more than ten distinct miRNAs. While the majority of the

248   miRNAs exhibited dynamic expression profiles in SLC accessions, we found that those in groups

249   1 and 7 exhibited an overall decrease while those in groups 5, 6, and 8 exhibited an overall increase

250   in expression over the course of tomato's recent evolution. The presence of eight distinct groups

251   also reflected the fluctuating expression patterns of miRNAs among different tomato accessions.

252   However, we did not observe the corresponding changes in the expression of their targets, which

253   will be further analyzed below.

254

255   **miRNAs and structure variants**

256   Despite the notion that most miRNA gene families had the same number of members mapped to

257   SL4.0 and the LA2093 genome, some underwent deletion or duplication events that changed the

258   number of members in each family. For example, there were two miR10535 genes in the LA2093

259   reference genome but only one copy in Heinz 1706 SL4.0 (**Fig. 4A**). Synteny analysis showed that

260   the loss of one miR10535 gene copy could be attributed to a deletion that had occurred during

261   tomato domestication (**Fig. 4A**).

262   Since a genome deletion caused the loss of one copy of miR10535 gene, we reasoned that

263   SVs might play important roles in determining the presence and the expression levels of miRNA

264   genes. To this end, we identified 25 miRNA genes associated with 32 SVs, including 19 conserved

265   and six novel miRNAs. Twenty-one out of 32 SVs were mapped to promoters of miRNA genes,

266   seven were mapped to miRNA gene bodies, and four were mapped to both promoters and gene

267   bodies of miRNAs (**Table S11**). Notably, most of these miRNAs do not have empirically

268   confirmed targets according to previous degradome studies in tomato [20, 39-41]. GO term

269   analysis showed that the computationally predicted target genes of SV-overlapping miRNAs were

270   mainly involved in the development/growth and responses to environmental stimuli (**Fig. 4B**).

271   SVs in promoter regions may affect miRNA expression. For example, a 17-bp deletion

272   (SV49979) in the promoter of the miR172a/b-2 gene was observed in genomes of most heirloom

