New Phytologist

Research

Bridging physiological and evolutionary time-scales in a gene regulatory network

Gwenaëlle Marchand[1,2], Vân Anh Huynh-Thu[3], Nolan C. Kane[4], Sandrine Arribat[5], Didier Varès[1,2], David Rengel[1,2], Sandrine Balzergue[5], Loren H. Rieseberg[6,7], Patrick Vincourt[1,2], Pierre Geurts[3], Matthieu Vignes[8] and Nicolas B. Langlade[1,2]

[1]INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, F-31326 Castanet-Tolosan, France; [2]CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, F-31326 Castanet-Tolosan, France; [3]Department of Electrical Engineering and Computer Science and GIGA-R, Systems and Modeling, University of Liège, Liège, Belgium; [4]Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO 80309, USA; [5]INRA, Unité de Recherche en Génomique Végétale (URGV), UMR1165 – Université d'Evry Val d'Essonne – ERL CNRS 8196, CP 5708, F-91057 Evry Cedex, France; [6]Department of Botany, University of British Columbia, Vancouver BC V6T 1Z4, Canada; [7]Department of Biology, Indiana University, Bloomington, IN 47405, USA; [8]INRA, Mathématiques et Informatique Appliquées (MIA), UPR875, F-31326 Castanet-Tolosan, France

Author for correspondence:
Nicolas B. Langlade
Tel: +33 5 61 28 54 58
Email: nicolas.langlade@toulouse.inra.fr

Received: *15 November 2013*
Accepted: *17 March 2014*

New Phytologist (2014)
doi: 10.1111/nph.12818

Key words: abscisic acid (ABA), CLC-A chloride channel protein, drought, F_{ST}, genetic differentiation, network inference, NITRATE TRANSPORTER 1.1 (NRT1.1).

Summary

- Gene regulatory networks (GRNs) govern phenotypic adaptations and reflect the trade-offs between physiological responses and evolutionary adaptation that act at different time-scales. To identify patterns of molecular function and genetic diversity in GRNs, we studied the drought response of the common sunflower, *Helianthus annuus*, and how the underlying GRN is related to its evolution.
- We examined the responses of 32 423 expressed sequences to drought and to abscisic acid (ABA) and selected 145 co-expressed transcripts. We characterized their regulatory relationships in nine kinetic studies based on different hormones. From this, we inferred a GRN by meta-analyses of a Gaussian graphical model and a random forest algorithm and studied the genetic differentiation among populations (F_{ST}) at nodes.
- We identified two main hubs in the network that transport nitrate in guard cells. This suggests that nitrate transport is a critical aspect of the sunflower physiological response to drought. We observed that differentiation of the network genes in elite sunflower cultivars is correlated with their position and connectivity.
- This systems biology approach combined molecular data at different time-scales and identified important physiological processes. At the evolutionary level, we propose that network topology could influence responses to human selection and possibly adaptation to dry environments.

Introduction

Phenotype is shaped during an organism's life by its physiological and developmental responses to environmental conditions and across generations through evolutionary genetic adjustments to new environments. On the time-scale of individual organisms, the phenotype can change rapidly as a consequence of gene regulatory networks (GRNs), which translate environmental and internal signals into physiological and developmental modifications. On an evolutionary time-scale, such phenotypic modifications are based on changes in the genes composing the network that may alter this network at the structural or functional level.

Relating phenotypic modifications occurring at physiological and evolutionary time-scales has been a major focus of evolutionary biologists for more than a century (Osborn, 1896) and Waddington (1942) as well as more recently Queitsch *et al.* (2002); Milo *et al.* (2007). Researchers have theorized (and later demonstrated) that physiological adaptation (e.g. via regulation of gene expression or biochemical characteristics) can be replaced by an evolutionary change that becomes constitutive and alleviates the fitness costs associated with plasticity. This paradigm can be revisited in the context of a gene network. While gene regulatory networks are products of evolution, similar to other biological objects, GRNs also shape and constrain the evolvability of phenotypic responses to the environment.

Systems biology approaches, such as GRN inference, provide a global view of the different pathways that respond to environmental variation. A GRN is a genetic network based on gene expression levels (Wilkins, 2005). It describes transcriptional interactions and dynamics in response to environmental stressors, and therefore the GRN is key to understanding how organisms such as plants adapt to their environment.

Responses to environmental signals are often mediated through hormones. For example, in plants, abscisic acid (ABA) is produced

during water stress in the vasculature and in the guard cells of the vegetative part of the plant (Boursiac *et al.*, 2013). Accordingly, the application of ABA induces the expression of genes involved in the response to dehydration and mimics drought stress. This interpretation has been confirmed by promoter analyses, which have demonstrated that these pathways share many targets (Shinozaki & Yamaguchi-Shinozaki, 1997). The signals of different hormones interact and are integrated to convey environmental signals through the plant (Wilkinson *et al.*, 2012), suggesting that hormones should share transcriptomic targets.

Drought stress is a major abiotic factor that drives dramatic phenotypic changes in plants, including *Helianthus*, in which drought stress appears to constrain the colonization of new environments in the arid regions of the southwestern USA (Seiler & Rieseberg, 1997). Therefore, the drought-stress GRN represents a tool for studying the interactions between organismal acclimation on the physiological time-scale and population adaptation on the evolutionary time-scale.

Several hormones mediate drought-stress responses; thus, the utilization of multiple hormonal treatments can elucidate the underlying GRN and highlight possible relationships between the genes involved. However, there are practical difficulties associated with the study of genetic networks. For example, the GRN identified could be biased toward interactions that have been previously detected in model species (Wilkins, 2005). To date, systems biology approaches, such as GRN inference, have been mostly restricted to model species, such as yeast (Dikicioglu *et al.*, 2011), *Drosophila* (Crombach *et al.*, 2012), or Arabidopsis (Ma *et al.*, 2007), and are typically performed under laboratory conditions. However, modeling dynamic biological processes requires time-series gene expression data that are relevant both to the biological process of interest and to the species targeted by the study. To understand genome function and evolutionary processes in an organism such as the sunflower (*Helianthus annuus*), it is important to infer the GRN for the gene sets that are actually involved in the responses to a given environmental stress and to avoid the pitfall of using nonadapted model species data.

In this study, we used inference methods on sunflower data complemented with knowledge from Arabidopsis. These methods were specifically designed for time-series gene expression data and allowed us to reconstruct a sunflower GRN. The inferred GRN provides us with a global view of the main physiological functions involved in the drought-stress responses occurring in the leaf, as well as their chronology.

On the evolutionary time-scale, studying the underlying GRN for responses to environmental stresses such as drought can help explain how plants evolved to become better suited to their environments. Knowledge of a gene's position in the GRN and its topological characteristics provides useful information about likely evolutionary constraints. For example, a highly connected gene is likely to be subject to many trade-offs, which would limit the accumulation of genetic diversity. Here, we identify correlations between network topology and genetic divergence between elite lines and landraces of sunflower and propose a mechanism to explain how sunflower genetic differentiation could be constrained in response to selective forces.

Materials and Methods

Plant material and growth conditions

Transcriptome interactions and dynamics were studied using the sunflower (*Helianthus annuus* L.) genotype XRQ. Plantlets were grown under hydroponic conditions in the previously described growth medium (Neumann *et al.*, 2000) in a growth chamber. After 14 d, the plantlets were treated by adding either mock solution (DMSO only in controls) or one of the following hormonal solutions: auxin (IAA); ethylene (ACC), gibberellic acid (GA3), salicylic acid (SA), methyl-jasmonate (MeJA), kinetin, ABA strigolactone (Stri) or Brassinol (Bras). Details of the hormonal solutions are provided in Supporting Information Methods S1. The first pair of leaves of each plant was harvested at 0 (just before treatment), 1, 3, 6, 9, 24 or 48h after treatment, immediately frozen in liquid nitrogen, and stored at −80°C. The whole procedure was repeated three times for ACC, Bras, GA3, IAA, kinetin, SA, and Stri and four times for ABA and MeJA.

Gene selection

To identify genes that are likely to play a role in the drought GRN, a global transcriptomic approach was employed using an Affymetrix chip (Affymetrix, Santa Clara, CA, USA) containing 32 423 probe sets corresponding to sequences expressed in *H. annuus* (Rengel *et al.*, 2012). Three different global transcriptomic data sets were analyzed and used to select genes. We selected genes that responded to at least two of the following conditions: drought stress under field conditions; drought stress under glasshouse conditions; and 10 μM ABA application under hydroponic conditions.

