Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato

Daniel Koenig[a,b,1], José M. Jiménez-Gómez[a,c,1], Seisuke Kimura[a,d,2], Daniel Fulop[a,2], Daniel H. Chitwood[a], Lauren R. Headland[a], Ravi Kumar[a], Michael F. Covington[a], Upendra Kumar Devisetty[a], An V. Tat[a], Takayuki Tohge[e], Anthony Bolger[f], Korbinian Schneeberger[b,g], Stephan Ossowski[b,h], Christa Lanz[b], Guangyan Xiong[i], Mallorie Taylor-Teeples[a,j], Siobhan M. Brady[a,j], Markus Pauly[i], Detlef Weigel[b,3], Björn Usadel[f,k,l], Alisdair R. Fernie[e], Jie Peng[m], Neelima R. Sinha[a], and Julin N. Maloof[a,3]

[a]Department of Plant Biology and [m]Department of Statistics, University of California, Davis, CA 95616; [b]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; [c]Department of Plant Breeding and Genetics and [g]Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, 50829 Köln, Germany; [d]Department of Bioresource and Environmental Sciences, Kyoto Sangyo University, Kyoto 603-8555, Japan; [e]Department of Molecular Physiology and [f]Department of Metabolic Networks, Max Planck Institute of Molecular Plant Physiology, 14476 Golm, Germany; [h]Genes and Disease Program, Centre for Genomic Regulation, Barcelona 08003, Spain; [i]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; [j]Genome Center, University of California, Davis, CA 95616; [k]Institute of Biology 1, Rheinisch-Westfaelische Technische Hochschule Aachen, 52056 Aachen, Germany; and [l]Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich, 52425 Jülich, Germany

Contributed by Detlef Weigel, May 30, 2013 (sent for review February 14, 2013)

Although applied over extremely short timescales, artificial selection has dramatically altered the form, physiology, and life history of cultivated plants. We have used RNAseq to define both gene sequence and expression divergence between cultivated tomato and five related wild species. Based on sequence differences, we detect footprints of positive selection in over 50 genes. We also document thousands of shifts in gene-expression level, many of which resulted from changes in selection pressure. These rapidly evolving genes are commonly associated with environmental response and stress tolerance. The importance of environmental inputs during evolution of gene expression is further highlighted by large-scale alteration of the light response coexpression network between wild and cultivated accessions. Human manipulation of the genome has heavily impacted the tomato transcriptome through directed admixture and by indirectly favoring nonsynonymous over synonymous substitutions. Taken together, our results shed light on the pervasive effects artificial and natural selection have had on the transcriptomes of tomato and its wild relatives.

domestication | biotic stress | abiotic stress

Domestication has long served as an important example of severe phenotypic divergence in response to selection. Darwin recognized the parallel between the processes of domestication and adaptation in the wild and used this analogy to emphasize the power of selection in generating phenotypic diversity (1). The genetic basis of domestication-associated phenotypes has been examined in several instances, most notably in maize, rice, tomato, and dogs (reviewed in refs. 2–5). The clear conclusion from these studies is that the rapid phenotypic divergence associated with domestication is often attributable to very few genetic loci (6). Improvements to DNA sequence technologies have allowed studies of the effect of domestication at the whole-genome level. Early population genetic analyses in maize found that very few genes (~5%) show evidence of positive selection during domestication of maize (7), and recent work using whole-genome resequencing has found a similar proportion of the genome was under positive selection (8). Evidence for strong selective sweeps at a limited number of loci has also been found in rice and dog genomes (9). Together with the previous genetic mapping work, these studies support the model that relatively few mutations experienced extremely strong selection by humans during domestication.

Although not the target of direct positive selection, the rest of the genome still experiences a dramatic shift in evolutionary pressures during domestication. Most characterized domestication events are associated with an extreme genetic bottleneck and alleviation of selective constraints in the original niche (10). These factors are predicted to increase the relative rate of nonsynonymous to synonymous (dN/dS) substitution, potentially resulting in the fixation of deleterious alleles (11). Previous studies comparing the distribution of polymorphisms between rice and dogs and their closest wild relatives have suggested that this may be the case (12, 13). However, the lack of genome-wide polymorphism data in multiple wild accessions has limited these comparisons because of ambiguous assignment of ancestral state. Evidence for changes at the transcriptional level during domestication have also been examined; for example, a recent study in maize has suggested widespread alteration of transcriptional networks during domestication (14). Although some of these changes are associated with genes that also show evidence of positive selection, changes in the topology of the gene-expression network might also result from fixation of mutations during the domestication bottleneck. Regardless, although absolute changes in gene expression or changes in network topology are thought to be important

Significance

One of the most important technological advances by humans is the domestication of plant species for the production of food. We have used high-throughput sequencing to identify changes in DNA sequence and gene expression that differentiate cultivated tomato and its wild relatives. We also identify hundreds of candidate genes that have evolved new protein sequences or have changed expression levels in response to natural selection in wild tomato relatives. Taken together, our analyses provide a snapshot of genome evolution under artificial and natural conditions.

Author contributions: D.K., J.M.J.-G., S.K., N.R.S., and J.N.M. designed research; D.K., J.M.J.-G., S.K., L.R.H., R.K., U.K.D., T.T., G.X., M.T.-T., S.M.B., M.P., B.U., and J.N.M. performed research; A.B., K.S., S.O., C.L., D.W., B.U., and A.R.F. contributed new reagents/analytic tools; D.K., J.M.J.-G., D.F., D.H.C., M.F.C., A.V.T., J.P., and J.N.M. analyzed data; and D.K., J.M.J.-G., D.F., N.R.S., and J.N.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE45774 and SRP019504).

[1]D.K. and J.M.J.-G. contributed equally to this work.

[2]S.K. and D.F. contributed equally to this work.

[3]To whom correspondence may be addressed. E-mail: weigel@tue.mpg.de or jnmaloof@ucdavis.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1309606110/-/DCSupplemental.

during domestication, genome-wide comparison of expression between domesticated and multiple wild species is lacking.

One of the most heavily studied domestication events is that of tomato. Tomato is a member of a complex of 13 interfertile species that occupy a wide range of habitats in South America (15). The exact date of tomato domestication is debated, but it is clear that domesticated lines existed in Mexico at the time of the arrival of Europeans, and were brought back to Europe as a novelty, only to be used for food there in the 17th or 18th century. Tomato cultivars were subsequently reintroduced to the Americas. Thus, cultivated tomato has undergone a series of sequential bottlenecks, resulting in extremely low intraspecific genetic diversity (15). The most obvious domestication associated trait in tomato is a dramatic increase in fruit size. This trait has been the subject of extensive genetic analysis, and is controlled by a relatively small number of loci (16) making it typical of most domestication-associated traits. The high phenotypic diversity among wild tomato relatives and the relatively recent domestication of tomato itself makes it an excellent system to compare the effects of artificial and natural selection.

We deeply sequenced the transcriptomes of six species to ascertain the effects of natural and artificial selection on gene expression and sequence diversity. Our panel included one accession of domesticated tomato (*Solanum lycopersicum* M82), two related red-fruited wild species (*Solanum pimpinellifolium* and *Solanum galapagense*) and three green-fruited wild accessions from vastly differing habitats (*Solanum habrochaites*, a high altitude-adapted, chilling-tolerant accession; a high altitude drought-tolerant accession, *Solanum chmielewskii*; and *Solanum pennellii*, a desert-adapted accession) (Fig. 1*A*). These five wild species were chosen because of their dramatic phenotypic variability, but also because of their widespread use as genetic donors during cultivated tomato improvement, allowing us to define sequence and expression-level polymorphisms relevant to breeding and natural variation (17). Our analysis provides ample evidence for evolution in response to environmental cues in tomato relatives, and suggests interesting differences between artificial and natural selection.