273   and modern tomatoes (**Fig. 5A**). It is worth noting that the allele frequency of SV49949
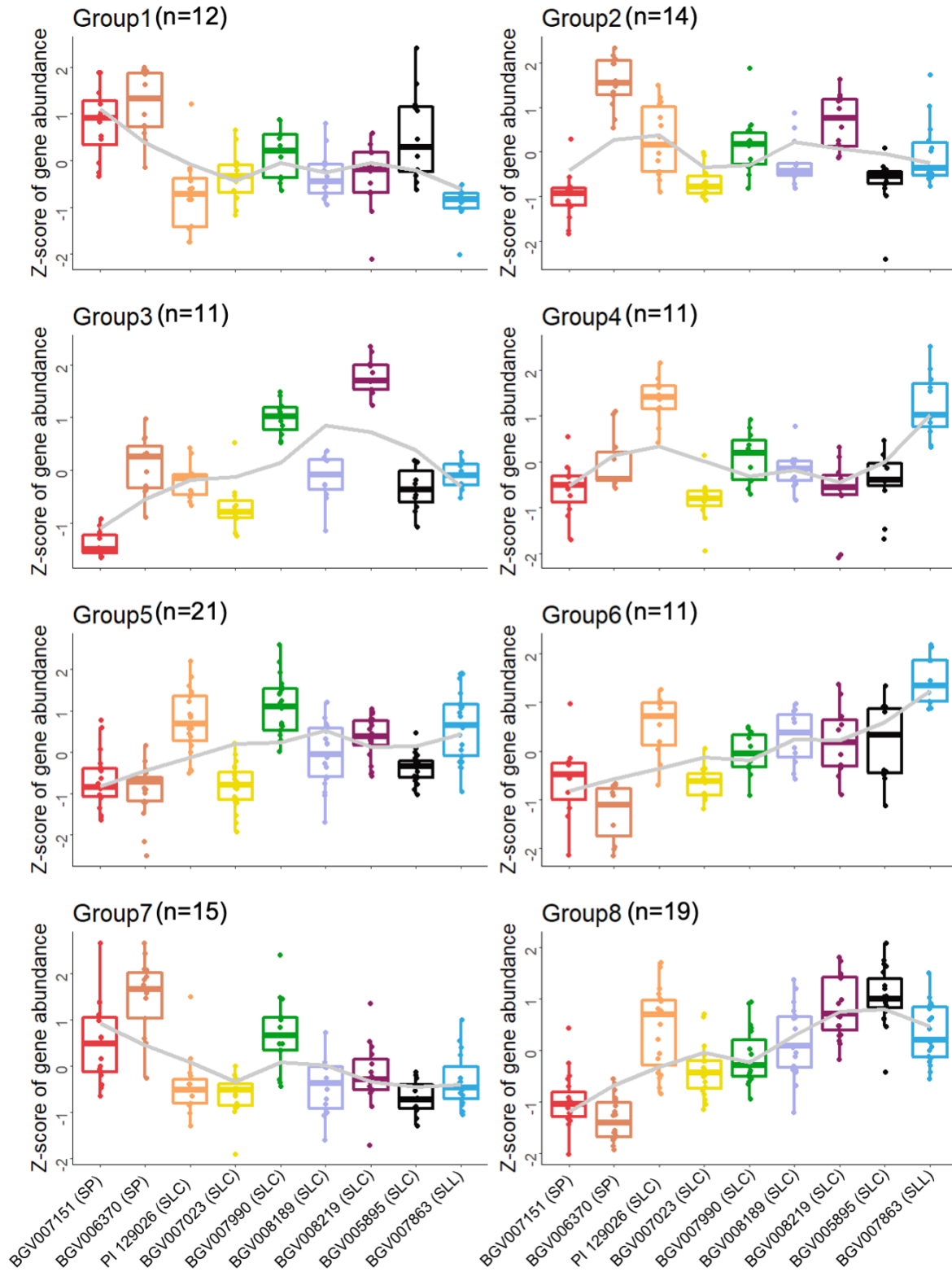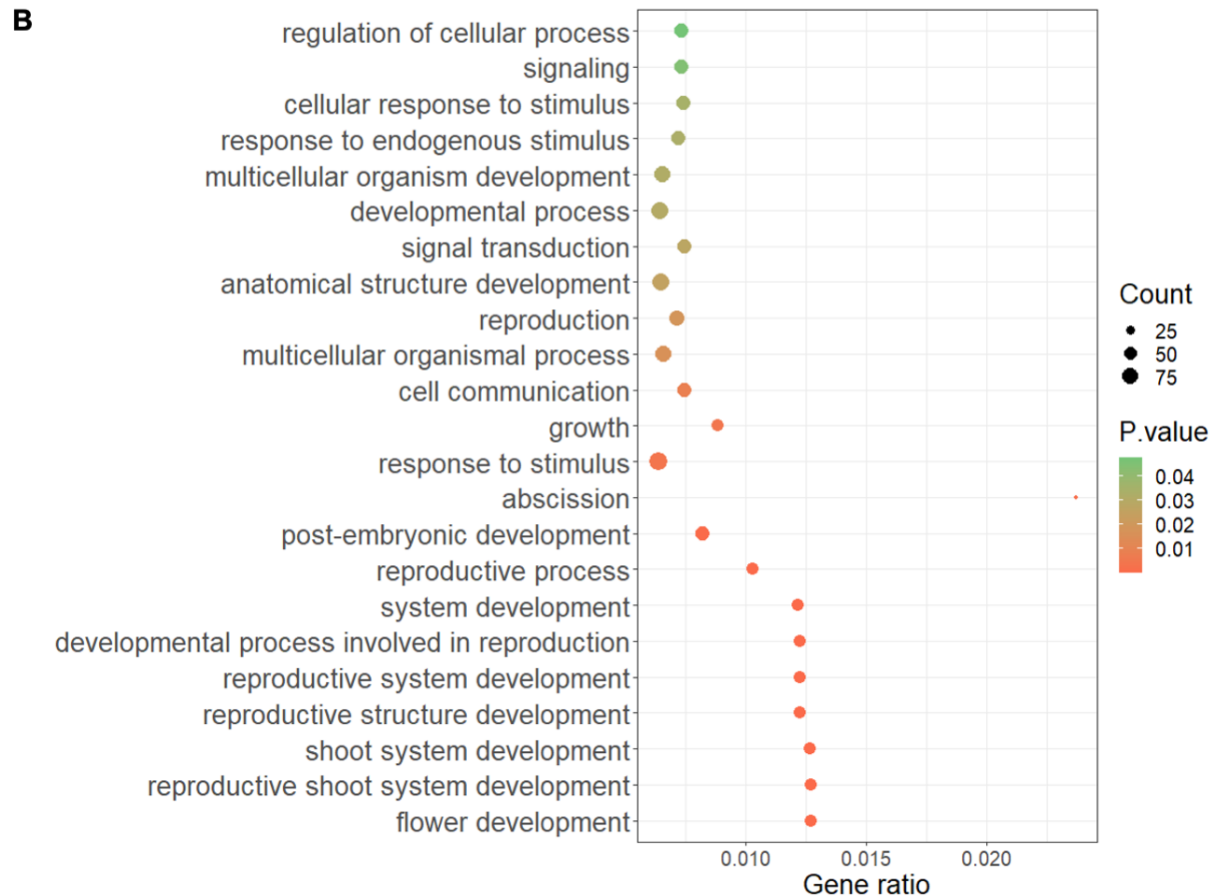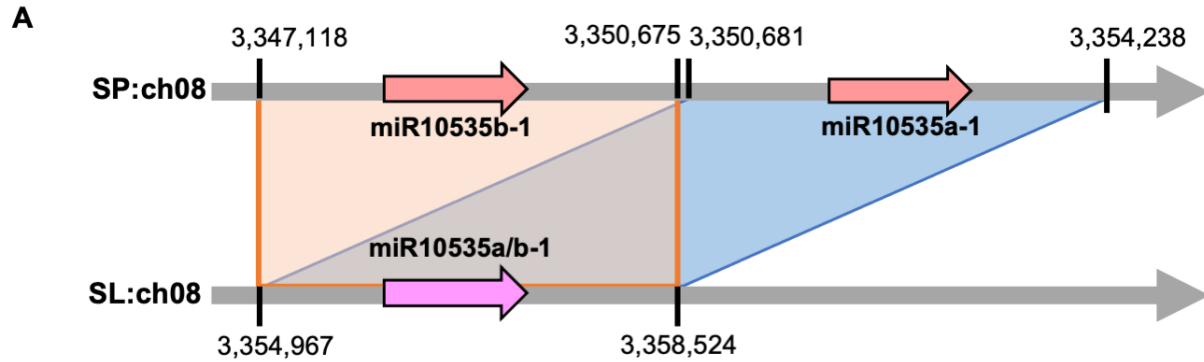
Figure 3. Distinct expression patterns of miRNAs. For each box plot, the lower and upper bounds of the box indicate the first and third quartiles, respectively, and the center line indicates the median. The whisker represents 1.5× interquartile range of the lower or upper quartile.

278
279 **Figure 4.** MiRNA genes overlapped with SVs. **A.** Synteny diagram of miR10535 gene(s) in SP and SLL
280 reference genomes. **B.** Enriched GO terms of predicted target genes of SV-related miRNAs.
281

282 significantly changed during domestication (**Fig. 5B**). Correspondingly, miR172a/b-2 had distinct

283 expression profiles during fruit development among the nine tomato accessions and showed an

284 overall increased expression pattern in the SLC and SLL accessions (**Fig. 5C**). Due to the low

13

285   abundance of miR172a/b-2 compared to the total miR172 abundance (**Tables S8 and S9**), this

286   expression change in miR172a/b-2 did not significantly affect the expression of any miR172

287   targets.

