Plant Cell Advance Publication. Published on May 9, 2018, doi:10.1105/tpc.18.00194

LARGE-SCALE BIOLOGY ARTICLE

Regulatory Divergence in Wound-Responsive Gene Expression between Domesticated and Wild Tomato

Ming-Jung Liu[a,b§*], Koichi Sugimoto[c§], Sahra Uygun[d], Nicholas Panchy[d], Michael S. Campbell[e], Mark Yandell[e,f], Gregg A. Howe[c,g,h] and Shin-Han Shiu[d,i,j*]

[a]Agricultural Biotechnology Research Center, Academia Sinica, Taipei 115, Taiwan
[b]Biotechnology Center in Southern Taiwan, Academia Sinica, Tainan 741, Taiwan
[c]Department of Energy-Plant Research Laboratory, Michigan State University, East Lansing, MI 48824, USA
[d]Genetics Program, Michigan State University, East Lansing, MI 48824
[e]Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA
[f]USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT 84112, USA.
[g]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA
[h]Plant Resilience Institute, Michigan State University, East Lansing, MI 48824, USA
[i]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA
[j]Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, MI 48824, USA
§ Joint first authors.
***Corresponding authors:** Ming-Jung Liu: mjliu@gate.sinica.edu.tw; Shin-Han Shiu: shius@msu.edu

Short title: Regulatory divergence of plant wound response
One-sentence summary: Profiling wound-responsive gene transcriptomes in wild *Solanum pennellii* and domesticated *S. lycopersicum* sheds light on the contribution of *cis*-regulatory variation to stress responsive gene expression divergence during species domestication.

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Ming-Jung Liu (mjliu@gate.sinica.edu.tw) and Shin-Han Shiu (shius@msu.edu).

ABSTRACT

The evolution of transcriptional regulatory mechanisms is central to how stress response and tolerance differ between species. However, it remains largely unknown how divergence in *cis*-regulatory sites and, subsequently, transcription factor (TF) binding specificity contribute to stress-responsive expression divergence, particularly between wild and domesticated species. By profiling wound-responsive gene transcriptomes in wild *Solanum pennellii* and domesticated *S. lycopersicum*, we found extensive wound-response divergence and identified 493 *S. lycopersicum* and 278 *S. pennellii* putative *cis*-regulatory elements (pCREs) that were predictive of wound-responsive gene expression. Only 24-52% of these wound-response pCREs (depending on wound-response patterns) were consistently enriched in the putative promoter regions of wound-responsive genes across species. In addition, between these two species, their differences in pCRE site sequences were significantly and positively correlated with differences in wound-responsive gene expression. Furthermore, ~11-39% of pCREs were specific to only one of the species and likely bound by TFs from different families. These findings indicate substantial regulatory divergence in these two plant species that diverged ~3-7 million years ago. Our study provides insights into the mechanistic basis of how the transcriptional response to wounding is regulated and, importantly, the contribution of *cis*-regulatory components to variation in wound-responsive gene expression between a wild and a domesticated plant species.

INTRODUCTION

Natural or artificial selection on diverse phenotypes leads to adaptation and domestication (Andersson, 2001; Doebley et al., 2006). Studies of the regulatory mechanisms underlying phenotypic diversity suggest that the variation in gene expression at the transcriptional level is one of the major contributing factors (Carroll, 2008; Romero et al., 2012). The divergent phenotypes between domesticated and wild plant species are the result of the domestication process in response to human selection (Doebley et al., 2006; Bauchet, 2012; Meyer and Purugganan, 2013; Chen et al., 2015). Comparisons of transcriptome profiles between domesticated and wild maize, carrot, cotton and tomato species have revealed that the extensive changes of gene expression are associated with phenotypic differences between closely related wild-domesticated species pairs (Swanson-Wagner et al., 2012; Koenig et al., 2013; Ichihashi et al., 2014; Rong et al., 2014). However, it remains unclear to what extent regulatory mechanisms have diverged between domesticated and wild species.

Two of the major components of the transcription regulatory program are *trans*-acting factors such as DNA-binding transcription factors (TFs) and *cis*-regulatory sites recognized by TFs (Kaufmann et al., 2010; Spitz and Furlong, 2012; Wittkopp and Kalay, 2012). The *cis*-regulatory sites are typically ~6-15 bp in length and located in close proximity to their target genes. A TF generally recognizes multiple, slightly different *cis*-regulatory sites that are collectively referred to as a *cis*-regulatory element (CRE), representing the binding specificity of TFs (Wittkopp and Kalay, 2012). Thus, variation in gene expression may result from the differences in the *cis*-regulatory sites and/or the TFs that regulate the genes in question. In cross-species studies, CREs have been shown to evolve much slower than individual *cis*-regulatory sites that have undergone extensive divergence (Doebley and Lukens, 1998; Wray et al., 2003; Carroll, 2008; Romero et al., 2012). For example, CREs amongst the orthologous TFs from fruit fly, mouse and human are highly conserved (Nitta et al., 2015). Similarly, by identifying sequence motifs resembling CREs from mouse and human based on DNase I footprints, >94% of the motifs are conserved (Stergachis et al., 2014). Because CREs are distinct TF binding motifs, these findings of CRE conservation indicate a high degree of conservation in *trans* regulatory mechanisms. Meanwhile, only ~20% of mouse DNase I footprints were co-localized with human footprints (Stergachis et al., 2014), suggesting extensive *cis*-regulatory site divergence. Since mouse and human were diverged ~100 million years ago, the mammalian regulatory mechanism has significantly diverged *cis*-regulatory sites but highly conserved CREs and thus *trans*-acting components (Stergachis et al., 2014).

In plants, studies have shown that the divergence of *cis*-regulatory sites affects the transcript levels of key developmental regulators of multiple domestication traits (Doebley et al., 2006; Ichihashi et al., 2014; Swinnen et al., 2016). In addition, because artificial selection for these domestication traits created bottleneck, genes relevant to biotic/abiotic tolerance could be eliminated in domesticated species (Rosenthal and Dirzo, 1997; Chaudhary, 2013; Chen et al., 2015), contributing to significant divergence in stress response. As a result, the wild species preserves much of the genetic variation and presumably regulatory mechanisms

underlying stress tolerance mechanisms (Hajjar and Hodgkin, 2007; Bauchet, 2012; Koenig et al., 2013; Bolger et al., 2014a). To understand how regulatory divergence contributes to stress tolerance traits, the response to wounding in domesticated and wild tomato species serves as a good model because of (1) their significant differences in stress tolerance (Bauchet, 2012; Koenig et al., 2013; Bolger et al., 2014a), (2) their divergence in transcriptional response to stress (Koenig et al., 2013; Bolger et al., 2014a), (3) the available information about the molecular underpinnings of responses to wounding in tomato (Howe and Jander, 2008; Howe and Schaller, 2008), and (4) knowledge of TFs and their corresponding *cis*-regulatory sites involved in regulating wound-responsive gene expression (Stankovic et al., 2000; Boter et al., 2004). Nonetheless, the identities of most CREs and their corresponding *cis*-regulatory sites underlying stress tolerance regulation in tomato and most other plant species have not been comprehensively examined. It also remains unclear how wound-induced patterns of gene expression differ between domesticated and wild tomato species such as *S. pennellii*, and how regulatory divergence contributes to divergence in wound transcriptional response between these species.

To assess the role of regulatory variation in gene expression divergence, one approach is to infer *cis*- and *trans*-regulatory divergence indirectly by comparing the differential gene expression of alleles between two parental lines and their F1 hybrid (Wittkopp et al., 2004; Emerson and Li, 2010). However, this strategy does not allow the exact CREs and the critical polymorphisms on binding sites to be evaluated. For this reason, we examined the regulatory mechanisms directly by identifying the CREs and CRE sites across species (Borneman et al., 2007; Sullivan et al., 2014; Nitta et al., 2015). To elucidate the regulatory mechanism divergence across species, we explored (1) to what extent the wound-responsive gene expression has diverged between *S. lycopersicum* and *S. pennellii*, (2) what CREs regulate differentially expressed genes between wound-treated and control samples from each species and between species, (3) to what degree CREs are relevant to wound-responsive gene expression conserved across species, and (4) to what extent differences in wound-induced transcriptional responses in these two tomato species are attributed to divergence in *cis*-regulatory sites.

RESULTS and DISCUSSION

Temporal and spatial expression profiles of wound-responsive genes in two Solanum species

To globally examine how the effects of wounding on gene expression differ in *S. pennellii* and *S. lycopersicum*, leaves were wounded mechanically to trigger the response in damaged (local) and undamaged (systemic) tissues and at 0.5 and 2 hour (h) time points post wounding, each condition with three biological replicates. Control leaf tissue was collected from unwounded plants. The data reproducibility was high among replicates of all conditions (**Supplemental Fig. S1A**). To evaluate the robustness of the wound-responsive gene expression profile revealed by RNA-seq (See Methods), the expression levels of several known

wound-responsive genes were further examined by reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR, **Supplemental Fig. S1B)**. We found that the RNA-seq and RT-qPCR results were generally consistent, suggesting the robust and reliable expression profiles. Thus, in subsequent analysis, we included all conditions that provide replicates and use RNA-seq analyses to identify wound-responsive genes.

A gene is defined as wound-responsive if it is either significantly up or down-regulated (multiple-testing adjusted $p<0.05$, $|\log_2(FC)|>2$, FC: Fold Change) in a wounded sample compared to the unwounded control. To increase the stringency of our analysis, we choose a FC threshold of four-fold instead of the conventional two-fold to emphasize robust changes in gene expression. This is also because we found that *cis*-element finding was more fruitful with more robustly differentially expressed genes. In both species, ~1,000 genes were significantly up-regulated by wounding (wound-induced) in local leaves during both time points (**Fig. 1A**). Interestingly, the pattern is very different for down-regulated genes where, at 0.5 h in the local tissue, there were only 59 down-regulated *S. lycopersicum* genes compared to 507 in *S. pennellii* (**Fig. 1A**). Similarly, at 2 h in the local tissue, 179 *S. lycopersicum* genes were down-regulated (left panel, **Fig. 1A**) compared to 983 in *S. pennellii* (right panel, **Fig. 1A**). Similar patterns were also observed for wound-responsive genes identified with the more conventional two-fold change (**Supplemental Fig. S2A**). In the systemic tissue, far fewer genes were differentially expressed in both species, with *S. pennellii* having more systemically-responsive genes than the cultivated species (353 in *S. lycopersicum* and 555 in *S. pennellii*) (**Fig. 1A**). Approximately 52% and 81% of these systemic wound-responsive genes in *S. lycopersicum* and *S. pennellii*, respectively, were a subset of the local wound-responsive genes, similar to previous microarray studies (Scranton et al., 2013), indicating similar wound responses between the local and systemic leaf. Taken together, these findings show that in response to wounding, both species have extensive changes in gene expression programs, but the extent of gene expression repression is more prominent in *S. pennellii*.

To assess in more detail how *S. lycopersicum* and *S. pennellii* differ in their wound response, orthologous genes that are wound-responsive (n=2,199) in any time point or tissue (i.e., local or systemic) in ≥1 species were compared. Hierarchical clustering of the overall expression patterns showed that the samples were clustered first based on the treatment location (local or systemic) and then by time points (0.5 or 2 h) and species (**Fig. 1B**), indicating that the spatial response has higher impact over the species origins or the duration of treatment on wound-responsive gene expression. Nonetheless, although the overall patterns of up and down-regulation are similar between species, there are important differences. In the local leaves at both time points, *S. pennellii* genes had higher amplitude of differential expression (higher absolute FC values) compared to their *S. lycopersicum* orthologs (dashed boxes, **Fig. 1B**). Thus, *S. pennellii* apparently responds to wounding earlier and stronger than *S. lycopersicum*, which is similar to the heightened tolerance to drought and salt in *S. pennellii* compared to *S. lycopersicum* (Tal and Shannon, 1983; Gong et al., 2010; Koenig et al., 2013; Bolger et al., 2014a).

Co-expression clustering and functions of wound-responsive genes

The overall transcript profile showed that wound-responsive genes differed significantly between species and could be classified into categories according to the time of treatment and spatial location of the response (**Fig. 1**). To further investigate how the wound response may have functionally diverged between species, we first categorized a wound-responsive gene from a species into one of 81 "wound-response clusters" based on whether the gene in question is up-regulated (“U”), non-regulated (“N”) and down-regulated (“D”) in response to wounding at a given time/location (major clusters shown in **Fig. 2A**; all clusters comprising <2% of wound-responsive genes in **Supplemental Data set 1**). For example, a gene is categorized in the UUDN cluster if it is up-regulated at both 0.5 h and 2 h in the local wounded leaf, down-regulated at the 0.5 h time point in the systemic undamaged leaf, and not changed significantly in the 2 h systemic response. Among the major wound-induced clusters (red, **Fig. 2A**), the UNNN, NUNN, and UUNN clusters were the largest with >250 genes in both species (**Fig. 2B**, **Supplemental Data set 1**). The number of up-regulated genes in these three major clusters was greater in *S. pennellii* than in *S. lycopersicum*. Similarly, the number of genes in the four major wound-repressed clusters (blue, **Fig. 2A**) was greater in *S. pennellii* (**Fig. 2B**). The same tendency was also observed when differential expression was defined as $|\log_2(FC)|>1$ (**Supplemental Fig. S2B,C**). Taken together, these findings suggest that *S. pennellii* has a more dynamic wound response, particularly in the case of down-regulated genes.

Considering the differences in wound-responsive gene expression between *S. lycopersicum* and *S. pennellii* (**Fig. 1,2**), we assessed the function of wound-responsive genes in each wound-response cluster with Gene Ontology (GO) and metabolic pathway annotations (see Methods). Wounding activates broad-spectrum defense responses in tomato (Green and Ryan, 1972; Howe and Jander, 2008; Howe and Schaller, 2008). Consistent with previous findings (Howe and Schaller, 2008; Scranton et al., 2013), the wound up-regulated genes in local leaves, especially those in the UNNN and UUNN clusters, were significantly enriched in genes responsive to multiple biotic and abiotic stresses, including those mediated by the stress hormones salicylic acid (SA) and abscisic acid (ABA) (**Fig. 2C**; also true for genes with $\log_2(FC)>1$, **Supplemental Fig. S2D**). Notably, most biological processes were more significantly enriched in *S. lycopersicum* than in *S. pennellii* for the genes with $\log_2(FC)>2$ (**Fig. 2C**), but not in genes with $\log_2(FC)>1$ (**Supplemental Fig. S2D**). This result suggests that, while the defense-related genes were wound-induced both in domesticated and wild species, wound stress results in higher degrees of gene induction and/or a proportionally higher number of defense-related genes in the domesticated tomato than that in the wild species.