Results

Characterization of Sequence Diversity in Wild and Cultivated Tomato. We conducted a series of experiments to define transcripts and identify sequence polymorphisms in our tomato panel. Two experiments (*Materials and Methods* and *SI Appendix*, Table S1) were conducted to ascertain interspecific variation in gene-expression levels. The first experiment compared gene expression in aerial seedling tissues of the species *S. pennellii*, *S. lycopersicum*, *S. habrochaites*, and *S. pimpinellifolium*. A second experiment compared six tissues collected from *S. lycopersicum* and *S. pennellii*. The remaining samples from either additional tissues or species were collected at separate times and used only for polymorphism discovery.

After alignment to the tomato reference genomic sequence (var. Heinz), our sequences covered an average of 67.4% of the annotated exonic gene space and allowed us to identify 1.5 million polymorphic sites among the 23.9 Mb covered in all samples (*SI Appendix*, Figs. S1–S6 and Tables S2 and S3, and Dataset S1). De novo contigs assembled from our reads covered 54% of the annotated genes and identified 34 transcripts not present in the current release of the Heinz reference genome (*SI Appendix*, Figs. S7 and S8, and Tables S4 and S5). Fewer than 20% (6 of 34) of these putative unique transcripts show homology with functionally annotated genes. Comparison of global patterns of nucleotide diversity across all accessions revealed a reduction in neutral divergence (dS) near the centromeres, but a relative increase in nonsynonymous substitution in the same regions (Fig. 1 *B* and *C*).

To initiate our evolutionary analysis, we used Bayesian inference methods to construct a phylogeny of the six species rooted with potato sequences (18) (*Solanum phureja*) (Fig. 1*A*).

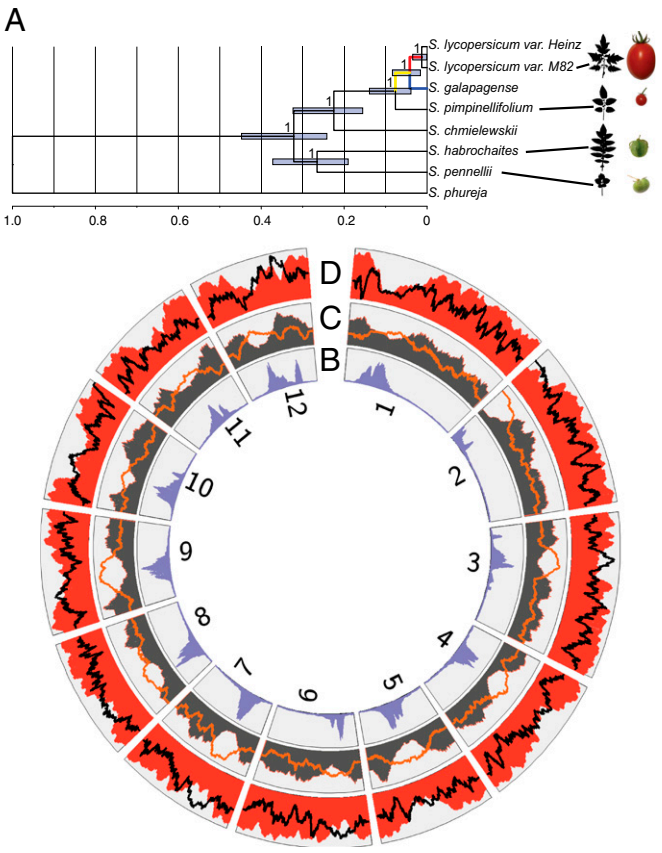

Fig. 1. Diversity in cultivated and wild tomatoes. (*A*) Bayesian relaxed-clock consensus chronogram, and examples of fruit and leaf divergence among tomato and wild relatives; nodes on the tree correspond to median branch lengths and blue bars represent 95% Bayesian confidence interval. (*B*) Distribution of mean distance to adjacent gene, larger distances are associated with centromeric sequences. (*C*) Single rate dS (gray) and single rate dN/dS (orange). (*D*) Frequency of expressed genes (red) and genes differentially expressed between tomato relatives (black). All plots reflect sliding windows (mean of 100 gene windows).

The resulting phylogeny is consistent with published tomato trees (19) resolving a monophyletic red/orange fruited clade and placing the green fruited *S. pennellii* and *S. habrochaites* in a sister clade. Like some previous studies but unlike others (19), the phylogeny places *S. galapagense* as the closest outgroup to domesticated samples. *S. pimpinellifolium*, *S. lycopersicum*, and *S. galapagense* have been shown to hybridize in the wild or through directed introgression (20) (in the case of *S. lycopersicum*) and it is possible that the difference in topologies results from incomplete lineage sorting in the three species and is specific to the particular accessions used in each study.

Consistent with previous studies, cultivated accessions are very similar to each other (< 1 SNP/kb), and a modest number of mutations separate cultivated tomato from its most closely related wild ancestors (< 5 SNP/kb) (*SI Appendix*, Fig. S3). By polarizing our data against the potato genome reference, we found that the spectrum of mutated sites varies between the lines. Mutations shared only by the cultivated tomato lines or unique to *S. galapagense* showed an increased ratio of nonsynonymous to synonymous substitutions (Fig. 2*A*). We directly tested whether the rate of nonsynonymous to synonymous substitution was elevated in cultivated tomato and *S. galapagense* by comparing the estimated tree-wide dN/dS to estimates for the terminal branches for each species and the branch leading to their most recent common ancestor (Fig. 2*B*). Each of the ter-

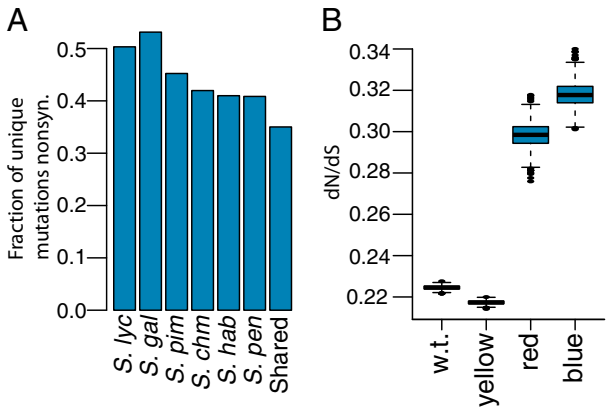

Fig. 2. Evidence for increased nonsynonymous substitution rate in *S. lycopersicum* and *S. galapagense*. (*A*) Fraction of species-specific derived mutations in the coding regions that are nonsynonymous. (*B*) Distributions of dN/dS estimates from 1,000 bootstraps of the transcriptome-wide alignment for the whole tree (w.t.) and the branches labeled with red, blue, and yellow in Fig. 1. *S. lyc*, *S. lycopersicum*; *S. gal*, *S. galapagense*; *S. pim*, *S. pimpinellifolium*; *S. chm*, *S. chmielewskii*; *S. hab*, *S. habrochaites*; *S. pen*, *S. pennellii*.

minal branches, but not the connecting branch, showed significant increases in dN/dS. Both *S. lycopersicum* and *S. galapagense* are thought to have experienced strong genetic bottlenecks (15, 21) (during domestication, and island colonization and recent adaptation, respectively). Our result is consistent with separate bottlenecks in these two species and increased accumulation of potentially deleterious mutations during cultivation and colonization. This change in mutation spectrum may be a result of relaxed purifying selection, fixation of mutations during the genetic bottleneck because of drift, or both.

