

GENOMICS OF HYBRIDIZATION

Insights into the Ecology and Evolution of Polyploid Plants through Network Analysis

JOSEPH P. GALLAGHER,^a CORRINNE E. GROVER,^a GUANJING HU and JONATHAN F. WENDEL

Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA

Abstract

Polyploidy is a widespread phenomenon throughout eukaryotes, with important ecological and evolutionary consequences. Although genes operate as components of complex pathways and networks, polyploid changes in genes and gene expression have typically been evaluated as either individual genes or as a part of broad-scale analyses. Network analysis has been fruitful in associating genomic and other 'omic'-based changes with phenotype for many systems. In polyploid species, network analysis has the potential not only to facilitate a better understanding of the complex 'omic' underpinnings of phenotypic and ecological traits common to polyploidy, but also to provide novel insight into the interaction among duplicated genes and genomes. This adds perspective to the global patterns of expression (and other 'omic') change that accompany polyploidy and to the patterns of recruitment and/or loss of genes following polyploidization. While network analysis in polyploid species faces challenges common to other analyses of duplicated genomes, present technologies combined with thoughtful experimental design provide a powerful system to explore polyploid evolution. Here, we demonstrate the utility and potential of network analysis to questions pertaining to polyploidy with an example involving evolution of the transgressively superior cotton fibres found in polyploid *Gossypium hirsutum*. By combining network analysis with prior knowledge, we provide further insights into the role of profilins in fibre domestication and exemplify the potential for network analysis in polyploid species.

Keywords: co-expression, gene expression, gene loss, gene retention, network analysis polyploidy

Received 24 November 2015; revision received 9 March 2016; accepted 22 March 2016

Polyploidy is a widespread phenomenon throughout eukaryotes, with important ecological and evolutionary consequences (Stebbins 1940; Levin 1983; Ramsey & Schemske 1998, 2002; Leitch & Leitch 2008; Van de Peer *et al.* 2009; Matsushita *et al.* 2012; Soltis *et al.* 2014). Although both recent and ancient polyploidy events have been identified in animals and fungi, modern polyploidy is not a particularly active process in these major clades (Muller 1925; Orr 1990; Albertin & Marullo 2012). Conversely, the importance of polyploidy to modern plant species is difficult to overstate. It is now

recognized that all flowering plants have experienced multiple rounds of polyploidy at some point in their ancestry (Bowers *et al.* 2003; Wood *et al.* 2009; Jiao *et al.* 2011) and that it remains an active evolutionary and ecological process in many lineages (Ramsey 2011; Soltis & Soltis 2012; Ramsey & Ramsey 2014; McAllister *et al.* 2015; Wendel 2015).

From a genomic perspective, the myriad consequences of polyploidy for the various 'omes' have been broadly evaluated for a number of model angiosperm genera (Wang *et al.* 2004; Pang *et al.* 2009; Chelaifa *et al.* 2010; Bao *et al.* 2011; Dong & Adams 2011; Ha *et al.* 2011; Xiong *et al.* 2011; Chester *et al.* 2012; Flagel *et al.* 2012; Koh *et al.* 2012; Ng *et al.* 2012; Paterson *et al.* 2012; Hu *et al.* 2013, 2015; Page *et al.*

Correspondence: Jonathan F. Wendel, Fax: +1 515-294-1337; E-mail: jfw@iastate.edu

^aThese two authors contributed equally to this work.

