Current Plant Biology 3-4 (2015) 30–39

Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb

Revealing gene regulation and associations through biological networks

Christophe Liseron-Monfils[a], Doreen Ware[a,b,*]

[a] Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, United States
[b] USDA-ARS-NEA, Robert W. Holley Center, Ithaca, NY 14853, United States

ARTICLE INFO

Article history:
Received 9 June 2015
Received in revised form 30 October 2015
Accepted 2 November 2015

Keywords:
Inference networks
QTL
In-vitro networks
Candidate genes
Gene prioritization
Co-expression

ABSTRACT

Traditionally, over the last 10,000 years, agriculture has relied on natural biological evolution and careful selection of plant varieties by farmers that was used as the founder material by plant breeders in the last 150 years for further genetic improvement. Plant breeders played an important role mainly by introgression of a trait of interest, through the transfer of genetic loci, into an elite crop line that exhibits high-yield performance across a wide range of conditions. Modern agriculture has relied on the use of biotechnologies and molecular biology to improve marker development and aid in the discovery of candidate loci or genes associated with desirable traits, thereby reducing the time required for selective breeding.

In this review, we briefly describe the evolution of the methods used to identify candidate leads (gene, loci or regulatory regions) for crop improvement, starting by quantitative genetic methods. The development of co-expression and molecular networks will be described. It will be shown how network analysis can reinforced the discovery of candidate genes/loci more rapidly and with higher confidence. These improvements will serve to accelerate genetic engineering and molecular breeding as modern agriculture confronts the challenging times ahead, with the increase of abiotic stresses for crops as drought, heat, soil high salinity or waterlogging.

Without taking in account possible losses due these growing stresses, the increase in crop yields needs to be significantly accelerated to feed the growing world population, following the FAO previsions.

1. Period preceding network analysis

1.1. Methods used before the advent of network analysis

To increase yield and stress tolerance, it is essential to understand how crop phenotypes were improved during the domestication process. To this end, researchers have investigated how early societies domesticated wild plants, e.g., by adapting teosinte into a high-yield crop like maize [1]. During domestication, diversity was dramatically reduced, as rare alleles in the population declined in favor of more frequent ones [2]. Awareness of the reduction in the size of the gene pool led to the idea that exploring the wild ancestors of crop plants could help to recover rare alleles. These rare alleles might be useful in improving modern crop varieties. According to simulation studies in maize, only 2–4% of the genome has been actively selected over the course of domestication [2]. This finding highlights the tremendous potential of using the wild varieties or ancestors of extant crop plants to recover lost genetic diversity. Such (re) discovery of lost alleles could help meet challenging breeding goals such as conferring tolerance to drought, salinity, and other stresses [1,2].

Historically, the methods of choice for identifying candidate loci/genes primarily involved identification of quantitative trait loci (QTLs). In such analyses, phenotypic observations are associated with genetic region segregation, taking advantage of linkage disequilibrium.

In crop species, one aim of QTL analysis is to define genetic regions responsible for architectural and stress-related phenotypes, e.g., the response to water deficit and the efficiency of nitrogen use [3–5]. However, these genetic regions can be very large, due to the size of the QTL confidence intervals. To discover the causal genes, it is required to manually parse ~100 candidate genes per locus. Sometimes, the number of candidate genes can be reduced using fine mapping, by finding additional makers within the QTL interval. Nevertheless, the process of narrowing a QTL is highly time-consuming [6]. Moreover, the outcome is not always guaranteed, as it depends upon the rate of crossing-over in the

* Corresponding author.
E-mail address: ware@cshl.edu (D. Ware).

http://dx.doi.org/10.1016/j.cpb.2015.11.001

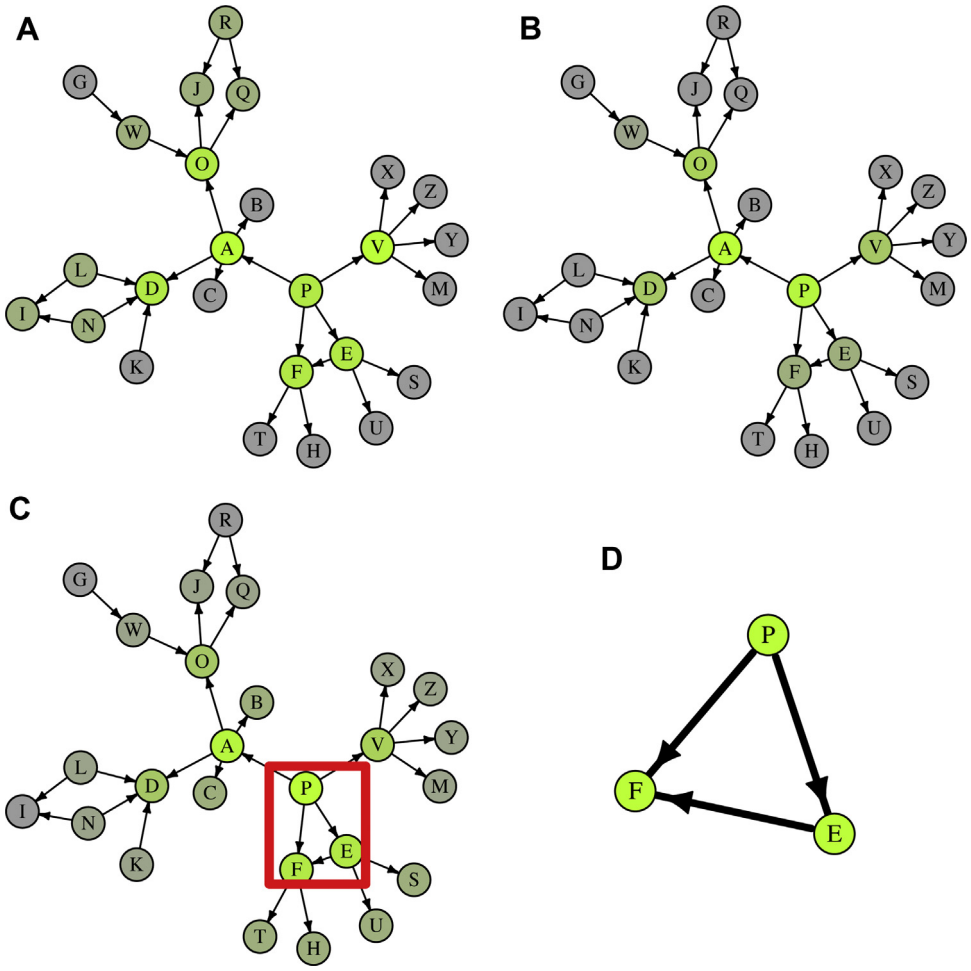

Fig. 1. Centrality and network motifs.
A random network is represented (A–C) mimicking a possible biological network. The greenest nodes represent the genes with the higher connectivity centrality (A), betweenness centrality (B) and Eigen vector centrality (C). Node A has the highest betweenness centrality (B) whereas node P has the highest Eigen vector centrality (C). It shows how a researcher can act on either node A or P in function as the desired action shown (e.g. Fig. 2). The grey nodes have a lower value for each of these respective centralities. A close-up from the red box (C) of a network building block called network motif is shown in D. It represents a Feed Forward Loop.

species under study, as well as the identification of new molecular markers.

1.2. Global expression data

1.2.1. Transcriptomic analysis

Following the emergence of microarrays in the mid-1990s, transcriptomic experiments have become a predominant method for discovering candidate genes related to phenotypes of interests [7]. Next-generation sequencing made the direct count of transcript possible based on the number of sequencing reads. Before, only relative expression levels were obtained in microarray experiments by converting level of fluorescence on chips [8,9]. RNA-seq also enabled the study of the whole transcriptome, independent from a fully sequenced genome, as well as detailed analysis of gene alternative splicing forms [10–16]. One major application of transcriptomic technologies is to reveal genes differentially expressed, potentially responsible for a phenotype change, for instance between mutant and wild type samples. As technology improved, the production of comprehensive gene expression resources across a large catalogue of tissues, developmental stages became possible. Named transcriptome atlases, these resources are used as a base to obtain insights into biological processes and gene function for *Arabidopsis* (e.g. AtGenExpress), maize (Maizeatlas), or rice (e.g. RiceXPro) [17–24]. At a single tissue level, the tissue and longitudinal root atlas was produced using a combination of microarrays, cell type-specific protoplasts and root cross sections [20]. Combining these 2D analyses gives a 3D expression map in *Arabidopsis* roots. Recent work in maize that has used RNA-seq in a combination of time-course and mutant analyses has provided greater insight into the regulation of maize meristems [25,26].