Evidence for Positive Selection in Wild and Cultivated Tomato. Although relaxed purifying selection is expected to elevate dN/dS by random substitution throughout the genome, positive selection is expected to increase dN/dS within specific loci. From comparison of gene-level estimates of dN/dS in all species (22, 23), we identified 51 genes that show statistically significant ($P < 0.05$) evidence of evolution under positive selection across the phylogeny (Dataset S2). Many of these genes have not been characterized in tomato, but annotated genes included the tomato homolog of the *Arabidopsis thaliana ARGONAUTE 2* and the known tomato-resistance gene *immunity to fusarium wilt-2C4* (24, 25), consistent with rapid evolution of protein sequences in response to pathogen pressure. Homologs of the aluminum transporter *ALUMINUM SENSITIVE 1* and the calcium uptake transporter *MID1-COMPLEMENTING ACTIVITY 1* also showed significantly elevated dN/dS pointing to positive selection in response to abiotic factors, such as soil chemistry (26, 27). This second set of genes is particularly interesting considering the high salt tolerance observed in wild tomato relatives (28).

Divergence in Gene Expression in Wild and Cultivated Tomato. We next searched for evidence for differential expression between aerial seedling tissues of *S. lycopersicum*, *S. habrochaites*, *S. pimpinellifolium*, and *S. pennellii*. For this process, seedling tissues were chosen to minimize the effects of developmental and environmental variation on gene expression. We detected expression of 25,012 transcripts in at least a single accession, and 20,389 in all surveyed accessions. Consistent with previous observations (29), gene expression was low in centromere proximal regions and higher in gene-dense chromosomal arms (Fig. 1*D* and *SI Appendix*, Fig. S9). We fit a generalized linear model (*SI Appendix*, *SI Materials and Methods* and Fig. S10) to our expression data to identify 7,903 genes showing evidence of differential expression among species (*SI Appendix*, Fig. S10). Gene ontology (GO) enrichment analysis of these revealed overrepresentation of genes involved in stress response, defense response, photosynthesis, response to high light, and redox pathways (*SI Appendix*, Table S6). Enrichment for these categories indicates that abiotic and biotic stresses have played a major role driving transcriptional variation among these species.

Interspecific comparisons based on nucleotide alignments and pairwise gene-expression differences revealed a general concordance in tree topology but a striking increase in the *S. pennellii* gene-expression branch length (Fig. 3 *A* and *B*). The number of expression changes specific to the *S. pennellii* lineage was much higher than any other lineage (Fig. 3*C* and *SI Appendix*, Tables S7 and S8), indicating that the transcriptional landscape of *S. pennellii* is highly diverged relative to the other three species. A small but significant increase in unique expression changes was also found in *S. lycopersicum* compared with *S. pimpinellifolium*, suggesting the possibility of accelerated divergence in expression in the domesticated lineage (141 and 91 genes, respectively, χ^2 P value = 0.0007). GO term enrichment analysis identified genes involved in salt stress in all comparisons with *S. pennellii* and modification to sucrose metabolism and starch metabolic process in all comparisons with *S. lycopersicum*; in addition, redox pathways were enriched in many comparisons (*SI Appendix*, Table S9

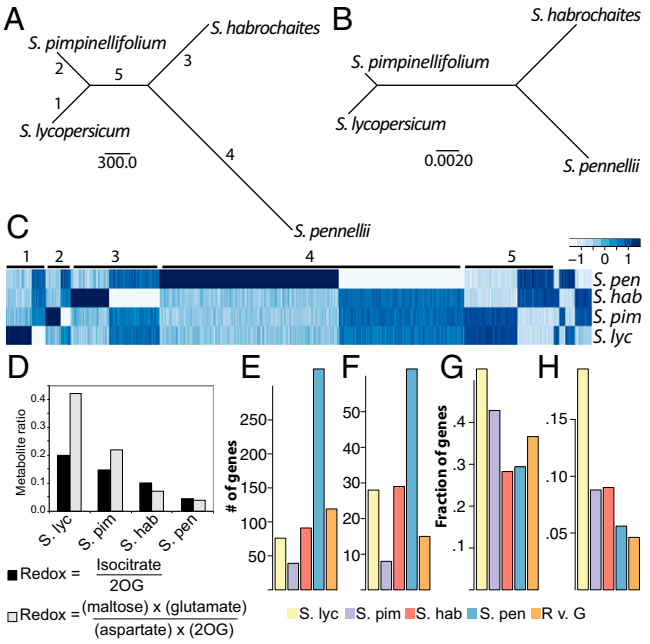

Fig. 3. Interspecific variation in expression. (*A*) Neighbor-joining tree built from the number of pairwise differentially expressed genes compared with (*B*) the unrooted genetic tree from Fig. 1*A*. The scale bar in *A* is for the number of differentially expressed genes and the scale bar in *B* is the expected number of substitutions per site. (*C*) Heatmap depicting scaled expression values of genes separated into two groups by significant contrasts (*SI Appendix, SI Materials and Methods*). The numbers correspond to the branch of the tree on which the changes are assumed to have occurred. (*D*) Product/substrate redox ratio of NAD(P)-linked reactions, calculated as described by refs. 74 and 75. Black and gray indicate redox value of isocitrate dehydrogenase and malate dehydrogenase reactions, respectively. 2OG: 2-oxoglutarate. (*E* and *F*) Number of differentially expressed genes showing evidence of accelerated expression divergence (two-rate Brownian motion fit better than one-rate Brownian motion and OU) at $\Delta AIC > 4$ or $\Delta AIC > 10$, respectively. (*G* and *H*) Proportion of differentially expressed genes unique to each lineage (as determined by pairwise contrasts) showing evidence of accelerated expression divergence at $\Delta AIC > 4$ or $\Delta AIC > 10$, respectively. RvG indicates genes that show contrasting rates of evolution in the red and green fruited lineages.

and S10). We validated that several of these expression changes are reflected in the metabolic state of the plants. Fructose levels were six- to ninefold lower in all wild species compared with *S. lycopersicum* (30). Analysis of existing GC-MS data (30, 31) revealed that the product-to-substrate ratio of redox-coupled NAD(P) reactions in *S. habrochaites* and *S. pimpinellifolium* were twofold and in *S. pennellii* more than 10-fold lower than in *S. lycopersicum*, indicating that the NAD(P) pool is in a more oxidized state in the three wild species (Fig. 3*D*). These metabolic changes combined with enrichment in transcriptional changes provide strong evidence that redox pathways are rapidly evolving among these species. Furthermore, the substantial shift in *S. pennellii* is consistent with adaptation to high light conditions. In summary, the pathways identified by these analyses are consistent with the expected selective pressures on each of these lineages, with strong natural selection for life in a desert environment for *S. pennellii* and artificial selection for palatable fruits during breeding of domesticated tomato.

Analysis of Selective Pressures on Gene Expression. Gene-expression variation can result from random genetic drift or changes in selective pressure. To identify genes that have potentially undergone a shift in selection regime, we compared the fit of three evolutionary models to the gene-expression levels in our dataset: a model of evolution under random drift (Brownian motion single rate), stabilizing selection (Ornstein-Uhlenbeck, OU), or a change in evolutionary rate along a particular lineage (Brownian motion two rate) (32–34) (*SI Appendix*, Table S11 and Dataset S3). Genes whose expression values showed a substantially better fit to the two-rate model and that had accelerated evolutionary rates in a particular lineage were considered candidates for alteration in selective regime in that lineage. Fit was assessed for each of the three models and then compared between the accelerated rate model and the other models using the change in Akaike information criteria (ΔAIC). Increased ΔAIC indicates stronger fit for the accelerated rate model compared with both of the other models (see *SI Appendix* for additional information). Among differentially expressed genes ($P < 0.01$), there was evidence for differing rates of evolution across the tree in 1,764 genes (22.3% of differentially expressed genes, ΔAIC > 4) and strong evidence in 428 genes (5.4% of differentially expressed genes; ΔAIC > 10) (*SI Appendix*, Table S11). The largest group of genes was evolving at a faster rate along the *S. pennellii* branch, but increasing the ΔAIC threshold increased the relative number of genes found in the other branches (Fig. 3 *E* and *F*). Furthermore, the proportion of differentially expressed genes with evidence of accelerated evolution of expression levels was higher in *S. lycopersicum* than in *S. pennellii* (or any of the other branches) (Fig. 3 *G* and *H*). These results indicate that much of the rapid divergence in gene expression that has occurred in *S. pennellii* can be explained by neutral processes. In contrast, relatively few genes have changed in *S. lycopersicum*, but these genes are more likely to show evidence for a *S. lycopersicum*-specific change in evolutionary rate.