Although there was a large number of wound down-regulated genes (**Fig. 1**), only two clusters (DNNN and NDNN) containing *S. pennellii* genes were significantly enriched in plant growth-related GO categories, including photosynthesis (**Fig. 2C**). This is consistent with previous studies showing the tradeoffs between growth and stress tolerance in wild species (Huot et al., 2014). The metabolic pathway analyses further showed that genes in the NDNN

clusters in *S. pennellii* were significantly enriched in phylloquinone biosynthesis (**Fig. 2D**). Phylloquinone is an integral part of the photosynthetic electron transport chain (Nowicka and Kruk, 2010). The reduction in the expression levels of genes associated with photosynthetic efficiency suggests an antagonistic relationship between defense response and plant growth in *S. pennellii* (**Fig. 2C**). In addition, photosynthesis-related functional categories were enriched in wound-repressed genes with $\log_2$(FC)<-1 in *S pennellii* (NDNN cluster in **Supplemental Fig. S2E**), further supporting the tradeoffs between growth and stress tolerance in wild tomato, a pattern that was not apparent in the domesticated species.

Taken together, our findings show that wound-response genes can be categorized into a few dominant clusters (**Fig. 2A,B**). Because some orthologs have differing responses to wounding (**Fig. 1B**), the identity and the enrichment test statistics of some GO categories and metabolic pathways also differ (**Fig. 2C,D**). Nonetheless, the number of GO categories and metabolic pathways enriched in genes up- or down-regulated in either species was small. This was particularly true for *S. pennellii* down-regulated genes. Since only the orthologs were included in the gene set enrichment analyses (see Methods), the small numbers of GO categories recovered may be due to the lower gene number in a cluster, which consequently decreases statistical power.

Divergence of wound responses among orthologous genes

Previous work in maize and tomato has suggested that the domestication process or the adaptation to extreme environments may result in extensive changes in the transcriptional regulation of genes controlling relevant morphological and physiological traits (Swanson-Wagner et al., 2012; Koenig et al., 2013). Our findings showed that there were substantial differences in the wound-responsive expression of *S. lycopersicum* and *S. pennellii* genes, as well as differences in the biological processes represented by these genes (**Fig. 2**). One immediate question is to what extent the orthologous genes in these two species differ in their wound response. To address this question, we first assessed which putative orthologous genes (see Methods) have consistent wound-response patterns (i.e. both orthologs are in the same wound-response cluster, **Fig. 3A**). These genes are referred to as "consistent genes". Interestingly, depending on the cluster (**Fig. 3A**), only 0-24% orthologs were considered consistent (**Fig. 3B**). These results showed that 76-100% of the wound-responsive orthologous genes were in different clusters and thus differentially regulated between species. Upon examination of the orthologous gene expression patterns side-by-side between species, some orthologous pairs had substantially different responses (cyan and orange bar, **Fig. 3C-F**). For example, in the UNNN cluster (**Fig. 3C**), in 47% cases the *S. pennellii* orthologous genes were either in the NNNN cluster (dotted rectangle *a*, **Fig. 3C**) or in the UUNN cluster (dotted *b*, **Fig. 3C**). The pattern of low consistency (<25%) in ortholog expression was also observed when genes with $|\log_2(FC)|>1$ were used (**Supplemental Fig. S2F,G**). These results suggest that wound responses have diverged among the majority of orthologs in the past 3-7 million years (Nesbitt and Tanksley, 2002; Kamenetzky et al., 2010).

To assess the extent to which the wound response differed between orthologous genes, we compared the wound-induced gene expression levels of "inconsistent orthologs", defined as orthologous gene pairs not in the same wound-response cluster, over the tested durations/tissues in the four largest clusters. In most cases, although the *S. lycopersicum* and the *S. pennellii* genes in inconsistent ortholog pairs belonged to different clusters, both orthologs were responsive but at different levels. For example, in the UNNN cluster in which only the *S. pennellii* genes were significantly up-regulated (above threshold) at 0.5 h in the local leaves (dotted rectangle *c*, **Fig. 3C**), the corresponding *S. lycopersicum* orthologs were also up-regulated but at levels below the threshold (dotted rectangle *d*, **Fig. 3C**). Similarly, in the NUNN cluster where only the *S. lycopersicum* orthologs were significantly up-regulated (dotted rectangle *a*, **Fig. 3E**), the expression of most corresponding *S. pennellii* orthologs was also induced but at levels below threshold (dotted rectangle *b*, **Fig. 3E**). This pattern was also true for down-regulated genes (cyan and orange bars, **Fig. 3F**). Given that most orthologs were wound-responsive but at different levels, the ancestral genes of these orthologs were likely wound-responsive as well. Thus, when the wound response of orthologous genes diverges, the divergence is not typically due to complete loss or gain of response but more likely due to diverging levels of responsiveness.

To this point our analysis focused on differential expression by comparing wounded leaves to unwounded, control leaves. Although induced gene expression is important for kick-starting defense systems in unfavorable environments (Green and Ryan, 1972; Howe and Jander, 2008; Howe and Schaller, 2008), constitutive defenses also contribute to plant resilience to environmental stress (Wittstock and Gershenzon, 2002). Using the *S. lycopersicum* gene expression level as a reference, we identified 374 and 219 *S. pennellii* genes that were expressed at significantly higher and lower levels, respectively, than their cultivated tomato orthologs (**Fig. 4A**). This finding indicates that significant differences in gene expression already exist between the two species prior to wounding, contributing to divergence in constitutive defense. For example, cuticular wax and cutin biosynthesis genes *CER6*, *CER8*, *MYB41* and *SICUS2* (Hooker et al., 2002; Cominelli et al., 2008; Lu et al., 2009) were expressed at higher levels in *S. pennellii* (**Fig. 4B**), consistent with findings of earlier studies (Bolger et al., 2014a). Given that expression levels are already different between the control samples, it is possible that a gene contributing to constitutive defense will have a consistently high expression level before and after wounding. To assess this, we also compared the gene expressions in wound-treated samples in both species against the *S. lycopersicum* unwounded control. A surprising pattern was that, if a *S. pennellii* gene had a significantly different (either higher or lower) expression level in unwounded control compared to that of its *S. lycopersicum* ortholog under control condition, the *S. pennellii* gene in question tended to remain significantly different in a consistent fashion after wounding in both time points and in both local and systemic tissues (**Fig. 4A**). This finding supports the hypothesis that the basal level of defense response is stronger in *S. pennellii* (Koenig et al., 2013; Bolger et al., 2014a).

Putative cis-regulatory sequences controlling wound-responsive gene regulation

The expression patterns in control and wounded tissue between *S. lycopersicum* and *S. pennellii* orthologous genes have diverged substantially, suggesting divergence of regulatory mechanisms central to controlling wound-responsive gene expression. Substitutions in *cis*-regulatory sites may lead to expression divergence due to the inability of orthologous TFs to bind to the site with substitutions. Alternatively, expression divergence may be due to substantial changes in *cis*-regulatory sites such that the orthologous gene is now bound by a different TF. To assess these two mechanisms, we first need to know what the *cis*-regulatory elements (CREs, representing the TF binding specificity) are and where they are located in the genome. We identified globally the CREs likely controlling wound-responsive gene expression for cross-species comparison with an enriched *k*-mer approach (an oligomer with the length $k \geq 5$ base-pairs, see Methods).

Since the sites of CREs may be located in both the promoter and 5'UTR regions of a gene (Sullivan et al., 2014), we queried whether an enriched *k*-mer sequence is located near the transcriptional start sites (TSS, see Methods) of member genes in each cluster. Zero to hundreds of *k*-mers were found to have significantly enriched numbers of sites among genes in wound-response clusters relative to nonresponsive genes (**Fig. 5A**). These enriched *k*-mers are referred to as putative CREs (pCREs). The pCREs identified include ones that resemble known CREs relevant to the wound response, including ABRE, W-box and G-box (Rushton and Somssich, 1998; Hobo et al., 1999; Siberil et al., 2001; Boter et al., 2004; Adie et al., 2007), as well as those that do not resemble known CREs (**Supplemental Fig. S3)**. To further assess how well these pCREs can jointly explain the wound response in each cluster, we applied a machine-learning algorithm, Support Vector Machine (SVM, see Methods), to predict wound-responsive expression of genes in each wound-response cluster based on identified pCREs. Among the ten clusters with pCREs in *S. lycopersicum* and/or *S. pennellii* (**Fig. 5A**), the wound-response prediction models based on pCREs performed significantly better than randomly expected (boxplots vs. gray spot, Wilcoxon signed rank test, all $p<0.01$; **Fig. 5B, Supplemental Fig. S4A**). In addition, our *k*-mer approaches led to a differential expression prediction model that outperformed the model built with motifs from the commonly used Multiple EM for Motif Elicitation (Bailey et al., 2009) (**Supplemental Fig. S3G,** see Methods). These results showed that our approach could efficiently identify short sequences resembling CREs because they are predictive of wound response in multiple clusters. In addition, the pCREs from clusters involving wound-induced expression (e.g. UNNN (red) and UUNN (orange), **Fig. 5C**) tend to be located within 500 bp upstream of the TSS, consistent with the finding that plant TFs tend to bind preferentially in the upstream region close to TSSs (Franco-Zorrilla et al., 2014; Heyndrickx et al., 2014).

In contrast to pCREs involved in up-regulation, the pCREs identified in wound down-regulated clusters (NDNN (blue) and DNNN (green), **Fig. 5C**) tend to be located downstream of TSSs, including 5'UTRs. This is similar to the 5'UTR of *excision repair cross complementation group-1* gene in human that contains binding sites for a transcription repressor (Yu et al.,

2001). Similarly, the cyclin D1 Inhibitory Element within the 5'UTRs represses the expression of the human *cyclic D1* gene in an-age-dependent manner (Berardi et al., 2003). Nonetheless, we discover no pCRE from the DDDD cluster, suggesting the potential role of post-transcriptional regulation such as transcript turnover (Narsai et al., 2007) in repression control of these genes. Taken together, the pCREs identified are predictive of wound-responsive gene expression in most clusters and have a position bias resembling the known TF binding sites, suggesting that they are authentic *cis*-elements in regulating gene expression.

Divergence of putative CREs between tomato species

To assess the degree of regulatory divergence across tomato species, we first examined if similar pCREs are regulating *S. lycopersicum* and *S. pennellii* genes with similar wound-response patterns. This was accomplished by asking whether a pCRE is consistently enriched in a wound-response cluster in both species. If a pCRE is consistently enriched, the pCRE in question is likely a component of a conserved wound-response regulatory program. We found that 24-52% of pCREs in UNNN, UUNN and NDNN clusters were consistently enriched between species (**Supplemental Data set 2**; for the UNNN cluster, pCREs in black, **Fig. 6A**), suggesting their conserved role in wound-response regulatory programs. This result also showed that the remaining 48-76% of pCREs, depending on the wound-response cluster, were species-specifically enriched (pCREs in blue or orange, **Fig. 6A,B**), indicating substantial divergence in regulatory programs. The presence of species-specific pCREs raises the question of whether they are (1) bound by the same sets of orthologous TFs that bind *cis*-regulatory sites with subtle differences between species (Zhang et al., 2006), or (2) bound by non-orthologous TFs between species. To assess the above possibilities, we first defined two sets of species-specific pCREs as those that were enriched only in *S. lycopersicum* and only in *S. pennellii* genes within a cluster, respectively. Next, we asked whether these two sets of species-specific pCREs could be bound by TFs from the same family. We adopted this conservative approach to ensure that we could provide a lower-bound estimate of the proportion of species-specific pCREs that are bound by distinct TFs across-species. We should also emphasize that the pCREs, including the species-specific ones, were identified first based on their enrichments in the putative promoters of genes in wound-response clusters relative to nonresponsive genes. Thus, these species-specific pCREs are likely relevant to species-specific wound-response regulation, a point supported based on modeling results in the next section.

Using *in vitro* TF binding data (see Methods), we divide pCREs into sub-groups where pCREs in a sub-group are likely bound by TFs of the same family (**Fig. 6A**, **Supplemental Fig. S5)**. For example, pCREs that were enriched in UNNN wound-responsive genes from ≥1 species could be divided into 33 pCRE sub-groups (**Fig. 6A**). A sub-group was defined as “dual-species” if it contained pCREs from both species. By contrast, if all pCREs in a sub-group came from only one species, this sub-group was then designated as “single-species” (asterisk, **Fig. 6A, Supplemental Fig. S5**). Together with whether a pCRE was en-

riched in the putative promoter regions of wound-responsive genes in one or both species, we classified pCREs into three types (**Fig. 7A**): (1) Type I: a pCRE is enriched in both species and belongs to a dual-species sub-group, (2) Type II: a pCRE is enriched only in one species but belongs to a dual-species sub-group, and (3) Type III: a pCRE is enriched only in one species and belong to a single-species sub-group. We should emphasize that Type I, II, and III pCREs are bound by TFs with increasingly divergent binding specificities. We have shown that 24-52% pCREs were Type I enriched in both species (**Supplemental Data set 2**). Type II pCREs were found in 32%, 52%, and 37% of sub-groups in UNNN, UUNN and NDNN clusters, respectively (**Fig. 6, Supplemental Fig. S5**). In the UNNN cluster, for example, the consensus sequence of six pCREs in the 8th sub-group is GTTGACT (yellow box, **Fig. 6**) similar to the W-box (TTGAC[C/T]) recognized by WRKY TFs that mediate biotic and abiotic stress responses (van Verk et al., 2008; Banerjee and Roychoudhury, 2015). Among these six pCREs, AGTCAAC and GTCAACT were enriched in both species, whereas the remaining pCREs were enriched specifically in *S. pennellii*. This indicates the conserved role of the same TF family across species in triggering wound responses but also implies the regulatory divergence at the level of individual TF-binding *cis*-regulatory elements.