However, these lists derived from transcriptomic analyses, do not immediately provide a clear insight into gene regulation. Rather, they allow selection of candidate genes that can be subsequently validated. Further reverse genetics approaches need to be used to identify mutant phenotypes. If the predicted phenotypes are observed in the mutant, novel variants in this location can be searched for in natural variation or germplasms for plant breeding improvement. Mutant analyses have some drawbacks: for example, non-conditional mutants exert their effects across all tissues and developmental stages, potentially resulting in undesirable interactions that are not directly linked to the phenotypes of interest [27,28].

2. Type of networks

2.1. Network theory

A network comprises all possible links existing between members of a community (Fig. 1). These members could be a group

of people within a social network, or genes contained within a biological network. Network members are called nodes or vertices. Interestingly, the organization of most of these networks relies on similar proprieties. The interactions linking nodes follow a non-random topographical model called scale free model [29]; the number of connections that each node has within the network generally obeys a power law distribution [29–31]. This means that few nodes have a high number of connections, when a majority of nodes have few connections. Additionally, biological networks are comprised of local modules, groups of genes that are locally highly interconnected. Both concepts of power law distribution and cluster modules are reconcilable by the fact that modules are connected to a whole network by node(s) with high connectivity [30,31]. The organization of node interconnectivity further defines the concept of centrality. The simplest centrality is connectivity, that is the number of interactions of a node with its neighbors (Fig. 1A).

One hypothesis from network analysis is that the propagation of information goes through the fastest path to minimize resource usage. One network parameter to measure the shortest path is betweenness centrality (Fig. 1B). The more a node is included in shortest paths between any 2 other nodes, the higher its betweenness centrality is within this network [32]. This is an important parameter to identify nodes for which the disruption could greatly affect a network response. Other types of centralities exists; for instance, the Eigen vector centrality measure the influence of a node on the network [33]. Nodes with high scoring Eigen vector values are linked to nodes that have high connectivity, as shown by the P node in Fig. 1C that affects highly-connected genes A and V. Consequently high-scoring Eigen-vector nodes potentially have influence on highly connected nodes and influence a large portion of the network. In a biological network, this could be for example, a transcription factor that acts on several pleiotropic genes.

2.2. Molecular interaction networks

Biological molecular networks can be categorized in several types. The notion of metagene can be introduced; a node represents at the same time a protein, its gene and its promoter. Using a metagene in network visualization allows a simplification of the network view, as it is possible to represent a transcription factor and its promoter in a single node. Visualization software for network view and analysis are important tools to interpret network behaviors. Among the network visualization tools, Cytoscape, Gephi or VisANT are highly accessible tools for non-computational biologists [34–36]. Other visualization tools can be accessible through command lines tools such as igraph, a library of the R programming languages [37]. Most of this visualization tools also contains network analyses applications to calculate for example network centralities or node clusters. One popular add-on from Cytoscape is NetworkAnalyzer [38]. Other tools have adopted another strategies and are almost completely dedicated to analysis as Network Analysis Tools (NeAT) [39].

Following the automation of interaction discovery (e.g., using the yeast two-hybrid system), a large catalogue of gene interaction networks is emerging [40]. Molecular interaction networks have also been generated using by summarizing large numbers of experiments from the literature [41]. Another category of molecular networks relies on a very large quantity of experimental and predicted data merged to form multi-source meta-networks such AraNet [42].

Protein-protein interaction networks are a major type of experimentally derived physical network. Several variants of such networks have been developed, including catalogues clock regulation signaling protein interactions in *Arabidopsis* [40,43].

Recently, the development of yeast-one-hybrid (Y1H) and chromatin immunoprecipitation (ChIP) methods has provided

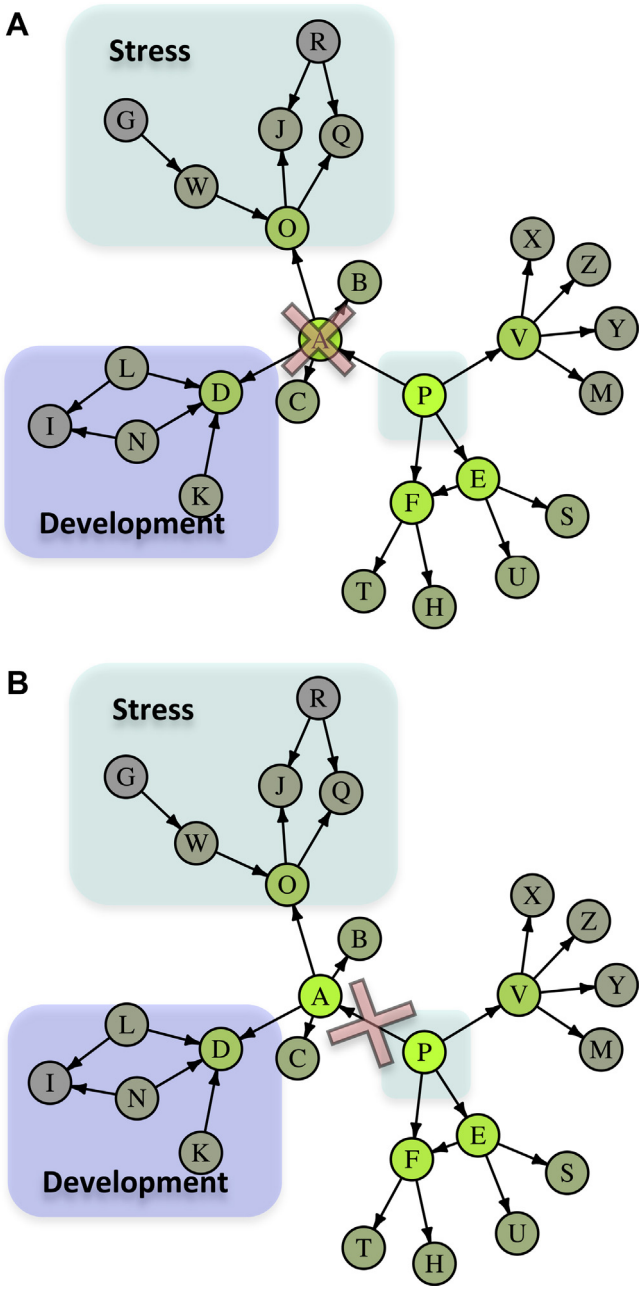

Fig. 2. Network and edgetics.
The random network from Fig. 2 is represented with the gene hypothetical functions and network modules. The green boxes are reacting to a stress. Gene P, the regulator is also expressed during this stress. The Blue box is expressed during a developmental process. Mutant approaches could consist in the disruption of the gene A (A). The edgetic mutant approach will only disrupt the interaction between gene P and gene A (B). The later will not affect the blue box (developmental processes) when the first approach (A) will have this pleiotropic effect on the blue box, which is putatively undesired.

a significant step towards a comprehensive understanding of gene-regulatory networks (GRNs). Here, GRNs are defined by the relationship between transcription factor and gene promoters. However, ChIP-seq data produces a high level of false positives and represents a full map of possible binding sites of a transcription factor without taking in account developmental stages or tissue specificity or gene regulation. To obtain a better picture at a tissue level, ChIP-seq could be combined with other high-throughput data. For instance, ChIP-seq in conjunction with transcriptome data helped to elucidate the regulation of determinacy in maize inflo-

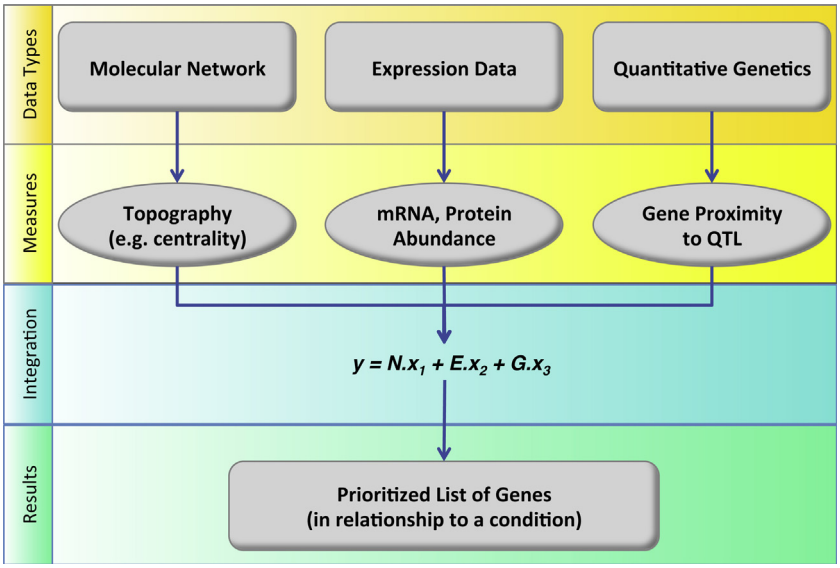

Fig. 3. Integration approach.
Here is represented a possible combination of a molecular network with its related omics and quantitative genetic data. This schematic view defines a methodology to prioritize gene candidates, for example, in relationship to a stress response.

rescence [26,44]. Tissue-specific chromatin remodeling technique is another type of data that can help to decompose the signal found in ChIP-seq.