The microarray data and analyses of the field and glasshouse conditions were previously reported by Rengel *et al.* (2012). Under field conditions, plants of the Melody genotype were harvested at the post-flowering stage at a stress intensity level of 0.63 and 0.22 (ratio between evapotranspiration and maximal evapotranspiration) for irrigated and nonirrigated plants, respectively. Under glasshouse conditions, we recorded data from Melody pre-flowering plants at a fraction of transpirable soil water (FTSW) of 0.83 and 0.03 for the irrigated and nonirrigated plants, respectively.

The global transcriptomic data for the application of 10 μM ABA are new results and were obtained using the 6-h treatment with ABA in the hydroponic experiment on the genotype XRQ (CATdb: AFFY_ABA_Sunflower or GEO accession: GSE22519). RNA quality verification, cDNA synthesis, and chip hybridization and washing were all performed using the Affymetrix platform at the INRA-URGV (Institut National de Recherche Agronomique – Unité de Recherches en Génomiques Végétales) in Evry, France, following the protocol described in Rengel *et al.* (2012). To identify the sunflower transcripts that were differentially regulated by ABA under our hydroponic conditions, the Affymetrix data were treated as previously described in Bazin *et al.* (2011).

This list was extended to 181 genes with genes known to respond to the application of ABA or other hormones (literature (Pastori & Foyer, 2002; Wang *et al.*, 2003; Bray, 2004; Kawaguchi *et al.*, 2004; Boudsocq & Lauriere, 2005; Li *et al.*, 2006; Rook *et al.*, 2006; Valliyodan & Nguyen, 2006; Seki *et al.*, 2007; Shinozaki & Yamaguchi-Shinozaki, 2007; Wasilewska *et al.*, 2008; Miller *et al.*, 2009; Sirichandra *et al.*, 2009; Hirayama & Shinozaki, 2010; Umezawa *et al.*, 2010) or gene ontology (GO) analysis).

Molecular analysis

The extraction of total RNA and cDNA synthesis were performed as described in Rengel *et al.* (2012). The expression levels of the 181 selected genes were analyzed in all samples by q-RT-PCR using the BioMark system (Fluidigm Corporation, San Francisco, CA, USA) as previously described (Spurgeon *et al.*, 2008). The q-RT-PCR results were analyzed following the 2^{ddC_t} method (Livak & Schmittgen, 2001). Gene expression levels were normalized to the mean of previously validated reference genes (Rengel *et al.*, 2012) and to the corresponding control sample with the mock treatment. A detailed description of the calculation of expression levels is provided in Methods S1.

Genetic differentiation among populations

Genetic polymorphisms of drought GRN genes were characterized in five different *Helianthus* populations, as described in a previous study (Renaut *et al.*, 2013): *Helianthus argophyllus* ($N = 28$), *Helianthus petiolaris* ($N = 25$), *H. annuus* elite lines ($N = 9$), *H. annuus* landrace lines ($N = 11$), and wild *H. annuus* ($N = 11$). Briefly, transcript sequences were obtained from young leaf tissues with two RNAseq technologies (Roche 454 FLX (454 Life Sciences, Roche, Brandford, CT, USA) and GAII Illumina (Illumina, San Diego, CA, USA) pair-end sequencing 2×100 bp). The transcript sequences were then aligned to the reference transcriptome using the Burros Wheeler Aligner (Li & Durbin, 2009). Single nucleotide polymorphisms (SNPs) were called using the program SAMTOOLS (Li *et al.*, 2009) with a minimum with Phred scaled genotype likelihoods of 30, corresponding to a genotyping accuracy of at least 99.9%. The population genetics statistic F_{ST} was calculated between these populations for 89 of the 181 candidate genes using the R package HIERSTAT (Goudet, 2005). F_{ST} is a widely used measure of genetic differentiation among populations.

GRN reconstruction

Missing values of gene expression (expressed as $\Delta\Delta C_t$) at time $t = 0$ were imputed as values of 1. Other missing values (< 1% of the values) were imputed with the R package IMPUTE by 10-nearest neighboring genes (Troyanskaya *et al.*, 2001).

After log transformation of the data, we performed an arithmetic mean over replicates to obtain a robust $\Delta\Delta C_t$ expression value for each gene under each condition (time × treatment). We obtained nine data sets corresponding to the nine hormonal treatments and containing expression values for 145 genes with robust expression data at seven different time-points. From these nine data sets, we inferred 10 GRNs: one GRN from each hormonal treatment and a global GRN taking into account all treatments. Two complementary inference methods were used to achieve GRN predictions. The first method represents an extension of GENIE3 (Huynh-Thu *et al.,* 2010) and was based on the random forest method (RF; Breiman, 2001). In summary, each gene expression at time $t + 1$ was successively considered as a target, and the method sought regulators of that gene via their expression at time t. Several regulator inclusion steps were successively performed: according to a variance reduction criterion in a regression tree framework, each step resulted in the inclusion in the model of the best regulator. The process was repeated on a randomized ensemble of trees, which made up the so-called random forest. This method allowed us to derive a ranking of the importance of all regulator expressions for the target by averaging the scores over all the trees of the random forest. The randomized subset of regulators allowed us to avoid the local minima of the global score, and the random subsample of the data used for each tree avoided over-fitting of the data and hence permitted more robust estimates. We tested on simulated data whether including auto-loops in the model improved the performance. Results are presented in Methods S1 and they show that no gain was obtained with such a modified version of our RF algorithm. Compared with previously developed tree ensemble methods, our method is novel because our modeling explicitly accounted for the dynamical and multi-condition aspects of the data.

The second method used a Gaussian graphical modeling (GGM) approach. In the GGM paradigm, an edge was inferred when a significant partial correlation was detected between the expressions profiles of two genes. Namely, the partial correlation between two genes is the correlation between the residuals of the expressions of these two genes after accounting for all other gene expressions patterns. A unique aspect of our approach is the combination of a temporal approach with a multiple graph structure inference scheme. The dynamic nature of the data allowed us to obtain directed edges between two genes (i.e. changes in the expression of gene p induced changes in the expression of gene q and not the converse). In addition, the multiple graph framework drove the inference of condition-specific networks. However, each of these hormonal networks took into account information from the others and therefore accounted for a coupled functioning of the biological mechanisms that they encoded. The details of the RF and GGM approaches are provided in Methods S1. For each of the ten GRNs, we selected only edges confirmed by both methods. The union of the nine hormonal consensus networks and the global consensus network formed a final unified network with hormone-specific edges and global edges.

Topological parameters

The topology of a GRN depicts the relative positions of the genes in the network and their importance in the structure of the network. The topological parameters for each node therefore represent quantitative measures of gene connectivity and network

position; these parameters are calculated from the oriented edges that connect one gene with another. The edge count, the indegree and the outdegree are three correlated parameters indicating the total number of edges (in and out) and the number of outgoing and ingoing edges, respectively. The average shortest path length of a node p is the average length of the shortest path between p and any other node. The closeness centrality is the reciprocal of the average shortest path length. The eccentricity is the maximum noninfinite length of the shortest path between p and another node in the network. As the network is directed, if p is a node without outgoing edges, the values of the average shortest path length, the closeness centrality, and the eccentricity could not be calculated. The betweenness centrality of a node p is the number of shortest paths from a node q to a node r (differents from p) divided by the number of shortest paths from q to r that pass through p. It reflects the amount of control that the node p exerts over the interactions of other nodes in the network. The stress centrality of a node p is the number of shortest paths passing through p. Finally, the neighborhood connectivity of a node p is the average connectivity of all neighbors of p. These different metrics were calculated for all genes with the NetworkAnalyzer plugin for Cytoscape (Assenov *et al.*, 2008).

Correlation between topological parameters and genetic differentiation

First, we performed a principal component analysis (PCA) on the topological parameters of the GRN to study the dependence of those parameters, with the function *princomp*. This allowed us to identify the components explaining the most variability of the parameters. From these PCA results, we selected the most representative topological parameters in order to avoid redundancy. The F_{ST} values were grouped into five subsets, each of them expressing the F_{ST} between one *Helianthus* population (wild *H. annuus*, landraces, elite, *H. argophyllus* and *H. petiolaris*) and the other populations. We performed a canonical correlation analysis (R function *cancor*) in order to identify the canonical correlations between the selected topological parameters on one side and each F_{ST} subset on the other side. We tested their significance with the test of Wilks as provided by the function *p.perm* of the R package CCP with 10 000 permutations.