2013b; Yoo *et al.* 2013; Chalhoub *et al.* 2014; Coate *et al.* 2014; Guan *et al.* 2014; International Wheat Genome Sequencing Consortium 2014; Li *et al.* 2014, 2015; Sehrish *et al.* 2014; Xu *et al.* 2014; Zhang *et al.* 2015b). Among the notable phenomena are intergenomic exchanges between genomes that once were isolated in divergent progenitor diploids (Salmon *et al.* 2010; Wang & Paterson 2011; Flagel *et al.* 2012; Guo *et al.* 2014); biased loss and/or fractionation of duplicated genes (hereafter 'homoeologs') (Ozkan *et al.* 2001; Thomas *et al.* 2006; Woodhouse *et al.* 2010; Tang *et al.* 2012); and a variety of types of changes in gene expression arising with and following polyploidy (Comai *et al.* 2000; Adams *et al.* 2003; Wang *et al.* 2004; Buggs *et al.* 2009, 2010; Rapp *et al.* 2009; Chelaifa *et al.* 2010; Flagel & Wendel 2010; Flagel *et al.* 2012; Shi *et al.* 2012; Coate *et al.* 2014). There also is deep interest in the relationship between these varied phenomena and biological diversification (Ramsey & Schemske 1998; Soltis & Soltis 1999, 2012; Ramsey 2011).

To date, alterations in genes and gene expression arising from polyploidy have been evaluated most commonly on a gene-by-gene basis (e.g. Adams *et al.* 2003), even when these are aggregated on a genome-wide scale (e.g. Yoo *et al.* 2013) for purposes of generalization. That is, comparisons have often been made between levels of gene expression between diploids and their polyploid derivatives, for any number of genes within the genomes in question. Because genes do not function in isolation, but as components of complex biological networks, it also is important to study gene expression of networks and pathways, as exemplified for plants in recent studies (Ni *et al.* 2009; Chang *et al.* 2010; Bekaert *et al.* 2011; Coate *et al.* 2013; Pfeifer *et al.* 2014). These analyses follow on the pioneering consideration of network responses to polyploidy in yeast (Wagner 2001; Conant & Wolfe 2006; van Hoek & Hogeweg 2009; Qian *et al.* 2011; Makino & McLysaght 2012).

Until recently, the analysis of biological networks was unrealistic for most nonmodel organisms; however, technological advances such as RNA-seq and other molecular genetic methodologies, combined with conceptual and computational advances in gene co-expression network reconstruction, have made network analysis an increasingly attractive tool for biological discovery. A gene co-expression network is a map of gene expression correlation among samples; nodes in the network represent genes or transcripts, while connections represent expression relationships (Albert 2005; Langfelder & Horvath 2008). An important advantage of gene co-expression networks is the statistical assignment of genes to clusters, called modules. These co-expression modules become more robust with a greater number and higher diversity of samples across tissues

and time points within an organism. With appropriate depth of sampling, comparisons of networks between species, ecotypes or genotypes may reveal changes in modules that explain interesting phenotypic differences. Consequently, we now have the potential to better understand the complex 'omic' underpinnings of phenotypic and ecological traits, as exemplified by recent studies of maize domestication (Swanson-Wagner *et al.* 2012), tomato leaf variability (Ichihashi *et al.* 2014) and environmental regulation of gene expression in pines (Cañas *et al.* 2015).

Modern high-throughput techniques also facilitate large-scale identification of other molecular components (e.g. genes, RNAs, proteins, metabolites, epigenetic marks), as well as characterization of their expression patterns and interactions. Consequently, various complex data sets are frequently generated to answer ecological, evolutionary and functional questions. The challenge then becomes how to distil these large multi-dimensional data sets into biologically informative conclusions. Network analysis is designed to capture interactions and dependencies among components, often independent of prior knowledge and the interactions of individual components. For nonmodel species, this is particularly attractive in that it permits genome-scale analysis of ecological or evolutionary traits among conditions, taxa or developmental stages, as well as the identification of gene modules that are likely to be functionally related. Although these correlations are based on the indirect evidence of relationship, strong expression correlation over multiple tissues/stages/etc. is often indicative of functional relationships (e.g. genes activated by promoters with similar regulatory elements). Furthermore, conservation of co-expression patterns among species could indicate the conservation of functional relationships (van Noort *et al.* 2003).