Because it is difficult to extract information from large networks, it is tempting to return to smaller networks to understand specific phenotypes. Small networks focus on specific phenotypes such as tissue development, allowing extraction of the putative regulators. Subsequently, these putative regulators can become targets for crop improvement, as illustrated by a recent study of secondary wall development [45,46], or the flowering formation [26,47]. These small networks can facilitate the identification of the different regulation layers in a GRN, which acts on the developmental program interconnectivity with stress response such as salinity [45]. As a result, a new candidate such as VASCULAR-RELATED NAC DOMAIN (VND7) that potentially regulates xylem formation and adjusts its synthesis in reaction to salinity stresses was discovered [45].

Metabolic networks are based either on the compilation of experimental results or predictions based on extrapolations from other organisms [48–50]. One daunting task for the near future is the experimental validation of the metabolic networks predicted to date. Validated metabolic networks are of great potential value: knowledge of such networks across plant species could help to explain plant evolution [51]. As an illustration of this concept, the numbers of enzymes allocated to each metabolic category are the consequences of past gene duplications and specializations, which contributed to the distinctive features of plant species. For example, plants with vascular tissues have mainly evolved specialized metabolic pathways, at the same time reducing the number of reactions involved in production of primary metabolites [51]. Gene duplication is the phenomenon important for gene copy number variation in a genome. One of the functions of gene duplication is to help in the conservation and protection of important functions. Additionally, this redundancy provides an organism with enough flexibility, to explore new functionality, potentially a base for evolution and speciation events [52,53]. Following large duplication events, the duplicated gene segments undergo gene losses to possibly reestablish equilibrium of the gene products and to reestablish the fitness of the complete biological network [54]. Certain genes in these duplication segments are retained while others are dropped by divergent gene retention phenomenon [52].

After a speciation event provoked in part by gene duplications, sets of retained gene regions between two species, will produce syntenic regions with almost the same order for orthologous regions [55]. At a single gene level, it was proposed that the most conserved genes involved in primary metabolism form a highly connected network. The consequence would be the conservation of primary metabolism networks across the plant species [56]. Logically, this conservation property should be preserved at the metabolic enzyme gene level [51]. As mentioned earlier, transcription factor families were expanded through gene duplication [57]. This confers “robustness and diversification” possibilities to the GRN [58] from these redundant genes [58]. For example, the *Arabidopsis* response regulators (ARR) transcription factor family has most of its members regulating the same set of genes. However, each single ARR transcription factor also regulates its own unique targets [58]. One consequence of these phenomenons is the increase in complexity of a GRN.

Several databases and repositories for molecular networks are available to plant researchers such as AGRIS, TAIR, BioGRID, IntAct, STRING, the Bio-Analytic Resource for Plant Biology or the *Arabidopsis* interactome consortium [40,41,59–62]. These databases, which cover different types of gene or protein interactions, could contribute to the identification of candidate metabolic genes, thereby providing greater insight into regulatory or enzymatic pathways.

2.3. *Network inference with co-expression patterns*

Using gene expression data, the similarity of gene expression patterns can be extracted to infer co-expression networks. Such methods measure co-expression patterns across tissues and treatments to predict gene interactions. Several metrics have been used to quantify the level of co-expression between genes. The most used one is Pearson Coefficient of correlation. Once all putative interactions are established using co-expression strength, the difficult part is the elimination of less reliable interactions with low interaction strength and putative false positives. Furthermore, the predicted regulators may in fact have similar expression patterns as their downstream targets and thus be grouped together within the same modules.

2.3.1. Types of network inference

Historically, parametric correlation metrics like the Pearson correlation coefficient have been most widely used in the creation of co-expression networks [63]. Weighted Gene Co-expression Network Analysis (WGCNA), a tool based on the weighted correlation, is able to identify gene modules that could be parts of metabolic pathways, e.g., in tomato or maize [64,65]. WGCNA also implements some non-parametric correlation methods, which are used when the correlation between genes is expected to be non-linear; popular examples include the Spearman and the Kendall tau coefficient of correlation. Non-parametric methods use the rank of gene expression, instead of relative expression values, to evaluate correlations. More recently, other non-parametric methods and hybrid methods have been applied to network inference analysis. The Kendall tau and the Gini correlation [66], which were first used in economics, have now been applied to biological data. Partial correlation methods, which seek to relax the assumption of linearity of co-expression, have been implemented in several tools such as Sparse PArtial Correlation Estimation (SPACE) and GeneNet [67,68].

Concerning probabilistic methods, the mutual information (MI) has been used in place of the Pearson correlation; motivated in part by findings that MI has more power than the Pearson correlation distance [69]. MI is based on the Shannon Entropy and implies discretization of the expression data. If two genes are statistically independent, their MI will be zero, as they will share no linear or non-linear correlation pattern [70]. Because the MI is overestimated for continuous variables, it must be corrected, e.g., using the shrinkage method, which is applicable to the type of small-sample datasets that are common in functional genomics [71]. MI forms the basis of network inference methods such as ARACNE [72,73]. MI methods generate robust networks, but they suffer from numerous drawbacks. First, the data need to be categorized by creating bins of level of gene expressions. The estimation of these bins is complex. Second, it is difficult to compare results from different datasets. To address these concerns, a new method was developed to fix all the pitfalls and obtain the maximum information coefficient. However, the methodology and the real advantages of this method are still being discussed [74–77].

Bayesian methods are another type of measurement used to generate networks inferred from co-expression data. Bayesian dynamic networks are particularly well suited for small networks, as these types of analysis are computationally expensive. Banjo is a tool for analysis of Bayesian and dynamic Bayesian networks [78]. In a small network of ~100 genes in loblolly pine, SND1 was discovered as a putative key regulator of lignin biosynthesis using a Bayesian algorithm [79].

2.3.2. Tools using network inferences

Allen et al. [80] compared different methods for inferring transcriptional networks. The results revealed that methods based on correlation (e.g., WGCNA) and mutual information [73] (e.g., ARACNE: algorithm for the Reconstruction of Accurate Cellular NEtworks) tend to outperform those based on partial correlation (e.g., GeneNet and SPACE). On the other hand, GeneNet was better at picking a few significant nodes within the network [80]; if a small number of true positives can be picked, it means that the network has a low rate of false positives. Bayesian inference methods (e.g., BNArray, Banjo) are the more difficult to use on large networks, largely because their computational costs result in prohibitive running times. Furthermore, they do not appear to have significant advantages compared to other methods, although they may be superior for network inferences based on very small networks.

After the network inference, the next steps, to find putative regulators and candidate gene linked to a trait, are to analyze the network dynamic. Several methods as ordinary differential expression or Boolean algorithms can be more predictive than co-expression networks with respect to this goal. An interaction of a transcription factor with a promoter is a kinetic process. This means that the transcription factor will need to reach a protein concentration or activity threshold to trigger the regulation of a promoter [81]. Boolean models aim to simplify this molecular kinetic by attributing states for mRNAs, protein or metabolite abundances. In most of the Boolean models, a molecule is either in active or inactive state. Furthermore, this state can evolve in function of clearly defined relationships of activation or inhibition attached to each network interaction. As a result, it gives the opportunity to establish more or less complex cascades of activity state, starting from a specific point to the end of the cascade. The effect of a regulator on a network could be measured by observing the state changes caused in downstream interactions. Using Boolean models in yeast and plants, the stability of the cell cycle network was established against possible perturbations [82]. It showed that the cell cycle network relies on the oscillatory interplays between cyclin-dependent kinases and transcription factors [83].