Genes accelerated in the green- and red-fruited lineages included *yellow-flesh*, a major locus controlling fruit color (35, 36). We also found many genes accelerated along the *S. pennellii* branch that are involved in responses to environmental stresses, such as salt, drought, heat, and oxidative damage, as well as genes in the abscisic acid pathway (Dataset S3). This finding is consistent with the results from differential expression and codon substitution models, and combined indicate that alteration in the pathways regulating stress responses has been important in the evolutionary history of this organism.

Evolution of the Tissue-Specific Expression in *S. pennellii* and *S. lycopersicum*. Natural variation has frequently been shown to involve tissue-specific gene expression alterations. We therefore examined whether gene-expression patterns might have been altered during domestication or in response to natural selection by contrasting *S. lycopersicum* var. M82 and the desert adapted *S. pennellii* (37–40). Gene-expression values between *S. lycopersicum* and *S. pennellii* were compared across a panel of six tissue types, including root, vegetative, and floral tissues.

We used principle components analysis (PCA) to identify major sources of variance in our transcriptome dataset (Fig. 4 *A* and *B*). Variation in expression across tissues explained the two largest principle components, but species-driven differences were also evident. Despite this substantial interspecific variation in gene expression, the pattern of gene expression across tissue types was positively correlated between species for the vast majority of genes (Fig. 4*B* and *SI Appendix*, Fig. S11). To examine the tissue specificity of expression differences between species, we applied PCA to between-species log fold-change values calculated for each tissue and found the majority (56%) of the variance was explained by global shifts in gene expression (Fig. 4 *C* and *D* and *SI Appendix*, Fig. S12). By fitting a statistical model accounting for species and tissue effects we identified 3,474 transcripts [false-discovery rate (FDR) < 0.01; 1,718 with log fold-change > 1] differentially expressed between species and 7,844 across tissues (FDR < 0.01) (Dataset S4). Only 166 transcripts were identified where the pattern of expression across tissues was significantly different between species, consistent with the general conservation in tissue-specific expression. We con-

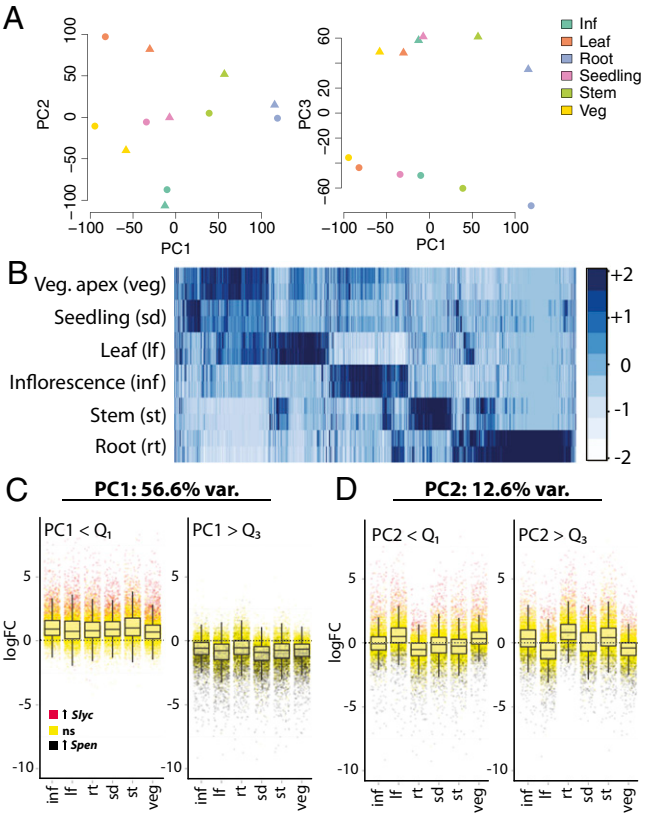

Fig. 4. Differential expression in *S. lycopersicum* and *S. pennellii*. (A) PCA factorial maps showing the largest components of variance, which separate samples by tissue (PC1 and PC2) and species (PC3). Triangles represent *S. pennellii* samples and circles *S. lycopersicum*. (*B*) Heat map comparing scaled expression values for *S. lycopersicum* (*Top*) and *S. pennellii* (*Bottom*) across tissues. Darker blue indicates higher expression. (*C*) PCA on log fold-changes by tissue. PC1 explains variance relating to global shifts in gene expression. (*D*) The remaining PCs describe tissue-specific shifts. Q3 and Q1 indicate the upper and lower quantile of the data, respectively (inf, inflorescence meristem; lf, leaf; rt, root; sd, aerial seedling; st, stem; veg, vegetative meristem).

firmed our relative expression estimates using quantitative RT-PCR and found strong correlation with our RNAseq data (ρ = 0.91) (*SI Appendix*, Fig. S13) validating our methodology.

Evolution of the Gene Coexpression Networks of *S. pennellii* and *S. lycopersicum*. To gain additional insight into the pattern of gene-expression changes between *S. pennellii* and *S. lycopersicum*, we built weighted gene coexpression networks for each species using genes significantly differentially expressed across tissues from our previous analysis (Fig. 5 and *SI Appendix*, Fig. S14). This approach allows us to compare the pattern of gene-expression correlations in both species, rather than the absolute level of gene expression, and has been shown to provide additional evolutionary insight (41). For both species, three major modules of highly coexpressed genes were identified (Fig. 5 *A* and *B*, *SI Appendix*, Fig. S14 and Table S12, and Dataset S5) (a fourth small module was also identified in *S. lycopersicum* but was not considered for the remainder of the analysis). The largest module (green; 852 genes found in both species) contained genes highly induced or repressed in photosynthetic tissues (leaf, vegetative shoot, and aerial seedling tissues) and was enriched for GO terms related to photosynthesis, carbon metabolism, and response to light (*SI Appendix*, Tables S13 and S14). A second module (purple; 272 genes found in both species) separated root tissues from all other tissues (*SI Appendix*, Tables S15 and S16, and Datasets S6–S9). The final large module (yellow; 144 genes in both species) differentiated vegetative and inflorescence shoot tissues from others and was enriched for GO terms related to cell division (*SI Appendix*, Tables S17 and S18). The overlap between these modules indicates extensive conservation of coexpression networks between the two species.

Although modules often overlapped between the two species, we noticed that characteristics of the two networks were not equivalent. In particular, the connectivity (as measured by the sum of the absolute values of the correlation coefficients of a focal gene with all other genes, see *SI Appendix, SI Materials and Methods*) of the *S. pennellii* network was on average higher than that of *S. lycopersicum* (Fig. 5*C* and *SI Appendix*, Fig. S14). This signal was primarily because of genes highly connected in both species but more highly connected in *S. pennellii*. Calculating connectivity for genes within each module gave similar values in both species for the purple and yellow modules, but connectivity was strongly reduced in the green module in *S. lycopersicum* (Wilcoxon test *P* value < 2e-16) (Fig. 5*D*). To further explore this finding, we identified species-specific connections (edges) between genes in each module and between genes not assigned to a module (Fig. 5*E*). If the networks had changed similarly since the two species diverged, one would expect equivalent numbers of gain/loss of edges in each network. In agreement, about the same number of species-specific edges were found between genes either not assigned to a module or in the yellow and purple modules. In contrast, a much higher number of *S. pennellii*-specific edges were identified in the green module. Taken together, these data demonstrate that photosynthetic tissue-specific gene expression is more tightly correlated in *S. pennellii* than in *S. lycopersicum*.