Analyses of biological networks hold promise for providing insights into the dynamics and resolution of polyploid genomes (Conant & Wolfe 2006; Bekaert *et al.* 2011; Pfeifer *et al.* 2014), particularly when network analyses are integrated with other 'omics' data sets. Network-based analyses in nonplant models have already demonstrated the utility of gene co-expression networks to reveal functional changes in genes and gene modules (van Noort *et al.* 2003; Conant & Wolfe 2006). Similar analyses can be used to provide further insight into the age-old question of how two (or more) diverged genomes function in a common nucleus subsequent to merger (at the time of hybridization) and following genome doubling, ultimately enhancing our understanding of polyploid ecology and evolution. We introduce this perspective here, by reviewing the application of network analysis to polyploid research and the layering of additional data types to yield novel

biological insights. We provide a brief empirical example of network analysis from our ongoing work in the cotton genus (*Gossypium*), drawing distinctions between the kinds of insights derived from traditional approaches to gene expression analysis and those derived from network tools.

Polyploidy and the prospects of gene network analysis

Network analysis in polyploid species has the capability of providing new perspectives into the interaction among duplicated genes and genomes and the changes that accompany polyploidy. These include patterns of genome-wide gene loss/retention and gene expression changes across the transcriptome. Ultimately, this analysis could disentangle some of the underlying forces that govern polyploid evolution and ecology. Here, we illustrate how network analysis could expand our current knowledge on polyploidy, focusing on gene retention, gene expression alterations and phenotypic changes.

Gene retention and loss following polyploidy

One outstanding question concerns the dynamics of duplicate gene loss vs. retention. This question has been considered from several perspectives, including selection related to broad functional categories (De Smet *et al.* 2013), gene dosage demands (Birchler & Veitia 2012) and preferential retention of homoeologs from the less fractionated of two genomes (Schnable *et al.* 2011). Each of these possibilities could be enlightened by a deeper consideration of genic interactions in biological pathways and networks, which necessarily entail a number of related functional or mechanistic constraints. Broad characterizations of the types of genes preferentially retained in duplicate, such as transcription and signalling-related genes in *Arabidopsis* (Blanc & Wolfe 2004; Seoighe & Gehring 2004; Maere *et al.* 2005), structural genes in the Compositae (Barker *et al.* 2008), or those returned to singleton status, such as photosynthesis-related genes (De Smet *et al.* 2013), provide evolutionary clues into the dynamics of duplicate gene expression and retention following polyploidy, but of necessity 'paint with a broad brush'. It seems likely that the compression of complex information into generalized categories conceals interactions among genes that influence retention and loss. We note that some explanations for patterns of duplicate gene loss invoke interaction among genes as important for gene retention or loss. The gene balance hypothesis, for example, posits that proteins that contribute to a multiunit complex are selectively maintained in the correct ratios, as imbalance among components of multiprotein complexes could be deleterious (Papp *et al.* 2003;

Birchler *et al.* 2005; Birchler & Veitia 2007, 2012; Conant *et al.* 2014). This concept has been extended to polyploid species, where the entire genic complement has been duplicated; that is, duplicate gene retention may be influenced by stoichiometric or 'balance' considerations for the *duplicated* members of multiprotein complexes. The application of the gene balance hypothesis, however, requires knowledge of interactions among the encoded proteins. For example, Makino & McLysaght (2012) showed that, in human, yeast and *Arabidopsis*, there is an enrichment of genes that have protein interactions with other genes in the same conserved block, as opposed to protein interactions across blocks. Their proximity and interaction suggest that these enriched genes might be coregulated, and are therefore in keeping with the gene balance hypothesis.