Based on these tools or co-expression metrics, several databases were generated to help understand gene regulation patterns and discover candidate genes that regulate particular phenotypes, such as ATTED II [84–86]. These databases employ visualization tools to help researchers make decisions regarding possible leads [87]. Analysis of co-expression networks helps in the selection of candidate genes linked the studied condition/trait. Co-expression networks give the possibility to shift focus from single candidate gene search to groups of related genes that are likely to operate together within a tissue, or response to a stress.

2.3.3. Co-expression network inference and limits

Network inference analysis depends on snapshots taken using transcriptomic or other omics data. Many groups are interested in understanding and predicting network behavior in order to strengthen the discovery of candidate genes. To this end, network databases have been created such as AraNet or ATTEDII that partly or completely based on co-expression analyses [42,88]. These inference networks have been used to build hypotheses: for instance, a model for the regulation of glucosinolates was generated using a combination of transcriptomics and metabolomics data [89,90].

However, co-expression networks can lead to misleading results, as the interactions are based similarity of expression patterns. Consequently, it is difficult to know if the expression interactions are due real connections between 2 genes, or a convergence of expression patterns, independent of any real physical or regulatory link. As a result, co-expression network generates a high level of false positive interactions, which could interfere with interpretations and the search for causal genes. Another problem is the low correlation between mRNA and protein levels under the condition studied [91]. On the other hand, one area in which inference networks seem to perform well is metabolic pathway prediction: factors involved in specialized metabolic pathways exhibit a high level of co-expression, suggesting that they are the most suitable targets for co-expression analysis [51].

2.4. Network motif

In order to understand regulation between groups of genes in biological networks, the notion of a network motif was introduced. A network motif is a pattern formed by subgroup varying from 2 to 10 genes linked through regulation interactions. The two most popular types of network motifs are 3-and 4-node motifs [92,93]. Network motifs help better characterize gene functions using the regulation knowledge about activation or repression between genes. Network complexity can be reduced using network motifs, demonstrating the interest of the field in studying them.

Interestingly, the occurrence of network motifs is clearly not random and certain motif types are overrepresented in biological, food chain or social networks [93]. It demonstrates that for unknown reasons, some regulatory "building blocks" are put in place by plants and other living organism in order to organize interactions between genes. What is the molecular abundance of each network motif component? How does this abundance level change within a network motif? Authors have tried to answer these difficult questions using simulation studies [94–96]. The simulations were made mostly on 3-node interaction types to discover the possible dynamics of network components such as the feed-forward loops shown in Fig. 1D [95]. Elucidating these regulatory dynamics could for instance determine if an interaction type produces an oscillatory behavior [97]. The number of network motifs increases with the size of a network, which makes the discovery of these motifs difficult. Consequently, network motif quantification and overrepresentation in a network are estimated using set of random motifs. Several tools tried to tackle this challenge as reviewed in the literature [98].

3. Determine candidate genes

3.1. Quantitative genetics and transcriptomics

A large portion of the genetic variation is not explained in most of the QTL studies, notably for complex traits such as yield [99,100]. It emphasizes the importance of understanding the whole spectrum of genetic variation identified including the low-significance QTLs as shown for maize flowering time [101]. The addition of these low-significance QTLs can affect high-significance QTLs. Because a large number of small-effect QTLs can make it difficult to discover causal loci, it is important to understand the biological processes underlying these small-effect QTLs to estimate the importance of each locus [100]. To this end, it would be useful to study the interactions between the putative genes underlying each QTL associated with a specific phenotype. Furthermore, the analysis of genetic networks can help identify false positives among the small-effect QTLs; by overlaying genes putative linked to a response part of interactions in a molecular network and gene physical location within QTLs. This method is conservative but will be robust if the QTL intervals do not contain a large number of genes.

One first step of integration of co-expression data was to merge them with other type of evidence such as QTL. Transcriptomics has been combined with quantitative genetics methods to facilitate the identification causal genes in participating in stresses responses. In *Arabidopsis*, causal regions of a phenotype response, as stress tolerance, are difficult to identify if only QTL analysis are used [102]. Microarrays have facilitated these associations. Another interesting method not discussed in this review is the eQTL analysis [103].

3.2. Discover candidate genes using network topography

The directionality of a network makes genes, with high betweenness centralities, bottlenecks for the circulation of the biological signals either between protein, protein interaction in signaling cascade or between metabolites in a metabolic network for instance. Associating Gene Ontology (GO) with network centrality is one of the possible outcomes to define hub genes that could be master regulators and candidate gene for plants [104]. Network topography has been mostly used in gene co-expression networks to discover or confirm essential genes within and between species as well as discover gene functions from associalogs [105–107].

3.3. Gene prioritization pipeline

Gene prioritization methods aim to identify genes that are relevant to specific stresses or conditions of interest. In particular, they aim to rank genes by combining different sources of information from the literature. Therefore, gene prioritization tools help to determine which candidate genes that should be assigned the highest priority in subsequent experimental validations following the large-scale (high-throughput) experiments. Unfortunately, most of these tools are not available for plant research, but for human diseases, as reviewed previously [108]. However, one could argue that some of the tools dedicated to plant network analysis are also partly gene prioritization tools [109,110]. Plant research could benefit from tools with minimal biases, as explained below.

Gene prioritization can take advantage of gene expression data as well as other sources of information. For example, text mining or GO annotation can be contributing parameters used to determine the importance and function of genes in a network dataset. GO annotation referred to standardized conventions to assign "molecular functions" to a gene product, as well as determining its connections to "biological processes" and its cell localization described as "cellular component". However, these data associations should be made carefully due to the risk of inaccurately assigning a function to the gene based on biased prior information. Gene annotation is the method of choice for identifying the putative function of a gene according to annotations of the surrounding genes in the network, referred to as the "guilt by association" method [105]. However, this method can generate biases because the genes used for functional annotations in GO are often pleiotropic, and do not necessary represent the functions of their associated genes in a particular analysis [111].

3.4. Tools for predicting candidate genes and perturbation of networks

A major challenge of large networks is determining how best to perform analysis in order to extract interesting information. Several new algorithms have attempted to simplify this process. For example, one recently reported algorithm could reduce the number of edges in a network by grouping them according to their pattern similarities [112]. Many of these analyses of plant interaction networks have been performed in *Arabidopsis*; hence, one important task is the on-going effort to transfer the knowledge accumulated in crops and apply it to their improvement. To facilitate this transfer, networks can be projected from one species to another one using orthologous analysis or associalogs [42,109,110]. An associalog is a functional orthologous association to generate a projection of network using potential functional orthologs to discover "orthologous" interactions. However, orthologs between two species is difficult to define well due to gene duplications. Furthermore, there is still a possibility of interaction rewiring of putative functional orthologs interactions within their respective gene networks, which could create false positive interactions in projected networks from associalogs.

Taking in account the conservation of an enzyme-centered metabolic network could be an approach of choice to transfer previous *Arabidopsis* molecular network knowledge. For instance, a projected protein–protein network could use this conserved network as a core. Then, extension of this projected network could be done using gene expression networks or molecular experiments in the new species to overcome the problem of less conserved genes. The less conserved genes between two species are subjected to have undergone more gene expansions and more interactions rewiring [113]. Transcription factor families are one example of possible gene expansions [52,58]. Large-scale molecular networks can contribute to our understanding of gene duplication phenomenon [40].

The rate of gene duplication and its consequence on the interaction rewiring rate has been subject to modeling in order to determine when and how they could occur [113]. This rewiring could be at the level of the entire duplication event meaning that the new genes copies part of a same network modules could evolve in a complete different sub-network not linked to the initial one. This was observed in *Arabidopsis* where duplicated gene branches had divergent pattern of gene expression [114].