Effect of Introgression on the Transcriptome of Domesticated Tomato. An important strategy in tomato improvement is the extensive use of wild germplasm during breeding. Previous work suggested the possibility of large introgressions in the tomato reference sequence var. Heinz (29). Such introgressions combine previously independently evolving alleles that may result in novel changes in expression. We searched for evidence of introgressions and found that SNPs differentiating the Heinz and M82 cultivars were nonrandomly distributed (Fig. 6*E*). These regions of high diversity showed increased allele sharing with *S. pimpinellifolium*, indicating recent introgression from this or a closely related species. Using this pattern of diversity (*Materials and Methods*), we defined 550 candidate introgressed genes in Heinz and 2,479 in M82. The large number of candidate loci introgressed in M82

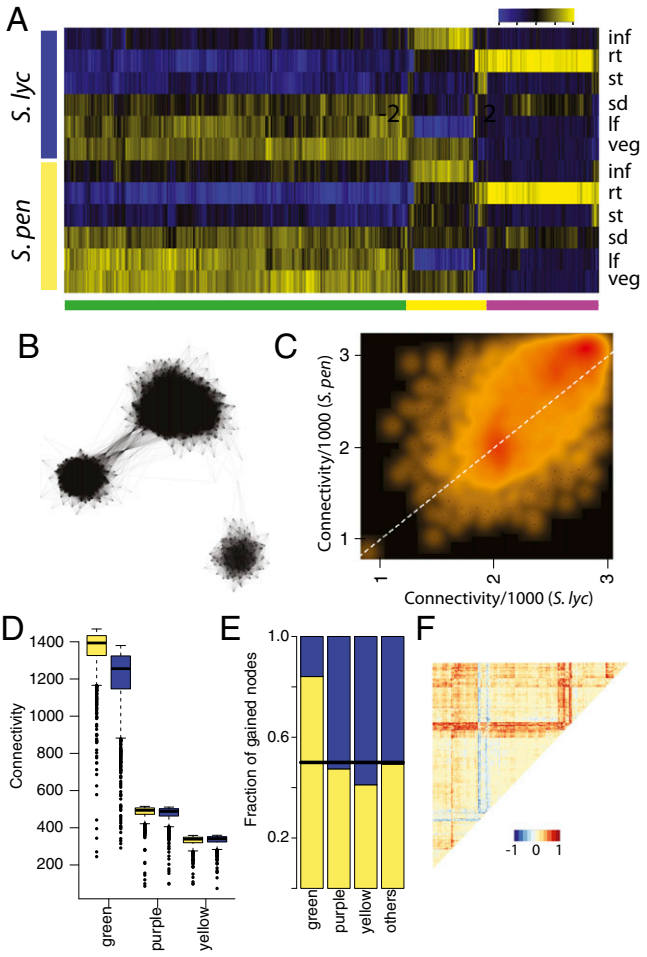

Fig. 5. Evolution of coexpression networks in *S. lycopersicum* and *S. pennellii*. (*A*) Heatmap depicting expression of genes assigned to the three modules in both species. Scaled $\log_2$ expression values are shown with yellow and blue indicating high and low expression respectively. Green, yellow and purple bars indicate membership in the three identified transcription modules. (*B*) Global depiction of conserved coexpression network components. Three clear clusters that correspond to the three major modules are evident. (*C*) Comparison of connectivity (sum of the absolute correlation of expression with all other genes) for genes in the two species. Black indicates a low density of points and red indicates a high density. Connectivity is positively correlated, but the highest values are increased in *S. pennellii*. White dashed line indicates a slope of 1. (*D*) Intramodule connectivity for each module in each species. Yellow boxes are *S. pennellii* values and blue *S. lycopersicum*. (*E*) Fraction of differential edges specific to *S. pennellii* (yellow) and *S. lycopersicum* (blue) for each module. (*F*) Heatmap showing the change in connectivity for all gene pairs in the green module. Red indicates correlations that are found only in *S. pennellii*, blue indicates correlations found only in *S. lycopersicum*, and yellow indicates correlations of approximately the same strength in both networks.

highlights the challenge of linkage drag during breeding using wild accessions, and may contribute to reduced genome-wide divergence in nucleotide sequence and divergence in gene expression between cultivated accessions and wild accessions.

Transgressive or nonparental expression phenotypes are a well-described characteristic of expression in hybrid lines (42), and thus introgression in tomato might result in new expression phenotypes. We examined whether introgressions of this size might contribute to expression divergence by comparing gene expression in *S. lycopersicum*, *S. pennellii*, and an introgression line (IL) where a portion of chromosome 4 from *S. pennellii* was introgressed into *S. lycopersicum* (Fig. 6 *A*–*D*)

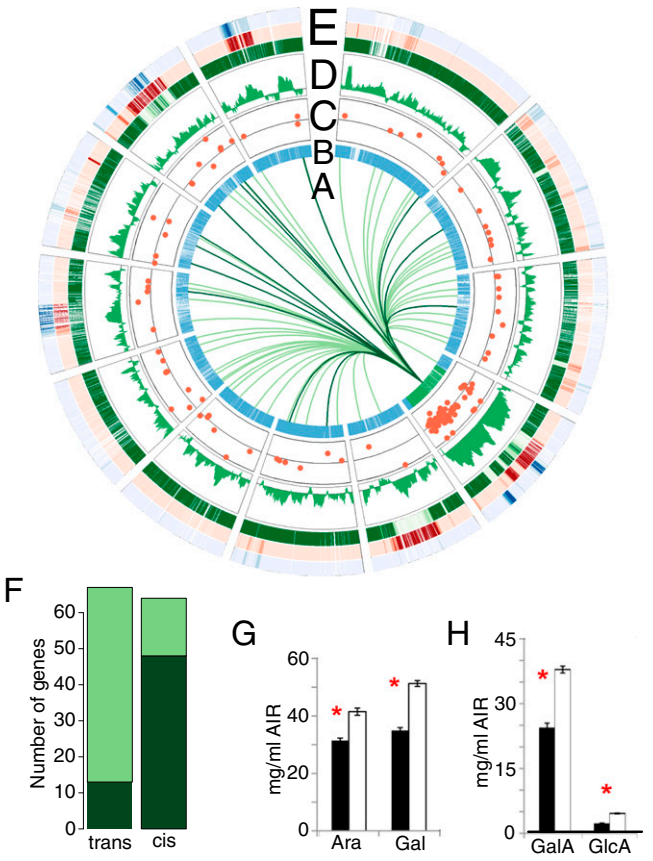

Fig. 6. *cis* and *trans* expression divergence. (*A*) Transgressive (light green) and *S. pennellii*-like (dark green) *trans* regulation relationships with genes on IL4-3. (*B*) Genotype by gene in IL4-3 (M82, blue and *S. pennellii*, green). (*C*) Log fold-change for differentially expressed genes between IL4-3 and M82. The black line indicates 0. (*D*) Correlation of log fold-change in expression between M82 and either the IL or PEN in sliding windows. A strong increase is seen on chromosome 4 indicating higher correspondence in gene expression between *S. pennellii* and the introgression line. (*E*) Heatmap showing median dissimilarity (polymorphisms per 1,000 bp) between M82 and Heinz (green), M82 and *S. pimpinellifolium* (red), and Heinz and *S. pimpinellifolium* (blue). Increasing dissimilarity is indicated by lighter color; increasing similarity is indicated by darker color. Introgressions into one of the two cultivated lines show increased polymorphism rate between Heinz and M82 and decreased polymorphism between *S. pimpinellifolium* and the acceptor cultivated line (see chromosome 5). Shared introgressed segments show decreased polymorphism between *S. pimpinellifolium* and both cultivated lines. Many of these introgressions are larger in M82 (see chromosome 11). Large introgressions were frequently associated with centromeres, likely resulting from the increased linkage drag in these regions during breeding. Sliding window size for *B*, *D*, and *E* is 100 genes. (*F*) Number of expression changes in *cis* or *trans* to the IL4-3 introgression. Light green indicated transgressive changes and dark green indicates *S. pennellii* like expression. (*G* and *H*) Abundance of pectic monosaccharides galacturonic acid (GalA), glucoronic acid (GlcA), arabinose (Ara), and galactose (Gal) present in the walls (AIR, alcohol insoluble residue) of roots from M82 (black) and *S. pennellii* (white). Red asterisk: $P < 0.001$.