288
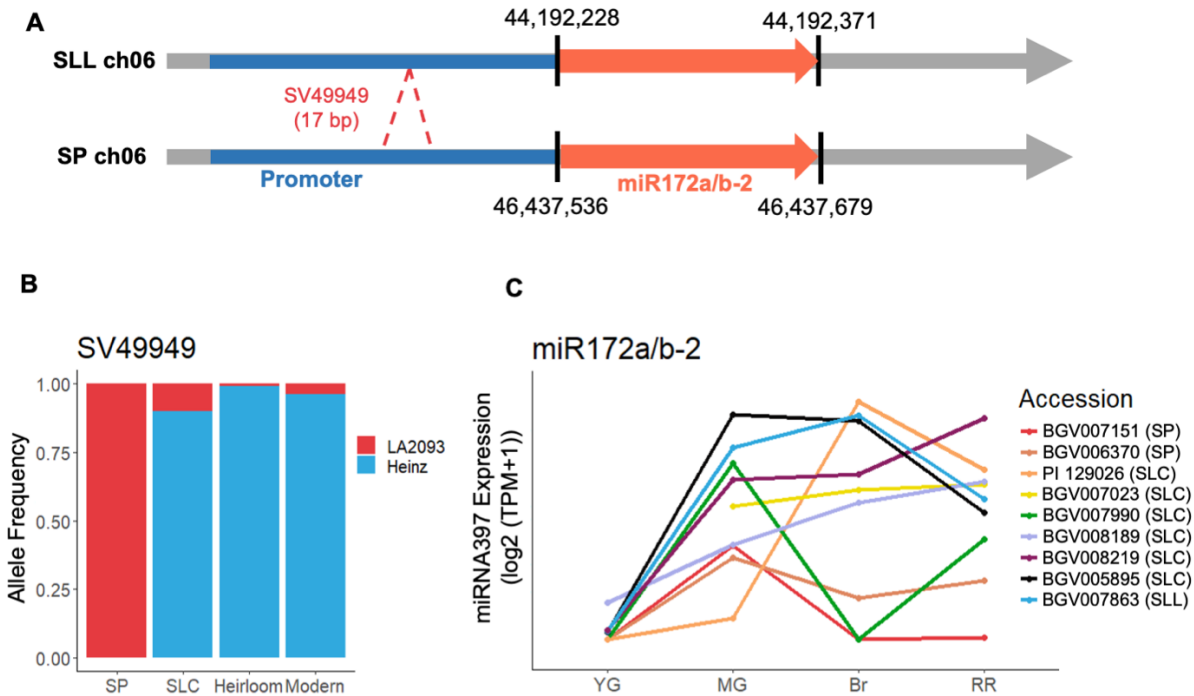
290   **Figure 5.** SV affects miRNA expression. **A.** Diagram showing SV49949 (a 17-bp indel) in the
291   promoter of miR172a/b-2 gene. **B.** Allele frequency of SV49949 in different tomato groups. **C.**
292   Expression profiles of miR172a/b-2 in different tomato accessions. YG, young green fruits. MG,
293   mature green fruits. Br, fruits at the breaker stage. RR, red ripe fruits.

294

Imprecisely processed miRNAs are differently selected

296   It is known that some miRNA precursors tend to generate a population of miRNA-miRNA* pairs

297   due to the imprecise processing by DCL1 [42]. The imprecise processing may affect miRNA

298   function that heavily relies on sequence complementarity between miRNAs and their targets [43].

299   We found eight out of 122 SLL miRNA genes and six out of 126 SP miRNA genes that generated

300   more than six variants in our dataset, among which six were shared in both SLL and SP reference

301   genomes (**Table S12**). The miR397 gene, a conserved miRNA across different plant lineages, had

302   six miR397-3P and seven miR397-5P variants in tomato (**Fig. 6A**). Notably, we observed a shift

303   in the most abundant product of the miR397 precursor: from only miR397-3P that was expressed

304   in SP accessions transitioning to miR397-5P (the conserved mature miR397 in plants) that became

14

305 more prevalent in most SLC and SLL accessions (**Fig. 6B**). Over-expression of miR397-5P can

306 enhance tomato response to drought [44], suggesting the beneficial function of miR397-5P in SLC

307 and SLL accessions in adaptation to adverse environments. Nevertheless, the target of miR397-5P

308 remains unclear in spite of extensive efforts of degradome analysis [20, 39-41].