3.5. Use of other high-throughput datasets

Proteomic data has also contributed to the identification of candidate genes for crop improvement. Several studies have reported the use of proteomic data to the investigation of plant stress responses. Moreover, combining several types of -omics data (transcriptomics, proteomics, and metabolomics) provides a more complete view of tissue dynamics; however, using multimodal data in combination necessarily increases the complexity of the analysis. Such analyses can change our perception of "normal regulation". For instance, metabolites are often seen as the end product of gene regulation, but an integrated omics approaches reveals that they can also serve as stimuli for the same gene-regulatory pathways [115].

3.6. How to use dynamic networks

Vital genes have a tendency to be enriched among hub genes [116]. A hub gene in a network is a gene with a large number of connections. These hub genes are important because they are often the more straightforward candidate genes for manipulation [117–119]. Two types of hub nodes are described in the literature, classified based on their co-expression with neighboring genes in the network [120]. Party hubs would be expressed only under specific conditions, and are co-expressed with linked genes within the network [120]. By contrast, date hubs would be expressed under different conditions; consequently, they can affect multiple processes [120]. Date hub genes spread information across the network but are not necessary the "driver nodes" (also called critical nodes) involved in regulation or perturbation of the entire network regulation. However, these notions of date/party hubs are still controversial and more robust studies are needed to clearly establish the veracity of these two hub types [121]. Another type of gene discovered in network analysis is the "flight hub gene," which the authors described as a putative "switch gene" involved in regulation of transcriptional processes in plants [122]. Consequently, local central nodes within modules could be more critical than nodes that are central to the network overall [123]. Based on simulations, it appears that the most important hubs are grouped in modules. These findings suggest that whole-network hub genes could be redundant if they are not critical for a network perturbation.

3.7. Forward and reverse "edgetics"

Hub genes in a network have a tendency to be pleiotropic. Therefore, it would be interesting to identify the tissue- and temporally-specific regulators of these hub genes. Observation of network data can help identify these types of regulators. From a statistical point of view, eigenvector centrality is also valuable for identifying such regulator [33].

Another approach used in the plant–microbe interaction field, is removal or modification of interactions (edges) rather than modifying or removing genes (nodes) from a network in order to affect a specific phenotype without deleterious effects on other conditions [124]. This method allows precise regulation of a specific hub gene by its regulators [125–127], assuming that another regulator does not redundantly influence the edge (Fig. 2).

As an example of this approach, let us consider the disruption of a transcription factor's action on a hub gene. The *cis*-acting elements can be identified through immuno-precipitation assay or yeast one-hybrid using partial promoter deletion of the hub gene, candidate target for the regulation of a trait. Once the binding sites are identified, one could modify them directly using "clustered, regularly interspaced, short palindromic repeat" (CRISPR) [128], or examine natural variants to determine whether a modification associated with the desired phenotype already exists in some varieties of the studied plant. The domain of genome editing is in wide expansion nowadays, as they can precisely change few nucleotides in a genome. Additionally to CRISPR-Cas9, some established methods as Transcription Activator-like Effector Nucleases (TALEN) or the zinc finger nucleases have gained some interests [129]. New methods as Cpf1, which could improve the endonuclease step of CRISPR, could become an alternative to Cas9 in the CRISPR/Cas9 protein complex showing the constant improvement of gene editing methods [130].

Forward edgetics starts with mutation linked to a phenotype, with the goal of identifying an interaction [126]. By contrast, reverse edgetics refers to approaches in which one starts from an interaction to arrive at a phenotype. Network analysis can help to connect reverse and forward edgetics by isolating important interactions linked to phenotypes of interest (e.g., stress responses based on transcriptomics data). Then, by removing a specific interaction, the phenotypic response can be confirmed under a specific condition.

4. Prospective and integrative networks

To fully integrate these different types of analyses, it will also be important to integrate prioritization strategies. However, in addition to focusing on networks inferred from expression data, it might be interesting to use molecular networks to study how the genes within a molecular network behave at the transcript or protein level in the genomic region of interest.

To use molecular networks, the challenge is in their creation. This requires substantial experimental efforts. A possible shortcut consists in transferring knowledge from one well-studied model such as *Arabidopsis* to crop species. Researchers could take advantage from the syntenic blocs and co-expression patterns as described earlier. In this scope, gene regulatory network, composed by transcription factors and their target genes, will be more complex to transfer from *Arabidopsis* to crop species. This is due to the fact that transcription factors have a higher rate of gene duplications and gene families expansions.

There is an urgent necessity to develop methods that take advantage of gene expression data without forgetting or ignoring all of the knowledge generated by conventional and molecular breeding. In our view, the future of gene/protein network analysis relies on the integration of these methods with other data sources. For example, molecular networks, transcriptomic data, and genetic map information from QTL and genome-wide association studies (GWAS) analyses can be integrated with each other, as already described in the literature [131]. This combination of data should significantly reduce the false positive rate inherent to each individual method.

In order to take advantage of small-effect loci found in QTL analyses, which are important parts of the responses to most abiotic stresses, it is important to connect loci selection based on genetic architecture with molecular regulatory network candidate genes. Linear models could constitute interesting methods to study the importance of a set of genes in the regulation of a complex process. Obviously to use such models, several data types needs to be available for a specific condition (e.g. drought stress). Otherwise, these

data need to be created to obtain quantitative genetic data, gene expression or molecular networks. Another point will be to define the weight to give in a linear model to each data type in order to define the final result. By connecting these different evidence forms linked to a trait or a stress response, possible regulators could be identified (Fig. 3). Gaining this level of knowledge from networks will also help to specify the genes that need to be investigated in the progenitor or wild varieties of each crop. The newly discovered genes/loci could then be used to identify new markers that will be useful in molecular breeding. Once in hand, interesting leads can be manipulated at the interaction level (edgetics), rather than by overexpressing or removing the genes altogether, enabling a more precise analysis and control of the biological phenomena of interest.