Compared to Type I and II, there are relatively fewer Type III pCREs. Among the largest clusters, 16%, 11%, and 39% pCREs in UNNN (**Fig. 6A**), UUNN (**Supplemental Fig. S5A**), and NDNN (**Supplemental Fig. S5C**) clusters were Type III (**Supplemental Data set 2**). In the UNNN cluster, for example, 14 Type III pCREs were enriched only in *S. lycopersicum* wound-responsive genes and belonged to a sub-group with only *S. lycopersicum* pCREs (blue and asterisk, **Fig. 6A, Supplemental Data set 2**), suggesting these pCREs are specific to *S. lycopersicum* and likely bound by specific *S. lycopersicum* TFs where their *S. pennellii* orthologs are either absent or do not bind. Note that the sub-groups were defined to ensure pCREs bound by TFs from the same family can be correctly identified but errs on the side of not calling pCREs truly regulated by distinct TFs. Thus the 16% represents the lower bound in terms of the degree of regulatory divergence involving pCREs bound by non-orthologous TFs in regulating the UNNN wound-response cluster between these two species.

To further assess the regulatory divergence of pCREs on wound response, we examined the enrichment of the species-specific pCREs (Type II and Type III) among inconsistent orthologs (**Fig. 3**). We found that the species-specific pCREs enriched within a wound-response cluster in a particular species were significantly enriched among inconsistent orthologous genes from the species in question but not in the other species (**Supplemental Fig. S6B,C**). This finding further supports the species-specific nature of these pCREs and their positive correlation with expression divergence. Taken together, while *S. lycopersicum* and *S. pennellii* may have similar pCREs to control wound-induced gene expression, there are distinct preferences of pCREs for wound response across species, supporting the presence of both regulatory conservation and divergence.

Relationship between pCRE conservation and gene regulation across species

We show that wound-response pCREs differ in their enrichment in genes between species and in whether they can be recognized by TFs from the same family (**Fig. 6, Supplemental Fig. S5**). Based on their enrichment and sub-group memberships, they can be classified into three types (**Fig. 7A**, **Supplemental Data set 2**). To assess which types of pCREs contribute more significantly to wound transcriptional response, we used the Type I, II, III pCREs to build machine learning models (see Methods) for predicting wound-responsive expression of genes in a wound-response cluster.

We found that models built with Type I pCREs were in most cases the best at predicting wound response in both species (red, **Fig. 7B, Supplemental Fig. S4B**), suggesting that these pCREs are components of conserved regulatory mechanisms across species. Type II pCREs predicted wound response well within species but not across species (compare blue and yellow, **Fig. 7B, Supplemental Fig. S4B**), supporting their roles in species-specific regulatory function. We should note that, except the NDNN clusters in *S. pennellii*, the prediction performance of Type II and III pCREs was not as accurate as the Type I pCREs (**Fig. 7B**). This suggests that the conserved *cis*-regulatory elements play a more central role in wound-responsive transcription in both tomato species, and that species-specific pCREs, to a lesser extent, contribute to differential gene expression species specifically.

Turnover of putative CRE sites between orthologous genes and their association with gene regulation

Our findings so far indicate substantial conservation of CREs between domesticated and wild tomato species and their association in predicting wound response (**Fig. 6,7**). In addition, we found extensive variation of wound-responsive gene expression among orthologous genes (**Fig. 3**). These differences may result from minor changes in CRE sequences, leading to differences in TF binding specificity (**Fig. 6**). Alternatively, the wound response divergence between orthologs may be the consequence of differential turnover (i.e. the gain and loss) of the *cis*-regulatory sites within orthologous regions (Carroll, 2008; Wittkopp and Kalay, 2012). To assess these possibilities, we next determined the extent to which these *cis*-regulatory sites were conserved or turned over across species and their association with gene expression divergence. Based on the relative position of the sites located in regulatory regions of orthologous gene pairs, the sites of a given pCRE were categorized into “shared”, “specific”, “compensatory” and “other” types (**Fig. 8A**; see Methods). Since the “compensatory” and “other” types accounted for small portions of the pCRE sites (**Supplemental Fig. S6A**), we focused on the “shared” and “specific” pCRE types.

To summarize the degree of conservation of the sites of each pCRE identified from various wound-response clusters (**Fig. 5A**), a conservation likelihood (L_c) for each pCRE was computed by calculating the log2 ratio between the proportion of sites that are shared and the proportion of sites that are specific (see Methods). Thus, a higher L_c indicates a higher degree of enrichment of shared sites relative to that of specific sites. A pCRE with a higher L_c was

considered more conserved than that with a lower L_c. First, to assess if the conservation of pCRE sites was correlated with the consistency of the wound response between orthologs, we compared the L_c values for the orthologs with consistent wound response and for those with inconsistent patterns. Using the UNNN cluster as an example (left panel, **Fig. 8B**), we found that the sites of pCREs in orthologous gene pairs with consistent wound-response patterns (median L_c=0.61) had significantly higher L_c values than sites in orthologous pairs with inconsistent patterns (median L_c=0.28, Mann-Whitney U test, p=2.5x10^{-3}). The same was true when comparing pCRE sites in genes with consistent patterns against sites found in the non-responsive orthologous genes (median L_c=-0.56; p<2.2x10^{-16}) (left panel, **Fig. 8B**). Similar results were also observed for the pCREs in the UUNN and NDNN cluster (middle and right panels, **Fig. 8B**). Taken together, these results imply that in UNNN, UUNN and NDNN clusters, the orthologs with consistent gene regulation tend to have more conserved pCRE sites, indicating that, as expected, conservation of pCREs sites contribute to a conserved wound up-regulated response across species.

Taken together, these findings suggest a positive correlation between the degrees of pCRE site conservation and the conservation of wound-regulated gene expression between wild and domesticated species.

CONCLUSION

In this study, we investigated the patterns and mechanisms of transcriptional divergence of environmental stress response in a wild and a domesticated tomato species. Specifically, our analyses focus on wound-responsive gene expression and the *cis*-regulatory components regulating wound responses. Despite the relatively recent divergence (~3-7 million years ago) between the wild *S. pennellii* and domesticated *S. lycopersicum* species (Nesbitt and Tanksley, 2002; Kamenetzky et al., 2010), the wound-responsive expression patterns of the orthologous genes have diverged significantly, which may be partly attributable to the combined action of natural and artificial selection. In addition, we characterized the pCREs significantly associated with wound-response regulation. pCREs identified in *S. lycopersicum* and *S. pennellii* were predictive of gene expression. In addition, Type I pCREs (enriched in both species) could better explain gene regulation between species than Type II and III pCREs (species-specifically enriched). This is in line with the conclusion in metazoan studies that the TF-binding specificity evolves slowly and is highly conserved amongst fruit fly, mouse and human (Stergachis et al., 2014; Nitta et al., 2015). Intriguingly, the Type II and III pCREs partially explain wound response within species, indicating the involvement of divergent TFs after speciation contributing to regulatory divergence. Our results based on the approaches of the differential enrichment of pCREs and whether they may be recognized by TFs from the same family suggest diverging binding preference of some TFs relevant to wound-response regulation across species. Further protein-DNA binding studies such as protein binding array and DNA affinity purification sequencing (Weirauch et al., 2014; O'Malley et al., 2016) should be useful to test the regulatory divergence hypothesized here.

Our finding of correlation between the turnover of the pCRE site and the expression divergence of orthologous genes further supports the evolutionary conservation of CREs for wound response in tomato. We should emphasize that, although the correlation is apparent, it is far from perfect. Specifically, some pCRE sites enriched among wound-responsive genes displayed high degrees of conservation between orthologous pairs with inconsistent wound-response patterns. One possibility is that these conserved pCREs in orthologs with inconsistent patterns are still regulating weaker wound responses. This is because the wound-response clusters were defined based on threshold differential expression – weaker wound responses may not pass the defined threshold. As a result, some orthologous genes were classified into different clusters despite a similar but significantly weaker response (**Fig. 3**). We should also point out that the conservation likelihood (L_c) distribution of some pCREs on orthologs with consistent wound responses may also be low (**Fig. 8B**), indicating that consistent expression patterns cannot be easily attributed to the pCREs analyzed. This highlights the complexity of the transcriptional regulatory systems and the need for studies to further ascertain the mechanistic basis of stress response conservation and divergence. Lastly, among these sites located in the regulatory regions, it is possible that only part of them are the *in vivo cis*-regulatory sites which can be further narrowed down based on chromatin state, GC-content or DNA structural properties on the surrounding regions (Raveh-Sadka et al., 2012; White et al., 2013; Tsai et al., 2015). Future studies aimed at reducing false positive identification of pCRE sites based on additional features and at identifying the combinatorial relationship between CREs will be helpful for further understanding the *cis*-regulatory codes and their evolution.

Our study provides global comparative analyses connecting the divergence of pCREs and turnover of *cis*-regulatory sites to gene expression divergence between species and orthologous genes. The comparison of pCREs predictive of the wound response revealed both *cis*-regulatory conservation and divergence. The correlation between the turnover of the *cis*-regulatory sites and the differential expression of orthologs uncovered *cis*-regulatory divergence underlying the gene expression variation. Collectively, these findings advance our understanding of the mechanistic basis underlying the stress-responsive gene expression divergence across a wild and a domesticated species.

METHODS

Plant materials and growth conditions

S. lycopersicum cv Castlemart was used as the domesticated species. Seeds for the wild species, *S. pennellii* (LA0716), were obtained from Tomato Genetic Resource Center (UC Davis) and grown on Jiffy-7 peat pot (Hummert International, Earth City, MO) in a growth chamber under 16 h light (6:00 to 22:00, 200 μmol m^{-2} s^{-1})/ 8 h dark cycle at 28ºC. Three- to four-week-old plants with 3 to 4 expanded true leaves were used for wound treatment as previously described (Li et al., 2004). For wound elicitation, the lower (older) two leaves were crushed with a hemostat across the midrib of all leaflets. All wounding was performed in the

morning (8:00-9:00), 2 h after the start of the light cycle. At the indicated time points, leaflets from multiple tomato seedlings were excised with a razor blade, pooled together and immediately frozen in liquid nitrogen. Damaged leaflets (local, older leaves) from the first and second leaves, and undamaged leaflets (systemic, younger leaves) from the third and fourth leaves were collected separately. Control leaves were harvested from a set of unwounded plants grown side-by-side with the set of wounded plants. Three biological replicates (i.e., three separate sets of plants sampled on different days) were harvested for each treatment and time point. Total RNA was isolated from frozen leaf tissue using a RNeasy Plant Mini Kit (Qiagen, Germantown, MD, USA). Except the locally 2-h wound-treated samples in *S. pennellii*, RNA sequencing (100-bp paired-end reads) of the samples was performed with the Illumina HiSeq2500 platform in the Michigan State University Research Technology Support Facility. RNA sequencing of the locally 2-h wound-treated samples in *S. pennellii* was performed with HiSeq4000 (150-bp paired-end reads).

Sequencing data processing

To map the RNA-seq reads and determine the gene expression level, the reference genome sequences and gene annotation of *Solanum lycopersicum* (ITAG2.4) and *S. pennellii* (Spenn_v2.0) were retrieved from Sol Genomics Network (https://solgenomics.net). The *S. pennellii* gene annotation was further re-annotated through Maker-P module (Campbell et al., 2014). The cumulative distribution plot of AED (Annotation Edit Distance), which provides the measure of how well the annotations are supported by the EST, protein and RNA-Seq evidence (Campbell et al., 2014), showed that the MAKER-mediated version performed better compared to the Spenn_v2.0 version (red vs. black lines, **Supplemental Fig. S7**). To further evaluate MAKER-P performance, we focused on genes annotated in both datasets (n=22,292) (**Supplemental Data set 3**). Among the genes with one-to-one relationship (i.e. overlapped genic regions) between Spenn_v2.0 and MAKER-P annotated version, the gene models annotated by MAKER-P had higher AED values than in Spenn_v2.0 version (green vs. gray lines, **Supplemental Fig. S7**). These results showed that MAKER-P improved the gene annotations of *S. pennellii*; thus, the MAKER-mediated version was adopted in this study.

The paired-end RNA reads were trimmed with Trimmomatic (default setting except leading=20, trailing=20 and minlen=20) (Bolger et al., 2014b) and mapped to the genome with Tophat2 (version 2.0.8) (Kim et al., 2013). Transcript levels of annotated genes were calculated with Cufflinks (version 2.1.1) (Trapnell et al., 2010) and shown as FPKM (Fragments Per Kilobase per Million fragments mapped). The numbers of raw, quality-filtered and mapped reads and the sequencing coverage are reported in **Supplemental Data set 4**. To evaluate the reproducibility of gene expressions among replicates and the similarity of gene expression profiles among treatments, the Spearman's rank correlation coefficient was determined by pairwise comparison of gene expression between samples. The distance (1- Spearman's rank correlation coefficient) was used to generate the dendrogram through hierarchical clustering function with "complete" method (**Supplemental Fig. S1A**). The three replicates of a given

564 treatment in one species were clustered together, showing the gene expression profiles were
565 similar and reproducible among replicates (**Supplemental Fig. S1A**).
566 To identify the significantly wound-responsive genes, only protein-coding genes with
567 the value of "FPKM" ≥1 in all replicates of any time point and tissue were considered
568 (n=17,945 in *S. lycopersicum* and 16,868 in *S. pennellii*). The transcript abundances of con-
569 trol and wound-treated samples were compared with EdgeR (Robinson et al., 2010). Genes
570 with false discovery rate adjusted $p<0.05$ (Benjamini, 1995) and with 4-fold difference in RNA
571 level between wound and control (unwounded) samples was considered to be wound-
572 responsive and included for the analyses. Note that the replicates of the local 2-h *S. pennellii*
573 sample were sequenced in a different Illumina platform that result in significantly higher num-
574 ber of reads (n=62-153 million) compared to those of the other samples (n=13-23 million)
575 (**Supplemental Data set 4**). Given between sample normalization was part of the modeling
576 process in EdgeR (Robinson et al., 2010), we expected that the DE gene call will not be sig-
577 nificantly influenced. Consistent with this, by down-sampling reads from the high coverage
578 replicate, we found that the difference of the input size of raw read numbers amongst samples
579 did not impact the identification of DE genes (**Fig. 1A**, **Supplemental Fig. S1C**).