309
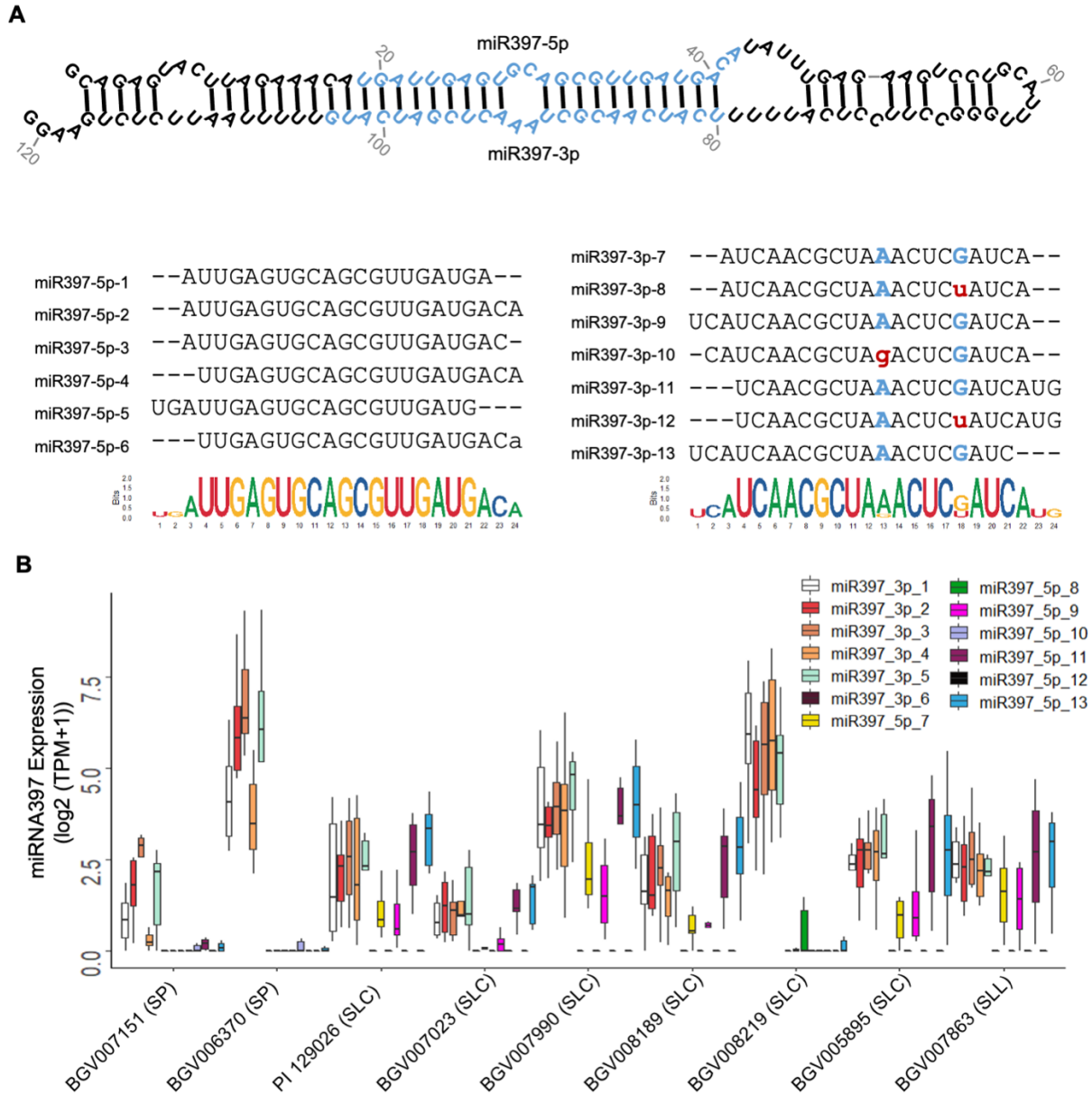
310

311 **Figure 6.** Changes in miR397 expression during tomato domestication. **A.** Diagram showing the six 5P and
312 seven 3P processing variants from the miR397 precursor. **B.** Abundance of each miR397 variant in different
313 tomato accessions.

314

315   We also observed a position shift of miRNA:miRNA* in the precursor of miR9472. The

316   miRNA:miRNA* duplex was closer to the terminal loop region in SP and SLC accessions but

317   resided at a more distant region away from the terminal loop in SLL (**Fig. S5**). This shift was

318   unlikely caused by any changes in recognition by DCL1 as the precursor sequences remained the

319   same. Notably, we found seven such examples as listed in **Table S13**. This observation indicates

320   that DCL1 recognition on miRNA precursors is likely flexible in plants, and the selection pressure

321   plays a role in determining the expression of the final miRNA:miRNA* duplexes.

322

323   **Discussion**

324   Small RNAs are critical regulators of gene expression underlying tomato growth and responses to

325   environmental cues [20]. sRNA abundance is known to directly impact their functions. To gain a

326   better understanding of sRNA dynamic expression during tomato domestication, we generated a

327   comprehensive sRNA dataset using nine representative tomato accessions spanning from the wild

328   SP progenitors, intermediate SLC accessions, and one domesticated accession, and covering

329   samples from leaf, flower, and fruits at four critical developmental stages. Our high-quality dataset

330   fulfills the immediate needs for high-resolution comparative analyses on sRNA expression and

331   inferring their functions in wild, semi-domesticated, and domesticated tomato plants, as well as

332   establishes a foundation for future exploration of sRNA functions.

333   Our dataset clearly demonstrates that all three classes of sRNAs (hc-siRNAs, phasiRNAs

334   and miRNAs) have significant changes in expression during tomato's recent evolution. Notably,

335   we found that SVs are an important driving force underlying the dynamic expression of these

336   sRNAs. Those SVs, particularly deletions and insertions, probably have a direct impact on the

337   production of all three major types of sRNAs. For example, deletions or insertions result in the

338   differential accumulation of hc-siRNAs in gene promoter regions and the gain or loss of PHASs.

339   SVs also contribute to the birth and death of miRNAs in domesticated tomatoes, as evidenced by

340   the deletion of miR10535 and miR482f. When SVs reside in promoter regions of miRNA genes,

341   they may influence miRNA expression, as evidenced in miR172a/b-2. In addition to SVs, we also

342   notice that imprecise processing of miRNA/miRNA* duplexes can lead to the dynamic expression

343   of miRNAs during tomato domestication.

344   Tomato has a complex history of domestication, selection and breeding [1-3, 18]. Tomato

345   domestication before cultivation possibly has a selection pressure on plant adaptation to new

346   environments and developmental processes. Domestication and re-domestication processes
347   possibly have posed a selection pressure on tomato flavor and yield. The modern breeding
348   processes possibly have posed a selection pressure on disease resistance. Interestingly, selection
349   pressures appear to have distinct impacts on different classes of sRNAs. The SV-related dynamic
350   expression of 24-nt hc-siRNAs and phasiRNAs predominantly impacts the expression of genes
351   related to stress responses and growth, implying that the selection pressure favors the regulation
352   of those trait-associated genes through hc-siRNA and phasiRNA pathways. In contrast, the
353   conserved miRNAs play a major role in plant development mostly through regulating transcription
354   factors [14, 15]. Some of those well-established regulations are conserved in most land plants [45-
355   47]. Therefore, the selection pressure is unlikely to favor changes in such critical regulations
356   related to plant growth in a relatively short timeframe during domestication. By contrast, many
357   less conserved miRNAs, mostly expressed at low levels and/or having no to very few confirmed
358   targets, exhibit highly dynamic expression during tomato domestication, reflecting little selection
359   pressure on those miRNAs [14, 15].

360   Our dataset serves as a foundation for future studies on sRNA function associated with
361   tomato growth, domestication, and beyond. Recent progress has demonstrated that miRNA gene
362   families may exert functions through developmentally-regulated expression of specific members
363   [48, 49]. Our dataset can help mining miRNA family members with developmentally-regulated
364   expression patterns. For example, miR390b, with a U20A substitution at position 20 in comparison
365   to miR390a, was specifically expressed in flowers (**Fig. S6**). The flower-specific expression of
366   miR390b tripled the amount of total miR390 in flowers, which markedly promoted production of
367   phasiRNAs from the *TAS3* locus as well as specifically suppressed the expression of *ARF3* and
368   *ARF4* in flowers (**Fig. S6**).