References

[1] J.F. Doebley, B.S. Gaut, B.D. Smith, The molecular genetics of crop domestication, Cell 127 (2006) 1309–1321.
[2] S.I. Wright, I.V. Bi, S.G. Schroeder, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, The effects of artificial selection on the maize genome, Science 308 (2005) 1310–1314.
[3] B.S. Zheng, L. Yang, W.P. Zhang, C.Z. Mao, Y.R. Wu, K.K. Yi, F.Y. Liu, P. Wu, Mapping QTLs and candidate genes for rice root traits under different water-supply conditions and comparative analysis across three populations, Theor. Appl. Genet. 107 (2003) 1505–1515.
[4] R. Tuberosa, S. Salvi, Genomics-based approaches to improve drought tolerance of crops, Trends Plant Sci. 11 (2006) 405–412.
[5] B. Hirel, P. Bertin, I. Quilleré, W. Bourdoncle, C. Attagnant, C. Dellay, A. Gouy, S. Cadiou, C. Retailliau, M. Falque, A. Gallais, Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize, Plant Physiol. 125 (2001) 1258–1270.
[6] S. Pflieger, V. Lefebvre, M. Causse, The candidate gene approach in plant genetics: a review, Mol. Breed. 7 (2001) 275–291.
[7] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science 270 (1995) 467–470.
[8] S. Marguerat, J. Bahler, RNA-seq: from technology to biology, Cell. Mol. Life Sci.: CMLS 67 (2010) 569–579.
[9] S.C. Schuster, Next-generation sequencing transforms today's biology, Nat. Methods 5 (2008) 16–18.
[10] V. Cahais, P. Gayral, G. Tsagkogeorga, J. Melo-Ferreira, M. Ballenghien, L. Weinert, Y. Chiari, K. Belkhir, V. Ranwez, N. Galtier, Reference-free transcriptome assembly in non-model animals from next-generation sequencing data, Mol. Ecol. Res. 12 (2012) 834–845.
[11] S.A. Filichkin, H.D. Priest, S.A. Givan, R. Shen, D.W. Bryant, S.E. Fox, W.K. Wong, T.C. Mockler, Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*, Genom. Res. 20 (2010) 45–58.
[12] H. Kudapa, S. Azam, A.G. Sharpe, B. Taran, R. Li, B. Deonovic, C. Cameron, A.D. Farmer, S.B. Cannon, R.K. Varshney, Comprehensive transcriptome assembly of Chickpea (*Cicer arietinum* L.) using sanger and next generation sequencing platforms: development and applications, PLoS One 9 (2014) e86039.
[13] A.E. Loraine, S. McCormick, A. Estrada, K. Patel, P. Qin, RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing, Plant Physiol. 162 (2013) 1092–1109.
[14] J.A. Martin, Z. Wang, Next-generation transcriptome assembly, Nat. Rev. Genet. 12 (2011) 671–682.
[15] Z. Xia, H. Xu, J. Zhai, D. Li, H. Luo, C. He, X. Huang, RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*, Plant Mol. Biol. 77 (2011) 299–308.
[16] X. Yang, X.Y. Yu, Y.F. Li, De novo assembly and characterization of the Barnyardgrass (*Echinochloa crus-galli*) transcriptome using next-generation pyrosequencing, PLoS One 8 (2013) e69168.
[17] M. Schmid, T.S. Davison, S.R. Henz, U.J. Pape, M. Demar, M. Vingron, B. Scholkopf, D. Weigel, J.U. Lohmann, A gene expression map of *Arabidopsis thaliana* development, Nat. Genet. 37 (2005) 501–506.
[18] J. Kilian, D. Whitehead, J. Horak, D. Wanke, S. Weinl, O. Batistic, C. D'Angelo, E. Bornberg-Bauer, J. Kudla, K. Harter, The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses, Plant J. 50 (2007) 347–363.
[19] M. Libault, A. Farmer, T. Joshi, K. Takahashi, R.J. Langley, L.D. Franklin, J. He, D. Xu, G. May, G. Stacey, An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants, Plant J. 63 (2010) 86–99.
[20] S.M. Brady, D.A. Orlando, J.Y. Lee, J.Y. Wang, J. Koch, J.R. Dinneny, D. Mace, U. Ohler, P.N. Benfey, A high-resolution root spatiotemporal map reveals dominant expression patterns, Science 318 (2007) 801–806.
[21] Y. Jiao, S.L. Tausta, N. Gandotra, N. Sun, T. Liu, N.K. Clay, T. Ceserani, M. Chen, L. Ma, M. Holford, H.Y. Zhang, H. Zhao, X.W. Deng, T. Nelson, A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies, Nat. Genet. 41 (2009) 258–263.
[22] Y. Sato, H. Takehisa, K. Kamatsuki, H. Minami, N. Namiki, H. Ikawa, H. Ohyanagi, K. Sugimoto, B.A. Antonio, Y. Nagamura, RiceXPro version 3.0: expanding the informatics resource for rice transcriptome, Nucleic Acids Res. 41 (2013) D1206–D1213.
[23] R.S. Sekhon, H. Lin, K.L. Childs, C.N. Hansey, C.R. Buell, N. de Leon, S.M. Kaeppler, Genome-wide atlas of transcription during maize development, Plant J. 66 (2011) 553–563.
[24] D.A. Cartwright, S.M. Brady, D.A. Orlando, B. Sturmfels, P.N. Benfey, Reconstructing spatiotemporal gene expression data from partial observations, Bioinformatics 25 (2009) 2581–2587.
[25] M. Pautler, A.L. Eveland, T. LaRue, F. Yang, R. Weeks, C. Lunde, B.I. Je, R. Meeley, M. Komatsu, E. Vollbrecht, H. Sakai, D. Jackson, FASCIATED EAR4 encodes a bZIP transcription factor that regulates shoot meristem size in maize, Plant Cell 27 (2015) 104–120.
[26] A.L. Eveland, A. Goldshmidt, M. Pautler, K. Morohashi, C. Liseron-Monfils, M.W. Lewis, S. Kumari, S. Hiraga, F. Yang, E. Unger-Wallace, A. Olson, S. Hake, E. Vollbrecht, E. Grotewold, D. Ware, D. Jackson, Regulatory modules controlling maize inflorescence architecture, Genom. Res. 24 (2014) 431–443.
[27] H. Darmency, Pleiotropic effects of herbicide-resistance genes on crop yield: a review, Pest Manage. Sci. 69 (2013) 897–904.
[28] S.S. Nadakuduti, M. Pollard, D.K. Kosma, C. Allen Jr., J.B. Ohlrogge, C.S. Barry, Pleiotropic phenotypes of the sticky peel mutant provide new insight into the role of CUTIN DEFICIENT2 in epidermal cell function in tomato, Plant Physiol. 159 (2012) 945–960.
[29] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[30] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.L. Barabasi, The large-scale organization of metabolic networks, Nature 407 (2000) 651–654.
[31] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.L. Barabasi, Hierarchical organization of modularity in metabolic networks, Science 297 (2002) 1551–1555.
[32] M.W. Hahn, A.D. Kern, Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks, Mol. Biol. Evol. 22 (2005) 803–806.
[33] P. Bonacich, P. Lloyd, Eigenvector-like measures of centrality for asymmetric relations, Soc. Netw. 23 (2001) 191–201.
[34] M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, T. Ideker, Cytoscape 2. 8: new features for data integration and network visualization, Bioinformatics 27 (2011) 431–432.
[35] Z. Hu, J.H. Hung, Y. Wang, Y.C. Chang, C.L. Huang, M. Huyck, C. DeLisi, VisANT 3. 5: multi-scale network visualization, analysis and inference based on the gene ontology, Nucleic Acids Res. 37 (2009) W115–W121.
[36] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, A.f.t.A.o.A. Intelligence (Ed.) Third International AAAI Conference on Weblogs and Social Media (2009).
[37] G. Csardi, T. Nepusz, The igraph software package for complex network research, InterJ., Complex Syst. (2006) 1695.
[38] Y. Assenov, F. Ramirez, S.E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, Bioinformatics 24 (2008) 282–284.
[39] S. Brohee, K. Faust, G. Lima-Mendez, G. Vanderstocken, J. van Helden, Network analysis tools: from biological networks to clusters and pathways, Nat. Protoc. 3 (2008) 1616–1629.
[40] *Arabidopsis* Interactome Mapping Consortium, Evidence for network evolution in an *Arabidopsis* interactome map, Science 333 (2011) 601-607.
[41] S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, C. Jandrasits, R.C. Jimenez, J. Khadake, U. Mahadevan, P. Masson, I. Pedruzzi, E. Pfeiffenberger, P. Porras, A. Raghunath, B. Roechert, S. Orchard, H. Hermjakob, The IntAct molecular interaction database in 2012, Nucleic Acids Res. 40 (2012) D841–D846.
[42] T. Lee, S. Yang, E. Kim, Y. Ko, S. Hwang, J. Shin, J.E. Shim, H. Shim, H. Kim, C. Kim, I. Lee, AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species, Nucleic Acids Res. 43 (2015) D996–1002.
[43] T. Wallach, K. Schellenberg, B. Maier, R.K. Kalathur, P. Porras, E.E. Wanker, M.E. Futschik, A. Kramer, Dynamic circadian protein–protein interaction networks predict temporal organization of cellular functions, PLoS Genet. 9 (2013) e1003398.
[44] N. Bolduc, A. Yilmaz, M.K. Mejia-Guerra, K. Morohashi, D. O'Connor, E. Grotewold, S. Hake, Unraveling the KNOTTED1 regulatory network in maize meristems, Genes Dev. 26 (2012) 1685–1690.
[45] M. Taylor-Teeples, L. Lin, M. de Lucas, G. Turco, T.W. Toal, A. Gaudinier, N.F. Young, G.M. Trabucco, M.T. Veling, R. Lamothe, P.P. Handakumbura, G. Xiong, C. Wang, J. Corwin, A. Tsoukalas, L. Zhang, D. Ware, M. Pauly, D.J. Kliebenstein, K. Dehesh, I. Tagkopoulos, G. Breton, J.L. Pruneda-Paz, S.E. Ahnert, S.A. Kay, S.P. Hazen, S.M. Brady, An *Arabidopsis* gene regulatory network for secondary cell wall synthesis, Nature 517 (2015) 571–575.
[46] J. Boruc, H. Van den Daele, J. Hollunder, S. Rombauts, E. Mylle, P. Hilson, D. Inze, L. De Veylder, E. Russinova, Functional modules in the *Arabidopsis* core cell cycle binary protein–protein interaction network, Plant Cell 22 (2010) 1264–1280.
[47] R.A. Chávez Montes, H. Herrera-Ubaldo, J. Serwatowska, S. de Folter, Towards a comprehensive and dynamic gynoecium gene regulatory network, Curr. Plant Biol. (2015) (in Press).