Results

Gene selection to infer the drought GRN

Gene identification using a global transcriptomic approach To identify genes that play a role in the drought GRN, a global transcriptomic approach was employed using an Affymetrix chip containing 32 423 probe sets, which corresponded to sequences expressed in *H. annuus*. The differential analysis identified 337 genes that responded to drought stress under field conditions and 447 genes that responded to drought stress under glasshouse conditions (Rengel *et al.*, 2012). Because ABA is the major plant hormone involved in the drought-stress response, we also identified genes displaying differential expression 6 h after ABA treatment at the plantlet stage under hydroponic conditions, using a similar global transcriptomic analysis. A total of 463 sunflower transcripts were found to be differentially expressed after ABA application (Table S1). The 463 ABA-regulated sunflower genes were validated by comparison with the expression of 226 homologues in Arabidopsis based on expression data from the Bio-Array Resource database or in projects from the AtGenExpress Consortium retrieved from the website http://www.weigelworld.org/resources/microarray/AtGenExpress/AtGe_Abiostress_gcRMA.zip. The authors employed a kinetic analysis of three time-points to assess the transcriptomic response to abiotic stresses such as cold, osmotic, salt, drought or heat stress in leaves using the Arabidopsis Affymetrix ATH1 microarray. This study was of particular interest because its kinetic approach imparts greater statistical power and avoids the issue of differences in kinetic parameters between sunflower and Arabidopsis. The Arabidopsis homologs of the sunflower genes in this study are all BLAST reciprocal best hits between *H. annuus* and Arabidopsis ESTs. The covariance analysis (ANCOVA) showed that the expression modulation by abiotic stresses over time of 27% of these Arabidopsis homologs (60 genes) exhibited a treatment effect or a treatment × time interaction effect. This proportion of Arabidopsis genes homologous to *H. annuus* genes responding to ABA corresponds to a significant enrichment in Arabidopsis genes responding to abiotic stresses (hypergeometric test giving $P = 1.10^{-4}$). The ANCOVA analysis, hypergeometric test and results are described in detail in Methods S1 and Table S2, respectively. This finding confirms that, at the transcriptomic level, ABA regulation and its role in abiotic stress responses are globally conserved between Arabidopsis and *H. annuus*, as has been documented in many plants; this conservation has occurred even though sunflowers are a very distantly related lineage separated by > 90 million yr of evolution (Chinnusamy *et al.*, 2004).

These three lists contain gene groups that respond to two drought-stress intensities and ABA application (mimicking a third drought-stress condition) at different developmental stages. Together, they provide complementary views of the drought-regulated genes in sunflower.

For inclusion in the GRN for drought stress, we stipulated that the genes must respond to at least two of the following conditions: drought stress under field conditions; drought stress under controlled glasshouse conditions; and/or ABA under hydroponic conditions (Fig. 1). As expected from the large variability of the biological material used to select the genes, the selected intersection was robust and should comprise the genes composing the core GRN for drought stress.

In addition to these groups of genes, we selected 56 genes that are known from the literature or GO analysis to be regulated in response to ABA or one of the other main plant hormones used for the treatment in our hydroponic experiment.

In all, 181 genes were selected (see complete list of sunflower transcripts, Arabidopsis homologs and annotations in Table S3).

Dissection of transcriptional regulation in the drought GRN by application of hormonal treatments The GRN of these drought-regulated genes was reconstructed from their expression

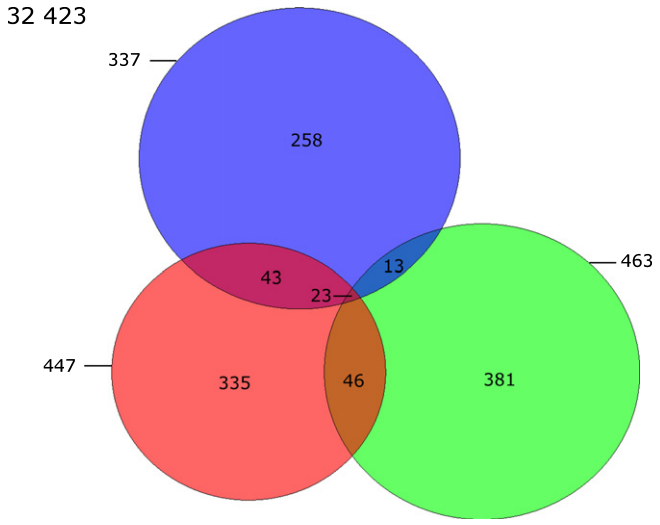

Fig. 1 Selection of genes likely to be involved in the drought gene regulatory network (GRN). Genes that responded to drought stress under field conditions, drought stress under glasshouse conditions, and abscisic acid (ABA) application under hydroponic conditions are indicated in blue, red, and green, respectively. The genes that were responsive under at least two of the different conditions were selected as part of the inferred GRN for drought-stress responses.

levels measured by q-RT-PCR. To perturb the network and identify regulatory relationships, leaf samples were harvested at seven different times after hormone treatment from hydroponically grown plants. A total of nine different hormones representing the main plant hormone groups were used. From the 181 selected candidate genes, we retained 145 robust genes based on technical filtering (efficiency equal to one, no missing data or only imputable missing data). The expression levels (expressed as $\Delta\Delta C_t$ in reference to five control genes and the mock control) before and after imputation of missing data are shown in Tables S4 and S5, respectively.

Inference of the drought GRN using the GGM and RF methods

Inferences of a global GRN and nine hormonal GRNs lead to the identification of a robust unified drought GRN To identify the final regulatory network between the 145 genes shown to be co-expressed during drought stress, we studied their regulation after several hormonal applications. This strategy was chosen because the environmental signal is transduced by different hormones whose regulatory pathways are very connected. The application of different hormones can reveal hormone-specific and global regulatory connections. Because we selected genes shown to respond to drought, the revealed regulatory connections are probably involved in drought-stress responses. We generated nine data sets corresponding to the nine hormonal treatments and containing expression values for the 145 robust genes at seven different time-points. From these nine data sets, we established 10 GRNs: one GRN from each hormonal treatment and one global GRN, which represents a consensus array of all hormonal treatments. The GRNs were inferred using two different inference methods: GGM and RF. These two approaches produce complementary predictions (Allouche *et al.*, 2013), and merging their results was shown to yield more reliable predictions than predictions obtained by any single method (Marbach *et al.*, 2012).

With the GGM method, we obtained between 112 and 158 edges for each hormonal network and a global network with 95 edges (Fig. 2).

With the RF method, the number of edges for each hormonal network was very different and varied from 11 to 174 edges. The global GRN with the RF inference was composed of 242 edges (Fig. 2).

Given the diversity in the inferred edges, we employed a very stringent approach to retain the core, most robust GRN. First, we discarded the results of SA treatment because the RF method inferred 629 edges. This number was far higher than that for the other hormones (49, 115, 38, 36, 94, 134, 147 and 16 when including SA). We chose not to take into account the SA edges in the final GRN to avoid an over-representation (> 25%) of specific edges for this hormone instead of drought edges. Secondly, for each GRN (hormonal or global), we considered an edge to be robust if it was selected by both the GGM and RF methods. This is a conservative approach that leads to high-quality edges; we chose to focus on a network with very reliable edges at the expense of potentially missing some weaker associations that might be relevant. This trade-off was confirmed in very different scenarios based on both simulated and real data sets (Vignes *et al.*, 2011; Marbach *et al.*, 2012). We validated both our models using simulated data that had the specific features of the data being studied (see Methods S1). Note that the numbers of robust edges were very different depending on the focal GRN. The final unified network, hereafter called the drought GRN, was formed by the union of all these robust edges (Fig. 2) and comprised 69 connected nodes, representing the genes linked by 79 unique edges. Among the 69 genes, 49 were differentially expressed in one of the three global transcriptomic experiments using the *Helianthus* Affymetrix chip, and only 20 came from the literature or GO analyses using BLAST reciprocal best hits to infer homology. Fig. S1 summarizes the origins of the 69 final genes of the network.