369

370   **Methods**

371   *Plant materials and RNA isolation*

372   Tomato plants were grown in a greenhouse at 25 °C and with a 16/8 hr light/dark cycle at Ohio
373   State University (Columbus, OH). For each accession, young leaves, anthesis-stage flowers, and
374   fruits at four different developmental stages (young green, mature green, breaker and red ripe)
375   were collected with three biological replicates. Total RNAs from tomato samples were isolated
376   and fractioned to >200 nt and <200 nt populations using the RNAzol RT reagent (Sigma-Aldrich,

17

377 St. Louis, MO). sRNA species were further purified using the miRVana miRNA isolation kit

378 (Thermo Fishier Scientific, Grand Island, NY) following the manufacturer's instructions. mRNA

379 populations were further purified using the Magnetic mRNA isolation kit (NEB, Ipswich, MA).

380

381 ***Library construction and sequencing***

382 sRNA libraries were constructed following the established protocol [50]. Briefly, 18-30 nt sRNA

383 populations purified on 15% (w/v) polyacrylamide/8M urea gel were ligated with 3'- and 5'-

384 adapters. sRNA populations with adapters were reverse transcribed, PCR amplified, and then

385 purified from the 8% native PAGE gel. Strand-specific RNA-Seq libraries were constructed using

386 the protocol described before [51]. All the constructed libraries were analyzed and quantified by

387 Bioanalyzer and sequenced on an Illumina HiSeq 2500 system.

388

389 ***sRNA sequence processing***

390 sRNA reads were processed to remove adaptors using the sRNA cleaning script provided in the

391 VirusDetect package [52]. The trimmed sRNA reads shorter than 15 nt were discarded. The

392 resulting sRNA reads were further cleaned by removing those that perfectly matched to the

393 sequences of tRNAs, snoRNAs, snRNAs (collected from GenBank) or rRNAs [53] using Bowtie

394 [54]. Raw counts for each unique sRNAs were derived and normalized into TPM (transcripts per

395 million).

396

397 ***Identification of miRNAs and differential expression analysis***

398 MiRNAs were identified using ShortStack [55] from each of the 159 samples, and a series of

399 filtering was applied to obtain high-confidence miRNAs. Briefly, the cleaned sRNA reads were

400 mapped to wild (LA2093) and cultivated tomato (Heinz 1706, SL4.0 and ITAG 4.1) genomes

401 separately, using ShortStack [55] with the parameter 'mmap' set to 'u'. Mature miRNAs and

402 corresponding pre-miRNAs were then identified by ShortStack [55]. The identified miRNAs from

403 these samples were collapsed if they were mapped to the exact locations in the genome. The

404 collapsed miRNAs that existed in at least three samples and expressed at more than 10 TPM were

405 considered as high-confidence miRNAs, which were compared with miRBase [56] to identify

406 conserved miRNAs, whereas miRNAs that showed no matches in the miRBase were considered

407 as novel miRNAs.

18

408   Raw counts of the identified miRNAs were processed using DESeq2 [57] to identify

409   differentially expressed miRNAs among accessions. MiRNAs with adjusted p values < 0.05 were

410   considered as differentially expressed. Differentially expressed miRNAs were further clustered

411   into groups according to their expression patterns using the DEGreport program [58]. Target genes

412   of differentially expressed miRNAs were predicted using the TargetFinder program [59]. GO

413   enrichment analysis was performed on the target genes using GO::TermFinder [60].

414

415   *Identification of candidate PHAS loci*

416   We used the previously described methods to identify PHAS loci [13]. In brief, the cleaned sRNA

417   reads were mapped to the wild (LA2093) and cultivated tomato (Heinz 1706, SL4.0 and ITAG

418   4.1) reference genomes using Bowtie [54] allowing no mismatch and no more than six hits. The

419   reference sequences were then scanned with a sliding window of 189 bp (nine 21-nt phase

420   registers). A positive window was considered to contain no less than 10 unique sRNAs, with more

421   than half of unique sRNAs being 21 nt in length and with no less than three 21-nt unique sRNAs

422   falling into the phase registers. Windows were combined if 1) they shared the same phase registers

423   and 2) fell into the same gene loci. P values and phasing scores for positive windows were

424   calculated following the methods described previously [32, 61]. The sequences of PHASs and the

425   flanking regions of 200 bp were retrieved and compared between wild and cultivated tomato

426   reference genomes using the BLAST program [62]. The BLAST results were then processed to

427   categorize candidate PHAS loci into three groups: PHAS loci shared by wild and cultivated

428   tomatoes, specific to wild tomato, and specific to cultivated tomato.

429

430   *Identification and analysis of 24-nt hc-siRNA hotspots*

431   Cleaned sRNA reads from all 159 samples were fed into ShortStack [55] to identify siRNA hotspot

432   regions that were defined by continuously covered sRNAs. The expression of 24-nt siRNA was

433   calculated by counting the number of 24-nt siRNA reads mapped to the corresponding regions.

434   Only a region with no less than ten mapped 24-nt siRNA reads was considered as a 24-nt siRNA

435   hotspot. The 24-nt siRNA expression was normalized to number of reads per kilobase of region

436   per million mapped reads (RPKM), based on all mapped reads.

437   To identify SV-related 24-nt siRNAs, pairwise comparisons were firstly applied between

438   stages or between tissues, and statistical analysis was performed using DESeq2 [57]. Only regions

439   with adjust P-values < 0.05 and fold change ≥ 2 were considered as significantly changed hotspots.

440   The significantly changed protein-coding genes and corresponding changed 24-nt siRNA hotspots

441   in promoter regions were treated as genes pairs involved in epigenetic regulation. To further

442   exclude potential false positive candidates, we only kept the pairs of 24-nt hotspots and the

443   corresponding protein-coding genes with a negative correlation in expression that occurred in at

444   least 10 samples. Previously reported SVs [5] that overlapped with significantly changed 24-nt

445   siRNA hotspots were then identified. The identified SVs were further filtered to keep those in

446   promoter regions of protein-coding genes that exhibited an opposite expression pattern compared

447   to the changes in abundance of the corresponding 24-nt siRNA hotspots.

