Evolution of two gene networks underlying adaptation to drought stress in the wild tomato *Solanum chilense*

Kai Wei*[1], Saida Sharifova*[2], Neelima Sinha[3], Hokuto Nakayama[4], Aurélien Tellier[#1], Gustavo A Silva-Arias[#1,5]

[1]Professorship for Population Genetics, Department of Life Science Systems, School of Life Sciences, Technical University of Munich, Liesel-Beckmann Strasse 2, 85354 Freising, Germany

[2]Department of Life Sciences, Graduate School of Science, Art and Technology, Khazar University, Mahsati 41, AZ1096, Baku, Azerbaijan

[3]Department of Plant Biology, University of California Davis, One Shields Avenue, Davis, CA 95616, USA

[4]Department of Biological Sciences, The University of Tokyo, 113-0033, Tokyo, Japan

[5]Instituto de Ciencias Naturales, Facultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Bogotá, Colombia

* these authors contributed equally

[#]Correspondence email: aurelien.tellier@tum.de, gasilvaa@unal.edu.co

Abstract

Drought stress is a key factor limiting plant growth and the colonization of arid habitats by plants. Here, we study the evolution of gene expression response to drought stress in a wild tomato, *Solanum chilense* naturally occurring around the Atacama Desert in South America. We conduct a transcriptome analysis of plants under standard and drought experimental conditions to understand the evolution of drought-response gene networks. We identify two main regulatory networks corresponding to two typical drought-responsive strategies: cell cycle and fundamental metabolic processes. We estimate the age of the genes in these networks and the age of the gene expression network, revealing that the metabolic network has a younger origin and more variable transcriptome than the cell-cycle network. Combining with analyses of population genetics, we found that a higher proportion of the metabolic network genes show signatures of recent positive selection underlying recent adaptation within *S. chilense,* while the cell-cycle network appears of ancient origin and is more conserved. For both networks, however, we find that genes showing older age of selective sweeps are the more connected in the network. Adaptation to southern arid habitats over the last 50,000 years occurred in *S. chilense* by adaptive changes core genes with substantial network rewiring and subsequently by smaller changes at peripheral genes.

Introduction

Drought stress is one of the major environmental constraints negatively influencing plant development and preventing plant growth, resulting in decreased yield in agriculture and as a constraining factor for colonization of arid or hyper-arid habitats (Ciais et al. 2005; Juenger 2013). Plants respond to water-insufficiency through multiple strategies underpinned by various physiological and developmental processes, such as storage of internal water to avoid tissue damage and tolerance (endurance) to drought stress to maintain the growth process (Basu et al. 2016). These strategies involve many biological functions such as increasing the metabolic activity of some tissues, i.e. root water uptake and closing stomata, or activation of metabolic pathways including phytohormone signaling, antioxidant and metabolite production in order to regulate osmotic processes (Rodrigues et al. 2019). Drought response involves numerous quantitative and polygenic traits governed by many genes acting in (complex) gene co-expression networks (GCN). To improve crops and predict the evolutionary potential of plant species under the current and predicted global water deficits, it is thus of interest to pinpoint and decipher the evolutionary history of the relevant GCNs underpinning the adaptation of wild plants to arid or hyper-arid habitats (Gehan et al. 2015).

Comparative transcriptomics involving the inference of gene co-expression patterns show that many GCNs are conserved along the tree of life (Stuart et al. 2003; Gerstein et al. 2014; Zarrineh et al. 2014; Crow et al. 2022). Moreover, phylogenetic and developmental studies have demonstrated that many physiological, structural, and regulatory innovations to cope drought stress have arisen throughout the history of plants, many of them even predating the emergence of land plants (Jill Harrison 2017; de Vries et al. 2018; de Vries and Archibald 2018; Mustafin et al. 2019; Wang et al. 2020; Bowles et al. 2021). Several conserved GCNs can be observed in fundamental biological processes such as protein metabolism, cell cycle, and photosynthesis and well as key traits such wood formation (Stuart et al. 2003; Ficklin and Feltus 2011; Zinkgraf et al. 2020).

A key question in functional and evolutionary genomics is thus to link GCN evolution and (relatively) short-scale evolutionary processes such as adaptation and population/species divergence in order to assess the relative importance of contingency, exaptation and evolution of novel genes (duplication, neofunctionalization) allowing colonization of novel habitats. Two main hypotheses are formulated. First, highly conserved sub-networks (so-called hubs or kernels) are under strong selective constraint to ensure the functionality of the GCNs (Papakostas et al. 2014; Josephs et al. 2017; Mähler et al. 2017; Masalia et al. 2017), so that variation can only be maintained at (less connected) genes at the periphery of the GCNs that may be the target of positive natural selection (Flowers et al. 2007; Kim et al. 2007; Luisi et al. 2015; Erwin 2020). However, this argument is likely based on the fact that the novel habitats may not differ much from the original one, so that only minor adjustments in the GCNs are enough to provide adaptation. This argument is in line with so-called developmental systems drift (DSD; True and Haag 2001), that predicts GCN rewiring only occurs in ‘flexible’ (sub-)modules with the accumulation of neutral variation that keep the network function intact until a new viable function (phenotype or developmental pathway) appears. Second, despite the general belief that genes with higher connectivity evolve at a slower rate, there is also evidence that changes at central genes (with higher connectivity) can be responsible for the short-term response to selection (Jovelin and Phillips 2009; Luisi et al. 2015) and promote rewiring of the GCN (Koubkova-Yu et al. 2018). Thus, highly connected genes may be targets of positive selection during environmental change, e.g. adaptation to novel arid habitats, even though these genes experience purifying selection in stable environments (Hämälä et al. 2020). Indeed, if the second hypothesis is correct, we expect a correlation between the age of positive selection and the connectivity of a gene in a network, but no correlation under the first hypothesis.

To test this hypothesis, we reveal the selective forces (positive versus purifying selection) acting on different components of the networks (hub vs peripheral genes) across species/lineages adapted to contrasting conditions, and correlate the presence of recent

positive selection with gene connectivity in the wild tomato species *Solanum chilense*. Wild tomatoes are a model of interest as their diversification is accompanied by the exploration of wide environmental gradients along the Pacific coast of South America (from tropical to subtropical, coastal to high mountain, and wet to extremely dry regions; Nakazato et al. 2010; Haak et al. 2014). In addition, the infra-specific diversification within *S. chilense* resulted in several lineages with strong environmental differentiation (Raduski and Igić 2021; Wei et al. 2022). Populations of *S. chilense* are challenged by prolonged drought, with the most severe drought conditions occurring in the southern part of the range (Figure 1A). Wild relative tomato species such *S. chilense*, *S. sitiens* and *S. pennellii* become well-established systems to study tolerance strategies to survive in extreme environments (Bolger et al. 2014; Martínez et al. 2014; Tapia et al. 2016; Kashyap et al. 2020; Molitor et al. 2021). In a previous study, we assayed for evidence of positive selection in 30 fully sequenced genomes of *S. chilense* to identify candidate genes underpinning adaptation along the species range. We found genes with putative functions related to root hair development and cell homeostasis as being likely involved in drought stress tolerance (Wei et al. 2022). However, to date, most research in *S. chilense* has focused on the evolution of a few genes potentially involved in abiotic stress response (Fischer et al. 2011; Mboup et al. 2012; Fischer et al. 2013; Böndel et al. 2015; Nosenko et al. 2016; Böndel et al. 2018), and we still lack information regarding the evolutionary mechanisms driving drought tolerance in this species.

Our aim is to study the GCN evolution underpinning *S. chilense* adaptation to arid and hyper-arid habitats. We identify drought stress responsive gene regulatory networks combining multiple analyses of transcriptome data of *S. chilense* and focus on two networks involved in cell-cycle and metabolic processes. Furthermore, we infer the evolutionary processes at these two networks across three different time scales by computation of transcriptome indices to explore the evolutionary age and sequence divergence of the drought responsive transcriptome. We then analyze the emergence of adaptive variation in

the identified drought-responsive genes of these networks and the association to gene connectivity.

Results

Transcriptome analyses

We analyze short-read transcriptome data from 16 libraries aligned to the reference genome of *S. chilense* (Dataset1 S1) and assess the consistency of the results by mapping to the reference genome of *S. lycopersicum* (ITAG 3.0) (Dataset2, S1). A total of 27,832 genes are identified to be expressed in the 16 libraries (Dataset1 S2), of which 1,536 genes are uniquely expressed in drought condition and 1,767 genes in control condition (Dataset1 S2). Using the ITAG 3.0 reference, we identify a total of 36,827 transcript isoforms corresponding to 15,697 genes (Dataset2 S2).

A principal component analysis (PCA) based on the gene expression profiles reveals consistent clustering primarily associated with the experimental conditions (control and drought) and secondarily to the developmental stages (leaf and shoot apex) (Figure 1A). PC1 accounts for 79% of the expression variability and separates the libraries from the two experimental conditions, indicating transcriptome remodeling between drought and control conditions. Libraries from different developmental stages are separated along the PC2 axis (accounting for 15% of the variance), supporting tissue age transcriptome specificity. Consistently, the transcriptome similarity analysis between libraries reveals that the watering conditions explain the major differences between treatments (Figure 1B). Therefore, we thereafter focus on comparing the transcriptome profiles of the drought and control experimental conditions.

The gene expression profiles based on reference genome of ITAG 3.0 show similar patterns on the PCA (Figure 1A and S1A). However, the analysis of expression correlation and library clustering based on the TPM values show reduced resolution in discriminating between experimental conditions and developmental stages using the *S. lycopersicum* reference (Figure S1B-D).

Identification of gene networks involved in drought stress

We identified gene networks involved in drought response in *S. chilense* based on differential expression analysis and weighted gene co-expression network analysis (WGCNA). First, three sets of DEGs are identified from three drought/control comparison groups (full data set, only leaf and only shoot apex tissues) (Figure 2A; Dataset1 S3; log2FoldChange ≥ 1, FDR P ≤ 0.001). A total of 4,905 DEGs are identified in three comparison groups, of which 2,484 DEGs (1,235 up-regulated and 1,249 down-regulated in drought transcriptome) are shared in three comparison groups (Figure 2B). We deduce that these shared DEGs correspond to a core functionally drought-responsive network. The consistency of these results is confirmed using gene expression data based on ITAG 3.0 reference genome. Similar DEGs can be identified between two reference genomes (Figure S2; Figure 2; Dataset2 S3). Although there is a notorious outlier sample (CL-A) when using reference genome of ITAG 3.0, in two shared DEG sets (2,484 DEGs based on *S. chilense* (Figure 2B), 2,585 DEGs based on ITAG 3.0 (Figure S2B)), 62.8% (1,560) DEGs still overlap (Figure S2C). This suggest that our greenhouse experiments, sequencing and parameters for differential expression analysis are highly robust. We use, thereafter, 2,484 shared DEGs based on the reference genome of *S. chilense* in the following analyses.

A set of 16,181 genes after filtering from all expressed genes were used in WGCNA, and clustered into seven co-expression modules named after different colors. The module sizes range from 183 up to 5,364 genes (Figure 2C, Dataset1 S4). Here, we do not directly used DEGs in WGCNA as suggested by the developer of WGCNA, because DEGs are invalid for assumption of the scale-free topology. Among the identified co-expression modules, the blue module (3,852 genes) shows significantly positive correlation with control condition and negative correlation with drought condition (Figure 2C, Kendall's test, P = 2.2e-11). In contrast, the turquoise module (5,364 genes) is significantly positively correlated with

drought condition and negatively correlated with control condition (Figure 2C, Kendall's test, P = 2.34e-13). In addition, the genes within blue and turquoise modules are observed to show higher connectivity than other modules (Figure S3, Kolmogorov-Smirnov test on connectivity measure, P = 2.41e-17), indicating higher interaction and closer correspondence in biological process among genes within each module in response to water deprivation.

We check the overlap between DEGs and modules to confirm that blue and turquoise modules are associated with drought stress in *S. chilense* (Table S1). DEGs share more genes with the blue and turquoise modules than with other co-expression modules. Almost all shared DEGs (2,302 genes out of 2,484) are found in the blue and turquoise modules. This confirms that blue and turquoise modules are two sets of co-expressed drought stress responsive genes. The overlapping DEGs and module genes are extracted to constitute the now refined two high-confidence subsets of the blue and turquoise modules and comprising 1,223 and 1,079 genes, respectively. The co-expression analysis with the ITAG 3.0 reference genome show consistent results and obtain eight co-expression modules (Figure S2D; Dataset2 S4), with two significantly correlated modules with drought stress (blue module shows negative correlation, and turquoise module shows positive correlation; Figure S2D), and show high overlapped ratio with DEGs based on ITAG 3.0 (Table S2).