[48] I. Thiele, B.O. Palsson, A protocol for generating a high-quality genome-scale metabolic reconstruction, Nat. Protoc. 5 (2010) 93–121.
[49] C.S. Henry, M. DeJongh, A.A. Best, P.M. Frybarger, B. Linsay, R.L. Stevens, High-throughput generation, optimization and analysis of genome-scale metabolic models, Nat. Biotechnol. 28 (2010) 977–982.
[50] S.M. Seaver, S. Gerdes, O. Frelin, C. Lerma-Ortiz, L.M. Bradbury, R. Zallot, G. Hasnain, T.D. Niehaus, B. El Yacoubi, S. Pasternak, R. Olson, G. Pusch, R. Overbeek, R. Stevens, V. de Crecy-Lagard, D. Ware, A.D. Hanson, C.S. Henry, High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 9645–9650.
[51] L. Chae, T. Kim, R. Nilo-Poyanco, S.Y. Rhee, Genomic signatures of specialized metabolism in plants, Science 344 (2014) 510–513.
[52] M. Freeling, Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition, Annu. Rev. Plant Biol. 60 (2009) 433–453.
[53] S.A. Teichmann, M.M. Babu, Gene regulatory network growth by duplication, Nat. Genet. 36 (2004) 492–496.
[54] J.S. Mattick, M.J. Gagen, Accelerating networks, Science 307 (2005) 856–858.
[55] H.M. Ku, T. Vision, J. Liu, S.D. Tanksley, Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 9121–9126.
[56] A. Wagner, D.A. Fell, The small world inside large metabolic networks, Proc. Biol. Sci. 268 (2001) 1803–1810.
[57] S.H. Shiu, M.C. Shih, W.H. Li, Transcription factor families have much higher expansion rates in plants than in animals, Plant Physiol. 139 (2005) 18–26.
[58] S.H. Choi, Y. Hyeon do, L.H. Lee, S.J. Park, S. Han, I.C. Lee, D. Hwang, H.G. Nam, Gene duplication of type-B ARR transcription factors systematically extends transcriptional regulatory structures in *Arabidopsis*, Sci. Rep. 4 (2014) 7197.
[59] A. Yilmaz, M.K. Mejia-Guerra, K. Kurz, X. Liang, L. Welch, E. Grotewold, AGRIS: the *Arabidopsis* gene regulatory information server, an update, Nucleic Acids Res. 39 (2011) D1118–D1122.
[60] P. Lamesch, T.Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D.L. Alexander, M. Garcia-Hernandez, A.S. Karthikeyan, C.H. Lee, W.D. Nelson, L. Ploetz, S. Singh, A. Wensel, E. Huala, The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools, Nucleic Acids Res. 40 (2012) D1202–D1210.
[61] A. Chatr-Aryamontri, B.J. Breitkreutz, S. Heinicke, L. Boucher, A. Winter, C. Stark, J. Nixon, L. Ramage, N. Kolas, L. O'Donnell, T. Reguly, A. Breitkreutz, A. Sellam, D. Chen, C. Chang, J. Rust, M. Livstone, R. Oughtred, K. Dolinski, M. Tyers, The BioGRID interaction database: 2013 update, Nucleic Acids Res. 41 (2013) D816–D823.
[62] K. Toufighi, S.M. Brady, R. Austin, E. Ly, N.J. Provart, The botany array resource: e-northerns, expression angling, and promoter analyses, Plant J. 43 (2005) 153–163.
[63] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinform. 9 (2008) 559.
[64] M.V. DiLeo, G.D. Strahan, M. den Bakker, O.A. Hoekenga, Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome, PLoS One 6 (2011) e26683.
[65] G.S. Downs, Y.M. Bi, J. Colasanti, W. Wu, X. Chen, T. Zhu, S.J. Rothstein, L.N. Lukens, A developmental transcriptional network for maize defines coexpression modules, Plant Physiol. 161 (2013) 1830–1843.
[66] C. Ma, X. Wang, Application of the Gini correlation coefficient to infer regulatory relationships in transcriptome analysis, Plant Physiol. 160 (2012) 192–203.
[67] J. Peng, P. Wang, N. Zhou, J. Zhu, Partial correlation estimation by joint sparse regression models, J. Am. Stat. Assoc. 104 (2009) 735–746.
[68] J. Schafer, K. Strimmer, An empirical Bayes approach to inferring large-scale gene association networks, Bioinformatics 21 (2005) 754–764.
[69] I. Priness, O. Maimon, I. Ben-Gal, Evaluation of gene-expression clustering via mutual information distance measure, BMC Bioinform. 8 (2007) 111.
[70] R. Steuer, J. Kurths, C.O. Daub, J. Weise, J. Selbig, The mutual information: detecting and evaluating dependencies between variables, Bioinformatics 18 (2002) S231–S240.
[71] J. Hausser, K. Strimmer, Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks, J. Mach. Learn Res. 10 (2009) 1469–1484.
[72] K. Basso, A.A. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, A. Califano, Reverse engineering of regulatory networks in human B cells, Nat. Genet. 37 (2005) 382–390.
[73] A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, A. Califano, ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, BMC Bioinform. 7 (Suppl. 1) (2006) S7.
[74] J.B. Kinney, G.S. Atwal, Equitability, mutual information, and the maximal information coefficient, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 3354–3359.
[75] J.B. Kinney, G.S. Atwal, Reply to Reshef et al.: falsifiability or bust, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) E3364.
[76] H.M. Liu, N. Rao, D. Yang, L. Yang, Y. Li, F. Ou, A novel method for identifying SNP disease association based on maximal information coefficient, Genet. Mol. Res.: GMR 13 (2014) 10863–10877.
[77] D.N. Reshef, Y.A. Reshef, M. Mitzenmacher, P.C. Sabeti, Cleaning up the record on the maximal information coefficient and equitability, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) E3362–E3363.
[78] M. Bansal, D. di Bernardo, Inference of gene networks from temporal gene expression profiles, IET Syst. Biol. 1 (2007) 306–312.
[79] S.R. Palle, C.M. Seeve, A.J. Eckert, W.P. Cumbie, B. Goldfarb, C.A. Loopstra, Natural variation in expression of genes involved in xylem development in loblolly pine (*Pinus taeda* L.), Tree Genet. Genom. 7 (2010) 193–206.
[80] J.D. Allen, Y. Xie, M. Chen, L. Girard, G. Xiao, Comparing statistical methods for constructing large scale gene networks, PLoS One 7 (2012) e29348.
[81] A. Garg, K. Mohanram, G. De Micheli, I. Xenarios, Implicit methods for qualitative modeling of gene regulatory networks, Methods Mol. Biol. 786 (2012) 397–443.
[82] F. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, The yeast cell-cycle network is robustly designed, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 4781–4786.
[83] D.A. Orlando, C.Y. Lin, A. Bernard, J.Y. Wang, J.E. Socolar, E.S. Iversen, A.J. Hartemink, S.B. Haase, Global control of cell-cycle transcription by coupled CDK and network oscillators, Nature 453 (2008) 944–947.
[84] S. Dash, J. Van Hemert, L. Hong, R.P. Wise, J.A. Dickerson, PLEXdb: gene expression resources for plants and plant pathogens, Nucleic Acids Res. 40 (2012) D1194–D1201.
[85] T. Hruz, O. Laule, G. Szabo, F. Wessendorp, S. Bleuler, L. Oertle, P. Widmayer, W. Gruissem, P. Zimmermann, Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes, Adv. Bioinform. 2008 (420747) (2008).
[86] T. Obayashi, Y. Okamura, S. Ito, S. Tadaka, Y. Aoki, M. Shirota, K. Kinoshita, ATTED-II in 2014: evaluation of gene coexpression in agriculturally important plants, Plant Cell Physiol. 55 (2014) e6.
[87] J. Geisler-Lee, N. O'Toole, R. Ammar, N.J. Provart, A.H. Millar, M. Geisler, A predicted interactome for *Arabidopsis*, Plant Physiol. 145 (2007) 317–329.
[88] T. Obayashi, K. Nishida, K. Kasahara, K. Kinoshita, ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants, Plant Cell Physiol. 52 (2011) 213–219.
[89] M.Y. Hirai, M. Klein, Y. Fujikawa, M. Yano, D.B. Goodenowe, Y. Yamazaki, S. Kanaya, Y. Nakamura, M. Kitayama, H. Suzuki, N. Sakurai, D. Shibata, J. Tokuhisa, M. Reichelt, J. Gershenzon, J. Papenbrock, K. Saito, Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics, J. Biol. Chem. 280 (2005) 25590–25595.
[90] M.Y. Hirai, K. Sugiyama, Y. Sawada, T. Tohge, T. Obayashi, A. Suzuki, R. Araki, N. Sakurai, H. Suzuki, K. Aoki, H. Goda, O.I. Nishizawa, D. Shibata, K. Saito, Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 6478–6483.
[91] J.J. Petricka, M.A. Schauer, M. Megraw, N.W. Breakfield, J.W. Thompson, S. Georgiev, E.J. Soderblom, U. Ohler, M.A. Moseley, U. Grossniklaus, P.N. Benfey, The protein expression landscape of the *Arabidopsis* root, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 6811–6818.
[92] S. Mangan, U. Alon, Structure and function of the feed-forward loop network motif, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 11980–11985.
[93] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (2002) 824–827.
[94] P.J. Ingram, M.P. Stumpf, J. Stark, Network motifs: structure does not determine function, BMC Genom. 7 (2006) 108.
[95] S. Widder, R. Sole, J. Macia, Evolvability of feed-forward loop architecture biases its abundance in transcription networks, BMC Syst. Biol. 6 (2012) 7.
[96] D. Siegal-Gaskins, M.K. Mejia-Guerra, G.D. Smith, E. Grotewold, Emergence of switch-like behavior in a large family of simple biochemical networks, PLoS Comput. Biol. 7 (2011) e1002039.
[97] M.B. Elowitz, S. Leibler, A synthetic oscillatory network of transcriptional regulators, Nature 403 (2000) 335–338.
[98] E. Wong, B. Baur, S. Quader, C.H. Huang, Biological network motif detection: principles and practice, Briefings Bioinform. 13 (2012) 202–215.
[99] J.B. Holland, Genetic architecture of complex traits in plants, Curr. Opin. Plant Biol. 10 (2007) 156–161.
[100] R. Bernardo, What if we knew all the genes for a quantitative trait in hybrid crops? Crop Sci. 41 (2001) 1–4.
[101] E.S. Buckler, J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J.C. Glaubitz, M.M. Goodman, C. Harjes, K. Guill, D.E. Kroon, S. Larsson, N.K. Lepak, H. Li, S.E. Mitchell, G. Pressoir, J.A. Peiffer, M.O. Rosas, T.R. Rocheford, M.C. Romay, S. Romero, S. Salvo, H. Sanchez Villeda, H.S. da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, M.D. McMullen, The genetic architecture of maize flowering time, Science 325 (2009) 714–718.
[102] K. Maruyama, Y. Sakuma, M. Kasuga, Y. Ito, M. Seki, H. Goda, Y. Shimada, S. Yoshida, K. Shinozaki, K. Yamaguchi-Shinozaki, Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems, Plant J. 38 (2004) 982–993.
[103] J.J. Michaelson, S. Loguercio, A. Beyer, Detection and interpretation of expression quantitative trait loci (eQTL), Methods 48 (2009) 265–276.
[104] C. Gene Ontology, The gene ontology project in 2008, Nucleic Acids Res. 36 (2008) D440–D444.
[105] S.P. Ficklin, F.A. Feltus, Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice, Plant Physiol. 156 (2011) 1244–1256.