448

449   ***RNA-Seq read processing and differential expression***

450   Single-end RNA-Seq reads were processed to remove adapters as well as low-quality bases using

451   Trimmomatic [63], and the trimmed reads shorter than 80 bp were discarded. The remaining high-

452   quality reads were subjected to rRNA sequence removal by aligning them to an rRNA database

453   [53] using Bowtie [54] allowing up to three mismatches. The cleaned RNA-Seq reads were aligned

454   to the cultivated tomato (Heinz 1706, SL4.0) reference genome using STAR [64] allowing up to

455   two mismatches. Gene expression was measured by counting the number of reads mapped to gene

456   regions (ITAG4.1), and then normalized to the number of reads per kilobase of exon per million

457   mapped reads (RPKM). Differential expression analysis was performed using DESeq2 [57]. To

458   obtain a global comparison among all samples, in particular to identify differentially expressed

459   genes in specific accessions or developmental stages, we followed a previously described linear

460   factorial modeling [65]. We also performed pairwise comparisons to identify differentially

461   expressed genes between stages for different accessions. Genes with adjusted P values <0.05 and

462   fold changes no less than two were considered differentially expressed.

463

464   **Accession numbers**

465   Raw sRNA and RNA-Seq reads have been deposited in the NCBI SRA under the accession number

466   SRP135718.

467

468   **Acknowledgements**

474

Author contributions

476   YZ, EvdK, YW, and ZF conceived the idea. YZ, YW, and ZF supervised the study. SM and YW

477   performed the plant analyses and sampling. YW constructed sRNA-Seq libraries. YQ, YZ, HNS,

478   XW, XZ, YW, and ZF analyzed the data. YZ and YW summarized the results. YZ, EvdK, YW,

479   and ZF wrote and revised the manuscript. All authors have read and approved the final manuscript.

480

Competing interests

482   The authors declare no competing interests.

483

References

1. Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Diez MJ, Francis D, Causse M, van der Knaap E, Canizares J: **Genomic variation in tomato, from wild ancestors to contemporary breeding accessions.** *BMC Genomics* 2015, **16:**257.

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

3. Razifard H, Ramos A, Della Valle AL, Bodary C, Goetz E, Manser EJ, Li X, Zhang L, Visa S, Tieman D, et al: **Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America.** *Mol Biol Evol* 2020, **37:**1118-1132.

4. Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C, et al: **Rewiring of the Fruit Metabolome in Tomato Breeding.** *Cell* 2018, **172:**249-261 e212.

5. Wang X, Gao L, Jiao C, Stravoravdis S, Hosmani PS, Saha S, Zhang J, Mainiero S, Strickler SR, Catala C, et al: **Genome of Solanum pimpinellifolium provides insights into structural variants during tomato breeding.** *Nat Commun* 2020, **11:**5817.

6. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al: **Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato.** *Cell* 2020, **182:**145-161 e123.

7. Axtell MJ: **Classification and comparison of small RNAs from plants.** *Annu Rev Plant Biol* 2013, **64:**137-159.

8. Borges F, Martienssen RA: **The expanding world of small RNAs in plants.** *Nat Rev Mol Cell Biol* 2015, **16:**727-741.

9. Vazquez F: **Arabidopsis endogenous small RNAs: highways and byways.** *Trends in plant science* 2006, **11:**460-468.

10. Baulcombe D: **RNA silencing in plants.** *Nature* 2004, **431:**356-363.

11. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136:**215-233.

12. Chavez Montes RA, de Fatima Rosas-Cardenas F, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, Marsch-Martinez N, Meyers BC, Green PJ, de Folter S: **Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs.** *Nat Commun* 2014, **5:**3722.

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

14. Cuperus JT, Fahlgren N, Carrington JC: **Evolution and functional diversification of MIRNA genes.** *The Plant cell* 2011, **23:**431-442.

15. Ma Z, Coruh C, Axtell MJ: **Arabidopsis lyrata small RNAs: transient MIRNA and small interfering RNA loci within the Arabidopsis genus.** *Plant Cell* 2010, **22:**1090-1103.

16. Baldrich P, Bélanger S, Kong S, Pokhrel S, Tamim S, Teng C, Schiebout C, Ranjan Gurazada SG, Gupta P, Patel P, et al: **The Evolutionary History of Small RNAs in the Solanaceae.** *bioRxiv* 2021**:**2021.2005.2026.445884.

17. Lunardon A, Johnson NR, Hagerott E, Phifer T, Polydore S, Coruh C, Axtell MJ: **Integrated annotations and analyses of small RNA-producing loci from 47 diverse plants.** *Genome Res* 2020, **30:**497-513.

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

19. Zhong S, Fei Z, Chen YR, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J, et al: **Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening.** *Nat Biotechnol* 2013, **31:**154-159.