To confirm regulatory relationships among genes in the two co-expression networks, we identify transcription factors (TFs) and transcription factor binding sites (TFBSs) for the two subsets of genes. Then, we extract the genes that can bind to one another (Table S3) from the two high-confidence subsets, which we hereafter name as sub-blue (686 genes) and sub-turquoise (948 genes), respectively (Dataset1 S5). The genes in the sub-blue and sub-turquoise networks not only show differential expression and specific co-expression patterns at the gene expression level, but we can also confirm that these interact as predicted by the DNA sequence and protein level. Subsequently, the co-expression network is reconstructed using the same steps for the set of genes of the sub-blue and sub-turquoise networks. Higher connectivity is observed in the sub-turquoise network (Figure S4,

Kolmogorov-Smirnov test on connectivity measure, P = 0.002), suggesting a closer regulatory relationship among genes in the sub-turquoise than in the sub-blue network. This may be due to more composite TF/TFBS relationships and functions in the sub-turquoise network (Table S3; Table S4). In a complementary analysis, we also obtain networks based on reference genome of ITAG 3.0 (Dataset2 S5) and check shared genes between final networks obtained based two different reference genomes, and the results show that more than half of the genes are overlapping in two networks, respectively (Figure S5). The two networks (Dataset1 S5) based on the reference genome of *S. chilense* are used for subsequent analysis.

Functional enrichment analysis of drought-responsive GCNs

We assess whether the two identified gene networks (sub-blue and sub-turquoise) show functional differences. The gene ontology (GO) enrichment reveals that sub-blue network is significantly enriched ($P < 0.05$) in cell cycle and regulation biological processes, including replication and modification of genetic information, ribosome production and assembly, cytoskeleton organization, among others (Figure 3A; Table S4). Conversely, the sub-turquoise network is enriched in biological processes related to response of physiological and metabolic processes to water shortage and heat, including some metabolic processes, signal pathways, changes of stomata and cuticle, amongst other processes (Figure 3A; Table S4). These differences of function thus suggest that genes in the two sub-networks are activated and expressed in different cellular compartments. Consistent with the biological process above mentioned, the sub-blue network genes are mainly enriched in cellular components in the nucleus, including nucleolus, chromosome, nuclear envelope, and ribosome (Figure 3B; Table S5). These cellular components are at the center of cell division processes. On the other hand, the sub-turquoise network is enriched in cellular components related to metabolism processes, such as complexes and membrane structures in the cell (Figure 3B; Table S5). We also check GO terms enriched in the two gene networks based on reference genome of ITAG 3.0. Although ITAG 3.0 does not share all drought-responsive

genes with gene networks based on reference genome of *S. chilense*, almost all of GO terms are consistent between networks based on two different reference genomes (Table S4; Dataset2 S6). This indicates that modulation in the cell cycle and fundamental metabolism are two main strategies in response to drought stress in *S. chilense*. We focus, thereafter, on these two sub-networks and from now on, the sub-blue network is referred to as the *cell-cycle network* and the sub-turquoise as the *metabolic network*.

Evolutionary age of drought-responsive transcriptome in *S. chilense*

To generate a comprehensive understanding of the emergence of the identified drought-responsive GCNs, we estimate the transcriptome ages of the identified cell cycle and fundamental metabolism networks. For that, we build phylostratigraphic profiles for all genes of the two GCNs, summarizing the gene emergence in 18 stages of plant evolution or phylostrata (PS): PS1 representing the emergence of oldest genes (at the time of the first cellular organisms) to PS18 for the most recent genes present only in *S. chilense*. The PS18 shares no homologue genes with any other species in the nr (non-redundant protein) databases of NCBI (Figure 4A and 4B, Dataset1 S6). Most genes in the two analyzed GCNs (76.79% in metabolic network and 65.45% in cell-cycle network) are assigned to three main PS, Cellular organisms (PS1), Land plants (Embryophyta; PS5) and Flowering plants (Magnoliopsida; PS8) (Figure 4A). This suggests that the two drought-responsive GCNs we identify have an ancient origin and the components are fairly conserved across the tree of life/plants. Therefore, many drought-responsive pathways likely emerged during the colonization of land by plants (PS5), but many others could derive from exaptation processes from GCNs involved in the core cell process (PS1) or reproductive organ differentiation of flowering plants (PS8). Interestingly, the cell-cycle network shows older origin ages (with more genes arising within the PS1-3), while the metabolic network presents a larger proportion of genes originating in PS8 (Figure 4A and 4B). Under drought conditions, we also find that cell-cycle network genes of almost all PS ages are down-regulated, while genes of the metabolic network are up-regulated (Figure S6).

Furthermore, we estimate the age of cell-cycle and metabolic GCNs using the transcriptome age index (TAI). The mean evolutionary ages of the transcriptomes are significantly different between drought and control conditions (Figure 4C; Kolmogorov-Smirnov test, P = 0.03). The TAI profile would be expected to be a flat horizontal line if the ages of genes are the same across the transcriptomes. In addition, a higher TAI value implies that evolutionary younger genes are preferentially expressed at the corresponding condition/developmental stage. We observe higher TAI in drought samples, supporting that the drought-responsive genes exhibit a younger transcriptome age than genes expressed under control conditions. Moreover, TAI of the metabolic GCN is significantly higher than the cell-cycle (Figure 4C; Kolmogorov-Smirnov test, P = 12.51e-7), supporting the previous result that transcriptome ages of the genes in the cell-cycle are older than in the metabolic GCNs.

The contributions of the different PS to the TAI profiles also show notable patterns between the cell-cycle and metabolic GCNs (Figure 4D and 4E). On the one hand, early divergent genes (PS1 to PS7) show more constant transcriptome age in all conditions and the genes with ages in PS1, PS5 and PS8 appeared as remarkably important in two GCNs. On the other hand, late-emerging genes (PS8 to PS18) contribute increasingly with their age to the differential expression patterns between control and drought samples, indicating that younger drought-responsive genes are differentially expressed under drought stress in both GCNs (as observed in Domazet-Lošo and Tautz 2010; Piasecka et al. 2013). Remarkably, the youngest genes in PS18, *i.e*. specific to *S. chilense*, also present a higher contribution in the metabolic GCN, suggesting that these genes are involved in either speciation or local adaptation of *S. chilense* to drought conditions. Note that younger genes (PS9 to PS18) in the cell-cycle GCN hardly contribute to the TAI profile (Figure 4D and 4E).

Divergence of the drought tolerance transcriptome in *S. chilense*

To drill down into the evaluation of the drought-response mechanisms over the time scale of *S. chilense* divergence, we calculate the TDI index, which represents the mean sequence

divergence of a transcriptome. A total of 10 divergence strata (DS) are constructed based on the sequence divergence between genes of *S. chilense* and the close relative *S. pennellii* by computing the Ka/Ks ratio (Figure 5A; Figure S7; Dataset1 S6). The distributions of the Ka/Ks ratio per gene for both GCNs indicate the action of purifying selection, which confirms the conservation of most of drought-responsive genes at the close-related species level. Consistent with the phylostratigraphic patterns, the purifying selection signals in the cell-cycle GCN (Ka/Ks = 0.279 ± 0.333) are stronger than in the metabolic GCN (Ka/Ks = 0.329 ± 0.331) (Figure 5A; Table S6). In addition, higher TDI values are observed in the drought samples (Figure 5B) suggesting that the expressed genes we identify in the two GCNs exhibit a more conserved transcriptome profile in control condition compared to drought conditions (Kolmogorov-Smirnov test, P = 0.04). This result supports that different selective pressures act on *S. chilense* GCNs across conditions. In accordance with the TAI results, the transcriptome of the metabolic GCN appears to exhibit a higher transcriptome divergence than the cell-cycle GCN (Figure 5B; Kolmogorov-Smirnov test, P = 2.25e-7). Moreover, the low TDI in the cell-cycle GCN and larger TDI differences between drought and control transcriptomes also suggest that regulation of the cell-cycle is likely an ancestral (older) strategy of stress response. Alternatively, it is also possible that the transcriptome of the cell-cycle GCN may have been evolving and changing in older times, and reached a conserved structure in recent times. Conversely, changes of metabolic pathways and rewiring of the metabolic GCN may appear to be more pronounced and/or common in recent times.

The contributions of the low divergence DS classes (low Ka/Ks in DS1 to DS5) in the cell-cycle GCN (~ 50% of the genes) are larger than in the metabolic GCN (DS1 to DS5 about 30%), especially in DS1 (lowest Ka/Ks ratio; Figure 5C and 5D). This indicates that purifying selection pressure is acting on genes of the cell-cycle GCN, possibly constraining further changes. In contrast, the metabolic network genes show about 70% contributions in high DS (higher Ka/Ks ratio in DS6 to DS10), especially in DS10 (highest Ka/Ks ratio), indicating that genes in the metabolic network evolve under weaker purifying selection

pressure and that recent evolutionary changes occurred. For the cell-cycle network, the TAI profile is almost entirely composed of older phylostrata (PS1 to PS8), while new genes contribute about 20% to the TAI profile of the metabolic network (Figure 4D and 4E). This indicates that the gene expression levels of the cell-cycle network have likely been optimized and fixed early on during evolution as an adaptive strategy to cope drought stress (Harrison et al. 2012). TDI profiles support this claim: conserved genes do contribute more to the TDI profiles in cell-cycle networks and show adaptive changes in expression for drought response (higher TDI difference between control and drought transcriptomes in cell-cycle network, Figure 6B). In contrast, drought-responsive genes in metabolism network appear unstable in their expression in response to drought stress, because this strategy may be linked to an initial response to severe water scarcity (Dubois and Inzé 2020).

Population genetics analysis of drought-responsive networks

We also study the selective forces acting on the identified drought-responsive gene networks at the intraspecific time scale. Using full genome sequences of six *S. chilense* populations (C_LA1963, C_LA3111, C_LA2931, SC_LA2932, SC_LA4107, and SH_LA4330; five plants each) recently reported in (Wei et al. 2022) aligned to the reference genome of *S. chilense*, we identify 45,208,263 high-quality single-nucleotide variants (SNPs), in which 111,606 SNPs are found in genes of the cell-cycle GCN and 167,334 SNPs in genes of the metabolic GCN. We first compare population structure between the whole-genome data and drought-responsive genes (Figure S8). The results corroborate the genetic structure revealed in (Wei et al. 2022) based on the sequence alignment to *S. pennellii* reference (Figure S8A and S8C). However, the structure exhibited by drought genes shows stronger differentiation among populations, especially to SH_LA4330, than the WGS data (especially for clustering of populations of the central region). Moreover, the strong differences from WGS data between the two south coastal populations (SC_LA2932 and SC_LA4107) is attenuated when analyzing SNPs from the drought-responsive genes (Figure S8B and S8D).

We find that the mean nucleotide diversity (π) per gene does not differ between the two GCNs (Figure S9A; Table S6; Kolmogorov-Smirnov test, $P = 0.15$). In addition, the π values of the promoter regions (2kb upstream of the transcription initiation site) are significantly higher than those of the gene (coding) regions (Figure S9A; Table S6; Kolmogorov-Smirnov test, $P = 0.03$). This result may be due to relaxed selective constrain in promoter regions while possibly explaining why few TFs can bind to multiple genes in the GCN (Table S3). TFs are indeed conserved at the coding sequence level, especially at the functional domains, but higher amount of polymorphism of TF binding sites in the promoter can be indicative of complex and diverse regulation, for example in response to stressful conditions (Spivakov 2014; Sato et al. 2016). Albeit, there is no difference in the nucleotide diversity at the promoter regions between the two GCNs (Figure S9A; Table S6).

Furthermore, the genes for the metabolic GCN show lower Tajima's D values than those of the cell-cycle GCN (Figure S9B; Table S6; Kolmogorov-Smirnov test, $P = 0.04$), suggesting more prevalent recent positive or negative selection pressure in the metabolic GCN. There is a very weak correlation between Tajima's D and Ka/Ks ratio for the cell-cycle GCN and absence of correlation for the metabolic GCN (Figure S10A and S10B). As a negative correlation between Tajima's D and Ka/Ks ratio is indicative of possible recent positive selection, the results could slightly hint to the occurrence of recent positive selection acting at more genes of the metabolic GCN (Figure S9B; Table S6).

We further find significant, but opposite, correlations between nucleotide diversity/Tajima's D and the contributions of the different DS for the two GCNs (Figure S10C and S10D). In the cell-cycle GCN, the contributions of different DS have significant positive correlation with nucleotide diversity and Tajima's D (Figure S10C and S10E). This indicates that DS of high contribution to TDI profiles show high nucleotide diversity (and positive Tajima's D), meaning that older genes are under stronger purifying selection than younger genes in this network because the sequence divergence of cell-cycle genes occurred at old time periods. In contrast, a negative correlation is observed between the contribution of each

DS and nucleotide diversity or Tajima's D in the metabolic network (Figure S10D and S10F). Hence, DS with high contribution show low nucleotide diversity and low Tajima's D, especially DS10. Therefore, it appears likely that the metabolic genes, which may be recently evolved, may be under recent positive selection due to the recent evolution of the drought response transcriptome.