The number of shared edges between the hormonal GRNs varied from 0 to 18 (Fig. S2, Table S6). The ethylene, cytokinin and auxin networks shared the largest number of edges, whereas the ABA, brassinosteroid, and strigolactone networks had no edges in common with the other hormonal networks.

Comparison of the drought GRN to Arabidopsis data and prior knowledge of biological networks We compared our sunflower drought GRN to the model plant Arabidopsis using expression data from the AtGenExpress Consortium (Goda *et al.*, 2008) (GEO accession: GSE39384 from AtGenExpress Consortium). This Arabidopsis data set was similar to the *Helianthus* data and includes seven hormonal treatments but is limited to only three time-points. As a consequence of this difference in the sampling frequency, we were unable to define a network from these data using the inference methods described in the section above. Therefore, we searched for gene expression correlations that were

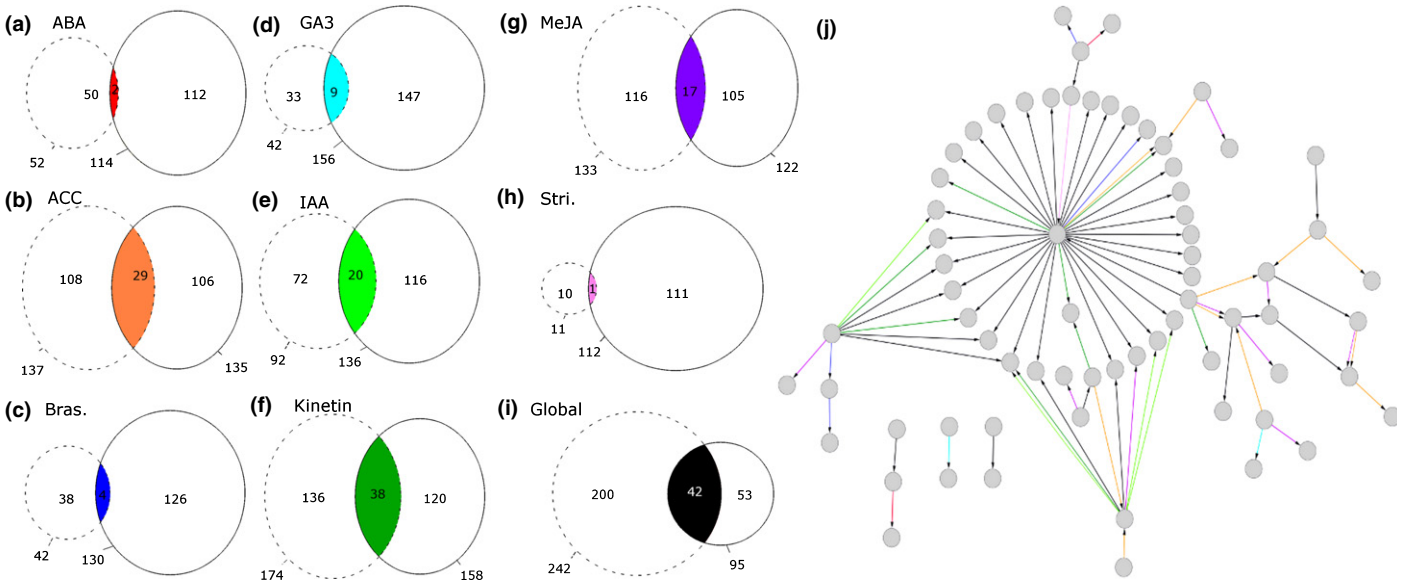

Fig. 2 Drought gene regulatory network (GRN) and selection of its edges. (a–i) The Venn diagrams for each hormonal GRN and global GRN represent the edges selected by the random forest (RF) method (dotted line) and the Gaussian graphical modeling (GGM) method (solid line). (a) Abscisic acid (ABA); (b) ethylene (ACC); (c) brassinosteroid (Bras.); (d) gibberellins (GA_3); (e) auxin (IAA); (f) kinetin; (g) methyl-jasmonate (MeJA); (h) strigolactone (Stri.); (i) global; (j) unified drought GRN representation. Gray circles represent the genes. Arrows represent the relationships between two genes (oriented edges), and their color represents the hormonal treatment that led to their identification: Red, ABA; orange, ethylene; dark blue, brassinosteroid; light blue, gibberellin; light green, IAA; dark green, kinetin; violet, MeJA; pink, strigolactone; black, global or non-hormone-specific edges.

consistent (or inconsistent) with the sunflower data. Among the 116 Arabidopsis genes that were homologous to the 145 sunflower genes that were initially used to develop the consensus drought GRN, significant correlations between gene pairs were more frequent for pairs corresponding to the network edges, according to an exact hypergeometric test ($P = 0.005$). The correlation analysis and hypergeometric test are described in Methods S1. This result demonstrated that the gene expression correlations identified from the Arabidopsis data were similar to the correlations identified in our sunflower drought GRN.

The topology of the drought GRN is consistent with what is known about biological networks. The degree distribution of the sunflower drought GRN followed a power law $y = 20.57x^{-1.98}$ with an R^2 of 0.72 (Fig. S3). This means that a few nodes had many connections and that the majority of the nodes had few edges, a finding that is a typical feature of the scale-free topology of biological networks (Barabasi & Oltvai, 2004).

Node connectivity defines different gene classes

Identification of two hubs sharing common targets The average value for the connectivity of a node (i.e. the number of outgoing or ingoing edges connecting a node to the others) in the inferred drought GRN was 2.3. However, we identified nodes with important connectivity; in particular, two nodes had the highest number of outgoing edges: eight and 32 (with a connectivity of nine and 32, respectively). These two genes were identified as important hubs in the inferred GRN. In addition, these genes shared seven common targets, while no common sources (i.e. a gene q that targets the studied gene p) between these genes were identified.

Relationship between connectivity and gene function Gene ontology annotations of the Arabidopsis genes homologous to the 69 *Helianthus* genes connected in the unified drought GRN were retrieved from TAIR (The Arabidopsis Information Resource) based on protein homology using the sunflower transcriptome web portal (www.heliagene.org/HaT13l). We observed that genes in the GO metabolism category accounted for the majority of the genes with low connectivity values: 40%, 80% and 60% of the genes with a connectivity of one, two and three, respectively; however, there was no significant enrichment using a hypergeometric test ($P = 0.190$). More interestingly, genes annotated as transcription factors and as having DNA-binding properties exhibited medium connectivity (i.e. four to five edges; $P = 0.002$), with the exception of one gene that had a single edge, possibly because its targets were filtered out during our analysis. Finally, the most highly connected genes were anion transporters. While the GO transporter included 20–30% of the genes with low connectivity, it also contained all the genes with high connectivity, including both hubs, which had nine and 32 edges (Fig. 3). The test showed that, despite the very low number of highly connected genes, this trend was significant ($P = 0.059$).

Canonical correlations between the topological parameters of the drought GRN and genetic differentiation statistics

To examine how the drought GRN might be related to the evolution of wild and domesticated sunflower populations, we looked for canonical correlations between nonredundant network topology parameters and the genetic differentiation statistics of the drought GRN nodes or genes. The topological parameters for each node represent quantitative measures of the gene position