580 ***Identification of putative cis-elements and prediction for wound response***

581 Wound-responsive genes are categorized into the different regulatory clusters de-
582 pending on the levels of differential gene expression in the indicated points as defined in **Fig.**
583 **2A**. Genes are regarded as nonresponsive genes if their fold-change (FC) values in all com-
584 parisons between wound treatments and controls are between 1.2 and 0.8 (n=3,548 in *S. ly-*
585 *copersicum*) and in *S. pennellii* (n=1,058). Note that replicates from a treatment were jointly
586 compared to control replicates in determining FC using EdgeR. The FC values were used for
587 pCRE identification (**Fig. 5A**) and pCRE site turnover analyses (**Fig. 8**). To identify the puta-
588 tive *cis*-regulatory elements (pCREs) associated with wound response (**Fig. 5A**), a *k*-mer (oli-
589 gomer with the length of k) pipeline was established by examining the frequency enrichment
590 of a *k*-mer sequence in the regulatory region among the genes of a given wound-response
591 cluster compared to the nonresponsive genes and determining the adjusted *p-values* through
592 Fisher's exact test and multiple testing (Benjamini-Hochberg method) (Benjamini, 1995). Here,
593 the regulatory region is defined as the region ranging from upstream 1 kb to downstream 0.5
594 kb of transcription start site (TSS).
595 Since the *cis*-regulatory elements range from 5 nt to ~30 nt (Stewart et al., 2012), this
596 *k*-mer pipeline includes several steps to discover the pCREs with various sequence lengths.
597 Step 1: a set of all possible 5-mer oligomers was evaluated for their enrichment among genes
598 in each wound-response cluster compared to nonresponsive genes. Only the 5-mers with
599 significant enrichment (adjusted *p-values*<0.05) were retained for the next step. Step 2: the
600 sequence of each significantly enriched 5-mer from step 1 was extended with 1 nt in either
601 direction, the resulting extended *k*-mer was examined for enrichment, and the significantly
602 enriched ones (adjusted *p-values*<0.05) were retained. The step was repeated until no ex-

603 tended *k*-mer sequence was found to be significantly enriched among the regulated genes.
604 Noted that if two *k*-mers were both significantly enriched and one *k*-mer sequence exactly
605 matched the other one, only the one with lower adjusted *p-value* was retained. Step 3: as de-
606 scribed in step 1, but starting with a set of all possible 6-mers. The significantly enriched 6-
607 mers were combined with the set of the *k*-mers identified from step 2. Step 4: as described in
608 step 2, but starting with the set of *k*-mers from step 3. Finally, the set of *k*-mers significantly
609 enriched in the indicated wound-response cluster was determined and considered as pCREs
610 (**Fig. 5A**). To compare the performance of our *k*-mer pipeline to the typical motif-finding ap-
611 proaches, we used MEME (Bailey et al., 2009) to identify pCREs in UNNN clusters in both
612 species. The prediction model employing MEME-derived pCREs performed significantly
613 worse than that employing identified *k*-mers (**Supplemental Fig. S3G),** suggesting that our
614 approach could more efficiently identify short sequences resembling CREs.
615 The support vector machine (SVM) method that allows predicting of wound response
616 of a gene based on a set of pCREs was performed using the LIBSVM implementation of the
617 SVM method through the Weka wrapper with the parameters described previously (Liu et al.,
618 2015). The pCREs were used as attributes whereas the binary status of genes with/without
619 wound regulation was the class we wanted to predict. For training the predictive models for
620 each regulatory pattern, the genes of the given clusters are positive examples whereas the
621 non-responsive genes are negative examples.

622 ***Sequence similarity of putative CREs between species and to the known TF binding***
623 ***motifs***

624 To identify the pCREs whose sequences are more significantly similar than expected
625 between TF families (thus the pCRE in question are likely bound by TF(s) from a family that
626 pCREs are similar to), the pairwise distances of known TF binding motifs (TFBM) across 30
627 TF families (Weirauch et al., 2014) were calculated and the 5th percentile of distance, 0.39,
628 (with a *p-value*=0.05) was set as a threshold (Liu et al., 2015).
629 To determine what pCREs identified in *S. lycopersicum* and *S. pennellii* for a given
630 cluster are likely bound by TF(s) of the same family, the pair-wise PCC (Pearson's correlation
631 coefficient) distance of the pCREs was generated with TAMO package (Gordon et al., 2005)
632 and used to construct the average linkage tree using UPGMA method in "cluster" package in
633 R (Maechler, 2016). The threshold of 0.39 value that corresponds to the distance of the motifs
634 among TF families was applied such that any pCREs within a branch length <0.39 are con-
635 sidered to be in the same sub-group and likely bound by TFs from the same family (**Fig. 6A,**
636 **Supplemental Fig. S5 and Data set 2**). The pCREs located in a given sub-group were
637 merged through STAMP with default settings (Mahony and Benos, 2007) to summarize the
638 sequence information of these pCREs since the pCREs within a sub-group are likely bound
639 by TFs from the same family but may have subtle nucleotide difference and various lengths
640 (**Supplemental Fig. S3**). Note that the presence of pCRE duplicates in **Fig. 6, Supplemental**

Fig. S5 is because some pCREs were identified from both *S. lycopersicum* and *S. pennellii*. In these cases, one copy was removed before merging.

The known TFBM dataset consists of 256 and 510 CREs from protein binding microarray (Weirauch et al., 2014) and DNA affinity purification sequencing approaches (O'Malley et al., 2016). The similarity between the merged pCREs and known TFBMs were determined with the threshold of PCC distance (p<0.05) as described previously (Liu et al., 2015) (**Supplemental Fig. S3**).

Identification of orthologous genes

Using the longest protein sequences for genes, an all vs. all comparison of protein sequences was run on a combined set of genes in *S. lycopersicum* and *S. pennellii* using BLAST. Custom python scripts were then used to extract reciprocal best matches between species. The set of the reciprocal best matches was divided into those which were the best overall match (the “Overall” set) and those where one of the two proteins had a better match within species (the “Reciprocal-Only” set). Initially, there were 19,657 Overall and 1,198 Reciprocal-Only best matches. For Reciprocal-Only best matches, the sequences of the better within species matches were obtained, creating a group of 3 or more protein sequences (i.e. the best match between species gene pairs and any genes that are better matches within species) for each Reciprocal-Only best match.

For each pair of Overall best matches and group of Reciprocal-Only best matches, protein sequences were aligned using MAFFT. Protein alignments were then back-aligned to the longest coding sequences for genes in each species using custom python scripts. The resulting aligned nucleotide sequences were used to determine the Ks of best-matches using PAML. The “yn00” algorithm was used on sequence pairs from Overall best matches and the “codeml” algorithm was used for sequence groups from Reciprocal-Only best matches. Next, we visualized the distribution of Ks values for the “Overall” set because they have a clear 1:1 relationship between *S. lycopersicum* and *S. pennellii*. Given the recent speciation event, we expected the Ks distribution to follow a normal distribution. We observed a roughly normal distribution with a long right tail. We theorize that the extremely large Ks value in the tail can be attributed to ancient duplication events which experienced reciprocal loss in both species. Therefore, to enrich the set of reciprocal-best matches for orthologs of the recent speciation event, a normal distribution was fit to set of Ks values for the “Overall” set in R using non-linear minimization. The 99th percentile of the fit distribution was determined and applied as a cutoff to both the “Overall” and “Reciprocal-Only” best matches. This resulted in a final set of 16,222 orthologous genes between *S. lycopersicum* and *S. pennellii*.

Gene ontology (GO) and metabolic pathway analyses

The datasets of GO annotation and metabolic pathways of genes in *S. lycopersicum* were retrieved from the Sol Genomics Network (https://solgenomics.net) and Plant Metabolic Network (http://www.plantcyc.org). To have comparable annotation set of GO and metabolic

pathways of genes across species, the annotations of genes from *S. lycopersicum* were inferred to the orthologous ones in *S. pennellii*. In the end, 10,091 and 2,006 orthologous genes with biological process and metabolic pathway were retrieved for the downstream analyses. The list of orthologous gene pairs between *S. lycopersicum* and *S. pennellii* was generated as mentioned above.

The enrichments of GO terms and metabolic pathways in the clusters and differentially regulated gene sets, compared to the total orthologous genes, were determined though Fisher's Exact test. A *p*-value obtained for each GO term and pathway comparison and was multiple-testing corrected (Bass et al., 2015).

Conservation and divergence of pCRE sites in orthologous gene pairs

The region of the 1 kb upstream and 500 bp downstream of transcript start sites in the orthologous gene pairs was defined to be regulatory regions and aligned with MUSCLE package (Edgar, 2004) (**Fig. 8A**). Based on the positions of the pCRE sites on the aligned sequences, these sites for each pCRE were assigned into 4 types: (1) "shared" (i.e. the site from each species was located on the same positions), (2) "specific" (i.e. the site was present only in one species), (3) "compensatory" (i.e. the site was present in both species but located in different location) and (4) "others" (i.e. any cases of pCRE sites were not assigned to the 3 types mentioned above). A likelihood score representing the conservation degree of pCRE sites for each pCRE was determined by taking the ratio of the pCRE site types (%) between the "shared" and "specific" ones. The orthologous gene pairs with consistent patter means the pairs are assigned to the same regulatory cluster as defined in **Fig. 2A**; otherwise, the OG pairs are considered to be with inconsistent patterns. Non-responsive orthologous genes are the orthologous genes if their fold-change values in all wound-treatment conditions, compared to the control one, are between 1.2 and 0.8 in both *S. lycopersicum* and *S. pennellii* (n=452).

Reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) analyses

Total RNA from 3 independent samples was reverse-transcribed with High Capacity cDNA Reverse Transcription Kit (Life Technologies, Carlsbad, CA, USA), respectively, according to the manufacturer's instructions. The resulting cDNA was subsequently used for quantification of transcripts with Power SYBR Green PCR Master Mix (Life Technologies) and analysis of products on an ABI 7500 Fast Real-Time PCR System (Life Technologies). The relative transcript abundances were calculated using the ΔCt (threshold cycle) method. The *ACTIN* gene was used as an internal control. Primers were designed to target the conserved regions of genes between *S. lycopersicum* and *S. pennellii* and listed in **Supplemental Table 1**.

ACCESSION NUMBERS

The RNA-seq data from this study have been submitted to the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE93556. The

716 names and accession numbers of genes described in this study can be found in **Supplemental Table 1**.

718

719 **SUPPLEMENTAL DATA**

720 **Supplemental Figure 1.** Similarity of gene expression profiles among replicates of each con-
721 dition and between RT-qPCR and RNA-seq.

722 **Supplemental Figure 2.** Comparison of wound-response clusters for genes with |log2FC)|>1
723 and their functional category enrichments.

724 **Supplemental Figure 3.** Properties of the putative *cis*-elements regulating wound response.

725 **Supplemental Figure 4.** Model performance employing pCREs in predicting wound response.

726 **Supplemental Figure 5.** Differential enrichment of pCREs in UUNN and NDNN cluster genes
727 between tomato species.

728 **Supplemental Figure 6.** The types of pCRE sites on the orthologous gene pairs and the dif-
729 ferential enrichments of species-specific pCREs in inconsistent orthologs.

730 **Supplemental Figure 7.** MAKER-P-mediated improvement on *S. pennellii* gene annotation

731 **Supplemental Table 1.** Primers used in this study

732 **Supplemental Data set 1.** Numbers of genes in different wound-responsive clusters in *S.*
733 *lycopersicum* (*Sl*) and *S. pennellii* (*Sp*).

734 **Supplemental Data set 2.** The species enrichment, types and the sub-groups of pCREs
735 identified in *S. lycopersicum* (*Sl*) and *S. pennellii* (*Sp*) as indicated in Fig. 6 and Supplemental
736 Fig. S5.

737 **Supplemental Data set 3.** Genes with one to one relationship (i.e. overlapped genic regions)
738 between Spenn_v2.0 and MAKER-P annotated versions and their corresponding gene name
739 (ID) in each dataset.

740 **Supplemental Data set 4.** Numbers of sequencing reads in each read-mapping step.

741 ## AUTHOR CONTRIBUTIONS

742 M.-J.L., K.S., M.Y., G.A.H. and S.-H.S. designed the research. K.S. performed the experi-
743 ments. M.-J.L., S.U., N.P. and M.S.C. analyzed the data. M.-J.L., G.A.H. and S.-H.S. wrote
744 the manuscript with contributions by all authors.

745 ## ACKNOWLEDGEMENTS

746 This work was in part supported by a grant from the Rackham Foundation and Michi-
747 gan AgBioResearch Project MICL02278 to G.A.H.; NSF IOS-1546617, DEB-1655386, and
748 DOE DE-SC0018409 to S.-H.S.; an MSU Discretionary Funding Initiative grant to S.-H.S; a
749 grant from Biotechnology Center in Southern Taiwan, Academia Sinica and MOST106-2311-
750 B-001-025-MY2 to M.-J.L; Japan Society for Promotion of Science, Research Fellowship for
751 Young Scientists (24.841) to K.S.

REFERENCES

Adie, B.A.T., Perez-Perez, J., Perez-Perez, M.M., Godoy, M., Sanchez-Serrano, J.J., Schmelz, E.A., and Solano, R. (2007). ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defenses in Arabidopsis. The Plant cell **19,** 1665-1681.

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. Nature reviews. Genetics **2,** 130-138.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research **37,** W202-208.

Banerjee, A., and Roychoudhury, A. (2015). WRKY proteins: signaling and regulation of expression during abiotic stress responses. ScientificWorldJournal **2015,** 807560.

Bass, A.J., Dabney, A., and Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. R package **version 2.6.0**.

Bauchet, G.a.C., M. (2012). Genetic diversity in tomato (Solanum lycopersicum) and its wild relatives. In Environmental Sciences (Çaliskan, M., ed.). Rijeka, Croatia: InTechOpen**,** 133–162.

Benjamini, Y.a.Y.H. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society **57,** 289-300.

Berardi, P., Meyyappan, M., and Riabowol, K.T. (2003). A novel transcriptional inhibitory element differentially regulates the cyclin D1 gene in senescent cells. J Biol Chem **278,** 7510-7519.