Drought-responsive genes under positive selection promote adaptive evolution in response to drought stress

Genome scan analyses have been recently used to detect candidate genes under positive selection in six populations of *S. chilense* (Wei et al. 2022). We search for overlap between genes of two drought-response GCNs studied here and our previously identified 799 candidate genes under positive selection (Wei et al. 2022). We find 74 and 126 drought-responsive genes in the cell-cycle and metabolic networks, respectively under the list of positive selection candidate genes (Figure 6A; Table S7). This indicates that drought stress is likely an important driver of adaptation and these drought-response genes may play key roles for colonization of new arid or hyper-arid habitats. Similar numbers of drought-responsive genes likely under selection are observed across different populations of *S. chilense* encompassing different parts of the range, except for SH_LA4330 (Wei et al. 2022). The number of candidate genes belonging to the metabolic or cell-cycle GCNs is similar in the three central populations (C_LA1963, C_LA3111 and C_LA2931) (Figure 6A; Table S7). The most recent diverged highland population (SH_LA4330) contains the largest number of positively selected drought-responsive genes (Figure 6A; Table S7) with a similar proportion of genes from both networks. Noticeably, in the two south-coast populations (SC_LA2932 and SC_LA4107) a large majority of genes under positive selection belong to the metabolic GCN (showing absence of cell-cycle genes in population SC_LA2932, Figure 6A; Table S7).

Previous studies have demonstrated that positively selected genes exhibit pleiotropy in local adaptation, and proposed connectivity of molecular networks for quantifying pleiotropic

effects (Wagner et al. 2007; Erwin and Davidson 2009; Hämälä et al. 2020). To address the role that (putatively) positively selected genes play in drought-responsive networks, we compare the gene connectivity of our candidate genes under selection in the two networks (Figure 6B; Table S8). In the metabolic network, the connectivity of positively selected genes (0.55 ± 0.10) is significantly higher than other drought-responsive genes (0.44 ± 0.12) (Figure S11A; Kolmogorov-Smirnov test, P = 0.017), but we do not observe such significant difference for the cell-cycle network (Figure S11A; Kolmogorov-Smirnov test, P = 0.43). Furthermore, the connectivity of positively selected genes of the metabolic network is much higher than those from the cell-cycle network in six populations (Figure 6B; Table S8; Kolmogorov-Smirnov test, P = 0.007). This result suggest that highly pleiotropic genes in the metabolic GCN may have facilitated the recent colonization of new habitats (Hämälä et al. 2020) during the divergence process of *S. chilense*. In contrast, the connectivity of positively selected genes in the cell-cycle network is significantly lower (Figure S11A). Therefore, we suggest that the two networks underwent different evolutionary selective pressures during the range expansion of *S. chilense*.

Finally, we compare the age of the selective sweep at the candidate genes of the two GCNs based on the results in Wei et al. (2022). We find that sweep ages at the cell-cycle genes are slightly younger than at those of the metabolic network, especially in the three highland populations (C_LA2931, C_LA3111 and SH_LA4330; Figure S11B and S11C; Table S8). This suggest that drought adaptation is in line with the inferences of demography and colonization found in our previous studies (Stam et al. 2019; Wei et al. 2022). Interestingly, we find significantly positive correlation between the age of the sweep and gene connectivity for both GCNs and across all six populations (Figure 6C). Figure 6D and 6E provide the visualizations of two networks and exhibit the relationship between sweep age and connectivity (with weighted connection strength greater than 0.65 between any two genes). In other words, it appears that selective sweeps tend to happen first at more connected genes and, subsequently at less connected genes, during the history of

colonization/adaptation of new arid habitats. To our knowledge, this is the first report of a correlation between the age of a selective sweep and the connectivity of genes in a network. To obtain more evidence to support this inference, we also calculate the tMRCA (time to most recent common ancestor) to estimate the age of drought-responsive genes based on allele frequency of SNPs. The positive correlation between tMRCA of drought-responsive genes under the positive selection and connectivity also is observed (Pearson's cor=0.69, P = 2.47e-5), consistent with the correlation with sweep age. Moreover, the low correlation (Pearson's cor=0.31, P = 0.14) is observed between tMRCA of other (non-sweep) drought-responsive genes and connectivity. This may indicate a pattern of polygenic adaptation in GCNs where the positive selection acts first on core genes (high connectivity) of networks, and subsequently on the marginal genes (less connectivity). These positively selected genes ultimately regulate the expression of other genes in the network.

Discussion

In this study, we identify two drought-responsive GCNs by analyzing gene expression profiles of plants growing under control and drought conditions. Two GCNs involved in cell-cycle and metabolic biological processes are detected and their structural relevance are confirmed by TF/TFBS predictions. These networks represent two different strategies for drought response (Farooq et al. 2009; Danilevskaya et al. 2019). We then demonstrate that the cell-cycle network is evolutionary older and more conserved than the metabolic network. Despite the ancient history of these two GCNs, we further show that both GCNs contribute to the recent history of adaptation to drought conditions (aridity) when *S. chilense* colonizes new habitats around the Atacama desert. The joint analyses of genomic and transcriptomic data indicates that 1) at the transcriptome level, metabolic GCN is more sensitive to mutation with younger selection events in response to new environments, 2) cell-cycle GCN is less evolvable, explaining the more divergent transcriptome between drought and normal conditions, and 3)

both networks still present signals of evolution under positive selection in core elements of the GCN, while peripheral genes of the network can be involved adaptation in later stages of the colonization processes.

Drought tolerance is mediated by regulation in cell proliferation and metabolism

The organ development can be roughly divided into cell proliferation and cell expansion, with water deficit being a limiting factor for both processes (Alves and Setter 2004; Verelst et al. 2013). Drought stress reduces the activity of the cell cycle and thus slows down the growth and development of plants. The down-regulated genes we find in the cell-cycle network also indicate that genes related to cell cycle are suppressed by drought stress possibly to restrict the cell division in *S.chilense.* Reduction of cell number due to mild drought stress is found in *A. thaliana* (Skirycz and Inzé 2010). This means that the cell-cycle response to drought may be very general and indirect. However, our speculations are mainly based on the aboveground tissues of *S. chilense*. The changes of fundamental metabolic activity may be a more early and variable drought-responsive strategy presumably related to an acclimation response (Harb et al. 2010). Plant water shortage is first reflected in changes in metabolic processes, such as accelerating the catabolism of macromolecules in order to regulate the penetration of tissues, to maintain physiological water balance, or slowing down metabolism to reduce energy and water consumption (Reddy et al. 2004; Gupta et al. 2020). In addition, the signaling pathways related to the metabolic gene network are also demonstrated to be a response to drought stress, for example, the abscisic acid (ABA) signaling pathway regulates the response to dehydration and optimizes water utilization (Harb et al. 2010; Wilkinson and Davies 2010). Although these two GCNs correspond to two different strategies of drought response, they are not isolated, but interact with one another in a time-dependent manner. Water deprivation and heat first change the metabolic processes leading to stomata closure, which leads then to cell cycle network to be affected under long-term lack of water. In return, the increased or decreased cell cycle gene expression affects the further physiology and metabolism of the plant (Gupta et al. 2020). Indeed, drought-responsive strategies regulating

the cell cycle appear to be activated later than metabolism processes, as glucose metabolism rapidly follows drought stress, whereas the accumulation of amino acids which is a crucial part of the cell cycle response starts at a later time in response to drought (Fàbregas and Fernie 2019).

Rewiring of ancient GCNs drives recent adaptation to dry environments

The phylostratigraphic analyses support that most drought-responsive genes in *S. chilense* originated in the early to middle stages of plant evolutionary history. This is consistent with the time of origin of multiple abiotic response genes in *Arabidopsis thaliana* (Mustafin et al. 2019). The divergence times of land plants and flowering plants are important periods for the origin drought-responsive genes. The divergence of main plant groups have been linked to recurrent whole genome duplication events that promoted gene family expansions, gene neo- and sub-functionalization and genome rearrangements (Wang et al. 2012; Clark and Donoghue 2018). Those genomic processes likely promote the enrichment of drought-responsive GCNs. For instance, key drought-response morphological traits such stomata are present in the ancestral land plant. However, stomatal genes predate the divergence of land plants showing multiple duplications along the evolution of the group, and their response to environmental cues such humidity, light, CO_2 and ABA are widely distributed and possibly ancestral to land plants (Clark et al. 2022). Therefore, we suggest that our two drought-responsive networks were mainly shaped before or along the divergence of land plants and expanded subsequently.

Previous studies show that TAI and TDI profiles across embryogenesis, seed germination and transition to flowering in *A. thaliana* exhibit a ‘hourglass pattern’ (older and conserved transcriptomes are preferentially active at the mid-development stages; Quint et al. 2012; Drost et al. 2016). Though, our TAI/TDI profiles for the two developmental stages remain stable under the same conditions (Figures 4C and 5B). The similar TAI/TDI between developmental stages we obtained is certainly because our analyses focused on two

modules (co-expressed genes) highly correlated to the differential expression between drought and control conditions (Figure S2D). Therefore, developmental stage-specific response genes are underrepresented in the two analyzed networks. However, increased TAI/TDI values under drought conditions suggest that stress response transcriptomes are composed of relatively more recently diverged genes, and therefore are more evolvable. We then highlight that this inference needs to be verified in other stress responsive transcriptomes (salt, heat, cold, etc.). We then speculate, that although abiotic stress response regulatory networks are mostly composed of highly ancient and conserved elements across species (Chen and Zhu 2004), networks retain the ability to change expression patterns to respond rapidly to environmental changes or explore new ecological niches. Moreover, given the pleiotropic nature of the abiotic stress-response traits, we can expect shared patterns of evolution (at the constitutive and expression components) of the networks for different stress conditions (and possible trade-offs between traits and GCNs).

Extensive network rewiring in relatively recent and short time-frames have been found in maize and tomato in response to domestication (Swanson-Wagner et al. 2012; Koenig et al. 2013). It is therefore not surprising to find signs of adaptive variation in core elements of rather conserved regulatory networks related to the colonization processes of new habitats. The genetic (and morphological) divergence of the southern populations, southern coastal and highland marginal, is recent but strong (Raduski and Igić 2021). It is congruent with theoretical results showing that gene networks with higher mutation sensitivity (more evolvable) can facilitate local adaptation, increasing gene expression and lead to accelerated range expansion processes in abiotic environmental gradients (Deshpande and Fronhofer 2022). Complementarily, our empirical approach shows the existence of two regulatory networks with different evolutionary tendencies, one more conserved than the other and with different gene expression responses. One GCN would exhibits a faster and more variable response (metabolic) and the other a later (delayed) but more constitutive response (cell-cycle) to drought. Despite the differences in gene age and variation between the networks,

our results show that both GCNs have undergone sufficient changes leading to their rewiring during the divergent process of colonization of *S. chilense* around the Atacama. Nevertheless, genes in the metabolic network show more recent evolution, with new genes members appearing in *S. chilense*, concomitantly with more variable expression in the drought transcriptome.

These drought-responsive genes to *S. chilense* likely facilitated the adaptation of this species to unique arid and hyper-arid habitats, especially when colonizing the southern part of the range. Indeed, population structure based on SNPs indicates that drought-responsive genes reflected adaptation/colonization to arid habitats in *S. chilense* (Figure S8). Importantly, we found about 200 drought-responsive genes previously identified as candidate genes under positive selection (i.e. located within sweep regions; Wei et al. 2022). This confirms that drought stress is an important driver of ecological divergence in *S. chilense.* We finally provide some indirect evidence that changes at central genes (with higher connectivity) can be responsible for the short-term response to selection (Jovelin and Phillips 2009; Luisi et al. 2015) and promote rewiring of the gene network (Koubkova-Yu et al. 2018). Thus, highly connected genes may be targets of positive selection during the first phase of the environmental change or colonization to contrasting environments, and may be keys for 'piggybacking', defined as the change in gene expression of a focal gene driving phenotypic change.

Limitations and further work

A limitation of our gene expression study is that our transcriptomic analyses are based on individuals from a single location (near the putative region of origin of the species; Wei et al. 2022), while variability in gene expression and phenotypic response has been observed between different populations (Mboup et al. 2012; Fischer et al. 2013; Nosenko et al. 2016). Further expression studies including plants from multiple locations would be useful to verify

that the identified GCNs are also present and expressed in other populations and study the possible variation in the most southern populations. More evidence based on multiple populations is needed to confirm the 'piggybacking' phenomenon of gene expression in *S. chilense*. Additional support on the variability of transcriptome evolution across populations as well as long read sequencing of more genomes will be beneficial in assessing the role of gene duplication and gene deletion yielding the evolution of the gene networks. Such studies would also allow the analysis of evolution of adaptive gene networks and polygenic selection occurring for complex traits such as drought tolerance. Finally, more detailed studies with a larger sample size from the field will help to discover other gene networks and their interactions related to abiotic stress and the evolution of the species. A detailed discussion of the potential biases associated with the use of multiplied accessions at TGRC (Tomato Genetics Resource Center, UC Davis, USA) compared to samples from natural populations is found in Wei et al. (2022). Sampling and experimental work in the field would improve the resolution of transcriptome and genomic studies, in order to assess phenotypic differences between organs and stages of development and thus extend the knowledge to other relevant characteristics such as secondary metabolism, which is known to have relevant influence on biotic and abiotic interactions (Mes et al. 2008; Bolger et al. 2014; Tapia et al. 2022).