(43). Of the 131 genes differentially expressed between the IL and M82, 61 (47%) exhibited nonparental expression levels and 70 showed expression similar to *S. pennellii*. Genes exhibiting *S. pennellii*-like expression were enriched in the introgressed fragment but the majority of genes showing nonparental expression patterns were found outside of the fragment (Fig. 6*F*). The enrichment for nonparental expression *in trans* to the introgression provides evidence of the existence of epistatically interacting mutations within each lineage that, when combined, result in unique expression phenotypes. An additional possible contributing factor is the recent discovery of transgressive siRNA expression patterns in tomato hybrids (44). These results point to introgression as a possible source of unique expression phenotypes in cultivated tomato.

Expression Divergence Correlates with Phenotypic Differences Among Wild and Cultivated Accessions. Our combined comparison of sequence and transcriptional diversity in cultivated and wild tomato relatives identified distinct footprints of selection under artificial and natural conditions. Adaptation to an extreme desert climate manifests as dramatic phenotypic shifts seen in *S. pennellii* compared with cultivated tomato. These phenotypes include up to 20-fold higher levels of epicuticular lipid deposition in *S. pennellii* leaves (45), amphistomic leaves with reduced stomatal pore size (~13% smaller, $P = 5.17 \times 10^{-6}$), and alterations in cell wall composition in the roots (*SI Appendix*, Fig. S15 and S16). Each of these phenotypes can be correlated with gene-expression profiles in our data.

A thick cuticle with an increased accumulation of epicuticular waxes is known to limit water loss and increase water-deficit tolerance (46–49). In the tomato relatives, epicuticular waxes account for up to 20% dry weight of *S. pennellii* leaves, whereas they make up only 0.9% of *S. lycopersicum* leaf dry weight (45). This striking increase in accumulation of epicuticular waxes is accompanied with marked differences in the expression of genes associated with wax deposition between *S. lycopersicum* and *S. pennellii* in our datasets (*SI Appendix*, Table S19). For example, the genes that encode the tomato orthologs of two enzymes involved in the production of aliphatic wax component precursors, *ECERIFERUM6* (CER6) and CER10 (50–52) are significantly higher in *S. pennellii* in comparison with *S. lycopersicum*. *FIDDLEHEAD*, which encodes a condensing enzyme involved in synthesis of cuticular lipids (53) and the genes encoding orthologs of CER1, CER2, and CER8, which are involved in conversion of very long-chain fatty acids to alkanes in *Arabidopsis* (54–57) are also higher in *S. pennellii.* Furthermore, *S. pennellii* has higher expression of a *CER5*-like gene, which might be involved in wax secretion (58) and the genes encoding the drought responsive nonspecific lipid transfer proteins (LTP) LTP1 and LTP2 (59, 60). Together, these results demonstrate a concerted up-regulation of candidate genes for wax accumulation in desert adapted *S. pennellii.*

S. pennellii leaves have several developmental features consistent with drought adaptation including reduced surface:volume ratio and changes in stomatal density (61). One developmental regulator that might be involved is *SCREAM1*, a positive regulator of stomatal index (the ratio of stomata to epidermal cells) (62, 63). *S. lycopersicum* shows almost twofold lower levels of *SCREAM1* ($P = 0.00018$). Consistent with these changes, *S. pennellii* has an increased adaxial stomatal index relative to *S. lycopersicum*, yielding a roughly even stomatal index on both leaf surfaces (*SI Appendix*, Fig. S16) (19, 64). Typical for a desert plant, *S. pennellii* has thick succulent leaves (1.46× the thickness of *S. lycopersicum*), (*SI Appendix*, Fig. S16), thus the relative increase in adaxial stomata may be required for efficient CO_2 diffusion in these thicker leaves (64, 65).

Previous studies have reported that root growth under drought conditions can be promoted by cell-wall modulation of glucoronoxylan and rhamnogalacturonan side chains in cell-wall components (66, 67). Correlating with these observations, genes expressed nearly exclusively in the root include many genes involved in cell-wall metabolism, such as multiple pectinesterases and polygalacturonases, several β-galactosidases, and a reversibly glycosylated protein involved in UDP-arabinofuranose production (68), the precursor for arabinan biosynthesis. To validate the relevance of the expression differences we examined root primary cell-wall composition in *S. lycopersicum* and *S. pennellii* and found that *S. pennellii* had higher levels of the abundant galactan and arabinan side-chains of rhamnogalacturonan I (Fig. 6 *G* and *H*) (69), consistent with its desert habitat.

Discussion

Tomato is one of our most important vegetable crops and its improvement is largely dependent on introgression of beneficial alleles from wild germplasm (15). Here we have identified hundreds of thousands of polymorphic positions that distinguish cultivated tomato from its wild relatives. All of these species have individual attributes that could be potentially valuable for tomato crop improvement, and our study provides the raw material necessary for marker-assisted introgression of such traits.

We have shown that domestication was associated with the fixation of many potentially deleterious protein and expression-level changes. The consequences of such changes are unknown, but it is possible that some have decreased vigor in domesticated lines. Adaptation to extreme environments among tomato relatives appears to have caused a broad alteration of transcriptional networks in parallel with positive selection at the sequence level for a number of genes related to environmental adaptation. This is particularly the case for the desert-adapted *S. pennellii*. Our finding that gene-expression changes in *S. pennellii* were highly accelerated relative to nucleotide divergence suggests that the previously noted importance of regulatory changes in morphological evolution (70, 71) is likely a genome-scale phenomenon. The signal of adaptation to extreme environments in the *S. pennellii* transcriptome is on par with that seen for biological processes classically thought to evolve at an accelerated rate, such as defense response and reproductive biology. Previous work in maize has suggested extensive transcriptional rewiring in response to domestication (14). The most extensive network rewiring that we discovered in *S. lycopersicum* relates to light responsiveness. Loss of connectivity in this network may reflect selection for reduced light response in *S. lycopersicum*, or may reflect a more robust response in the desert-adapted *S. pennellii*; this hypothesis is amenable to future genetic experimentation. In contrast to adaptation to pressures emanating from the natural environment, as deduced from differences between wild tomato species, we have found artificial selection and domestication to be associated with a relatively small number of changes at both at the sequence and transcriptional level. Taken together, our studies highlight both parallels and contrasts between natural and artificial selection and their effects on genome evolution.