533   20.   Karlova R, van Haarst JC, Maliepaard C, van de Geest H, Bovy AG, Lammers M, Angenent GC,
534   de Maagd RA: **Identification of microRNA targets in tomato fruit development using high-**
535   **throughput sequencing and degradome analysis.** *J Exp Bot* 2013, **64:**1863-1878.
536   21.   Pereira L, Sapkota M, Alonge M, Zheng Y, Zhang Y, Razifard H, Taitano NK, Schatz MC, Fernie
537   AR, Wang Y, et al: **Natural Genetic Diversity in Tomato Flavor Genes.** *Front Plant Sci* 2021,
538   **12:**642828.
539   22.   Li B, Yadav RK, Jeong GS, Kim HR, Chae HJ: **The characteristics of Bax inhibitor-1 and its**
540   **related diseases.** *Curr Mol Med* 2014, **14:**603-615.
541   23.   Xu G, Wang S, Han S, Xie K, Wang Y, Li J, Liu Y: **Plant Bax Inhibitor-1 interacts with ATG6**
542   **to regulate autophagy and programmed cell death.** *Autophagy* 2017, **13:**1161-1175.
543   24.   Chae HJ, Ke N, Kim HR, Chen S, Godzik A, Dickman M, Reed JC: **Evolutionarily conserved**
544   **cytoprotection provided by Bax Inhibitor-1 homologs from animals, plants, and yeast.** *Gene*
545   2003, **323:**101-113.
546   25.   Lu PP, Yu TF, Zheng WJ, Chen M, Zhou YB, Chen J, Ma YZ, Xi YJ, Xu ZS: **The Wheat Bax**
547   **Inhibitor-1 Protein Interacts with an Aquaporin TaPIP1 and Enhances Disease Resistance in**
548   **Arabidopsis.** *Front Plant Sci* 2018, **9:**20.
549   26.   Hernandez-Lopez A, Diaz M, Rodriguez-Lopez J, Guillen G, Sanchez F, Diaz-Camino C:
550   **Uncovering Bax inhibitor-1 dual role in the legume-rhizobia symbiosis in common bean roots.**
551   *J Exp Bot* 2019, **70:**1049-1061.
552   27.   Scotton DC, Azevedo MDS, Sestari I, Da Silva JS, Souza LA, Peres LEP, Leal GA, Jr., Figueira
553   A: **Expression of the Theobroma cacao Bax-inhibitor-1 gene in tomato reduces infection by**
554   **the hemibiotrophic pathogen Moniliophthora perniciosa.** *Mol Plant Pathol* 2017, **18:**1101-
555   1112.
556   28.   Fei Q, Xia R, Meyers BC: **Phased, secondary, small interfering RNAs in posttranscriptional**
557   **regulatory networks.** *Plant Cell* 2013, **25:**2400-2415.
558   29.   Jiang P, Lian B, Liu C, Fu Z, Shen Y, Cheng Z, Qi Y: **21-nt phasiRNAs direct target mRNA**
559   **cleavage in rice male germ cells.** *Nat Commun* 2020, **11:**5191.
560   30.   Pokhrel S, Huang K, Meyers BC: **Conserved and non-conserved triggers of 24-nucleotide**
561   **reproductive phasiRNAs in eudicots.** *Plant J* 2021.
562   31.   Zhang YC, Lei MQ, Zhou YF, Yang YW, Lian JP, Yu Y, Feng YZ, Zhou KR, He RR, He H, et al:
563   **Reproductive phasiRNAs regulate reprogramming of gene expression and meiotic**
564   **progression in rice.** *Nat Commun* 2020, **11:**6031.
565   32.   Xia R, Meyers BC, Liu Z, Beers EP, Ye S, Liu Z: **MicroRNA superfamilies descended from**
566   **miR390 and their roles in secondary small interfering RNA Biogenesis in Eudicots.** *Plant Cell*
567   2013, **25:**1555-1572.
568   33.   Canto-Pastor A, Santos B, Valli AA, Summers W, Schornack S, Baulcombe DC: **Enhanced**
569   **resistance to bacterial and oomycete pathogens by short tandem target mimic RNAs in**
570   **tomato.** *Proc Natl Acad Sci U S A* 2019, **116:**2755-2760.
571   34.   Moxon S, Jing R, Szittya G, Schwach F, Rusholme Pilcher R, Moulton V, Dalmay T: **Deep**
572   **sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit**
573   **ripening.** *Genome Research* 2008, **18:**1602.
574   35.   Wang Y, Itaya A, Zhong X, Wu Y, Zhang J, van der Knaap E, Olmstead R, Qi Y, Ding B: **Function**
575   **and evolution of a MicroRNA that regulates a Ca2+-ATPase and triggers the formation of**
576   **phased small interfering RNAs in tomato reproductive growth.** *Plant Cell* 2011, **23:**3185-3203.
577   36.   Axtell MJ, Meyers BC: **Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big**
578   **Data.** *Plant Cell* 2018, **30:**272-284.
579   37.   Moxon S, Jing R, Szittya G, Schwach F, Rusholme Pilcher RL, Moulton V, Dalmay T: **Deep**
580   **sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit**
581   **ripening.** *Genome Res* 2008, **18:**1602-1609.
582   38.   Tomato Genome C: **The tomato genome sequence provides insights into fleshy fruit evolution.**
583   *Nature* 2012, **485:**635-641.

23

584 39. Zheng Y, Ding B, Fei Z, Wang Y: **Comprehensive transcriptome analyses reveal tomato plant**
585 **responses to tobacco rattle virus-based gene silencing vectors.** *Sci Rep* 2017, **7:**9771.
586 40. Zheng Y, Wang Y, Ding B, Fei Z: **Comprehensive transcriptome analyses reveal that potato**
587 **spindle tuber viroid triggers genome-wide changes in alternative splicing, inducible trans-**
588 **acting activity of phasiRNAs and immune responses.** *J Virol* 2017.
589 41. Pan C, Ye L, Zheng Y, Wang Y, Yang D, Liu X, Chen L, Zhang Y, Fei Z, Lu G: **Identification**
590 **and expression profiling of microRNAs involved in the stigma exsertion under high-**
591 **temperature stress in tomato.** *BMC Genomics* 2017, **18:**843.
592 42. Cuperus JT, Fahlgren N, Carrington JC: **Evolution and functional diversification of MIRNA**
593 **genes.** *Plant Cell* 2011, **23:**431-442.
594 43. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R: **Fast and effective prediction of**
595 **microRNA/target duplexes.** *RNA* 2004, **10:**1507-1517.
596 44. Huang S, Zhou J, Gao L, Tang Y: **Plant miR397 and its functions.** *Funct Plant Biol* 2021, **48:**361-
597 370.
598 45. Xia R, Xu J, Meyers BC: **The Emergence, Evolution, and Diversification of the miR390-TAS3-**
599 **ARF Pathway in Land Plants.** *Plant Cell* 2017, **29:**1232-1247.
600 46. Millar AA, Lohe A, Wong G: **Biology and Function of miR159 in Plants.** *Plants (Basel)* 2019,
601 **8**.
602 47. Huijser P, Schmid M: **The control of developmental phase transitions in plants.** *Development*
603 2011, **138:**4117-4129.
604 48. Lian H, Wang L, Ma N, Zhou CM, Han L, Zhang TQ, Wang JW: **Redundant and specific roles**
605 **of individual MIR172 genes in plant development.** *PLoS Biol* 2021, **19:**e3001044.
606 49. O'Maoileidigh DS, van Driel AD, Singh A, Sang Q, Le Bec N, Vincent C, de Olalla EBG,
607 Vayssieres A, Romera Branchat M, Severing E, et al: **Systematic analyses of the MIR172 family**
608 **members of Arabidopsis define their distinct roles in regulation of APETALA2 during floral**
609 **transition.** *PLoS Biol* 2021, **19:**e3001043.
610 50. Chen YR, Zheng Y, Liu B, Zhong S, Giovannoni J, Fei Z: **A cost-effective method for Illumina**
611 **small RNA-Seq library preparation using T4 RNA ligase 1 adenylated adapters.** *Plant*
612 *Methods* 2012, **8:**41.
613 51. Zhong S, Joung JG, Zheng Y, Chen YR, Liu B, Shao Y, Xiang JZ, Fei Z, Giovannoni JJ: **High-**
614 **throughput illumina strand-specific RNA sequencing library preparation.** *Cold Spring Harb*
615 *Protoc* 2011, **2011:**940-949.
616 52. Zheng Y, Gao S, Padmanabhan C, Li R, Galvez M, Gutierrez D, Fuentes S, Ling KS, Kreuze J, Fei
617 Z: **VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of**
618 **small RNAs.** *Virology* 2017, **500:**130-138.
619 53. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: **The SILVA**
620 **ribosomal RNA gene database project: improved data processing and web-based tools.**
621 *Nucleic Acids Res* 2013, **41:**D590-596.
622 54. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of**
623 **short DNA sequences to the human genome.** *Genome Biol* 2009, **10:**R25.
624 55. Axtell MJ: **ShortStack: comprehensive annotation and quantification of small RNA genes.**
625 *RNA* 2013, **19:**740-751.
626 56. Kozomara A, Griffiths-Jones S: **miRBase: annotating high confidence microRNAs using deep**
627 **sequencing data.** *Nucleic Acids Res* 2014, **42:**D68-73.
628 57. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-**
629 **seq data with DESeq2.** *Genome Biol* 2014, **15:**550.
630 58. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto
631 L, Girke T, et al: **Orchestrating high-throughput genomic analysis with Bioconductor.** *Nat*
632 *Methods* 2015, **12:**115-121.
633 59. Fahlgren N, Carrington JC: **miRNA Target Prediction in Plants.** *Methods Mol Biol* 2010, **592:**51-
634 57.