Material and methods

Plant material and drought stress experiment

Seeds of *S. chilense* accession LA1963 were acquired from Tomato Genetics Resource Center (TGRC), University of California at Davis. Seeds were soaked in 50% household bleach (2.7% sodium hypochlorite) for 30 minutes and rinsed thoroughly with water according to instructions provided by TGRC. The rinsed seeds were sown into pots containing sterilized soil with perlite and sand (1:2) and grown under controlled conditions (22C day/20C night, 16h light/8h dark photoperiod). On the 24th day after sowing, all plants

were randomly distributed into two groups and watered with a sufficient volume to reach the bottom of containers (30-40 ml). The first group of plants were maintained under normal watering condition, watered with a sufficient volume of water (50-55 ml) on 4, 7 and 11 days after start of the experiment (day 24). A moderate water stress regime was imposed for second group of plants by stopping irrigation for 7 days followed by re-watering with 25 ml of water. On day 12, newly expanded leaf (1-1.5 cm length) and shoot apices with immediately surrounding leaf primordia (shoot apices and P1-P5 leaf primordia) from each group were dissected carefully using razor blades and immediately grounded into fine powder in liquid nitrogen for RNA extraction. Four biological replicates were used for all RNA-Seq experiments from each tissue type. Each replicate of leaf and shoot apex included the pooled tissues from five and six plants, respectively.

RNA extraction and cDNA library construction

Libraries were constructed and named as follows: leaves under control (optimal watering) condition (CL-A to D), shoot apices under control condition (CSA-E to H), leaves under drought condition (DL-I to L), and shoot apices under drought condition (DSA-M to P). Tissues were lysed using zircon beads in Lysate Binding Buffer containing Sodium Dodecyl Sulfate. mRNA was isolated from 200 µl of lysate per sample with streptavidin coated magnetic beads for indexed non-strand specific RNA-Seq library preparation according to the method described by (Kumar et al. 2012). 1 µl of 12.5 µM of 5-prime biotinylated polyT oligonucleotide and streptavidin-coated magnetic beads were used to capture mRNA and isolate captured mRNAs from the lysate, respectively. Equal amount of mRNA of each experimental group were used to construct 16 libraries. For library construction the rapid version of Kumar et al. (2012) RNA-sequencing method (Townsley et al. 2015) was used. Each sample was barcoded using standard Illumina adaptors 1-16 to allow up to 16 samples to be pooled in one lane of sequencing on Illumina HiSeq4000. The libraries were eluted from the pellet with 10 µl 10 mM Tris pH 8.0 and pooled as described by Kumar et al. (2012). Quantification and quality assessment of resulting libraries were performed on Fragment

Analyzer (FGL_DNF-474-2- HS NGS Fragment 1-6000bp.mthds) and sequenced using the Illumina HiSeq 4000 platform to generate 100 bp single-end reads at the Vincent J. Coates Genomic Sequencing Facility at UC Berkeley.

Transcriptome and genome data processing and mapping

For transcriptome data, the adapters were removed from raw reads by two consecutive rounds using BBDuk in BBTools v38.90 (Bushnell 2014). Two sets of parameters were used in two rounds respectively: first round 'ktrim=r k=21 mink=11 hdist=2 tpe tbo minlength=21 trimpolya=4'; second round 'ktrim=r k=19 mink=9 hdist=1 tpe tbo minlength=21 trimpolya=4'. Then Low-quality reads were also removed with BBDuk using parameters 'k=31 hdist=1 qtrim=lr trimq=10 maq=12 minlength=21 maxns=5 ziplevel=5'. The clean reads of each sample were mapped to the *S. chilense* reference genome (Silva-Arias et al. submited) using BBMap in BBTools. The SAM files were then converted and sorted to BAM files using Samtools v1.11 (Wysoker et al. 2009). The number of reads were mapped to each gene were counted via featureCounts v2.0.1 in each sample (Liao et al. 2014). To eliminate the differences between samples, the gene expression level was normalized using the TPM (Transcripts Per Kilobase Million) method (Wagner et al. 2012). In addition, the transcriptome data also was processed and mapped using CLC Genomics Workbench 10 (Liu and Di 2020) based on reference genome of *S. lycopersicum* (ITAG 3.0; The Tomato Genome Consortium 2012). Parameters considered for filtration were adapter trimming, removing the low-quality reads (Q < 25), removal of bases of the start of a read and the end of a read (Score < 25). Reads were mapped to the ITAG 3.0 reference genome using Large Gap Read Mapping tool of Transcript Discovery Plugins of CLC Genomics Workbench 10. Annotated reference sequences predicted by CLC bio Transcript Discovery Tool were extracted and used as reference in a subsequent RNA-Seq analysis. Finally, TPM value was also calculated.

The relationships among transcriptome samples were evaluated using the TPM values. The correlation coefficient between two samples was calculated to evaluate repeatability

between samples using Pearson's test. Principal component analysis (PCA) was performed using the *plotPCA()* function in DESeq2 R package (Love et al. 2014).

Identification of differentially expressed genes and gene co-expression analysis

Differential expression analysis of groups among the different conditions and tissues was performed using the DESeq2 R package. The raw read counts were inputted to detect Differential Expressed Genes (DEGs). The *P*-value ≤ 0.001, the absolute value of log2FoldChange ≥ 1 and a false discovery rate (FDR) adjusted *P* ≤ 0.001 were classified as differentially expressed genes.

To identify the gene co-expression networks, weighted gene correlation network analysis (WGCNA) was constructed using TPM values to identify specific modules of co-expressed genes associated with drought stress (Langfelder and Horvath 2008). We first checked for genes and samples with too many missing values using *goodSamplesGenes()* function in WGCNA R package. We then removed the offending genes (the last statement returns 'FALSE'). To construct an approximate scale-free network, a soft thresholding power of five was used to calculate adjacency matrix for a signed co-expression network. Topological overlap matrix (TOM) and dynamic-cut tree algorithm were used to extract network modules. We used a minimum module size of 30 genes for the initial network construction and merged similar modules exhibiting > 75% similarity. To discover modules of significantly drought-related, module eigengenes were used to calculate correlation with samples with different conditions. The visualization of networks were created using Cytoscape v3.8.2 (Su et al. 2014).

Identification of transcript factor families and transcript factor binding sites

The protein sequences were obtained from the reference genome and annotation 'gff' file with GffRead (Pertea and Pertea 2020), and were used to identify TF families using online tool PlantTFDB v5.0 (Guo et al. 2007). Furthermore, the upstream 2000 bp sequences of the

transcription start sites (TSS) were extracted as the gene promoter from the reference genome to detect TFBS. The TFBS dataset of relative species *S. pennellii* was also downloaded from Plant Transcriptional Regulatory Map (PlantRegMap, http://plantregmap.gao-lab.org/) as background of TFBS identification (Tian et al. 2020). Then, the TFBS of *S. chilense* was identified using FIMO program in motif-based sequence analysis tools MEME Suit v5.3.2 (Bailey et al. 2015). The TFBS was extracted with $p < 1e-5$ and $q < 0.01$.

Gene ontology (GO) analysis

We first constructed the dataset of assigned GO terms for all genes used protein sequence by PANTHER v16.0 (Mi et al. 2021). Then, the GO enrichment analysis of drought-responsive genes was performed using clusterProfilter v3.14.2 (Yu et al. 2012). Benjamini–Hochberg method was used to calibrate *P* value, and the significant GO terms were selected with *P*-value below to 0.05.

Construction of phylostratigraphic map

We performed phylostratigraphic analysis based on the following steps. First, the phylostrata (PS) was defined according to the full linkage of *S. chilense* from NCBI taxonomy database. The similar PS was merged and finally 18 PS were generated (Figure 4A). Second, the protein sequences were blast to a database of non-redundant (nr) proteins downloaded from NCBI (https://ftp.ncbi.nlm.nih.gov/blast/db/) with a minimum length of 30 amino acids and an E-value below 10^{-6} using blastp v2.9.0 (Camacho et al. 2009). Third, each gene was assigned to its PS by the following criterion: if no blast hit or only one hit of *S. chilense* with an E-value below 10^{-6} was identified, we assigned the gene to the youngest PS18. When multiple blast hits were identified, we computed lowest common ancestor (LCA) for multiple hits using TaxonKit v0.8.0 (Shen and Ren 2021) and then assigned LCA to specific PS.

Construction of divergence map

We performed divergence stratigraphy analysis to construct sequence divergence map of *S. chilense* using function *divergence_stratigraphy()* of R package 'orthologr' (Drost et al. 2015) following four steps: 1) the coding sequences for each gene of *S. chilense* and *S. pennellii* (NCBI assembly SPENNV200) were extracted from their reference and annotation files. 2) We identified orthologous gene pairs of both species by choosing the best blast hit for each gene using blastp. We only considered a gene pair orthologous when the best hit has an E-value below 10^{-6}, the gene pair is considered orthologous; otherwise, it is discarded. 3) Codon alignments of the orthologous gene pairs were performed using PAL2NAL (Suyama et al. 2006). Then, Ka/Ks values of the codon alignments were calculated using Comeron's method (Comeron 1995). And 4) all genes were sorted according to Ka/Ks values into discrete deciles, which are called divergence stratum (DS).

Estimation of transcriptome age index and transcriptome divergence index

The TAI is computed based on phylostratigraphy and expression profile, which assign each gene to different phylogenetic ages by identification of homologous sequences in other species (Domazet-Lošo et al. 2007). The evolutionary age of each gene was quantified combining its PS and expression level to obtain weighted evolutionary age. Finally, weighted ages of all genes are averaged to yield TAI, which is defined as the mean evolutionary age of a transcriptome (Domazet-Lošo and Tautz 2010). A lower value of TAI describes an older mean evolutionary age, whereas a higher value of TAI denotes a younger mean evolutionary age and implies that evolutionary younger genes are preferentially expressed in the corresponding sample or condition (Domazet-Lošo and Tautz 2010; Piasecka et al. 2013). The TDI represents the mean sequence divergence of a transcriptome quantified by divergence strata (DS) and gene expression profile (Quint et al. 2012). The genes are assigned to different DS and then weighted by their expression level to yield the TDI. A lower value of TDI describes a more conserved transcriptome (in terms of sequence dissimilarity), whereas a higher value of TDI denotes a more variable transcriptome. Here, we calculate TAI and TDI profiles in different samples using *PlotSignature()* function of the myTAI R package.

Population genetics analysis and detection of positive selection on drought-responsive genes

Whole-genome sequence data from six populations *S. chilense* (five individuals each) previously analyzed in (Wei et al. 2022; BioProject PRJEB47577) were used to calculate population genetics statistics for coding and promoter region sequences for all genes identified in the GCNs. Single nucleotide variants (SNPs) based on the short-read alignment to the new reference genome for *S. chilense* (Silva-Arias et al. submitted) using the same methods in Wei et al. (2022). Population genetics statistics namely, nucleotide diversity (π) and Tajima's D were calculated with ANGSD v0.937 (Korneliussen et al. 2014) over gene and promoter regions. These statistics first were calculated at per site in gene and promoter regions, and then we used a R script (https://gitlab.lrz.de/population_genetics/s.chilense-drought-transcriptome) to obtain statistics in each gene and promoter regions. PCA on SNP data from 30 whole genomes was also performed using GCTA (v1.91.4; Yang et al. 2011). The genetic structure inference was performed using ADMIXTURE v1.3.0 (Alexander et al. 2009).

Drought-responsive genes under positive selection were extracted by blast (e-value < 1e-6) between drought-responsive genes identified in this study and the genes located inside sweep regions in our previous study using *S. pennellii* as the reference genome. We also use the sweep ages obtained in Wei et al. (2022).

Estimation of allele age

We implemented in GEVA (Genealogical Estimation of Variant Age; Albers and McVean 2020) to dating genomic variants in the drought-responsive genes. We generated input for GEVA based on the recombination rate 3.24 x 10^{-9} per site per generation (based on the overall recombination density in *S. lycopersicum* [1.41 cM/Mb] Anderson and Stack 2002; Nieri et al. 2017; and within the possible range of rates in Wei et al. 2022). We used population size (N_e)

20,000 and mutation rate 5.1 x 10^{-9} (Roselius et al. 2005; Wei et al. 2022), and then relied on the recombination clock to estimate the age of alleles (tMRCA).

Supplementary material

Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgements

KW was funded by the Chinese Scholarship Council. SS was funded by the Fulbright Visiting Scholar Program of the U.S. Department of State and by German Academic Exchange Service (DAAD) within the framework of Research Stays for University Academics and Scientists. GAS-A was funded by the Technical University of Munich. AT acknowledges funding from DFG (Deutscheforschungsgemeinschaft) Grant Number: 317616126 (TE809/7-1). NS acknowledges funding from U.S. National Science Foundation (NSF/ IOS-1238243). This work was supported by grants from JSPS KAKENHI (JP19K23742, JP20K06682, JP20KK0340 to H.N.), and NSF PGRP IOS-1856749 and IOS-211980 to NS. The laboratory work of H.N. was supported by a Grant-in-Aid for Scientific Research on Innovative Areas (JP19H05672). We thank the Tomato Genetics Resource Center (TGRC) of the University of California, Davis for generously providing us with the seeds of the accession included in this study.