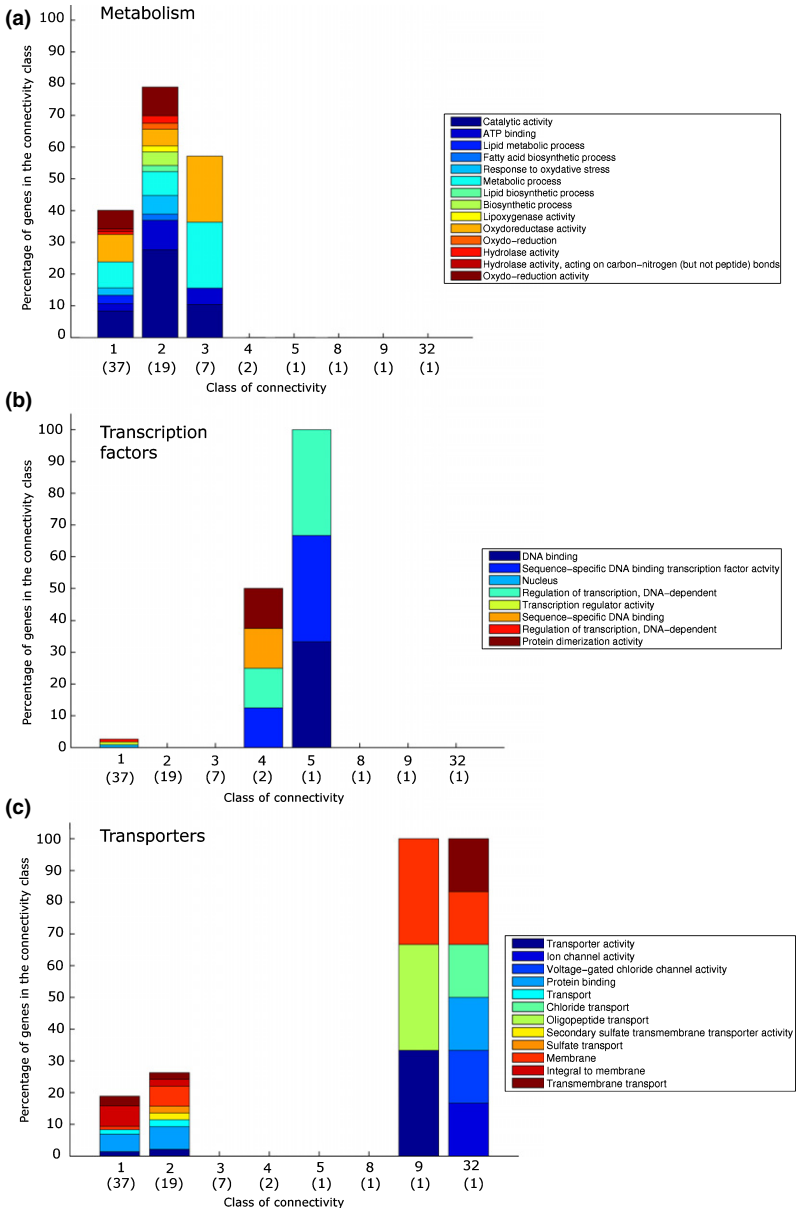

Fig. 3 Percentage of gene ontology (GO) terms for each connectivity class of genes in the drought gene regulatory network (GRN), where the GOs are represented by different colors. (a) Metabolism; (b) transcription factor or DNA binding; (c) transporters. The number of genes in each connectivity class is indicated in brackets. Note that a unique gene belongs to the connectivity class of 8. This gene does not belong to any of the three main classes of GO represented here (metabolism, transcription factor and transporters).

and relationships to others in the network. They are calculated from the number of oriented edges that connect one gene with another and are not independent by construction. In our GRN, edges are oriented; thus, we only considered genes with outgoing edges to compare the predictive values of the topological parameters. In addition, we were able to calculate F_{ST} for 15 of these genes among five populations of *Helianthus*: wild *H. annuus*, landrace lines of *H. annuus,* elite lines of *H. annuus*, *H. petiolaris* and *H. argophyllus*.

In a first step we used results from the PCA (cf Fig. S4 and Table S7) with topological parameters to reduce dimensionality and to obtain independent variables. The first and second components explained 67% of the variance. Regarding their loadings on the first two principal components, we selected ASPL and EdgeCount (cf Table S7). Genetic differentiation was analyzed using five distinct F_{ST} subsets, each of them expressing the F_{ST} between one *Helianthus* population and the other populations. Canonical correlation analysis (Tables 1, S8) between each of these five F_{ST} subsets on one side, and the two topological variables selected on the other side allowed us to detect significant canonical correlations only for the elite F_{ST} subset (Wilks's test, $P = 2.00 \times 10^{-3}$) and for the landrace F_{ST} subset ($P = 1.00 \times 10^{-4}$). As the intersection between these two subsets was F_{ST} between elite and landrace, this suggests that this variable in particular is correlated to the topological properties of the GRN. This was confirmed by the comparison of the canonical correlation analyses including only the F_{ST} value between landraces and elite lines (Wilks's test, $P = 1.90 \times 10^{-3}$) or the F_{ST} value between landraces and wild (Wilks's test, $P = 0.26$). More specifically, we found a significant Pearson correlation between the F_{ST} value between landraces and elite lines and ASPL ($R = 0.74$, $P = 0.003$).

Table 1 Coefficients of canonical correlations between, on the one hand, topological parameter values of the drought gene regulatory network (GRN) nodes and, on the other hand, their genetic differentiation measured as F_{ST} and grouped into five subsets

	Rho correlation coefficient 1	Rho correlation coefficient 2
F_{ST} subset of *H. argophyllus*	0.672 (*P*-value = 0.299)	0.524 (*P*-value: ns)
F_{ST} subset of *H. petiolaris*	0.493 (*P*-value = 0.818)	0.369 (*P*-value: ns)
F_{ST} subset of *H. annuus* wild	0.728 (*P*-value = 0.362)	0.292 (*P*-value: ns)
F_{ST} subset of *H. annuus* landraces	0.976 (*P*-value = 1×10^{-4})	0.299 (*P*-value: ns)
F_{ST} subset of *H. annuus* elite lines	0.946 (*P*-value = 0.002)	0.280 (*P*-value: ns)

Each subset of F_{ST} compares genetic differentiation of one *Helianthus* population to the four other populations of *Helianthus*. Correlations were tested for significance with Wilks's test with the function *p.perm* of the R software computing 10 000 permutations. ns, nonsignificant.

Discussion

In this study, we reconstructed a GRN based on gene expression that represents the transcriptional regulations that occur within a plant organ in response to environmental cues. As such, this drought GRN is not based on physical interactions between gene products and promoters and thus is not a molecular cell biology model. Instead, this GRN provides a more physiological view based on transcriptional events involved in drought-stress responses, similarly to the study of Hannah *et al.* (2006) on freezing tolerance in Arabidopsis. In addition, as a consequence of the temporal approach, the network edges are oriented and can be interpreted as dependent relationships. Together, these characteristics produce a network based on molecular regulations that also integrates physiological processes with their chronology at the organ level. This provides a representation of plant physiological responses to dry conditions and therefore of the fitness in such an environment.

Network inference highlights the importance of nitrate transport in guard cells

Drought GRN hubs are nitrate transporters and drive transcriptional regulation In the inferred network, two genes had many outgoing connections compared with other genes and could therefore be considered hubs. The first hub (HaT13l030730) is homologous to the transcript of the Arabidopsis gene chloride channel A (CLC-A; AT5G40890). CLC family members are involved in anion compartmentalization in intracellular organelles and in stomatal guard cell vacuoles (Jossier *et al.*, 2010). More precisely, CLC-A and CLC-C are expressed in stomata and control their opening through translocation of NO_3^- and Cl^-, respectively. This difference in anion selectivity among the CLC family members is a result of an amino acid change in the selectivity filter (Wege *et al.*, 2010). The sunflower transcript HaT13l030730, which is homologous to Arabidopsis CLC-A, possesses the same amino acid conferring nitrate specificity. This suggests that the main hub identified in the drought GRN is probably a nitrate channel involved in stomatal aperture control and, therefore, transpiration.

The second hub (HaT13l003541) is homologous to the transcript of the Arabidopsis gene nitrate transporter 1 (NRT1.1; AT1G12110), which encodes a dual-affinity nitrate transporter in Arabidopsis. Guo *et al.* (2003) demonstrated that this gene is expressed in guard cells of stomata and that transpiration is affected in mutants in an ABA-independent manner. The reduction of the stomatal aperture in mutants appeared to be attributable to nitrate uptake in guard cells. The control of stomatal transpiration by anion channels and transporters in guard cells was further confirmed (De Angeli *et al.*, 2013) in Arabidopsis.

Our approach identified the key role of two sunflower homologs of Arabidopsis anion transporters. This strongly suggests that this process is important for the regulation of the sunflower drought response. However, the two hubs do not directly regulate the expression of their target as transcription factors do; instead, the hubs drive downstream signaling cascades through indirect physiological and distant regulation.