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sorensen, I., Lichtenstein, G., Fich, E.A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., Aflitos, S.A., Schijlen, E., Jimenez-Gomez, J.M., Ryngajllo, M., Kimura, S., Kumar, R., Koenig, D., Headland, L.R., Maloof, J.N., Sinha, N., van Ham, R.C., Lankhorst, R.K., Mao, L., Vogel, A., Arsova, B., Panstruga, R., Fei, Z., Rose, J.K., Zamir, D., Carrari, F., Giovannoni, J.J., Weigel, D., Usadel, B., and Fernie, A.R. (2014a). The genome of the stress-tolerant wild tomato species Solanum pennellii. Nat Genet **46,** 1034-1038.

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

Borneman, A.R., Gianoulis, T.A., Zhang, Z.D., Yu, H., Rozowsky, J., Seringhaus, M.R., Wang, L.Y., Gerstein, M., and Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species. Science **317,** 815-819.

Boter, M., Ruiz-Rivero, O., Abdeen, A., and Prat, S. (2004). Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. Genes & development **18,** 1577-1591.

Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C.J., Ware, D., Shiu, S.H., Childs, K.L., Sun, Y., Jiang, N., and Yandell, M. (2014). MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol **164,** 513-524.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell **134,** 25-36.

Chaudhary, B. (2013). Plant domestication and resistance to herbivory. Int J Plant Genomics **2013,** 572784.

Chen, Y.H., Gols, R., and Benrey, B. (2015). Crop domestication and its impact on naturally selected trophic interactions. Annu Rev Entomol **60,** 35-58.

Cominelli, E., Sala, T., Calvi, D., Gusmaroli, G., and Tonelli, C. (2008). Over-expression of the Arabidopsis AtMYB41 gene alters cell expansion and leaf surface permeability. Plant J **53,** 53-64.

Doebley, J., and Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. The Plant cell **10,** 1075-1082.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell **127,** 1309-1321.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research **32,** 1792-1797.

Emerson, J.J., and Li, W.H. (2010). The genetic basis of evolutionary change in gene expression levels. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **365,** 2581-2590.

Franco-Zorrilla, J.M., Lopez-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. Proceedings of the National Academy of Sciences of the United States of America **111,** 2367-2372.

Gong, P.J., Zhang, J.H., Li, H.X., Yang, C.X., Zhang, C.J., Zhang, X.H., Khurram, Z., Zhang, Y.Y., Wang, T.T., Fei, Z.J., and Ye, Z.B. (2010). Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. J Exp Bot **61,** 3563-3575.

Gordon, D.B., Nekludova, L., McCallum, S., and Fraenkel, E. (2005). TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs. Bioinformatics **21,** 3164-3165.

Green, T.R., and Ryan, C.A. (1972). Wound-Induced Proteinase Inhibitor in Plant Leaves: A Possible Defense Mechanism against Insects. Science **175,** 776-777.

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. Euphytica **156,** 1-13.

Heyndrickx, K.S., Van de Velde, J., Wang, C., Weigel, D., and Vandepoele, K. (2014). A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. The Plant cell **26,** 3894-3910.

Hobo, T., Asada, M., Kowyama, Y., and Hattori, T. (1999). ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. Plant J **19,** 679-689.

Hooker, T.S., Millar, A.A., and Kunst, L. (2002). Significance of the expression of the CER6 condensing enzyme for cuticular wax production in Arabidopsis. Plant Physiol **129,** 1568-1580.

Howe, G.A., and Jander, G. (2008). Plant immunity to insect herbivores. Annu Rev Plant Biol **59,** 41-66.

Howe, G.A., and Schaller, A. (2008). Direct Defenses in Plants and Their Induction by Wounding and Insect Herbivores. (Springer Netherlands).

Huot, B., Yao, J., Montgomery, B.L., and He, S.Y. (2014). Growth-defense tradeoffs in plants: a balancing act to optimize fitness. Mol Plant **7,** 1267-1287.

Ichihashi, Y., Aguilar-Martinez, J.A., Farhi, M., Chitwood, D.H., Kumar, R., Millon, L.V., Peng, J., Maloof, J.N., and Sinha, N.R. (2014). Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. Proceedings of the National Academy of Sciences of the United States of America **111,** E2616-2621.

Kamenetzky, L., Asis, R., Bassi, S., de Godoy, F., Bermudez, L., Fernie, A.R., Van Sluys, M.A., Vrebalov, J., Giovannoni, J.J., Rossi, M., and Carrari, F. (2010). Genomic analysis of wild tomato introgressions determining metabolism- and yield-associated traits. Plant Physiol **152,** 1772-1786.

Kaufmann, K., Pajoro, A., and Angenent, G.C. (2010). Regulation of transcription in plants: mechanisms controlling developmental switches. Nat Rev Genet **11,** 830-842.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14,** R36.

Koenig, D., Jimenez-Gomez, J.M., Kimura, S., Fulop, D., Chitwood, D.H., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., Tat, A.V., Tohge, T., Bolger, A., Schneeberger, K., Ossowski, S., Lanz, C., Xiong, G., Taylor-Teeples, M., Brady, S.M., Pauly, M., Weigel, D., Usadel, B., Fernie, A.R., Peng, J., Sinha, N.R., and Maloof, J.N. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proceedings of the National Academy of Sciences of the United States of America **110,** E2655-2662.

Li, L., Zhao, Y., McCaig, B.C., Wingerd, B.A., Wang, J., Whalon, M.E., Pichersky, E., and Howe, G.A. (2004). The tomato homolog of CORONATINE-INSENSITIVE1 is required for the maternal control of seed maturation, jasmonate-signaled defense responses, and glandular trichome development. The Plant cell **16,** 126-143.

Liu, M.J., Seddon, A.E., Tsai, Z.T., Major, I.T., Floer, M., Howe, G.A., and Shiu, S.H. (2015). Determinants of nucleosome positioning and their influence on plant gene expression. Genome Res **25,** 1182-1195.

Lu, S., Song, T., Kosma, D.K., Parsons, E.P., Rowland, O., and Jenks, M.A. (2009). Arabidopsis CER8 encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1) that has overlapping functions with LACS2 in plant wax and cutin synthesis. Plant J **59,** 553-564.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2016). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.4.

Mahony, S., and Benos, P.V. (2007). STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic acids research **35,** W253-258.

Meyer, R.S., and Purugganan, M.D. (2013). Evolution of crop species: genetics of domestication and diversification. Nature reviews. Genetics **14,** 840-852.

Narsai, R., Howell, K.A., Millar, A.H., O'Toole, N., Small, I., and Whelan, J. (2007). Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. The Plant cell **19,** 3418-3436.

Nesbitt, T.C., and Tanksley, S.D. (2002). Comparative sequencing in the genus Lycopersicon. Implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics **162,** 365-379.

Nitta, K.R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E.E., and Taipale, J. (2015). Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. Elife **4**.

Nowicka, B., and Kruk, J. (2010). Occurrence, biosynthesis and function of isoprenoid quinones. Biochim Biophys Acta **1797,** 1587-1605.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **166,** 1598.

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat Genet **44,** 743-750.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Romero, I.G., Ruvinsky, I., and Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. Nature reviews. Genetics **13,** 505-516.

Rong, J., Lammers, Y., Strasburg, J.L., Schidlo, N.S., Ariyurek, Y., de Jong, T.J., Klinkhamer, P.G., Smulders, M.J., and Vrieling, K. (2014). New insights into domestication of carrot from root transcriptome analyses. BMC Genomics **15,** 895.

Rosenthal, J.P., and Dirzo, R. (1997). Effects of life history, domestication and agronomic selection on plant defence against insects: Evidence from maizes and wild relatives. Evolutionary Ecology **11,** 337-355.

Rushton, P.J., and Somssich, I.E. (1998). Transcriptional control of plant genes responsive to pathogens. Curr Opin Plant Biol **1,** 311-315.

Scranton, M.A., Fowler, J.H., Girke, T., and Walling, L.L. (2013). Microarray analysis of tomato's early and late wound response reveals new regulatory targets for Leucine aminopeptidase A. PLoS One **8,** e77889.

Siberil, Y., Doireau, P., and Gantet, P. (2001). Plant bZIP G-box binding factors. Modular structure and activation mechanisms. Eur J Biochem **268,** 5655-5666.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews. Genetics **13,** 613-626.

Stankovic, B., Vian, A., Henry-Vian, C., and Davies, E. (2000). Molecular cloning and characterization of a tomato cDNA encoding a systemically wound-inducible bZIP DNA-binding protein. Planta **212,** 60-66.

Stergachis, A.B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A.P., Zhang, M., Byron, R., Canfield, T., Stelhing-Sun, S., Lee, K., Thurman, R.E., Vong, S., Bates, D., Neri, F., Diegel, M., Giste, E., Dunn, D., Vierstra, J., Hansen, R.S., Johnson, A.K., Sabo, P.J., Wilken, M.S., Reh, T.A., Treuting, P.M., Kaul, R., Groudine, M., Bender, M.A., Borenstein, E., and Stamatoyannopoulos, J.A. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. Nature **515,** 365-370.

Stewart, A.J., Hannenhalli, S., and Plotkin, J.B. (2012). Why transcription factor binding sites are ten nucleotides long. Genetics **192,** 973-985.

Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., Stergachis, A.B., Vernot, B., Johnson, A.K., Haugen, E., Sullivan, S.T., Thompson, A., Neri, F.V., 3rd, Weaver, M., Diegel, M., Mnaimneh, S., Yang, A., Hughes, T.R., Nemhauser, J.L., Queitsch, C., and Stamatoyannopoulos, J.A. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Rep **8,** 2015-2030.

Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M.B., Ross-Ibarra, J., Myers, C.L., Tiffin, P., and Springer, N.M. (2012). Reshaping of the maize transcriptome by domestication. Proceedings of the National Academy of Sciences of the United States of America **109,** 11878-11883.

Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. Trends Plant Sci **21,** 506-515.

Tal, M., and Shannon, M.C. (1983). Salt Tolerance in the Wild Relatives of the Cultivated Tomato - Responses of Lycopersicon-Esculentum, Lycopersicon-Cheesmanii, Lycopersicon-Peruvianum, Solanum-Pennellii and F1 Hybrids to High Salinity. Aust J Plant Physiol **10,** 109-117.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28,** 511-515.

Tsai, Z.T., Shiu, S.H., and Tsai, H.K. (2015). Contribution of Sequence Motif, Chromatin State, and DNA Structure Features to Predictive Models of Transcription Factor Binding in Yeast. PLoS computational biology **11,** e1004418.

van Verk, M.C., Pappaioannou, D., Neeleman, L., Bol, J.F., and Linthorst, H.J. (2008). A Novel WRKY transcription factor is required for induction of PR-1a

gene expression by salicylic acid and bacterial elicitors. Plant Physiol **146,** 1983-1995.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J.C., Galli, M., Lewsey, M.G., Huang, E., Mukherjee, T., Chen, X., Reece-Hoyes, J.S., Govindarajan, S., Shaulsky, G., Walhout, A.J., Bouget, F.Y., Ratsch, G., Larrondo, L.F., Ecker, J.R., and Hughes, T.R. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell **158,** 1431-1443.

White, M.A., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2013). Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. Proceedings of the National Academy of Sciences of the United States of America **110,** 11952-11957.

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature reviews. Genetics **13,** 59-69.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in cis and trans gene regulation. Nature **430,** 85-88.

Wittstock, U., and Gershenzon, J. (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. Current Opinion in Plant Biology **5,** 300-307.

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol **20,** 1377-1419.

Yu, J.J., Thornton, K., Guo, Y., Kotz, H., and Reed, E. (2001). An ERCC1 splicing variant involving the 5'-UTR of the mRNA may have a transcriptional modulatory function. Oncogene **20,** 7694-7698.

Zhang, C., Xuan, Z., Otto, S., Hover, J.R., McCorkle, S.R., Mandel, G., and Zhang, M.Q. (2006). A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome. Nucleic acids research **34,** 2238-2246.

FIGURE LEGEND

Figure 1. Similarities and differences in wound-responsive gene expression between tomato species

(A) Number of significantly differentially regulated genes ($|\log_2 FC)|>2$, FC: fold change) upon mechanical wounding in local and systemic leaves of *S. lycopersicum* (*Sl*) and *S. pennellii* (*Sp*) for the indicated time points (hour(s)) post wounding. (B) Differential gene expression values of orthologous genes (rows) in different location/species/time points (columns). Only orthologous genes significantly up- or down-regulated in ≥1 sample were included (n=2,199). Dashed boxes and arrows indicate clusters of orthologous genes with inconsistent regulatory patterns across species in local tissues.

Figure 2. Numbers of genes and functional category enrichments in wound-response clusters

(A) Definitions of wound-response clusters. U (red): up-regulation, $(\log_2 FC)>2$. N (gray): no significant change, $2>(\log_2 FC)>-2$. D (blue): down-regulation, $(\log_2 FC)<-2$. Only clusters with >40 genes in ≥1 species were shown. (B) Numbers of wound-responsive genes in the clusters shown in (A) for *S. lycopersicum* (left) and *S. pennellii* (right). Red and Blue: up- and down-regulated clusters. (C) Gene ontology biological process categories significantly enriched in wound up- (adjusted *p*-values <1e-03) and down-regulated (adjusted *p*-values <1e-02) cluster genes from *S. lycopersicum* (*Sl*) or *S. pennellii* (*Sp*). (D) Metabolic pathways significantly enriched (adjusted *p*-values <5e-02) in *S. lycopersicum* and *S. pennellii* genes from wound up- and down-regulated clusters. Deeper shades of blue indicate higher $-\log_{10}$(adjusted *p*-value).

Figure 3. Divergence of wound responses among orthologous genes

(A) Number of orthologous genes with $|\log_2 FC)|>2$ in the wound-response clusters as defined in **Fig. 2A**. Gray: orthologous genes from both species were in the same cluster. Cyan: the *S. lycopersicum* (*Sl*) ortholog is in the indicated cluster but not the *S. pennellii* one. Orange: the *S. pennellii* (*Sp*) ortholog is in the indicated cluster but not the *S. lycopersicum* one. (B) Percentage of the orthologous genes that are considered to have consistent regulatory patterns (in the same cluster) in each cluster. (C,D,E,F) Heatmaps showing the differential expression levels ($\log_2$(FC), FC: fold change) of orthologous genes in UNNN (C), UUNN (D), NUNN (E) and NDNN (F) clusters. The bars on the left of each heatmap are colored the same way as in (A). The dotted rectangles highlight differential expression patterns discussed in the main text.