635 60. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G: **GO::TermFinder--open**
636   **source software for accessing Gene Ontology information and finding significantly enriched**
637   **Gene Ontology terms associated with a list of genes.** *Bioinformatics* 2004, **20:**3710-3715.
638 61. De Paoli E, Dorantes-Acosta A, Zhai J, Accerbi M, Jeong DH, Park S, Meyers BC, Jorgensen RA,
639   Green PJ: **Distinct extremely abundant siRNAs associated with cosuppression in petunia.** *RNA*
640   2009, **15:**1965-1970.
641 62. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+:**
642   **architecture and applications.** *BMC Bioinformatics* 2009, **10:**421.
643 63. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.**
644   *Bioinformatics* 2014, **30:**2114-2120.
645 64. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
646   TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29:**15-21.
647 65. Clevenger J, Chu Y, Arrais Guimaraes L, Maia T, Bertioli D, Leal-Bertioli S, Timper P, Holbrook
648   CC, Ozias-Akins P: **Gene expression profiling describes the genetic regulation of Meloidogyne**
649   **arenaria resistance in Arachis hypogaea and reveals a candidate gene for resistance.** *Sci Rep*
650   2017, **7:**1317.

651

652 **Figure legend**

653 **Figure 1.** Functional analysis of SV-related 24-nt hc-siRNA regions among wild and cultivated

654 tomatoes. **A.** Circos plot of the densities of 24-nt hc-siRNAs, genes, repeat sequences, and SVs

655 across the tomato genome. **B.** Differentially expressed genes showing a negative correlation with

656 the corresponding SV-related 24-nt hc-siRNA clusters in their promoters in at least one pairwise

657 comparison. YG, young green fruit; MG, mature green fruit; Br, fruit at the breaker stage; RR, red

658 ripe fruit. **C.** GO term enrichment analysis of differentially expressed genes in (**B**). **D.** SV-related

659 24-nt hc-siRNA (Cluster_71406) enrichment leading to the repression of *Solyc06g084130*

660 expression in SLL (*Solanum lycopersicum* var. *lycopersicum*) and SLC (*S. lycopersicum* var.

661 *cerasiforme*) compared to SP (*S. pimpinellifolium*).

662

663 **Figure 2.** SV-related PHASs. **A.** Distribution of PHAS loci and SV-related PHAD loci across the

664 12 tomato chromosomes. The number depicts the distinct PHASs at each locus. **B.** SV90890 causes

665 gain/loss of PHASs and protein-coding genes. **C.** Allele frequency of SV90890 in different tomato

666 groups. **D.** Abundance of phasiRNAs in the genome region of SLL Heinz 1706 containing

667 SV90890 in the nine tomato accessions. For each box plot, the lower and upper bounds of the box

668 indicate the first and third quartiles, respectively, and the center line indicates the median. The

669 whisker represents 1.5× interquartile range of the lower or upper quartile.

670

671 **Figure 3.** Distinct expression patterns of miRNAs. For each box plot, the lower and upper bounds

672 of the box indicate the first and third quartiles, respectively, and the center line indicates the median.

673 The whisker represents 1.5× interquartile range of the lower or upper quartile.

674

675 **Figure 4.** MiRNA genes overlapped with SVs. **A.** Synteny diagram of miR10535 gene(s) in SP

676 and SLL reference genomes. **B.** Enriched GO terms of predicted target genes of SV-related

677 miRNAs.

678

679 **Figure 5.** SV affects miRNA expression. **A.** Diagram showing SV49949 (a 17-bp indel) in the

680 promoter of miR172a/b-2 gene. **B.** Allele frequency of SV49949 in different tomato groups. **C.**

681 Expression profiles of miR172a/b-2 in different tomato accessions. YG, young green fruits. MG,

682 mature green fruits. Br, fruits at the breaker stage. RR, red ripe fruits.

683

684   **Figure 6.** Changes in miR397 expression during tomato domestication. **A.** Diagram showing the

685   six 5P and seven 3P processing variants from the miR397 precursor. **B.** Abundance of each

686   miR397 variant in different tomato accessions.