Data Availability

The raw single-end sequencing RNA data is available in PRJDB15063. The raw pair-end sequencing genomic data can be accessed at the European Nucleotide Archive (ENA) project accession PRJEB47577. All codes used in this study and other previously published

genomic data are available at the sources referenced. The code for implementing the analyses used in this paper can be found on our GitLab repository: https://gitlab.lrz.de/population_genetics/s.chilense-drought-transcriptome

References

Albers PK, McVean G. 2020. Dating genomic variants and shared ancestry in population-scale sequencing data. *PLOS Biol.* 18:e3000586.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.

Alves AA, Setter TL. 2004. Response of cassava leaf area expansion to water deficit: cell proliferation, cell expansion and delayed development. *Ann. Bot.* 94:605–613.

Anderson L, Stack S. 2002. Meiotic recombination in plants. *Curr. Genomics* 3:507–525.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. *Nucleic Acids Res.* 43:W39–W49.

Basu S, Ramegowda V, Kumar A, Pereira A. 2016. Plant adaptation to drought stress. *F1000Research* 5.

Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S, Sørensen I, Lichtenstein G, et al. 2014. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* 46:1034–1038.

Böndel KB, Lainer H, Nosenko T, Mboup M, Tellier A, Stephan W. 2015. North–south colonization associated with local adaptation of the wild tomato species *Solanum chilense*. *Mol. Biol. Evol.* 32:2932–2943.

Böndel KB, Nosenko T, Stephan W. 2018. Signatures of natural selection in abiotic stress-responsive genes of *Solanum chilense*. *R. Soc. Open Sci.* 5:171198.

Bowles AMC, Paps J, Bechtold U. 2021. Evolutionary origins of drought tolerance in Spermatophytes. *Front. Plant Sci.* 12.

Bushnell B. 2014. BBTools. A suite of bioinformatics tools used for DNA and RNA sequence data analysis. *DOE Jt. Genome Inst.* [Internet]. Available from: https://sourceforge.net/projects/bbmap/

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:1–9.

Chen WJ, Zhu T. 2004. Networks of transcription factors with roles in environmental stress response. *Trends Plant Sci.* 9:591–596.

Ciais P, Reichstein M, Viovy N, Granier A, Ogée J, Allard V, Aubinet M, Buchmann N, Bernhofer C, Carrara A, et al. 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* 437:529–533.

Clark JW, Donoghue PCJ. 2018. Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* 23:933–945.

Clark JW, Harris BJ, Hetherington AJ, Hurtado-Castaño N, Brench RA, Casson S, Williams TA, Gray JE, Hetherington AM. 2022. The origin and evolution of stomata. *Curr. Biol.* 32:R539–R553.

Crow M, Suresh H, Lee J, Gillis J. 2022. Coexpression reveals conserved gene programs that co-vary with cell type across kingdoms. *Nucleic Acids Res.* 50:4302–4314.

Danilevskaya ON, Yu G, Meng X, Xu J, Stephenson E, Estrada S, Chilakamarri S, Zastrow-Hayes G, Thatcher S. 2019. Developmental and transcriptional responses of maize to drought stress under field conditions. *Plant Direct* 3:e00129.

Deshpande JN, Fronhofer EA. 2022. Genetic architecture of dispersal and local adaptation drives accelerating range expansions. *Proc. Natl. Acad. Sci.* 119:e2121858119.

Domazet-Lošo T, Brajković J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* 23:533–539.

Domazet-Lošo T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468:815–818.

Drost H-G, Bellstädt J, Ó'Maoiléidigh DS, Silva AT, Gabel A, Weinholdt C, Ryan PT, Dekkers BJW, Bentsink L, Hilhorst HWM, et al. 2016. Post-embryonic hourglass patterns mark ontogenetic transitions in plant development. *Mol. Biol. Evol.* 33:1158–1163.

Drost H-G, Gabel A, Grosse I, Quint M. 2015. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol. Biol. Evol.* 32:1221–1231.

Dubois M, Inzé D. 2020. Plant growth under suboptimal water conditions: early responses and methods to study them.Mittler R, editor. *J. Exp. Bot.* 71:1706–1722.

Erwin DH. 2020. Chapter Thirteen - Evolutionary dynamics of gene regulation. In: Peter IS, editor. Current Topics in Developmental Biology. Vol. 139. Gene Regulatory Networks. Academic Press. p. 407–431. Available from: https://www.sciencedirect.com/science/article/pii/S0070215320300326

Erwin DH, Davidson EH. 2009. The evolution of hierarchical gene regulatory networks. *Nat. Rev. Genet.* 10:141–148.

Fàbregas N, Fernie AR. 2019. The metabolic response to drought. *J. Exp. Bot.* 70:1077–1085.

Farooq M, Wahid A, Kobayashi N, Fujita D, Basra S. 2009. Plant drought stress: effects, mechanisms and management. *Sustain. Agric.*:153–188.

Ficklin SP, Feltus FA. 2011. Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiol.* 156:1244–1256.

Fischer I, Camus-Kulandaivelu L, Allal F, Stephan W. 2011. Adaptation to drought in two wild tomato species: the evolution of the *Asr* gene family. *New Phytol.* 190:1032–1044.

Fischer I, Steige KA, Stephan W, Mboup M. 2013. Sequence evolution and expression regulation of stress-responsive genes in natural populations of wild tomato. *PLoS One* 8:e78182–e78182.

Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, Eanes WF. 2007. Adaptive evolution of metabolic pathways in *Drosophila*. *Mol. Biol. Evol.* 24:1347–1354.

Gehan MA, Greenham K, Mockler TC, McClung CR. 2015. Transcriptional networks—crops, clocks, and abiotic stress. *Curr. Opin. Plant Biol.* 24:39–46.

Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ. 2014. Comparative analysis of the transcriptome across distant species. *Nature* 512:445–448.

Guo A-Y, Chen X, Gao G, Zhang H, Zhu Q-H, Liu X-C, Zhong Y-F, Gu X, He K, Luo J. 2007. PlantTFDB: a comprehensive plant transcription factor database. *Nucleic Acids Res.* 36:D966–D969.

Gupta A, Rico-Medina A, Caño-Delgado AI. 2020. The physiology of plant responses to drought. *Science* 368:266–269.

Haak DC, Kostyun JL, Moyle LC. 2014. Merging ecology and genomics to dissect diversity in wild tomatoes and their relatives. In: Landry CR, Aubin-Horth N, editors. Ecological Genomics. Vol. 781. Dordrecht: Springer Netherlands. p. 273–298. Available from: http://link.springer.com/10.1007/978-94-007-7347-9_14

Hämälä T, Gorton AJ, Moeller DA, Tiffin P. 2020. Pleiotropy facilitates local adaptation to distant optima in common ragweed (*Ambrosia artemisiifolia*). *PLoS Genet.* 16:e1008707.

Harb A, Krishnan A, Ambavaram MMR, Pereira A. 2010. Molecular and physiological analysis of drought stress in *Arabidopsis* reveals early responses leading to acclimation in plant growth. *Plant Physiol.* 154:1254–1271.

Harrison PW, Wright AE, Mank JE. 2012. The evolution of gene expression and the transcriptome–phenotype relationship. *Semin. Cell Dev. Biol.* 23:222–229.

Jill Harrison C. 2017. Development and genetics in the evolution of land plant body plans. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20150490.

Josephs EB, Wright SI, Stinchcombe JR, Schoen DJ. 2017. The relationship between selection, network connectivity, and regulatory variation within a population of *Capsella grandiflora*. *Genome Biol. Evol.* 9:1099–1109.

Jovelin R, Phillips PC. 2009. Evolutionary rates and centrality in the yeast gene regulatory network. *Genome Biol.* 10:1–10.

Juenger TE. 2013. Natural variation and genetic constraints on drought tolerance. *Curr. Opin. Plant Biol.* 16:274–281.

Kashyap SP, Prasanna HC, Kumari N, Mishra P, Singh B. 2020. Understanding salt tolerance mechanism using transcriptome profiling and *de novo* assembly of wild tomato *Solanum chilense*. *Sci. Rep.* 10:1–20.

Kim PM, Korbel JO, Gerstein MB. 2007. Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. *Proc. Natl. Acad. Sci.* 104:20274–20279.

Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, et al. 2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci.* 110:E2655–E2662.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356.

Koubkova-Yu TC-T, Chao J-C, Leu J-Y. 2018. Heterologous Hsp90 promotes phenotypic diversity through network evolution. *PLoS Biol.* 16:e2006450.

Kumar R, Ichihashi Y, Kimura S, Chitwood DH, Headland LR, Peng J, Maloof JN, Sinha NR. 2012. A high-throughput method for Illumina RNA-Seq library preparation. *Front. Plant Sci.* 3:202.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.

Liu C-H, Di YP. 2020. Analysis of RNA sequencing data using CLC genomics workbench. In: Molecular Toxicology Protocols. Springer. p. 61–113.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Luisi P, Alvarez-Ponce D, Pybus M, Fares MA, Bertranpetit J, Laayouni H. 2015. Recent positive selection has acted on genes encoding proteins with more interactions within the whole human interactome. *Genome Biol. Evol.* 7:1141–1154.

Mähler N, Wang J, Terebieniec BK, Ingvarsson PK, Street NR, Hvidsten TR. 2017. Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet.* 13:e1006402.

Martínez JP, Antúnez A, Araya H, Pertuzé R, Fuentes L, Lizana XC, Lutts S. 2014. Salt stress differently affects growth, water status and antioxidant enzyme activities in *Solanum lycopersicum* and its wild relative *Solanum chilense*. *Aust. J. Bot.* 62:359–368.

Masalia RR, Bewick AJ, Burke JM. 2017. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. *PLoS One* 12:e0182289.

Mboup M, Fischer I, Lainer H, Stephan W. 2012. Trans-species polymorphism and allele-specific expression in the *CBF* gene family of wild tomatoes. *Mol. Biol. Evol.* 29:3641–3652.

Mes PJ, Boches P, Myers JR, Durst R. 2008. Characterization of tomatoes expressing anthocyanin in the fruit. *J. Am. Soc. Hortic. Sci.* 133:262–269.

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* 49:D394–D403.

Molitor C, Kurowski TJ, Fidalgo de Almeida PM, Eerolla P, Spindlow DJ, Kashyap SP, Singh B, Prasanna HC, Thompson AJ, Mohareb FR. 2021. *De novo* genome assembly of *Solanum sitiens* reveals structural variation associated with drought and salinity tolerance.Valencia A, editor. *Bioinformatics* 37:1941–1945.

Mustafin ZS, Zamyatin VI, Konstantinov DK, Doroshkov AV, Lashin SA, Afonnikov DA. 2019. Phylostratigraphic analysis shows the earliest origination of the abiotic stress associated genes in *A. thaliana*. *Genes* 10:963.

Nakazato T, Warren DL, Moyle LC. 2010. Ecological and geographic modes of species divergence in wild tomatoes. *Am. J. Bot.* 97:680–693.

Nieri D, Di Donato A, Ercolano MR. 2017. Analysis of tomato meiotic recombination profile reveals preferential chromosome positions for NB-LRR genes. *Euphytica* 213:206.

Nosenko T, Böndel KB, Kumpfmüller G, Stephan W. 2016. Adaptation to low temperatures in the wild tomato species *Solanum chilense*. *Mol. Ecol.* 25:2853–2869.

Papakostas S, Vøllestad LA, Bruneaux M, Aykanat T, Vanoverbeke J, Ning M, Primmer CR, Leder EH. 2014. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nat. Commun.* 5:1–9.

Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare. Available from: https://f1000research.com/articles/9-304

Piasecka B, Lichocki P, Moretti S, Bergmann S, Robinson-Rechavi M. 2013. The hourglass and the early conservation models—co-existing patterns of developmental constraints in vertebrates. *PLoS Genet.* 9:e1003476.

Quint M, Drost H-G, Gabel A, Ullrich KK, Bönn M, Grosse I. 2012. A transcriptomic hourglass in plant embryogenesis. *Nature* 490:98–101.

Raduski AR, Igić B. 2021. Biosystematic studies on the status of *Solanum chilense*. *Am. J. Bot.* 108:520–537.

Reddy AR, Chaitanya KV, Vivekanandan M. 2004. Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J. Plant Physiol.* 161:1189–1202.

Rodrigues J, Inzé D, Nelissen H, Saibo NJ. 2019. Source–sink regulation in crops under water deficit. *Trends Plant Sci.* 24:652–663.

Roselius K, Stephan W, Städler T. 2005. The relationship of nucleotide polymorphism, recombination rate and selection in wild Tomato species. *Genetics* 171:753–763.