Figure 4. Genes differentially expressed between species prior to wounding

(A) Heatmap showing differential expression where fold change (FC) values of all samples were calculated using the *S. lycopersicum* unwounded control (time point 0) expression values as the denominator. Only genes in the unwounded control in *S. pennellii* with significant FC values in comparison to the unwounded control in *S. lycopersicum* were shown (n=593,

1041 $|\log_2(FC)|>2$). (B) Differential expression values and test statistics contrasting *S. pennellii* and
1042 *S. lycopersicum* unwounded controls between orthologous gene pairs from both species in-
1043 volved in biosynthesis of cuticular wax and cutin in this and an earlier study (*, Bolger et al.,
1044 2014a).
1045
1046 **Figure 5. Evidence indicating biological relevance of putative CREs**
1047 (A) Number of putative CREs (pCREs) identified through the *k*-mer pipeline (see Methods)
1048 for each wound-response cluster in *S. lycopersicum* (blue) or *S. pennellii* (orange). Only the
1049 clusters with pCREs in ≥1 species are shown. (B) Boxplot showing the wound-response pre-
1050 diction performance (F-measure) based on a model using pCREs identified from genes in a
1051 wound-response cluster. F-measure: the harmonic mean of precision (proportion predicted
1052 correctly) and recall (proportion true positives predicted). The maximum F-measure is 1, indi-
1053 cating a perfect model. For each wound-response cluster, 10 F-measures were calculated
1054 from 10-fold cross validation and are shown as a boxplot. Gray dot: the average F-measure of
1055 10,000 random predictions indicating the performance of a meaningless model. NA: not appli-
1056 cable since no pCRE was found in the cluster. (C) Enrichment of sites of pCREs identified
1057 from four different clusters. For each pCRE, the degree of enrichment of its sites around tran-
1058 scription start sites (TSSs) was represented as the $\log_2$ ratio between pCRE site frequencies
1059 of genes in a cluster and frequencies of the same pCREs in genes not responsive to wound-
1060 ing. This log ratio was generated for each pCRE in the region from 1 kb upstream to 0.5 kb
1061 downstream of transcription start sites (TSSs) with a sliding window of 100 bp and a step size
1062 of 25 bp. For each cluster, the median log ratios of all pCREs identified from the cluster in
1063 question was shown.
1064
1065 **Figure 6. Differential enrichment of pCREs in UNNN cluster genes from two tomato**
1066 **species**
1067 (A) Dendrogram showing the distances between the pCREs identified from UNNN cluster
1068 genes and enriched in *S. lycopersicum* only (blue), *S. pennellii* only (orange), or both species
1069 (black). The dotted line indicates the threshold distance defined based on the 95th percentile
1070 distances between binding motifs of TFs from distinct families and defines multiple pCRE sub-
1071 groups (numbered) where each sub-group contains pCREs likely bound by TFs of the same
1072 family (distance threshold=0.39). Single-species subgroups with pCREs from only one spe-
1073 cies are labeled with asterisks. Note that some pCRE duplicates were due to their identifica-
1074 tion from both *S. lycopersicum* and *S. pennellii*. (B) Degrees of pCRE site enrichment in *S.*
1075 *lycopersicum* (blue) and *S. pennellii* (orange) UNNN genes. Adjusted *p*-value: multiple-testing
1076 corrected *p*-value. Dashed line, adjusted $p < 0.05$. Yellow box: pCREs similar to the W-box
1077 element.
1078
1079 **Figure 7. Performance of the Type I, II, III pCREs in predicting wound response**

(A) Numbers of pCREs that were consistently enriched in both species and belong to a dual-species sub-group (red, Type I), belong to “dual-species” subgroup but were specifically enriched in *S. lycopersicum* (blue, Type II from *Sl*) or in *S. pennellii* (orange, Type II from *Sp*), and belong to “single-species” subgroup and were specifically enriched in *S. lycopersicum* (purple, Type III from *Sl*) or in *S. pennellii* (green, Type III from *Sp*) in three example wound-response clusters. (B) Boxplot showing the wound-response prediction performance (F-measure) based on a model using the pCRE sets in (A). For each wound-response cluster, 10 F-measures were calculated from 10-fold cross validation and shown as a boxplot. Gray dot: the average F-measure of 10,000 random predictions indicating the performance of a meaningless model.

Figure 8. Relationships of pCRE site turnover and wound-responsive gene expression between orthologs

(A) Types of pCRE sites. Shared: the sites of a pCRE are present in both orthologs and located at the same position. Specific: the site of a pCRE is present only in one ortholog but not the other. Compensatory: the sites are present in both species but in different locations. Others: any situation that does not belong to the previous three types. Gray line: the defined regulatory regions from the orthologous gene pairs (See Methods). (B) The conservation likelihood (*Lc*) of a pCRE in the UNNN (left panel), the UUNN (middle panel), and NDNN (right panel) clusters. For a pCRE, its *Lc* is defined as the log ratio between the proportions of sites that shared and those that are specific (see Methods). The *Lc* for each pCRE was evaluated using orthologous gene pairs with consistent (belong to the same wound-response cluster, orange) and inconsistent (belong to different clusters, blue) wound responses, as well as orthologous genes that are not responsive to wounding (Non-responsive, gray). *P*-values: testing whether the likelihood scores generating based the blue or gray datasets differ from the orange one (one-sided Mann-Whitney U test).

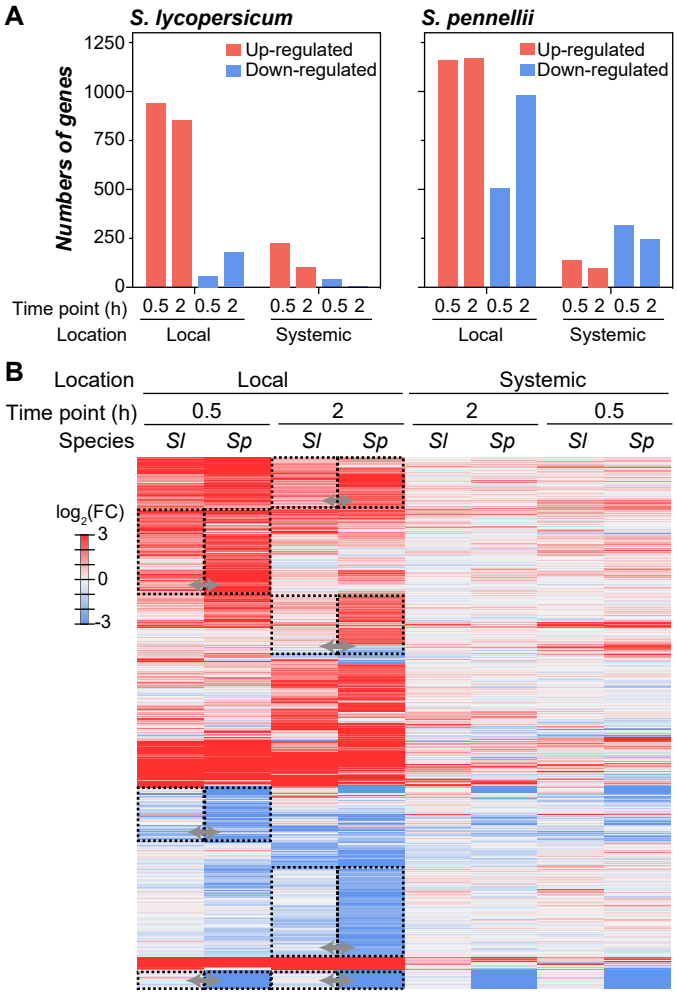


Figure 1. Similarities and differences in wound-responsive gene expression between tomato species
(A) Number of significantly differentially regulated genes ($|\log_2 FC|>2$, FC: fold change) upon mechanical wounding in local and systemic leaves of *S. lycopersicum* (*Sl*) and *S. pennellii* (*Sp*) for the indicated time points (hour(s)) post wounding. (B) Differential gene expression values of orthologous genes (rows) in different location/species/time points (columns). Only orthologous genes significantly up- or down-regulated in ≥1 sample were included (n=2,199). Dashed boxes and arrows indicate clusters of orthologous genes with inconsistent regulatory patterns across species in local tissues

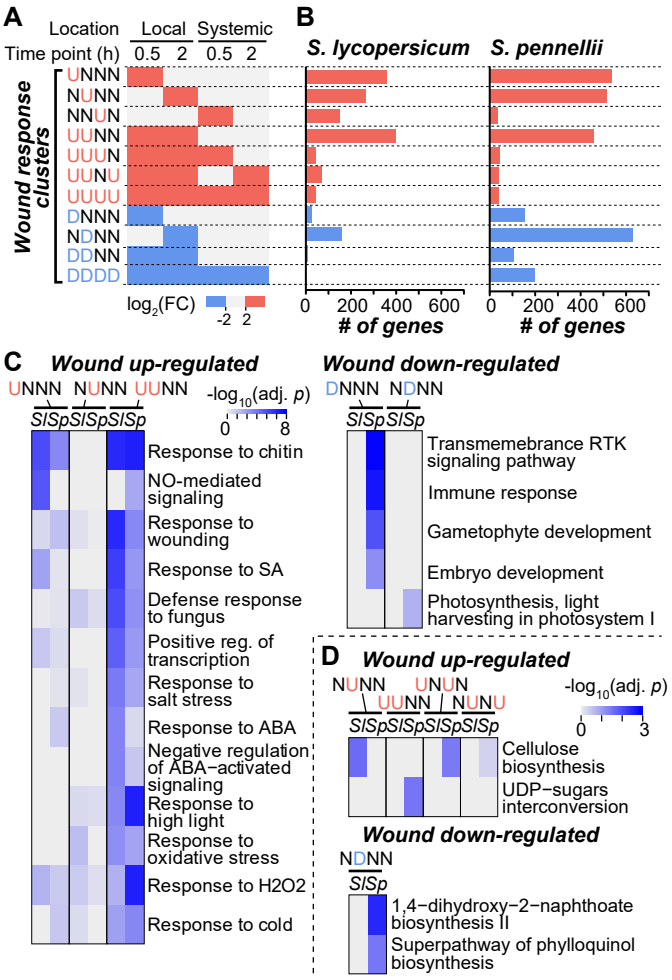


Figure 2. Numbers of genes and functional category enrichments in wound response clusters
(A) Definitions of wound-response clusters. U (red): up-regulation, (log_2FC)>2. N (gray): no significant change, 2>(log_2FC)>-2. D (blue): down-regulation, (log_2FC)<-2. Only clusters with >40 genes in ≥1 species were shown. (B) Numbers of wound-responsive genes in the clusters shown in (A) for *S. lycopersicum* (left) and *S. pennellii* (right). Red and Blue: up- and down-regulated clusters. (C) Gene ontology biological process categories significantly enriched in wound up- (adjusted *p*-values <1e-03) and down-regulated (adjusted *p*-values <1e-02) cluster genes from *S. lycopersicum* (*Sl*) or *S. pennellii* (*Sp*). (D) Metabolic pathways significantly enriched (adjusted *p*-values <5e-02) in *S. lycopersicum* and *S. pennellii* genes from wound up- and down-regulated clusters. Deeper shades of blue indicate higher -log_{10}(adjusted *p*-value).

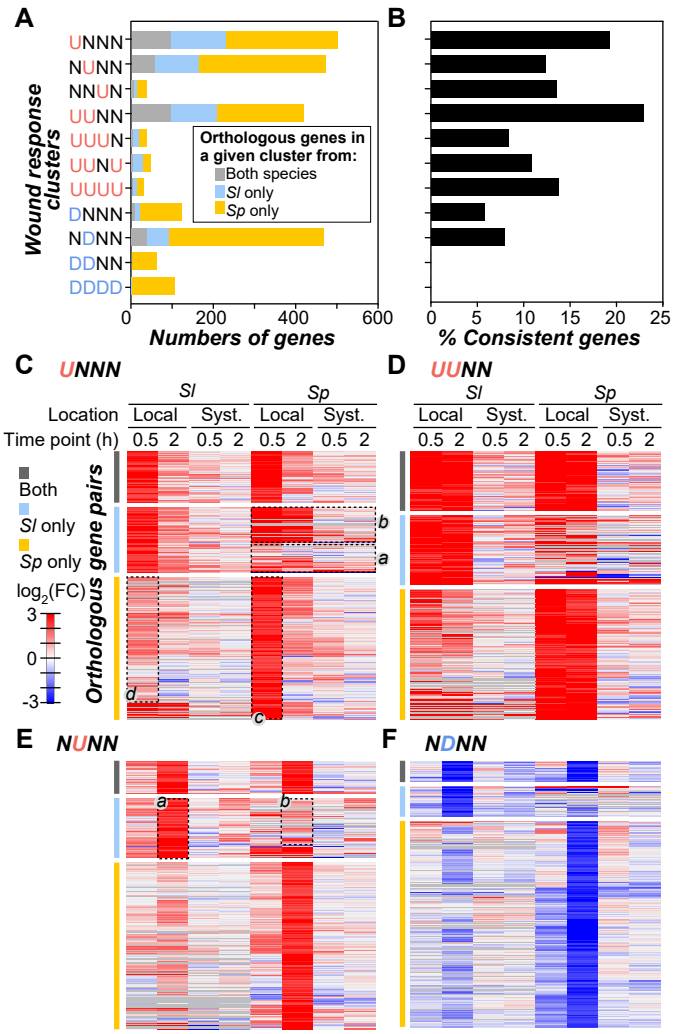
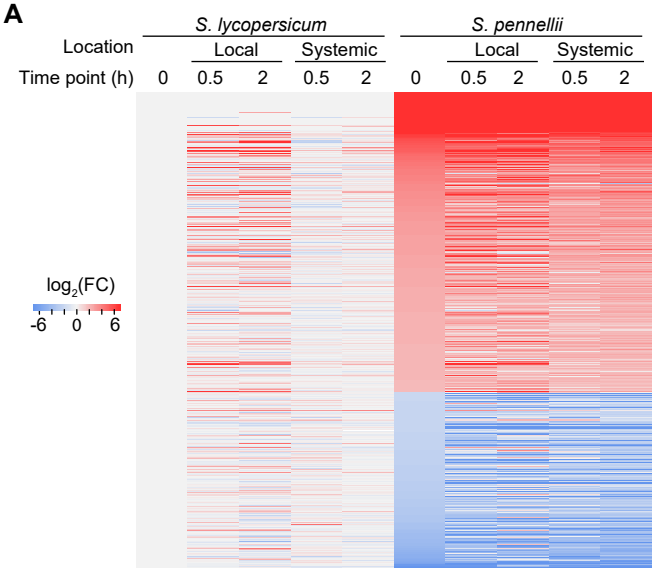


Figure 3. Divergence of wound responses among orthologous genes

(A) Number of orthologous genes with $|\log_2 FC)|>2$ in the wound-response clusters as defined in **Fig. 2A**. Gray: orthologous genes from both species were in the same cluster. Cyan: the *S. lycopersicum* (*Sl*) ortholog is in the indicated cluster but not the *S. pennellii* one. Orange: the *S. pennellii* (*Sp*) ortholog is in the indicated cluster but not the *S. lycopersicum* one. (B) Percentage of the orthologous genes that are considered to have consistent regulatory patterns (in the same cluster) in each cluster. (C,D,E,F) Heatmaps showing the differential expression levels ($\log_2$(FC), FC: fold change) of orthologous genes in UNNN (C), UUNN (D), NUNN (E) and NDNN (F) clusters. The bars on the left of each heatmap are colored the same way as in (A). The dotted rectangles highlight differential expression patterns discussed in the main text.