The drought GRN identifies connections between ABA-dependent and ABA-independent pathways In the inferred network, both hubs had seven common targets but no common source. This suggests that the NRT1.1 and CLC-A sunflower homologs could represent two pathways controlling drought stress responses. However, we could not exclude cross-talk between NRT1.1 and CLC-A with an upstream regulator absent from our initial data set. By inferring sunflower gene function based on Arabidopsis homology and the analogous expression response to drought, we could tentatively investigate the molecular pathways characterized in the sunflower drought GRN. Functional annotation of the targets of the two hubs revealed genes that are directly involved in cell protection and stress tolerance, such as the reactive oxygen species (ROS) scavenger (ascorbate peroxidase 1 (APX1)) and two enzymes involved in synthesis of an osmo-protectant, choline (phosphoethanolamine n-methyltransferase (PMEAMT) and phosphorylcholine cytidylyltransferase2 (CCT2)). Interestingly, we also identified genes involved in signal transduction, such as kinases (HaT13l074901 and embryo defective 1075 (emb1075)), phosphatases (hypersensitive to ABA-deficient 1 (HAB1)), calmodulin-binding proteins (calmodulin-domain protein kinase 5 (CPK5)), and transcriptional regulators (MYC2 and arm repeat protein interacting with ABF2 (ARIA)), downstream of the anion transporters, as described in Fig. 4.

CLC-A and NRT1.1 define an ABA-dependent and an ABA-independent, respectively, pathway in our experimental results, as well as in Arabidopsis (Guo *et al.*, 2003; Jossier *et al.*, 2010). Both sources of CLC-A, sulfate transporter 1 (SULTR1) and ABF2 (ABRE-binding bZip factor), are regulated by ABA in Arabidopsis (Fujita *et al.*, 2005; Ernst *et al.*, 2010) and also in our experiment for ABF2. In addition, specific targets of CLC-A are part of the ABA signaling cascade in Arabidopsis. HAB1 is a protein phosphatase that is strongly up-regulated by ABA (Rodriguez, 1998) and functions in ABA signaling. ABA1 is known to catalyze the

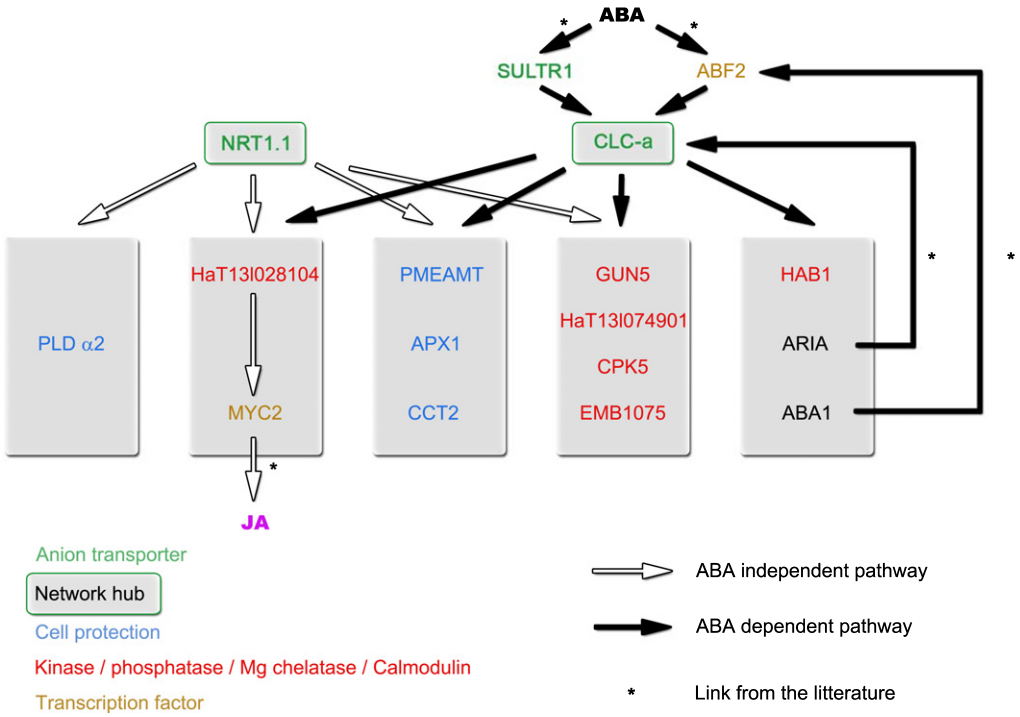

Fig. 4 Functional network involving the two hubs of the inferred drought gene regulatory network (GRN), their sources, and their targets. Blank edges represent the abscisic acid (ABA)-dependent pathway, including chloride channel A (CLC-A). Solid edges represent the ABA-independent pathway, including nitrate transporter 1 (NRT1.1). Common targets involved in signal transduction are indicated in red, those involved in transcriptional regulation are shown in orange, and those involved in cell protection are shown in blue. PLDα2, phospholipase D alpha2; MYC2, MYC-related transcriptional activator; PMEAMT, phosphoethanolamine n-methyltransferase; APX1, ascorbate peroxidase 1; CCT2, phosphorylcholine cytidylyltransferase2; GUN5, genomes uncoupled 5; CPK5, calmodulin-domain protein kinase 5; emb1075, embryo defective 1075; HAB1, hypersensitive to ABA1; ARIA, arm repeat protein interacting with ABF2; ABA1, ABA-deficient 1; SULTR1, sulfate transporter 1; JA, jasmonate.

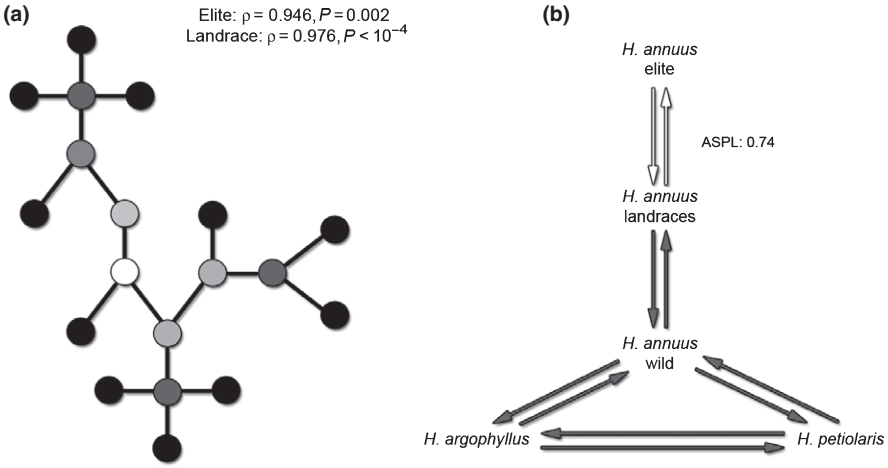

Fig. 5 (a) Representation of genetic differentiation between *Helianthus annuus* landraces and *H. annuus* elite lines in function of the gene positions, i.e. central or peripheral genes, in a schematic gene regulatory network. The colors of the genes in the schematic gene regulatory network (GRN) represent the difference in heterozygocity between the two populations for the considered gene. The node color represents differentiation between elite and landrace sunflowers: darker nodes appeared more differentiated compared with lighter nodes. Canonical coefficients (ρ) and *P*-values of Wilks's test for the correlations between network topological parameters and F_{ST} values of one population compared with the others are indicated for elite lines and landraces. (b) Hypothesis concerning differences in genetic differentiation between the five *Helianthus* populations. Note that only the five comparisons representing the selective history of the sunflower are shown. Black edges indicate no variability in the genetic differentiation within genes network between the two populations. White edges indicate changes in the genetic differentiation between populations as observed for the 15 genes in the drought GRN analyzed in the canonical correlation analysis (CCA). The coefficient of the Pearson's correlation between the topological parameter average shortest path length (ASPL) and F_{ST} between *H. annuus* elite lines and landraces is indicated.

first step of ABA synthesis (Rock & Zeevaart, 1991), and ARIA is an armadillo repeat protein that is known to interact with the transcription factor ABF2 (Kim *et al.*, 2004). Together, these regulatory connections identified in Arabidopsis form a loop involving ABA synthesis (in vascular cells) (Boursiac *et al.*, 2013) and a signaling pathway across the different cell types (including

guard cells) throughout the leaf (Fig. 4). In the drought GRN, we were able to partially identify the corresponding regulatory loop between sunflower homologs. These results suggest that the same ABA regulatory loop exists in the sunflower drought GRN and therefore could be largely shared across the plant kingdom.