Sato MP, Makino T, Kawata M. 2016. Natural selection in a population of *Drosophila melanogaster* explained by changes in gene expression caused by sequence variation in core promoter regions. *BMC Evol. Biol.* 16:1–12.

Shen W, Ren H. 2021. TaxonKit: A practical and efficient NCBI taxonomy toolkit. *J. Genet. Genomics*.

Skirycz A, Inzé D. 2010. More from less: plant growth under limited water. *Curr. Opin. Biotechnol.* 21:197–203.

Spivakov M. 2014. Spurious transcription factor binding: non-functional or genetically redundant? *Bioessays* 36:798–806.

Stam R, Silva-Arias GA, Tellier A. 2019. Subsets of NLR genes show differential signatures of adaptation during colonization of new habitats. *New Phytol.* 224:367–379.

Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global discovery of conserved genetic modules. *science* 302:249–255.

Su G, Morris JH, Demchak B, Bader GD. 2014. Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinforma.* 47:8.13.1-8.13.24.

Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM. 2012. Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci.* 109:11878–11883.

Tapia G, Castro M, Gaete-Eastman C, Figueroa CR. 2022. Regulation of anthocyanin biosynthesis by drought and UV-B radiation in wild tomato (*Solanum peruvianum*) fruit. *Antioxidants* 11:1639.

Tapia G, Méndez J, Inostroza L. 2016. Different combinations of morpho-physiological traits are responsible for tolerance to drought in wild tomatoes *Solanum chilense* and *Solanum peruvianum*. *Plant Biol.* 18:406–416.

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

Tian F, Yang D-C, Meng Y-Q, Jin J, Gao G. 2020. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* 48:D1104–D1113.

Townsley BT, Covington MF, Ichihashi Y, Zumstein K, Sinha NR. 2015. BrAD-seq: Breath Adapter Directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. *Front. Plant Sci.* 6.

True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3:109–119.

Verelst W, Bertolini E, De Bodt S, Vandepoele K, Demeulenaere M, Pè ME, Inzé D. 2013. Molecular and physiological analysis of growth-limiting drought stress in *Brachypodium distachyon* leaves. *Mol. Plant* 6:311–322.

de Vries J, Archibald JM. 2018. Plant evolution: landmarks on the path to terrestrial life. *New Phytol.* 217:1428–1434.

de Vries J, Curtis BA, Gould SB, Archibald JM. 2018. Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proc. Natl. Acad. Sci.* 115:E3471–E3480.

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* 131:281–285.

Wagner GP, Pavlicev M, Cheverud JM. 2007. The road to modularity. *Nat. Rev. Genet.* 8:921–931.

Wang S, Li L, Li H, Sahu SK, Wang H, Xu Y, Xian W, Song B, Liang H, Cheng S, et al. 2020. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants* 6:95–106.

Wang Y, Wang X, Paterson AH. 2012. Genome and gene duplications and gene expression divergence: a view from plants. *Ann. N. Y. Acad. Sci.* 1256:1–14.

Wei K, Silva-Arias GA, Tellier A. 2022. Selective sweeps linked to colonization of novel habitats and climatic changes in a wild tomato species. *New Phytol.*:nph.18634.

Wilkinson S, Davies WJ. 2010. Drought, ozone, ABA and ethylene: new insights from cell to plant to community. *Plant Cell Environ.* 33:510–525.

Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25:2078–2079.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* 16:284–287.

Zarrineh P, Sánchez-Rodríguez A, Hosseinkhan N, Narimani Z, Marchal K, Masoudi-Nejad A. 2014. Genome-scale co-expression network comparison across *Escherichia coli* and *Salmonella enterica* serovar *Typhimurium* reveals significant conservation at the regulon level of local regulators despite their dissimilar lifestyles. *PLoS One* 9:e102871.

Zinkgraf M, Zhao S-T, Canning C, Gerttula S, Lu M-Z, Filkov V, Groover A. 2020. Evolutionary network genomics of wood formation in a phylogenetic survey of angiosperm forest trees. *New Phytol.* 228:1811–1823.

Figure legends

Figure 1. Exploratory analyses of RNA-seq differential expression patterns in 16 libraries of *Solanum chilense*. (A) PCA reveals stronger clustering associated with the experimental conditions. (B) Heatmap plot of sample correlation (Pearson's test) reveals exact drought specificity. RNAseq libraries abbreviations, CL-A to CL-D: leaves in control condition, CSA-E to CSA-H: shoot apex in control condition, DL-I to DL-L: leave in drought condition, DSA-M to DSA-P: shoot apex in drought condition. Color scale indicates correlation coefficients from high values in red to low values in white.

Figure 2. Identification of drought-response networks in *Solanum chilense*. (A) Differentially Expressed Genes (DEGs) identified from three comparison groups from left to right: 8 control versus 8 drought samples, 4 control leaves versus 4 drought leaves, 4 control shoot apices versus 4 drought shoot apices. Red indicates significantly upregulated genes, and green indicates significantly downregulated genes between control and drought samples using fold change higher than two ($P \leq 0.001$). (B) Venn diagram show 2,484 shared DEGs in three comparison groups. (C) The correlation between samples expression patterns for the eight modules. Color scale indicates correlation coefficients from high positive coefficient in red to high negative coefficient in green. No correlation is indicated in white.

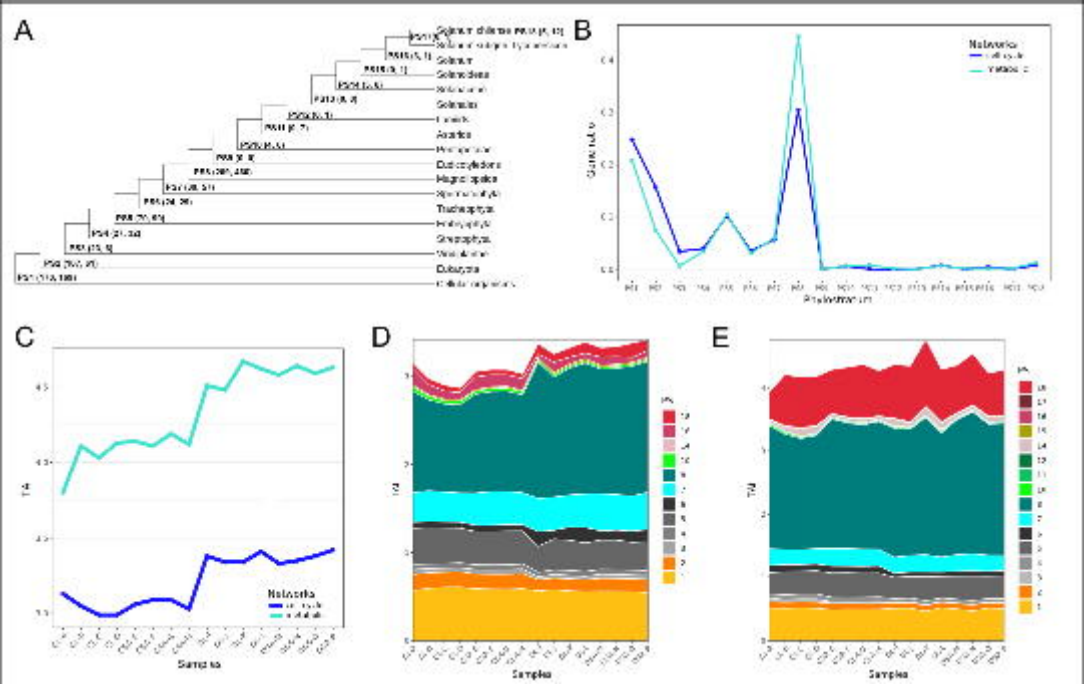
Figure 3. Gene ontology (GO) term enrichment in the cell-cycle and metabolic drought-response networks. (A) Top 20 terms of biological process. (B) Top 20 terms of cellular component.

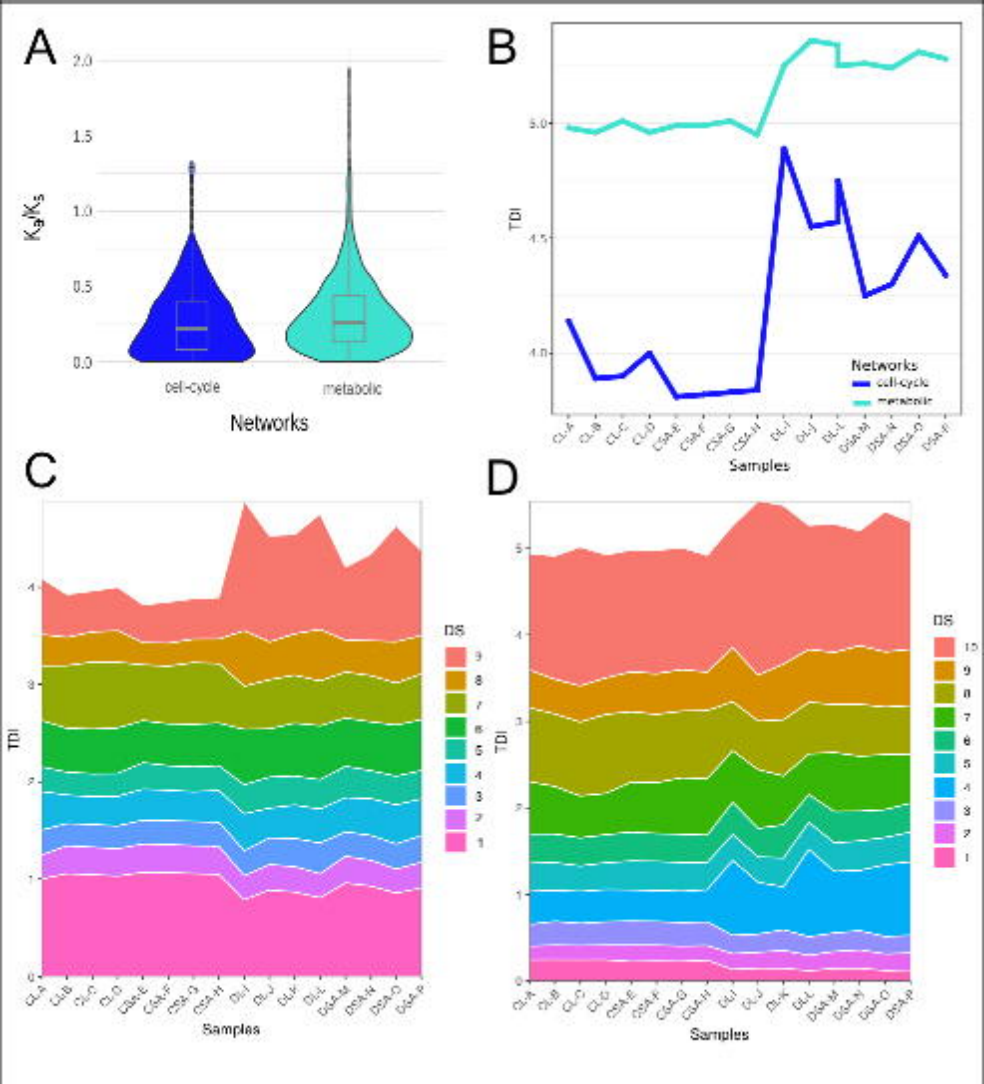
Figure 4. Transcriptome age index (TAI) profiles of cell-cycle and metabolic networks. (A) Phylostratigraphic map of two networks and phylogeny used in the search for the evolutionary origin of *Solanum chilense* genes. Numbers in parentheses denote the number of genes assigned to each phylostratum (PS) in cell-cycle and metabolic network, respectively. (B) Gene ratio in each PS for two networks. (C) TAI profiles of two networks across samples. (D) TAI contributions split according to different PS in cell-cycle network. (E) TAI contributions split according to different PS in metabolic network.

Figure 5. Transcriptome divergence index (TDI) profiles of cell-cycle and metabolic networks. (A) Distribution of Ka/Ks ratio of genes in two networks, respectively. (B) TDI profiles of two networks across samples. (C) TDI contributions split according to different DS in cell-cycle network. (D) TDI contributions split according to different DS in metabolic network.