A



B

Gene	*Sl* ID	*Sp* ID	log_2(FC)	Adj. *p*	log_2(FC)*
SlCUS2	*Solyc04g050730*	*Spenn04g02142*	1.01	$1.38x10^{-25}$	1.34
CER6	*Solyc02g085870*	*Sopen02g030550*	0.73	$1.55x10^{-13}$	1.57
CER8	*Solyc01g099100*	*Sopen01g042660*	0.78	$1.08x10^{-13}$	0.89
MYB41	*Solyc02g079280*	*Spenn02g01871*	2.74	$1.46x10^{-10}$	3.51

Figure 4. Genes differentially expressed between species prior to wounding
(A) Heatmap showing differential expression where fold change (FC) values of all samples were calculated using the *S. lycopersicum* unwounded control (time point 0) expression values as the denominator. Only genes in the unwounded control in *S. pennellii* with significant FC values in comparison to the unwounded control in *S. lycopersicum* were shown (n=593, $|log_2(FC)|>2$). (B) Differential expression values and test statistics contrasting *S. pennellii* and *S. lycopersicum* unwounded controls between orthologous gene pairs from both species involved in biosynthesis of cuticular wax and cutin in this and an earlier study (*, Bolger et al., 2014a).

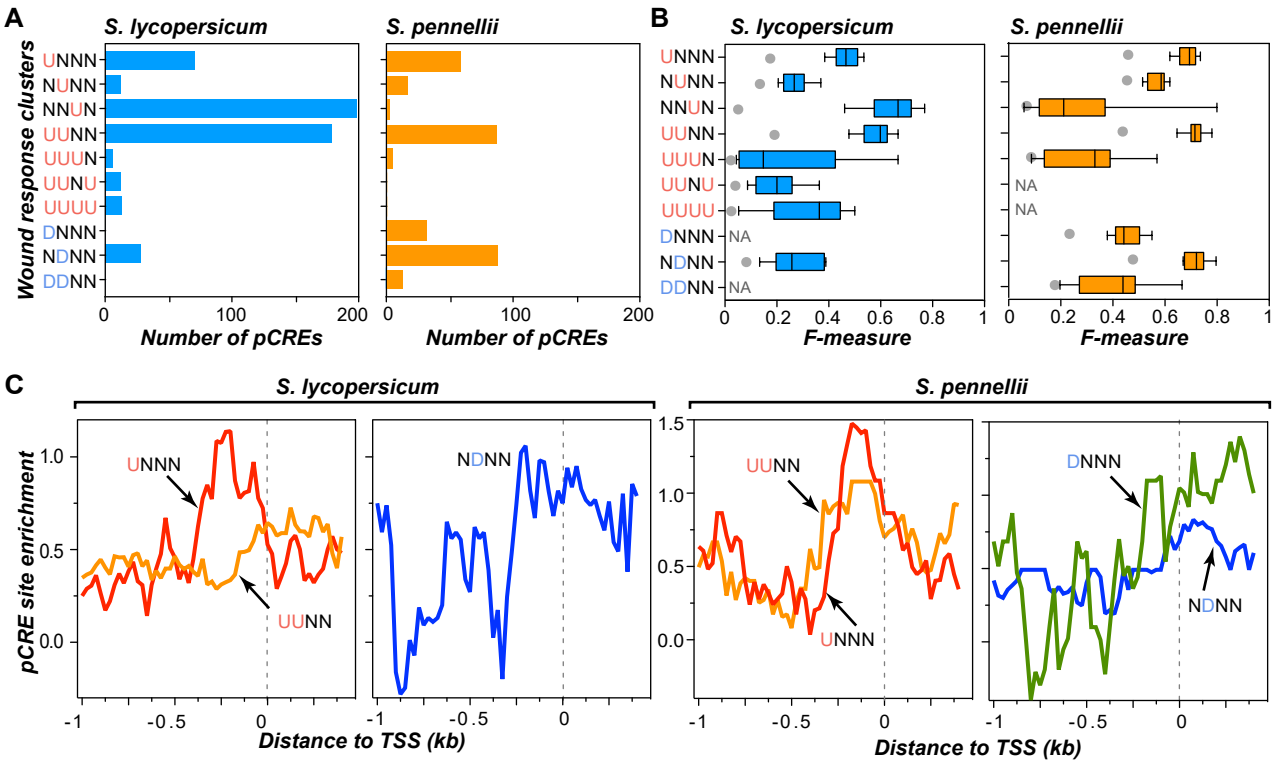


Figure 5. Evidence indicating biological relevance of putative CREs

(A) Number of putative CREs (pCREs) identified through the k-mer pipeline (see Methods) for each wound-response cluster in *S. lycopersicum* (blue) or *S. pennellii* (orange). Only the clusters with pCREs in ≥1 species are shown. (B) Boxplot showing the wound-response prediction performance (F-measure) based on a model using pCREs identified from genes in a wound response cluster. F-measure: the harmonic mean of precision (proportion predicted correctly) and recall (proportion true positives predicted). The maximum F-measure is 1, indicating a perfect model. For each wound-response cluster, 10 F-measures were calculated from 10-fold cross validation and are shown as a boxplot. Gray dot: the average F-measure of 10,000 random predictions indicating the performance of a meaningless model. NA: not applicable since no pCRE was found in the cluster. (C) Enrichment of sites of pCREs identified from four different clusters. For each pCRE, the degree of enrichment of its sites around transcription start sites (TSSs) was represented as the $\log_2$ ratio between pCRE site frequencies of genes in a cluster and frequencies of the same pCREs in genes not responsive to wounding. This log ratio was generated for each pCRE in the region from 1 kb upstream to 0.5 kb downstream of transcription start sites (TSSs) with a sliding window of 100 bp and a step size of 25 bp. For each cluster, the median log ratios of all pCREs identified from the cluster in question was shown.

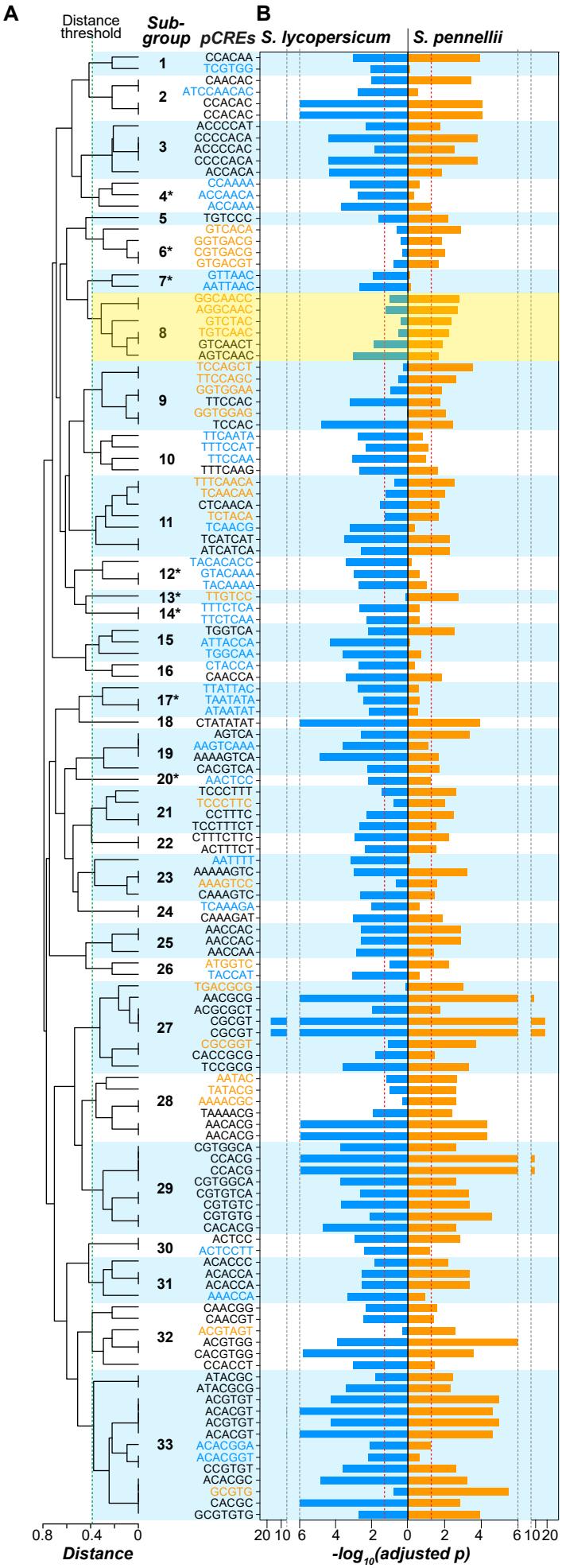


Figure 6. Differential enrichment of pCREs in UNNN cluster genes from two tomato species

(A) Dendrogram showing the distances between the pCREs identified from UNNN cluster genes and enriched in *S. lycopersicum* only (blue), *S. pennellii* only (orange), or both species (black). The dotted line indicates the threshold distance defined based on the 95th percentile distances between binding motifs of TFs from distinct families and defines multiple pCRE sub-groups (numbered) where each sub-group contains pCREs likely bound by TFs of the same family (distance threshold=0.39). Single-species subgroups with pCREs from only one spe-cies are labeled with asterisks. Note that some pCRE duplicates were due to their identifica-tion from both *S. lycopersicum* and *S. pennellii*. (B) Degrees of pCRE site enrichment in *S. lycopersicum* (blue) and *S. pennellii* (orange) UNNN genes. Adjusted *p*-value: multiple-testing corrected *p*-value. Dashed line, adjusted p <0.05. Yellow box: pCREs similar to the W-box element.

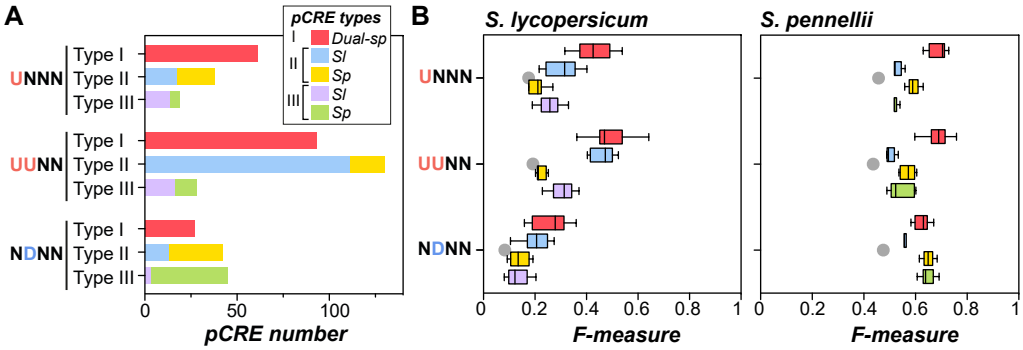


Figure 7. Performance of the Type I, II, III pCREs in predicting wound response
(A) Numbers of pCREs that were consistently enriched in both species and belong to a dual-species sub-group (red, Type I), belong to "dual-species" subgroup but were specifically enriched in *S. lycopersicum* (blue, Type II from *Sl*) or in *S. pennellii* (orange, Type II from *Sp*), and belong to "single-species" subgroup and were specifically enriched in *S. lycopersicum* (purple, Type III from *Sl*) or in *S. pennellii* (green, Type III from *Sp*) in three example wound-response clusters. (B) Boxplot showing the wound response prediction performance (F-measure) based on a model using the pCRE sets in (A). For each wound-response cluster, 10 F-measures were calculated from 10-fold cross validation and shown as a boxplot. Gray dot: the average F-measure of 10,000 random predictions indicating the performance of a meaningless model.

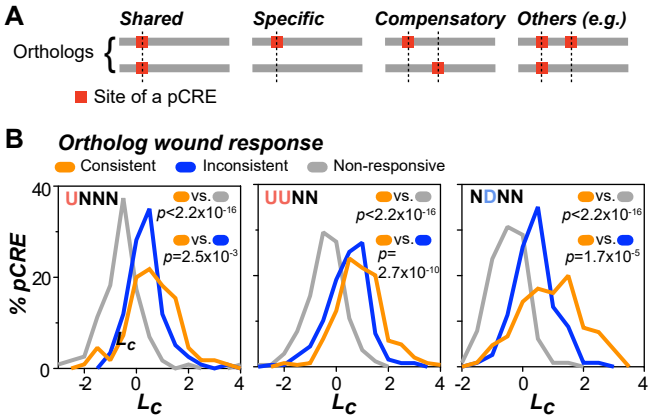


Figure 8. Relationships of pCRE site turnover and wound-responsive gene expression between orthologs
(A) Types of pCRE sites. Shared: the sites of a pCRE are present in both orthologs and located at the same position. Specific: the site of a pCRE is present only in one ortholog but not the other. Compensatory: the sites are present in both species but in different locations. Others: any situation that does not belong to the previous three types. Gray line: the defined regulatory regions from the orthologous gene pairs (See Methods). (B) The conservation likelihood (*Lc*) of a pCRE in the UNNN (left panel), the UUNN (middle panel), and NDNN (right panel) clusters. For a pCRE, its *Lc* is defined as the log ratio between the proportions of sites that shared and those that are specific (see Methods). The *Lc* for each pCRE was evaluated using orthologous gene pairs with consistent (belong to the same wound-response cluster, orange) and inconsistent (belong to different clusters, blue) wound responses, as well as orthologous genes that are not responsive to wounding (Non-responsive, gray). *P*-values: testing whether the likelihood scores generating based the blue or gray datasets differ from the orange one (one-sided Mann-Whitney U test).