Materials and Methods

Plant Materials. *S. lycopersicum* var. M82 (LA3475), *S. pennellii* (LA0716), and the *S. pennellii* introgression line IL4-3 (LA4051) were donated by Dani Zamir, The Hebrew University of Jerusalem, Jerusalem, Israel (43). *S. habrochaites* (LA1777) and *S. pimpinellifolium* (LA1589) were obtained from the C. M. Rick Tomato Genetics Resource Center, University of California at Davis. *S. chmielewskii* (LA1840) was donated by Keygene, Wageningen, The Netherlands. *S. galapagense* (LA0530) was donated by Maria Asins at the Instituto Valenciano de Investigaciones Agrarias, Valencia, Spain. The obligate outcrossing *S. habrochaites* line was maintained by growth of 10 or more plants and cross-pollinated by hand. All other accessions were maintained by selfing.

RNA Isolation. In the transcriptome experiment, total RNA from all tissues except fruits were extracted using TRIzol (Invitrogen) according to the manufacturer's standard protocol. The following modifications were included in the protocol to extract RNA from fruits: Total RNA from the aqueous phase in the chloroform extraction step was precipitated with 0.25 volume isopropanol and 0.25 volume of 1.2 M sodium chloride/0.8 M sodium citrate buffer, washed with 70% (vol/vol) ethanol, and resuspended in water. Another precipitation step with 0.8 volume lithium chloride and 3 volumes 100% (vol/vol) ethanol, was done if the 260/230 absorbance ratio of the total RNA was less than 1.5.

The RNeasy plant mini kit (Qiagen) was used to extract total RNA for the seedling experiment.

Sequencing and Read Filtering. A total of 57 libraries from *S. lycopersicum* var. M82, *S. pimpinellifolium*, *S. pennellii*, *S. habrochaites* were sequenced in 14 lanes from seven different 84 cycle runs of the Illumina GA II, returning 406,874,298 paired-end and 169,290,821 single-end reads. Additionally, single libraries from *S. galapagense* and *S. chmielewskii* were sequenced in a Hiseq2000 to obtain 67,504,782 and 53,873,978 100-bp paired end reads, respectively. After separating reads by barcode, removing Illumina adapter sequences, and trimming low-quality bases, we used in our analysis 547,612,718 reads with a minimum length of 50 bp (average of 81 bp).

RNAseq Read Alignment. Three different strategies were used for RNAseq read analysis. For polymorphism detection and total coverage calculations we aligned the reads against the *S. lycopersicum* genomic reference. For quantification of gene expression we created a matched set of contigs that were used as a reference. De novo assembly was performed on reads obtained from *S. lycopersicum* var. M82 and *S. pennellii* to identify unannotated transcripts and transcripts not synthesized by *S. lycopersicum* var. Heinz.

SNP Calling. A custom Bioperl script was used to detect SNPs and indels between each sequenced species and the reference sequence (72). Homozygous SNPs/indels were called in positions with a minimum coverage of four reads and an allele frequency higher than 0.66 for SNPs and 0.33 for indels. Heterozygous SNPs were called in positions with at least four reads per allele and a frequency of at least 25% in both alleles. To avoid calling polymorphisms from the ends of the reads that span exon-intron junctions, we divided the reads into five equal regions and discarded SNPs and indels covered only by a single region of the reads. All polymorphisms from all species were merged in a matrix and their positions genotyped in all species where the polymorphism was not present. These genotypes were called using the same allele frequency thresholds as above but no coverage threshold.

Statistical Analysis. All statistical analysis was done using the R statistical programming environment (73).

ACKNOWLEDGMENTS. We thank Stacey Harmer and Danelle Seymour for helpful comments on the analysis and manuscript; the Tomato Genetics Resource Center, Keygene, Dr. Maria Asins, and Dr. Dani Zamir for providing seed; the Tomato Sequencing Consortium for prepublication access to the tomato genome sequence; Niels Müller and Kerstin H Richau for technical assistance with *Solanum galapagense* and *Solanum chmielewskii* sequencing; Heike Keller (Max Planck Institute for Developmental Biology), Fernando Carrari, and Gabriel Lichtenstein (Instituto Nacional de Tecnología Agropecuaria, Argentina) for helping with the *Solanum pennellii* genome build; Brian Moore for access to the AutoParts software prior to publication; Liam Revell, Carl Boettiger, and Luke Harmon for discussions on the phylogenetic modeling of gene expression; Sergei Kosakovsky Pond for assistance with HyPhy analyses; and the iPlant Collaborative (http://www.iplantcollaborative.org) for providing computational resources and data storage. RNA-seq reads and SNPs from this study can be viewed in a genome browser at http://phytonetworks.ucdavis.edu/tomato and have been submitted to the sol genomics network (http://solgenomics.net) for distribution. Promoter sequences, de novo contigs with no match to the tomato 2.40 references chromosomes, and the matched *S. lyc*/*S. pen.* CDS set used in this study can be downloaded from http://phytonetworks.ucdavis.edu/Download. This work was funded through a National Science Foundation Grant (IOS-0820854) to N.R.S., J.N.M., and J.P., and a Human Frontier Science Program Fellowship (LT000783) to D.K.

1. Darwin C (1868) *The Variation of Animals and Plants Under Domestication* (J. Murray, London).
2. Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127(7):1309–1321.
3. Parker HG, Shearin AL, Ostrander EA (2010) Man's best friend becomes biology's best in show: genome analyses in the domestic dog. *Annu Rev Genet* 44:309–336.
4. Izawa T, Konishi S, Shomura A, Yano M (2009) DNA changes tell us about rice domestication. *Curr Opin Plant Biol* 12(2):185–192.
5. Paran I, van der Knaap E (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J Exp Bot* 58(14):3841–3852.
6. Gross BL, Olsen KM (2010) Genetic perspectives on crop domestication. *Trends Plant Sci* 15(9):529–537.
7. Wright SI, et al. (2005) The effects of artificial selection on the maize genome. *Science* 308(5726):1310–1314.
8. Hufford MB, et al. (2012) Comparative population genomics of maize domestication and improvement. *Nat Genet* 44(7):808–811.
9. Huang X, et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490(7421):497–501.
10. Gepts P (2004) Crop domestication as a long-term selection experiment. *Plant Breed Rev* 24(2):1–44.