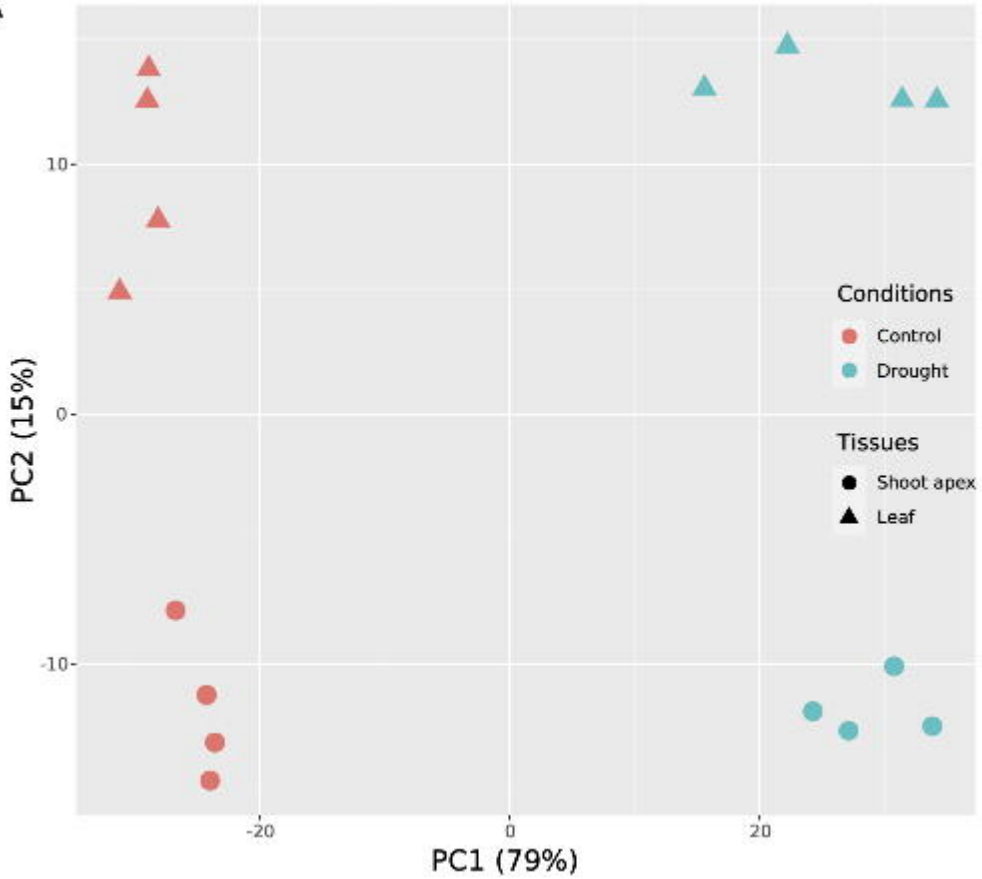
Figure 6. Drought-responsive genes under positive selection. (A) The candidate genes under positive selection were identified in our previous study (Wei et al. 2022). The bar plot shows that number of positive selection genes in the cell-cycle and metabolic networks, the pie charts denote positive selection genes of two networks in six populations. The size of pie represents number of genes (see also Table S6). (B) The connectivity of drought-responsive genes under positive selection in the two

networks and six populations. (C) The correlations between connectivity and age of drought-responsive genes under positive selection in the two networks and six populations. (D) The visualization of cell-cycle network. (E) The visualization of metabolic network. Red-to-orange colored dots denote candidate genes under positive selection identified in Wei *et al*. (2022). The red-to-orange scale denote the ages of selective sweeps. The location of the dots closer to the center of the networks indicates that the gene exhibits higher connectivity.


A
B
C
D
E
Sample
Sample
Sample


A
2.0
1.5
1.0
0.5
0.0
Ka/Ks
cell-cycle
metabolic
Networks
B
5.0
4.5
4.0
TDI
Networks
cell-cycle
metabolic
CLA
CLB
CLC
CLD
CSA-E
CSA-F
CSA-G
CSA-H
DLI
DLJ
DLK
DLL
DSA-M
DSA-N
DSA-O
DSA-P
Samples
C
4
3
2
1
0
TDI
DS
9
8
7
6
5
4
3
2
1
CLA
CLB
CLC
CLD
CSA-E
CSA-F
CSA-G
CSA-H
DLI
DLJ
DLK
DLL
DSA-M
DSA-N
DSA-O
DSA-P
Samples
D
5
4
3
2
1
0
TDI
DS
10
9
8
7
6
5
4
3
2
1
CLA
CLB
CLC
CLD
CSA-E
CSA-F
CSA-G
CSA-H
DLI
DLJ
DLK
DLL
DSA-M
DSA-N
DSA-O
DSA-P
Samples