Parsed Citations

Adie, B.A.T., Perez-Perez, J., Perez-Perez, M.M., Godoy, M., Sanchez-Serrano, J.J., Schmelz, E.A., and Solano, R. (2007). ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defenses in Arabidopsis. The Plant cell 19, 1665-1681.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. Nature reviews. Genetics 2, 130-138.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research 37, W202-208.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Banerjee, A., and Roychoudhury, A. (2015). WRKY proteins: signaling and regulation of expression during abiotic stress responses. ScientificWorldJournal 2015, 807560.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bass, A.J., Dabney, A., and Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. R package version 2.6.0.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bauchet, G.a.C., M. (2012). Genetic diversity in tomato (Solanum lycopersicum) and its wild relatives. In Environmental Sciences (Çaliskan, M., ed.). Rijeka, Croatia: InTechOpen, 133–162.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Benjamini, Y.a.Y.H. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57, 289-300.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Berardi, P., Meyyappan, M., and Riabowol, K.T. (2003). A novel transcriptional inhibitory element differentially regulates the cyclin D1 gene in senescent cells. J Biol Chem 278, 7510-7519.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sorensen, I., Lichtenstein, G., Fich, E.A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., Aflitos, S.A., Schijlen, E., Jimenez-Gomez, J.M., Ryngajllo, M., Kimura, S., Kumar, R., Koenig, D., Headland, L.R., Maloof, J.N., Sinha, N., van Ham, R.C., Lankhorst, R.K., Mao, L., Vogel, A., Arsova, B., Panstruga, R., Fei, Z., Rose, J.K., Zamir, D., Carrari, F., Giovannoni, J.J., Weigel, D., Usadel, B., and Fernie, A.R. (2014a). The genome of the stress-tolerant wild tomato species Solanum pennellii. Nat Genet 46, 1034-1038.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

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

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Borneman, A.R., Gianoulis, T.A., Zhang, Z.D., Yu, H., Rozowsky, J., Seringhaus, M.R., Wang, L.Y., Gerstein, M., and Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species. Science 317, 815-819.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Boter, M., Ruiz-Rivero, O., Abdeen, A., and Prat, S. (2004). Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. Genes & development 18, 1577-1591.

Pubmed: Author and Title
CrossRef: Author and Title

Google Scholar: Author Only Title Only Author and Title

Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C.J., Ware, D., Shiu, S.H., Childs, K.L., Sun, Y., Jiang, N., and Yandell, M. (2014). MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol 164, 513-524.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134, 25-36.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chaudhary, B. (2013). Plant domestication and resistance to herbivory. Int J Plant Genomics 2013, 572784.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chen, Y.H., Gols, R., and Benrey, B. (2015). Crop domestication and its impact on naturally selected trophic interactions. Annu Rev Entomol 60, 35-58.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cominelli, E., Sala, T., Calvi, D., Gusmaroli, G., and Tonelli, C. (2008). Over-expression of the Arabidopsis AtMYB41 gene alters cell expansion and leaf surface permeability. Plant J 53, 53-64.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Doebley, J., and Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. The Plant cell 10, 1075-1082.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell 127, 1309-1321.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32, 1792-1797.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Emerson, J.J., and Li, W.H. (2010). The genetic basis of evolutionary change in gene expression levels. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 365, 2581-2590.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Franco-Zorrilla, J.M., Lopez-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. Proceedings of the National Academy of Sciences of the United States of America 111, 2367-2372.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Gong, P.J., Zhang, J.H., Li, H.X., Yang, C.X., Zhang, C.J., Zhang, X.H., Khurram, Z., Zhang, Y.Y., Wang, T.T., Fei, Z.J., and Ye, Z.B. (2010). Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. J Exp Bot 61, 3563-3575.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Gordon, D.B., Nekludova, L., McCallum, S., and Fraenkel, E. (2005). TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs. Bioinformatics 21, 3164-3165.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Green, T.R., and Ryan, C.A. (1972). Wound-Induced Proteinase Inhibitor in Plant Leaves: A Possible Defense Mechanism against Insects. Science 175, 776-777.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. Euphytica 156, 1-13.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Heyndrickx, K.S., Van de Velde, J., Wang, C., Weigel, D., and Vandepoele, K. (2014). A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. The Plant cell 26, 3894-3910.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hobo, T., Asada, M., Kowyama, Y., and Hattori, T. (1999). ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. Plant J 19, 679-689.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hooker, T.S., Millar, A.A., and Kunst, L. (2002). Significance of the expression of the CER6 condensing enzyme for cuticular wax production in Arabidopsis. Plant Physiol 129, 1568-1580.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Howe, G.A., and Jander, G. (2008). Plant immunity to insect herbivores. Annu Rev Plant Biol 59, 41-66.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Howe, G.A., and Schaller, A. (2008). Direct Defenses in Plants and Their Induction by Wounding and Insect Herbivores. (Springer Netherlands).
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Huot, B., Yao, J., Montgomery, B.L., and He, S.Y. (2014). Growth-defense tradeoffs in plants: a balancing act to optimize fitness. Mol Plant 7, 1267-1287.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ichihashi, Y., Aguilar-Martinez, J.A., Farhi, M., Chitwood, D.H., Kumar, R., Millon, L.V., Peng, J., Maloof, J.N., and Sinha, N.R. (2014). Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. Proceedings of the National Academy of Sciences of the United States of America 111, E2616-2621.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kamenetzky, L., Asis, R., Bassi, S., de Godoy, F., Bermudez, L., Fernie, A.R., Van Sluys, M.A., Vrebalov, J., Giovannoni, J.J., Rossi, M., and Carrari, F. (2010). Genomic analysis of wild tomato introgressions determining metabolism- and yield-associated traits. Plant Physiol 152, 1772-1786.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kaufmann, K., Pajoro, A., and Angenent, G.C. (2010). Regulation of transcription in plants: mechanisms controlling developmental switches. Nat Rev Genet 11, 830-842.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Koenig, D., Jimenez-Gomez, J.M., Kimura, S., Fulop, D., Chitwood, D.H., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., Tat, A.V., Tohge, T., Bolger, A., Schneeberger, K., Ossowski, S., Lanz, C., Xiong, G., Taylor-Teeples, M., Brady, S.M., Pauly, M., Weigel, D., Usadel, B., Fernie, A.R., Peng, J., Sinha, N.R., and Maloof, J.N. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proceedings of the National Academy of Sciences of the United States of America 110, E2655-2662.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li, L., Zhao, Y., McCaig, B.C., Wingerd, B.A., Wang, J., Whalon, M.E., Pichersky, E., and Howe, G.A. (2004). The tomato homolog of CORONATINE-INSENSITIVE1 is required for the maternal control of seed maturation, jasmonate-signaled defense responses, and glandular trichome development. The Plant cell 16, 126-143.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Liu, M.J., Seddon, A.E., Tsai, Z.T., Major, I.T., Floer, M., Howe, G.A., and Shiu, S.H. (2015). Determinants of nucleosome positioning and their influence on plant gene expression. Genome Res 25, 1182-1195.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lu, S., Song, T., Kosma, D.K., Parsons, E.P., Rowland, O., and Jenks, M.A. (2009). Arabidopsis CER8 encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1) that has overlapping functions with LACS2 in plant wax and cutin synthesis. Plant J 59, 553-564.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2016). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.4.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Mahony, S., and Benos, P.V. (2007). STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic acids research 35, W253-258.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Meyer, R.S., and Purugganan, M.D. (2013). Evolution of crop species: genetics of domestication and diversification. Nature reviews. Genetics 14, 840-852.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Narsai, R., Howell, K.A., Millar, A.H., O'Toole, N., Small, I., and Whelan, J. (2007). Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. The Plant cell 19, 3418-3436.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Nesbitt, T.C., and Tanksley, S.D. (2002). Comparative sequencing in the genus Lycopersicon. Implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics 162, 365-379.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Nitta, K.R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E.E., and Taipale, J. (2015). Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. Elife 4.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Nowicka, B., and Kruk, J. (2010). Occurrence, biosynthesis and function of isoprenoid quinones. Biochim Biophys Acta 1797, 1587-1605.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 166, 1598.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat Genet 44, 743-750.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Romero, I.G., Ruvinsky, I., and Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. Nature reviews. Genetics 13, 505-516.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Rong, J., Lammers, Y., Strasburg, J.L., Schidlo, N.S., Ariyurek, Y., de Jong, T.J., Klinkhamer, P.G., Smulders, M.J., and Vrieling, K. (2014). New insights into domestication of carrot from root transcriptome analyses. BMC Genomics 15, 895.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Rosenthal, J.P., and Dirzo, R. (1997). Effects of life history, domestication and agronomic selection on plant defence against insects: Evidence from maizes and wild relatives. Evolutionary Ecology 11, 337-355.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Rushton, P.J., and Somssich, I.E. (1998). Transcriptional control of plant genes responsive to pathogens. Curr Opin Plant Biol 1, 311-315.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Scranton, M.A., Fowler, J.H., Girke, T., and Walling, L.L. (2013). Microarray analysis of tomato's early and late wound response reveals new regulatory targets for Leucine aminopeptidase A. PLoS One 8, e77889.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Siberil, Y., Doireau, P., and Gantet, P. (2001). Plant bZIP G-box binding factors. Modular structure and activation mechanisms. Eur J Biochem 268, 5655-5666.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews. Genetics 13, 613-626.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Stankovic, B., Vian, A., Henry-Vian, C., and Davies, E. (2000). Molecular cloning and characterization of a tomato cDNA encoding a systemically wound-inducible bZIP DNA-binding protein. Planta 212, 60-66.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Stergachis, A.B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A.P., Zhang, M., Byron, R., Canfield, T., Stelhing-Sun, S., Lee, K., Thurman, R.E., Vong, S., Bates, D., Neri, F., Diegel, M., Giste, E., Dunn, D., Vierstra, J., Hansen, R.S., Johnson, A.K., Sabo, P.J., Wilken, M.S., Reh, T.A., Treuting, P.M., Kaul, R., Groudine, M., Bender, M.A., Borenstein, E., and Stamatoyannopoulos, J.A. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. Nature 515, 365-370.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Stewart, A.J., Hannenhalli, S., and Plotkin, J.B. (2012). Why transcription factor binding sites are ten nucleotides long. Genetics 192, 973-985.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., Stergachis, A.B., Vernot, B., Johnson, A.K., Haugen, E., Sullivan, S.T., Thompson, A., Neri, F.V., 3rd, Weaver, M., Diegel, M., Mnaimneh, S., Yang, A., Hughes, T.R., Nemhauser, J.L., Queitsch, C., and Stamatoyannopoulos, J.A. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Rep 8, 2015-2030.
Pubmed: Author and Title
CrossRef: Author and Title

Google Scholar: Author Only Title Only Author and Title

Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M.B., Ross-Ibarra, J., Myers, C.L., Tiffin, P., and Springer, N.M. (2012). Reshaping of the maize transcriptome by domestication. Proceedings of the National Academy of Sciences of the United States of America 109, 11878-11883.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. Trends Plant Sci 21, 506-515.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Tal, M., and Shannon, M.C. (1983). Salt Tolerance in the Wild Relatives of the Cultivated Tomato - Responses of Lycopersicon-Esculentum, Lycopersicon-Cheesmanii, Lycopersicon-Peruvianum, Solanum-Pennellii and F1 Hybrids to High Salinity. Aust J Plant Physiol 10, 109-117.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511-515.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Tsai, Z.T., Shiu, S.H., and Tsai, H.K. (2015). Contribution of Sequence Motif, Chromatin State, and DNA Structure Features to Predictive Models of Transcription Factor Binding in Yeast. PLoS computational biology 11, e1004418.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

van Verk, M.C., Pappaioannou, D., Neeleman, L., Bol, J.F., and Linthorst, H.J. (2008). A Novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. Plant Physiol 146, 1983-1995.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J.C., Galli, M., Lewsey, M.G., Huang, E., Mukherjee, T., Chen, X., Reece-Hoyes, J.S., Govindarajan, S., Shaulsky, G., Walhout, A.J., Bouget, F.Y., Ratsch, G., Larrondo, L.F., Ecker, J.R., and Hughes, T.R. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158, 1431-1443.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

White, M.A., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2013). Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. Proceedings of the National Academy of Sciences of the United States of America 110, 11952-11957.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature reviews. Genetics 13, 59-69.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in cis and trans gene regulation. Nature 430, 85-88.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wittstock, U., and Gershenzon, J. (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. Current Opinion in Plant Biology 5, 300-307.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 20, 1377-1419.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Yu, J.J., Thornton, K., Guo, Y., Kotz, H., and Reed, E. (2001). An ERCC1 splicing variant involving the 5'-UTR of the mRNA may have a transcriptional modulatory function. Oncogene 20, 7694-7698.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zhang, C., Xuan, Z., Otto, S., Hover, J.R., McCorkle, S.R., Mandel, G., and Zhang, M.Q. (2006). A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome. Nucleic acids research 34, 2238-2246.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Regulatory Divergence in Wound-Responsive Gene Expression between Domesticated and Wild Tomato

Ming-Jung Liu, Koichi Sugimoto, Sahra Uygun, Nicholas Panchy, Michael S. Campbell, Mark Yandell, Gregg A. Howe and Shin-Han Shiu

Plant Cell; originally published online May 9, 2018;
DOI 10.1105/tpc.18.00194