11. Charlesworth B, Charlesworth D (2010) *Elements of Evolutionary Genetics* (Roberts and Co, Greenwood Village, CO), pp xxvii, 734 pp.
12. Lu J, et al. (2006) The accumulation of deleterious mutations in rice genomes: A hypothesis on the cost of domestication. *Trends Genet* 22(3):126–131.
13. Cruz F, Vilà C, Webster MT (2008) The legacy of domestication: Accumulation of deleterious mutations in the dog genome. *Mol Biol Evol* 25(11):2331–2336.
14. Swanson-Wagner R, et al. (2012) Reshaping of the maize transcriptome by domestication. *Proc Natl Acad Sci USA* 109(29):11878–11883.
15. Atherton JG, Rudich J (1986) *The Tomato Crop: A Scientific Basis for Improvement* (Chapman and Hall, London, New York), pp xv, 661 pp.
16. Grandillo S, Tanksley SD (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theor Appl Genet* 92(8):935–951.
17. Mattoo A, Razdan MK, eds (2007) *Genetic Improvement of Solanaceous Crops Volume 2. Tomato.* (Science, Enfield, NH), pp xx, 637.
18. Xu X, et al.; Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355):189–195.
19. Spooner DM, Peralta IE, Knapp S (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes *Solanum* L. section Lycopersicon (Mill.) Wettst. *Taxon* 54(1):43–61.
20. Darwin SC, Knapp S, Peralta IE (2003) Taxonomy of tomatoes in the Galápagos Islands: Native and introduced species of *Solanum* section Lycopersicon (Solanaceae). *Syst Biodivers* 1(1):29–53.
21. Nuez F, Prohens J, Blanca JM (2004) Relationships, origin, and diversity of Galapagos tomatoes: Implications for the conservation of natural populations. *Am J Bot* 91(1): 86–99.
22. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11(5):725–736.
23. Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11(5):715–724.
24. Zhang X, et al. (2011) *Arabidopsis* Argonaute 2 regulates innate immunity via miRNA393(*)-mediated silencing of a Golgi-localized SNARE gene, MEMB12. *Mol Cell* 42(3):356–366.
25. Ori N, et al. (1997) The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. *Plant Cell* 9(4):521–532.
26. Larsen PB, Cancel J, Rounds M, Ochoa V (2007) *Arabidopsis* ALS1 encodes a root tip and stele localized half type ABC transporter required for root growth in an aluminum toxic environment. *Planta* 225(6):1447–1458.
27. Yamanaka T, et al. (2010) MCA1 and MCA2 that mediate Ca2+ uptake have distinct and overlapping roles in *Arabidopsis*. *Plant Physiol* 152(3):1284–1296.
28. Frary A, et al. (2010) Salt tolerance in *Solanum pennellii*: Antioxidant response and related QTL. *BMC Plant Biol* 10:58.
29. Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400):635–641.
30. Schauer N, Zamir D, Fernie AR (2005) Metabolic profiling of leaves and fruit of wild species tomato: A survey of the *Solanum lycopersicum* complex. *J Exp Bot* 56(410): 297–307.
31. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* 1(1):387–396.
32. Bedford T, Hartl DL (2009) Optimization of gene expression by natural selection. *Proc Natl Acad Sci USA* 106(4):1133–1138.
33. Hansen TF, Martins EP (1996) Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data. *Evolution* 50(4):1404–1417.
34. O'Meara BC, Ané C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60(5):922–933.
35. Fray RG, Grierson D (1993) Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant Mol Biol* 22(4):589–602.
36. Ronen G, Cohen M, Zamir D, Hirschberg J (1999) Regulation of carotenoid biosynthesis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. *Plant J* 17(4):341–351.
37. Frary A, et al. (2000) fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. *Science* 289(5476):85–88.
38. Chen KY, Cong B, Wing R, Vrebalov J, Tanksley SD (2007) Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. *Science* 318(5850): 643–645.
39. Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 305(5691):1786–1789.
40. Martin B, Thorstenson YR (1988) Stable carbon isotope composition (deltaC), water use efficiency, and biomass productivity of *Lycopersicon esculentum*, *Lycopersicon pennellii*, and the F(1) hybrid. *Plant Physiol* 88(1):213–217.
41. Oldham MC, Horvath S, Geschwind DH (2006) Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci USA* 103 (47):17973–17978.
42. Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci USA* 102(5):1572–1577.
43. Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141(3):1147–1162.
44. Shivaprasad PV, Dunn RM, Santos BA, Bassett A, Baulcombe DC (2012) Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J* 31(2):257–266.
45. Fobes JF, Mudd JB, Marsden MP (1985) Epicuticular lipid accumulation on the leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. *Plant Physiol* 77(3):567–570.
46. Kosma DK, et al. (2009) The impact of water deficiency on leaf cuticle lipids of *Arabidopsis*. *Plant Physiol* 151(4):1918–1929.
47. Zhang JY, et al. (2005) Overexpression of WXP1, a putative *Medicago truncatula* AP2 domain-containing transcription factor gene, increases cuticular wax accumulation and enhances drought tolerance in transgenic alfalfa (*Medicago sativa*). *Plant J* 42(5): 689–707.
48. Zhang JY, Broeckling CD, Sumner LW, Wang ZY (2007) Heterologous expression of two *Medicago truncatula* putative ERF transcription factor genes, WXP1 and WXP2, in *Arabidopsis* led to increased leaf wax accumulation and improved drought tolerance, but differential response in freezing tolerance. *Plant Mol Biol* 64(3):265–278.
49. Aharoni A, et al. (2004) The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. *Plant Cell* 16(9):2463–2480.
50. Millar AA, et al. (1999) CUT1, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *Plant Cell* 11(5):825–838.
51. Fiebig A, et al. (2000) Alterations in CER6, a gene identical to CUT1, differentially affect long-chain lipid content on the surface of pollen and stems. *Plant Cell* 12(10): 2001–2008.
52. Zheng H, Rowland O, Kunst L (2005) Disruptions of the *Arabidopsis* Enoyl-CoA reductase gene reveal an essential role for very-long-chain fatty acid synthesis in cell expansion during plant morphogenesis. *Plant Cell* 17(5):1467–1481.
53. Yephremov A, et al. (1999) Characterization of the FIDDLEHEAD gene of *Arabidopsis* reveals a link between adhesion response and cell differentiation in the epidermis. *Plant Cell* 11(11):2187–2201.
54. Aarts MG, Keijzer CJ, Stiekema WJ, Pereira A (1995) Molecular characterization of the CER1 gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* 7(12):2115–2127.
55. Negruk V, Yang P, Subramanian M, McNevin JP, Lemieux B (1996) Molecular cloning and characterization of the CER2 gene of *Arabidopsis* thaliana. *Plant J* 9(2):137–145.
56. Xia Y, Nikolau BJ, Schnable PS (1996) Cloning and characterization of CER2, an *Arabidopsis* gene that affects cuticular wax accumulation. *Plant Cell* 8(8):1291–1304.
57. Lü S, et al. (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(4):553–564.
58. Pighin JA, et al. (2004) Plant cuticular lipid export requires an ABC transporter. *Science* 306(5696):702–704.
59. Thoma S, et al. (1994) Tissue-specific expression of a gene encoding a cell wall-localized lipid transfer protein from *Arabidopsis*. *Plant Physiol* 105(1):35–45.
60. Trevino MB, OConnell MA (1998) Three drought-responsive members of the nonspecific lipid-transfer protein gene family in *Lycopersicon pennellii* show different developmental patterns of expression. *Plant Physiol* 116(4):1461–1468.
61. Kebede H, Martin B, Nienhuis J, King G (1994) Leaf anatomy of two *Lycopersicon* species with contrasting gas exchange properties. *Crop Sci* 34(1):108–113.
62. Kanaoka MM, et al. (2008) SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. *Plant Cell* 20(7): 1775–1785.
63. Abrash EB, Bergmann DC (2010) Regional specification of stomatal production by the putative ligand CHALLAH. *Development* 137(3):447–455.
64. Nakazato T, Warren DL, Moyle LC (2010) Ecological and geographic modes of species divergence in wild tomatoes. *Am J Bot* 97(4):680–693.
65. Mott KA, Gibson AC, O'Leary JW (1982) The adaptive significance of amphistomatic leaves. *Plant Cell Environ* 5(6):455–460.
66. Leucci MR, Lenucci MS, Piro G, Dalessandro G (2008) Water stress and cell wall polysaccharides in the apical root zone of wheat cultivars varying in drought tolerance. *J Plant Physiol* 165(11):1168–1180.
67. Keppler BD, Showalter AM (2010) IRX14 and IRX14-LIKE, two glycosyl transferases involved in glucuronoxylan biosynthesis and drought tolerance in *Arabidopsis*. *Mol Plant* 3(5):834–841.
68. Konishi T, et al. (2007) A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. *Glycobiology* 17(3):345–354.
69. Mohnen D (2008) Pectin structure and biosynthesis. *Curr Opin Plant Biol* 11(3): 266–277.
70. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* 134(1):25–36.
71. Zhao Q, et al. (2008) The role of regulatory genes during maize domestication: Evidence from nucleotide polymorphism and gene expression. *Genetics* 178(4): 2133–2143.
72. Stajich JE, et al. (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* 12(10):1611–1618.
73. R Development Core Team (2011) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
74. Heineke D, et al. (1991) Redox transfer across the inner chloroplast envelope membrane. *Plant Physiol* 95(4):1131–1137.
75. Sies H (1982) Nicotinamide nucleotide compartmentation. *Metabolic Compartmentation*, ed Sies H (Academic, London), pp 235–257.