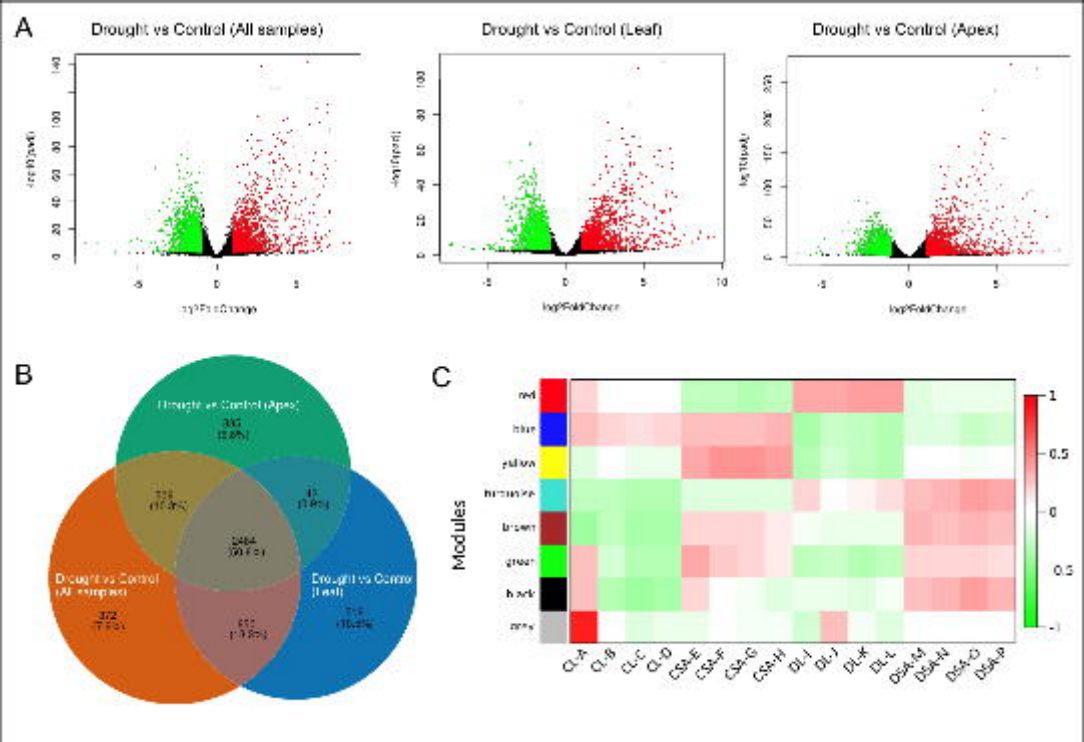
A

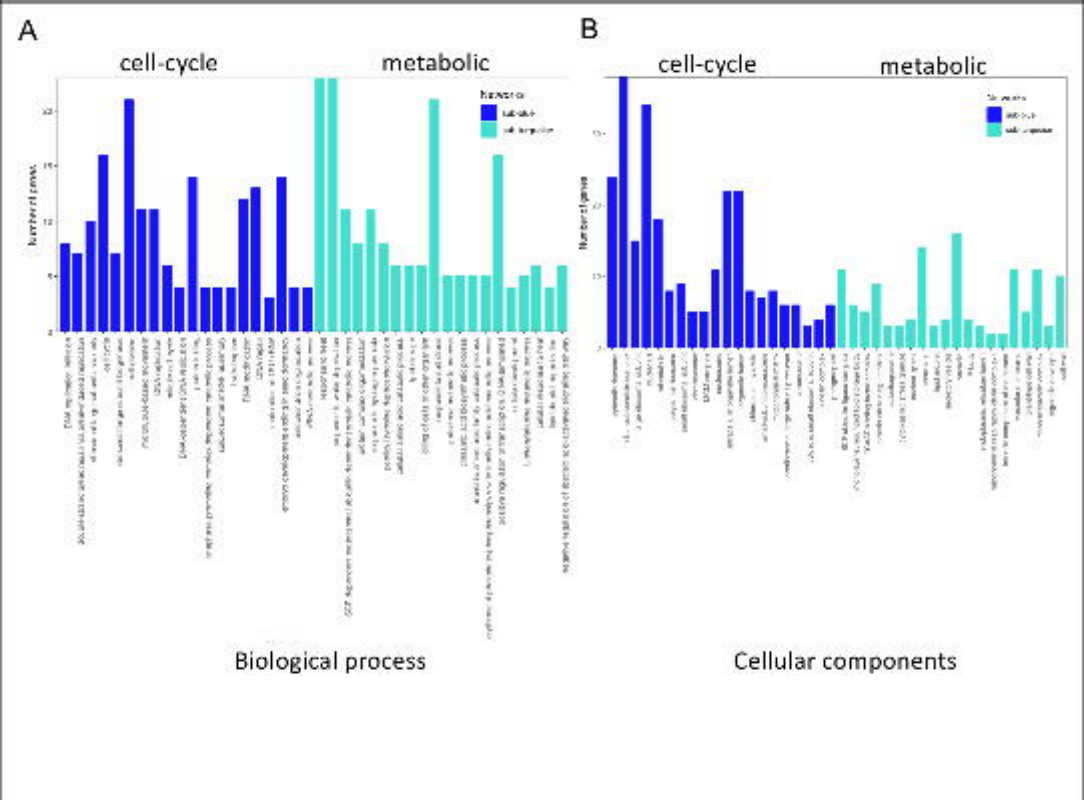


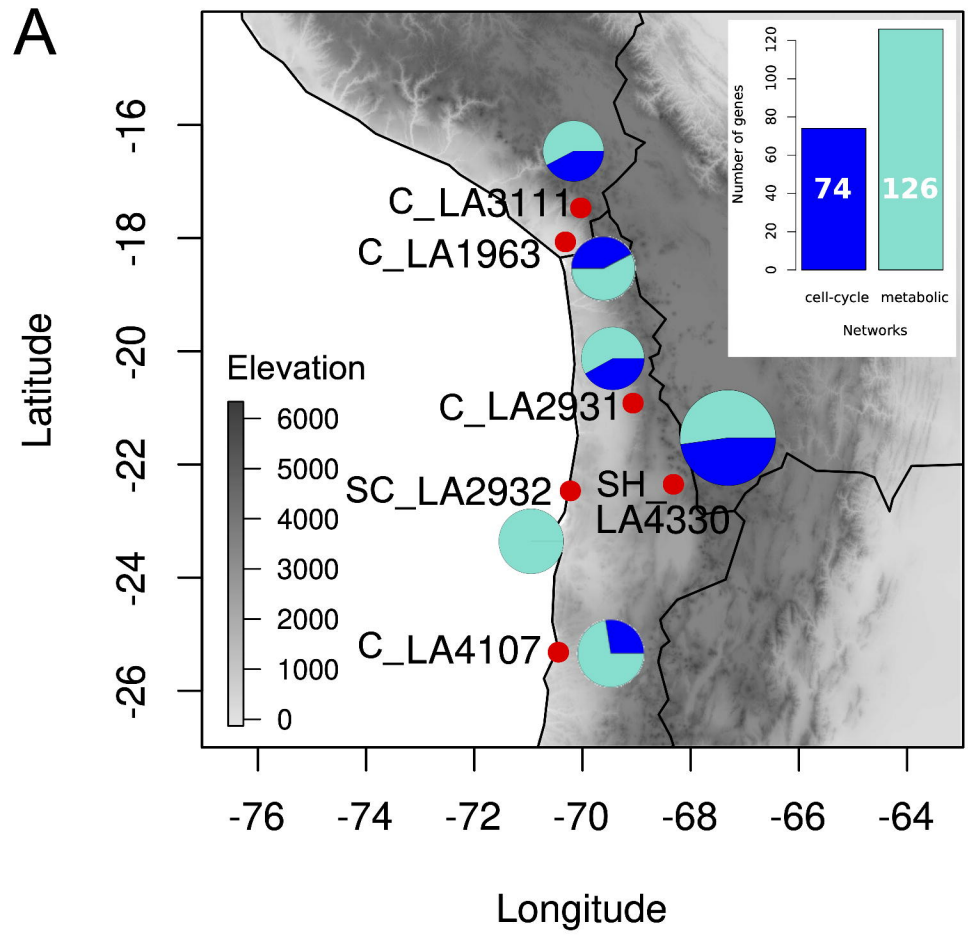
B

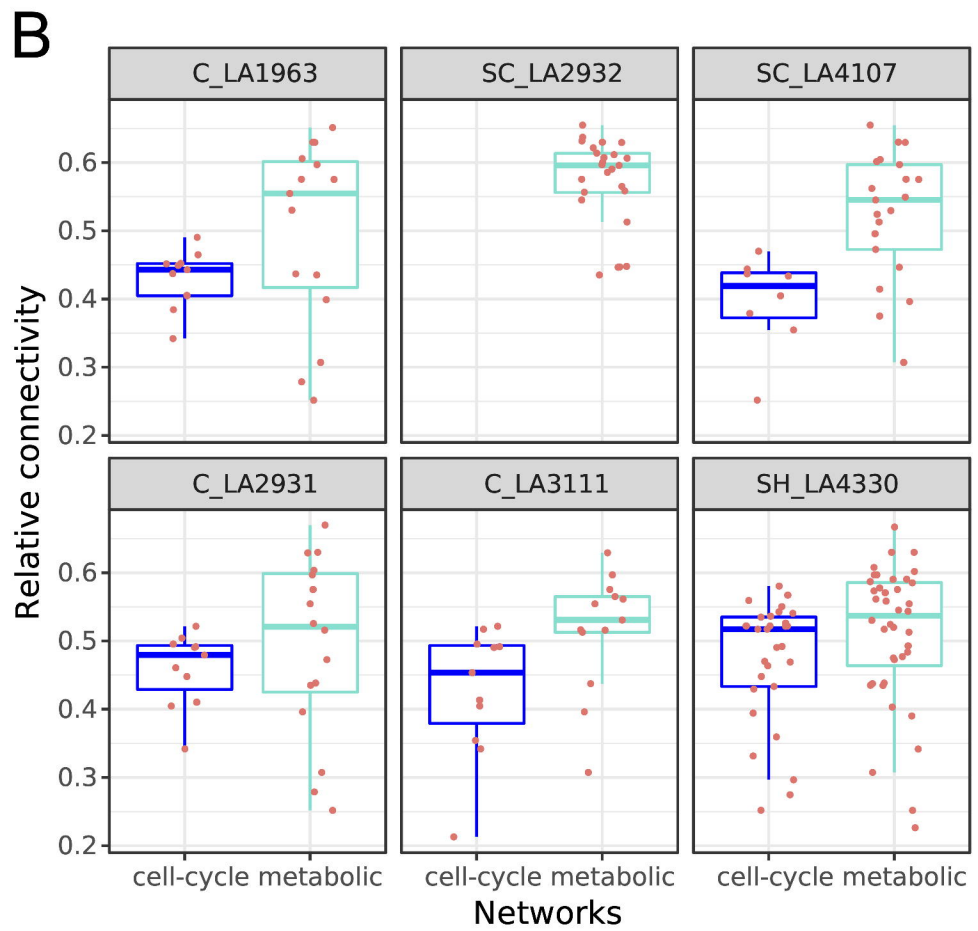
	CL-A	CL-B	CL-C	CL-D	CSA-E	CSA-F	CSA-G	CSA-H	DL-I	DL-J	DL-K	DL-L	DSA-M	DSA-N	DSA-O	DSA-P
CL-A	1.00	0.89	0.88	0.90	0.77	0.75	0.77	0.77	0.44	0.71	0.68	0.48	0.45	0.48	0.34	0.40
CL-B	0.89	1.00	0.99	0.99	0.88	0.87	0.89	0.89	0.50	0.79	0.78	0.55	0.53	0.56	0.40	0.47
CL-C	0.88	0.99	1.00	0.99	0.83	0.81	0.84	0.85	0.51	0.79	0.80	0.55	0.51	0.54	0.38	0.45
CL-D	0.90	0.99	0.99	1.00	0.84	0.81	0.84	0.85	0.49	0.78	0.78	0.54	0.49	0.52	0.37	0.44
CSA-E	0.77	0.88	0.83	0.84	1.00	0.99	0.98	0.98	0.47	0.71	0.68	0.48	0.63	0.65	0.48	0.54
CSA-F	0.75	0.87	0.81	0.81	0.99	1.00	0.99	0.98	0.46	0.69	0.67	0.47	0.64	0.66	0.49	0.55
CSA-G	0.77	0.89	0.84	0.84	0.98	0.99	1.00	0.99	0.47	0.71	0.68	0.48	0.64	0.66	0.49	0.55
CSA-H	0.77	0.89	0.85	0.85	0.98	0.98	0.99	1.00	0.45	0.68	0.66	0.46	0.60	0.62	0.45	0.51
DL-I	0.44	0.50	0.51	0.49	0.47	0.46	0.47	0.45	1.00	0.81	0.84	0.98	0.89	0.86	0.86	0.88
DL-J	0.71	0.79	0.79	0.78	0.71	0.69	0.71	0.68	0.81	1.00	0.96	0.83	0.83	0.84	0.77	0.79
DL-K	0.68	0.78	0.80	0.78	0.68	0.67	0.68	0.66	0.84	0.96	1.00	0.83	0.79	0.81	0.71	0.74
DL-L	0.48	0.55	0.55	0.54	0.48	0.47	0.48	0.46	0.98	0.83	0.83	1.00	0.87	0.85	0.86	0.90
DSA-M	0.45	0.53	0.51	0.49	0.63	0.64	0.64	0.60	0.89	0.83	0.79	0.87	1.00	0.98	0.96	0.96
DSA-N	0.48	0.56	0.54	0.52	0.65	0.66	0.66	0.62	0.86	0.84	0.81	0.85	0.98	1.00	0.95	0.96
DSA-O	0.34	0.40	0.38	0.37	0.48	0.49	0.49	0.45	0.86	0.77	0.71	0.86	0.96	0.95	1.00	0.98
DSA-P	0.40	0.47	0.45	0.44	0.54	0.55	0.55	0.51	0.88	0.79	0.74	0.90	0.96	0.96	0.98	1.00

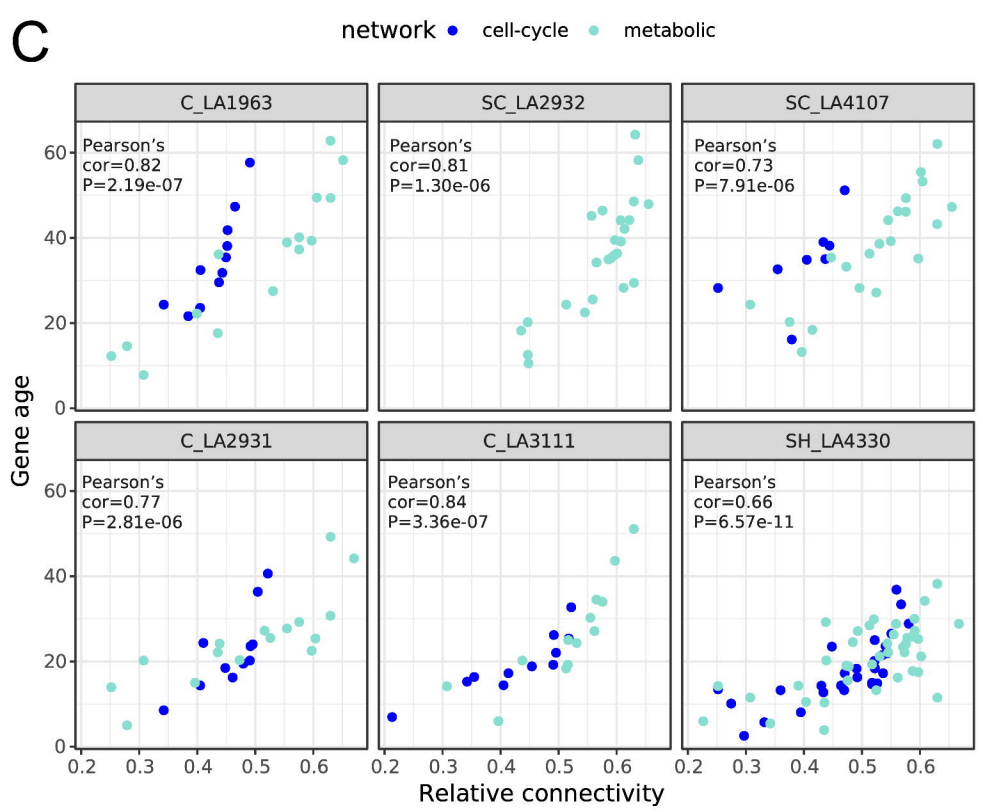
1
0.9
0.8
0.7
0.6
0.5
0.4


A
Drought vs Control (All samples)
Drought vs Control (Leaf)
Drought vs Control (Apex)
B
Drought vs Control (Apex)
Drought vs Control (All samples)
Drought vs Control (Leaf)
C
Modules
red
blue
yellow
turquoise
brown
green
black
grey
CL-A
CL-B
CL-C
CL-D
CSA-E
CSA-F
CSA-G
CSA-H
DL-I
DL-J
DL-K
DL-L
DSA-M
DSA-N
DSA-O
DSA-P
1
0.5
0
0.5
-1

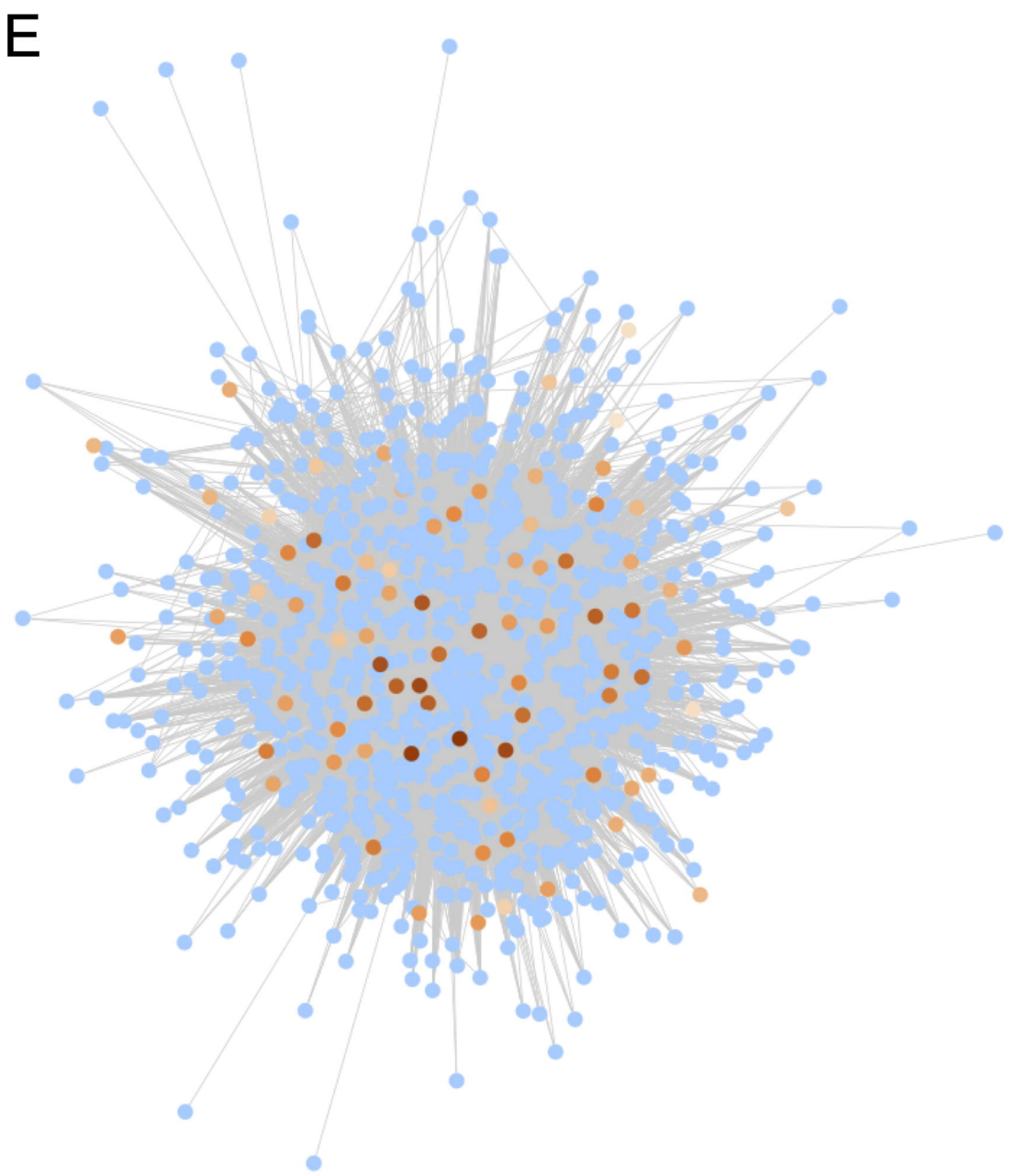

A
cell-cycle
metabolic
Number of genes
Biological process
B
cell-cycle
metabolic
Number of genes
Cellular components


A
Latitude
Longitude
Elevation
6000
5000
4000
3000
2000
1000
0
C_LA3111
C_LA1963
C_LA2931
SC_LA2932
SH_LA4330
C_LA4107
Number of genes
74
126
cell-cycle
metabolic
Networks


B
C_LA1963
SC_LA2932
SC_LA4107
C_LA2931
C_LA3111
SH_LA4330
Relative connectivity
cell-cycle metabolic
Networks


C
network cell-cycle metabolic
C_LA1963
Pearson's cor=0.82 P=2.19e-07
SC_LA2932
Pearson's cor=0.81 P=1.30e-06
SC_LA4107
Pearson's cor=0.73 P=7.91e-06
C_LA2931
Pearson's cor=0.77 P=2.81e-06
C_LA3111
Pearson's cor=0.84 P=3.36e-07
SH_LA4330
Pearson's cor=0.66 P=6.57e-11
Gene age
Relative connectivity

D
Sweep age (kya)
6
30
57


E