Similar to the shared targets of CLC-A and NRT1.1, specific targets of NRT1.1 are also involved in cell protection (phospholipase D alpha2 (PLDα2)) and signal transduction (HaT13l028104). An interesting downstream target is MYC2, which is a central regulator of the hormone jasmonate, which is mostly involved in plant defense and the development and integration of many hormonal signals (Kazan & Manners, 2013).

Across the sunflower drought GRN, several different pathways show some conservation across plant species, such as Arabidopsis. Therefore, the GRN inference approach developed in this study appears to be robust, and we can propose the reasonable hypothesis that the main regulatory pathways and hubs identified in the drought GRN are conserved among distant plant species and therefore also across the *Helianthus* genus. Although from our data we were not able to demonstrate network conservation across the *Helianthus* populations (this would require inferring the network for each one, which would be too laborious with the present technologies), this hypothesis allows us to explore new questions about how the GRN could constrain plant adaptation to dry environments.

Drought GRN topology and *Helianthus* evolution

Network topology constrains genetic variation of the gene network Gene networks are the products of evolution, like other biological objects, but gene network relationships can also constrain evolutionary changes, such as adaptations to new environments and responses to selective pressure during domestication or breeding. For example, Rausher *et al.* (1999) demonstrated different evolutionary histories for upstream and downstream genes in the anthocyanin biosynthetic pathway.

The evolution of the GRN architecture can lead to new nodes, potentially introducing new functions and new edges between these nodes. Previous researchers (Hinman *et al.*, 2003) examined GRN evolution in echinoderms and demonstrated that some features of developmental GRNs were conserved and that others were specific to each taxon. Network architecture is known to affect evolutionary rates (Ramsay *et al.*, 2009), and we expect evolutionary changes to the nodes to be constrained by their connectivity and the number of neighbors. A hub in the network is involved in several pathways. The functional trade-offs for such genes are higher than those for peripheral genes that are involved neither in regulatory processes nor in the interaction with partners.

To understand how populations and species evolve and adapt to a new environment, we examined the putative constraints of the network architecture on the genetic differentiation between populations of *H. annuus*, and two wild species that are cross-compatible with *H. annuus*: *H. argophyllus* and *H. petiolaris*.

No evidence of network topology constraints during the divergence of *H. argophyllus* and *H. petiolaris* *Helianthus argophyllus* is native to the dry, sandy soils of southern Texas, an arid environment that imposes strong selection for tolerance to drought stress. Indeed, *H. argophyllus* is considered the most drought-tolerant sunflower species because its pubescent leaves reflect sunlight, reduce water loss, and exhibit low transpiration (Seiler & Rieseberg, 1997). However, network topology and F_{ST} values between *H. argophyllus* and other populations were not significantly correlated. This could be because the adaptation of *H. argophyllus* to dry environments involved physiological mechanisms that are not captured in our GRN or because the network topology has itself evolved and the topological parameters in *H. argophyllus* are too dissimilar to those in *H. annuus*. Interestingly, the highest value of F_{ST} between *H. argophyllus* and other populations was for the network hub, NRT1.1, which is involved in transpiration. This result is consistent with positive selection acting on NRT1.1 during adaptation of *H. argophyllus* to dry environments. Keeping in mind the overall nonsignificant correlation, the high value of F_{ST} suggests that NRT1.1 could be an example of the fore-mentioned hypothesis, i.e. the positive selection.

In *H. petiolaris*, we observed no correlation between the GRN topology and F_{ST} for comparisons with other populations. Because *H. petiolaris* has a large geographic range that overlaps with that of *H. annuus* in the Great Plains of the USA, drought stress might not be the major selective force separating these species. This could explain the similar divergence patterns within the drought network genes between these two populations, as illustrated in Fig. 5b.

Genetic diversity within the GRN was modified during modern breeding The network topological parameters and the F_{ST} between the landraces and elite lines of *H. annuus* were correlated (Fig. 5a). This reflects a difference in genetic differentiation between these two populations between the center and the periphery of the network. We did not observe this correlation for F_{ST} between wild *H. annuus* and landraces. This suggests that the position and connectivity of genes in the drought GRN influenced the response to selection during the last century of genetic improvement but not during the initial domestication of *H. annuus*. This difference in selective responses could be attributable to the fact that highly connected genes are subjected to more trade-offs as they are master regulators with involvement in several genetic pathways, in contrast to less connected terminal genes (Fig. 5a). Drought tolerance is considered to be a long-standing goal of sunflower breeders. We would expect that the selection they exert had led to a global reduction of genetic diversity in the drought GRN. However, we observed a higher divergence of terminal genes compared with central ones, which implies a stabilizing selection acting on the network hubs. Interestingly, our F_{ST} studies in *H. argophyllus* suggest that a different selective pressure acted on one of the network hubs (Fig. 5b). This highlights our lack of global understanding of how evolutionary forces and functional relationships interacted to produce contemporary phenotypic diversity and suggests a potentially important way of improving the breeders' methods, through the integration of regulatory networks in quantitative genetics models such as genomic selection.

In conclusion, this work investigated the interaction between physiological and evolutionary processes in the context of a genetic network for the drought-stress response. Interactions between physiological and evolutionary time-scales could be revealed in the future through global transcriptomic studies, although some limitations of network inference methods remain to be overcome. This type of work will facilitate the study of responses to other environmental factors and clarify whether physiological mechanisms and evolutionary adaptation, which are reciprocally constrained in the gene regulatory network, are similar in abiotic and biotic interactions.

Acknowledgements

This work was part of the OLEOSOL project funded by French public funds for competitiveness clusters (FUI), the European Regional Development Fund (ERDF), and the Government of the Région Midi-Pyrénées. The study benefited from the support of the Genoscope project AP09/10 and a PhD grant co-funded by Syngenta Seeds and the Région Midi-Pyrénées. This work was part of the 'Laboratoire d'Excellence (LABEX), entitled TULIP (ANR-10-LABX-41). We thank the Platform GeT PlaGe from INRA Toulouse for their technical support and the sunflower team members and the common services of the LIPM.

References

Allouche D, Cierco-Ayrolles C, de Givry S, Guillermin G, Mangin B, Schiex T, Vandel J, Vignes M 2013. A panel of learning methods for the reconstruction of gene regulatory networks in a systems genetics context. In: de la Fuente A, ed. *Verification of methods for gene network inference from Systems Genetics data*. Heidelberg, Germany: Springer, 9–31.

Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M. 2008. Computing topological parameters of biological networks. *Bioinformatics* **24**: 282–284.

Barabasi AL, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5**: 101–113.

Bazin J, Langlade N, Vincourt P, Arribat S, Balzergue S, El-Maarouf-Bouteau H, Bailly C. 2011. Targeted mRNA oxidation regulates sunflower seed dormancy alleviation during dry after-ripening. *Plant Cell* **23**: 2196–2208.

Boudsocq M, Lauriere C. 2005. Osmotic signaling in plants. Multiple pathways mediated by emerging kinase families. *Plant Physiology* **138**: 1185–1194.

Boursiac Y, Léran S, Corratgé-Faillie C, Gojon A, Krouk G, Lacombe B. 2013. ABA transport and transporters. *Trends in Plant Science* **18**: 1878–4372.

Bray E. 2004. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. *Journal of Experimental Botany* **55**: 2331–2341.

Breiman L. 2001. Random forests. *Machine Learning* **45**: 5–32.

Chinnusamy V, Schumaker K, Zhu JK. 2004. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *Journal of Experimental Botany* **55**: 225–236.

Crombach A, Wotton KR, Cicin-Sain D, Ashyraliyev M, Jaeger J. 2012. Efficient reverse-engineering of a developmental gene regulatory network. *Plos Computational Biology* **8**: 21.

De Angeli A, Zhang JB, Meyer S, Martinoia E. 2013. AtALMT9 is a malate-activated vacuolar chloride channel required for stomatal opening in Arabidopsis. *Nature Communications* **4**: 1804.

Dikicioglu D, Karabekmez E, Rash B, Pir P, Kirdar B, Oliver SG. 2011. How yeast re-programmes its transcriptional profile in response to different nutrient impulses. *BMC Systems Biology* **5**: 148.

Ernst L, Goodger JQD, Alvarez S, Marsh EL, Berla B, Lockhart E, Jung J, Li PH, Bohnert HJ, Schachtman DP. 2010. Sulphate as a xylem-borne chemical signal precedes the expression of ABA biosynthetic genes in maize roots. *Journal of Experimental Botany* **61**: 3395–3405.

Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K, Yamaguchi-Shinozaki K. 2005. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. *Plant Cell* **17**: 3470–3488.

Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li WQ, Ogawa M, Yamauchi Y, Preston J, Aoki K *et al.* 2008. The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant Journal* **55**: 526–542.

Goudet J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5**: 184–186.

Guo FO, Young J, Crawford NM. 2003. The nitrate transporter AtNRT1.1 (CHL1) functions in stomatal opening and contributes to drought susceptibility in arabidopsis. *Plant Cell* **15**: 107–117.

Hannah MA, Wiese D, Freund S, Fiehn O, Heyer AG, Hincha DK. 2006. Natural genetic variation of freezing tolerance in arabidopsis. *Plant Physiology* **142**: 98–112.

Hinman VF, Nguyen AT, Cameron RA, Davidson EH. 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proceedings of the National Academy of Sciences, USA* **100**: 13356–13361.

Hirayama T, Shinozaki K. 2010. Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant Journal* **61**: 1041–1052.

Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. 2010. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**: e12776.

Jossier M, Kroniewicz L, Dalmas F, Le Thiec D, Ephritikhine G, Thomine S, Barbier-Brygoo H, Vavasseur A, Filleur S, Leonhardt N. 2010. The Arabidopsis vacuolar anion transporter, AtCLCc, is involved in the regulation of stomatal movements and contributes to salt tolerance. *Plant Journal* **64**: 563–576.

Kawaguchi R, Girke T, Bray E, Bailey-Serres J. 2004. Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant Journal* **38**: 823–839.

Kazan K, Manners JM. 2013. MYC2: the master in action. *Molecular Plant* **6**: 686–703.

Kim S, Choi HI, Ryu HJ, Park JH, Kim MD, Kim SY. 2004. ARIA, an Arabidopsis arm repeat protein interacting with a transcriptional regulator of abscisic acid-responsive gene expression, is a novel abscisic acid signaling component. *Plant Physiology* **136**: 3639–3648.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data P. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li S, Assmann S, Albert R. 2006. Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biology* **4**: 1732–1748.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**: 402–408.

Ma SS, Gong QQ, Bohnert HJ. 2007. An Arabidopsis gene network based on the graphical Gaussian model. *Genome Research* **17**: 1614–1625.

Marbach D, Costello J, Kuffner R, Vega N, Prill R, Camacho D, Allison K, Kellis M, Collins J, Stolovitzky G, Consortium D. 2012. Wisdom of crowds for robust gene network inference. *Nature Methods* **9**: 796–804.

Miller G, Suzuki N, Ciftci-Yilmaz S, Mittler R. 2009. Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant, Cell & Environment* **33**: 453–467.

Milo R, Hou JH, Springer M, Brenner MP, Kirschner MW. 2007. The relationship between evolutionary and physiological variation in hemoglobin. *Proceedings of the National Academy of Sciences, USA* **104**: 16998–17003.

Neumann G, Massonneau A, Langlade N, Dinkelaker B, Hengeler C, Romheld V, Martinoia E. 2000. Physiological aspects of cluster root function and development in phosphorus-deficient white lupin (*Lupinus albus* L.). *Annals of Botany* **85**: 909–919.

Osborn HF. 1896. A mode of evolution requiring neither natural selection nor the inheritance of acquired characters. *Transactions of the New York Academy of Sciences* **15**: 141–142.

Pastori G, Foyer C. 2002. Common components, networks, and pathways of cross-tolerance to stress. The central role of "redox" and abscisic acid-mediated controls. *Plant Physiology* **129**: 460–468.

Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**: 618–624.

Ramsay H, Rieseberg LH, Ritland K. 2009. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. *Molecular Biology and Evolution* **26**: 1045–1053.

Rausher M, Miller R, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Molecular Biology and Evolution* **16**: 266–274.

Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* **4**: 1827.

Rengel D, Arribat S, Maury P, Martin-Magniette ML, Hourlier T, Laporte M, Vares D, Carrere S, Grieu P, Balzergue S *et al.* 2012. A gene-phenotype network based on genetic variability for drought responses reveals key physiological processes in controlled and natural environments. *PLoS ONE* **7**: e45249.

Rock CD, Zeevaart JAD. 1991. The *aba* mutant of *Arabidopsis thaliana* is impaired in epoxy-carotenoid biosynthesis. *Proceedings of the National Academy of Sciences, USA* **88**: 7496–7499.

Rodriguez PL. 1998. Protein phosphatase 2C (PP2C) function in higher plants. *Plant Molecular Biology* **38**: 919–927.

Rook F, Hadingham S, Li Y, Bevan M. 2006. Sugar and ABA response pathways and the control of gene expression. *Plant, Cell & Environment* **29**: 426–434.

Seiler GJ, Rieseberg LH 1997. Systematics, origin, and germplasm resources of the wild and domesticated sunflower. In: Schneiter AA, ed. *Sunflower science and technology*. Madison, WI, USA: ASA, CSSA, and ASSA, 21–66.

Seki M, Umezawa T, Urano K, Shinozaki K. 2007. Regulatory metabolic networks in drought stress responses. *Current Opinion in Plant Biology* **10**: 296–302.

Shinozaki K, Yamaguchi-Shinozaki K. 1997. Gene expression and signal transduction in water-stress response. *Plant Physiology* **115**: 327–334.

Shinozaki K, Yamaguchi-Shinozaki K. 2007. Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany* **58**: 221–227.

Sirichandra C, Wasilewska A, Vlad F, Valon C, Leung J. 2009. The guard cell as a single-cell model towards understanding drought tolerance and abscisic acid action. *Journal of Experimental Botany* **60**: 1439–1463.

Spurgeon SL, Jones RC, Ramakrishnan R. 2008. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS ONE* **3**: e1662.

Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**: 520–525.

Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K. 2010. Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. *Plant and Cell Physiology* **51**: 1821–1839.

Valliyodan B, Nguyen H. 2006. Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Current Opinion in Plant Biology* **9**: 189–195.

Vignes M, Vandel J, Allouche D, Ramadan-Alban N, Cierco-Ayrolles C, Schiex T, Mangin B, de Givry S. 2011. Gene regulatory network reconstruction using Bayesian networks, the Dantzig Selector, the Lasso and their meta-analysis. *PLoS ONE* **6**: e29165.

Waddington CH. 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150**: 563–565.

Wang W, Vinocur B, Altman A. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* **218**: 1–14.

Wasilewska A, Vlad F, Sirichandra C, Redko Y, Jammes F, Valon C, Frey N, Leung J. 2008. An update on abscisic acid signaling in plants and more. *Molecular Plant* **1**: 198–217.

Wege S, Jossier M, Filleur S, Thomine S, Barbier-Brygoo H, Gambale F, De Angeli A. 2010. The proline 160 in the selectivity filter of the Arabidopsis NO_3^-/H^+ exchanger AtCLCa is essential for nitrate accumulation *in planta*. *Plant Journal* **63**: 861–869.

Wilkins AS. 2005. Recasting developmental evolution in terms of genetic pathway and network evolution . . . and the implications for comparative biology. *Brain Research Bulletin* **66**: 495–509.

Wilkinson S, Kudoyarova GR, Veselov DS, Arkhipova TN, Davies WJ. 2012. Plant hormone interactions: innovative targets for crop breeding and management. *Journal of Experimental Botany* **63**: 3499–3509.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Origin of the selection for the inferred genes of the drought GRN.

Fig. S2 Hormonal and global networks.

Fig. S3 Degree distribution.

Fig. S4 Bi-plot of the effects of the topological parameters in a principal component analysis.

Table S1 Results of a *t*-test demonstrating the differential expression of genes upon application of 10 μM ABA

Table S2 Results of ANCOVA showing the validation of the data set of ABA genes

Table S3 Description of the genes selected for GRN inference

Table S4 Raw gene expression

Table S5 Gene expression after log transformation and missing data imputation

Table S6 Number of edges detected for each hormone

Table S7 Complete results of the principal component analysis on the topological parameters for the drought GRN

Table S8 Complete results of the canonical correlation analysis between the topological parameters and F_{ST} subsets

Methods S1 Supporting methods providing detailed information about plant material, molecular biology procedures and GRN inference.

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.