[106] E. Prifti, J.D. Zucker, K. Clement, C. Henegar, Interactional and functional centrality in transcriptional co-expression networks, Bioinformatics 26 (2010) 3083–3089.
[107] K. Mochida, Y. Uehara-Yamaguchi, T. Yoshida, T. Sakurai, K. Shinozaki, Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops, Plant Cell Physiol. 52 (2011) 785–803.
[108] L.C. Tranchevent, F.B. Capdevila, D. Nitsch, B. De Moor, P. De Causmaecker, Y. Moreau, A guide to web tools to prioritize candidate genes, Briefings Bioinform. 12 (2011) 22–32.
[109] D. Warde-Farley, S.L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C.T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G.D. Bader, Q. Morris, The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function, Nucleic Acids Res. 38 (2010) W214–W220.
[110] S. De Bodt, J. Hollunder, H. Nelissen, N. Meulemeester, D. Inze, CORNET 2.0: integrating plant coexpression, protein–protein interactions, regulatory interactions, gene associations and functional annotations, New Phytol. 195 (2012) 707–720.
[111] J. Gillis, P. Pavlidis, Guilt by association is the exception rather than the rule in gene networks, PLoS Comput. Biol. 8 (2012) e1002444.
[112] S.E. Ahnert, Generalised power graph compression reveals dominant relationship patterns in complex networks, Sci. Rep. 4 (4385) (2014).
[113] R. Pastor-Satorras, E. Smith, R.V. Solé, Evolving protein interaction networks through gene duplication, J. Theor. Biol. 222 (2003) 199–210.
[114] G. Blanc, K.H. Wolfe, Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution, Plant Cell 16 (2004) 1679–1691.
[115] Y. Gibon, B. Usadel, O.E. Blaesing, B. Kamlage, M. Hoehne, R. Trethewey, M. Stitt, Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in *Arabidopsis* rosettes, Genom. Biol. 7 (2006) R76.
[116] S.L. Carter, C.M. Brechbuhler, M. Griffin, A.T. Bond, Gene co-expression network topology provides a framework for molecular characterization of cellular state, Bioinformatics 20 (2004) 2242–2250.
[117] H. Yu, P.M. Kim, E. Sprecher, V. Trifonov, M. Gerstein, The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics, PLoS Comput. Biol. 3 (2007) e59.
[118] H. Jeong, S.P. Mason, A.L. Barabasi, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001) 41–42.
[119] J. Song, M. Singh, From hub proteins to hub modules: the relationship between essentiality and centrality in the yeast interactome at different scales of organization, PLoS Comput. Biol. 9 (2013) e1002910.
[120] J.D. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J. Walhout, M.E. Cusick, F.P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein–protein interaction network, Nature 430 (2004) 88–93.
[121] S. Agarwal, C.M. Deane, M.A. Porter, N.S. Jones, Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks, PLoS Comput. Biol. 6 (2010) e1000817.
[122] M.C. Palumbo, S. Zenoni, M. Fasoli, M. Massonnet, L. Farina, F. Castiglione, M. Pezzotti, P. Paci, Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development, Plant Cell 26 (2014) 4617–4635.
[123] P. Langfelder, P.S. Mischel, S. Horvath, When is hub gene selection better than standard meta-analysis? PLoS One 8 (2013) e61505.
[124] C.C. Garbutt, P.V. Bangalore, P. Kannar, M.S. Mukhtar, Getting to the edge: protein dynamical networks as a new frontier in plant-microbe interactions, Front. Plant Sci. 5 (2014) 312.
[125] N. Sahni, S. Yi, Q. Zhong, N. Jailkhani, B. Charloteaux, M.E. Cusick, M. Vidal, Edgotype: a fundamental link between genotype and phenotype, Curr. Opin. Genet. Dev. 23 (2013) 649–657.
[126] B. Charloteaux, Q. Zhong, M. Dreze, M.E. Cusick, D.E. Hill, M. Vidal, Protein–protein interactions and networks: forward and reverse edgetics, Methods Mol. Biol. 759 (2011) 197–213.
[127] T. Nepusz, T. Vicsek, Controlling edge dynamics in complex networks, Nat. Phys. 8 (2012) 568–573.
[128] W. Jiang, H. Zhou, H. Bi, M. Fromm, B. Yang, D.P. Weeks, Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice, Nucleic Acids Res. 41 (2013) e188.
[129] K. Chen, C. Gao, Targeted genome modification technologies and their applications in crop improvements, Plant Cell Rep. 33 (2014) 575–583.
[130] B. Zetsche, J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, S.E. Volz, J. Joung, J. van der Oost, A. Regev, E.V. Koonin, F. Zhang, Cpf1 is a single RNA-guided endonuclease of a class 2CRISPR-Cas system, Cell 163 (2015) 759–771.
[131] T. Lee, H. Kim, I. Lee, Network-assisted crop systems genetics: network inference and integrative analysis, Curr. Opin. Plant Biol. 24 (2015) 61–70.