An added layer of complexity arises when the duplicated genes are nonidentical, as is the case for many or most genes in allopolyploids, polyploids derived from two divergent genomes. In this case, there may be sub-optimal interactions among the protein products from the newly combined genomes, leading to fitness differentials among multi-subunit proteins derived from homo-genomic vs. hetero-genomic protein assembly. In principle, this may result in preferential removal or silencing of specific homoeologs or even a cascade of gene loss sufficient to reduce maladaptive interactions and maintain stoichiometric (gene) balance (Chang *et al.* 2010; Birchler & Veitia 2012). Insights into these possibilities are best facilitated when genes and their protein products are viewed in the context of their functional connections, or networks. An example of such an approach is provided by Bekaert *et al.* (2011), who used the *Arabidopsis thaliana* metabolic network, a map of functional interactions, to look at gene retention and loss following ancient polyploidy events; one of their findings was that genes retained in duplicate from the most recent whole-genome duplication were clustered in the network, that is interacting with each other, as would be expected according to the gene balance hypothesis.

Duplicate gene expression after polyploidy

One of the important realizations about polyploids is that gene expression is often massively altered relative to diploid progenitors (see Introduction). While gene *retention* is fundamentally binary (retained or not, the latter including the full spectrum of silencing mutations), gene *expression* is quantitative. Moreover, the aggregate expression of both homoeologs is of interest, not just the expression of individual members of each duplicate gene pair. In this respect, duplicate gene expression may variously be transgressive relative to the progenitor diploids, additive (or average), or in some sense

mimic only one of the two parents (Yoo *et al.* 2014). Broad patterns of expression alteration may hint at mechanistic causes of these changes, given predicted functional information of up- or downregulated genes, homoeolog usage, etc.; however, here again information regarding interactions among genes holds promise for providing insight into mechanisms of expression alteration.

Classically, the causes of duplicate gene retention have entailed some form of 'subfunctionalization' or 'neofunctionalization' (Force *et al.* 1999; Lynch & Force 2000), although these terms are not mutually exclusive nor wholly sufficient (Conant *et al.* 2014). Neofunctionalization traditionally invokes a novel and presumably adaptive function for one homoeolog arising post-duplication, while the other homoeolog(s) maintains its ancestral function (Stephens 1951; Ohno 1970), whereas subfunctionalization traditionally occurs as regulatory divergence, such that the ancestral aggregate expression becomes partitioned among homoeologs in the relevant tissues and/or stages (Force *et al.* 1999; Lynch & Force 2000). In the context of polyploidy, an important early observation was that of Adams *et al.* (2003), who discovered reciprocal silencing of alternate homoeologs in various tissues from tetraploid cotton, demonstrating that subfunctionalization may be rapidly established following polyploidization. Similar work has since been conducted on a genome-wide scale for a number of species (Duarte *et al.* 2006; Roulin *et al.* 2013; Hughes *et al.* 2014; Renny-Byfield *et al.* 2014); in some cases, the data are more suggestive of neofunctionalization, although sub- and neofunctionalization may be hard to disentangle in many cases (Conant & Wolfe 2008; Conant *et al.* 2014). In the context of the present study the connectivity of individual homoeologs of different gene pairs may be an important consideration, one that only becomes evident using network approaches. For example, Pfeifer *et al.* (2014) generated a co-expression network for bread wheat, grouping all homoeologs into a single gene node to construct the network, and subsequently evaluating the contribution of each homoeolog to the expression pool for various cell types and developmental stages. Interestingly, they found certain network modules exhibit unbalanced expression bias, that is an overall favouring of expression of homoeologs derived from one parent (Grover *et al.* 2012), which could be associated with function (based on Gene Ontology categorization) and tissue type. This observation, that functional modules may become biased towards a specific homoeologous genome, has broad implications for understanding the evolution of polyploid species. For example, there may be phenotypic consequences that arise from preferential utilization of functionally related genes derived from only one homoeologous genome (e.g. phenotypic

similarity to one parent for a given trait). This preferential utilization or functional differentiation of individual homoeologous networks may be revealed by conducting separate network analyses for each constituent set of homoeologs, rather than on data where gene expression for each duplicate gene pair has been lumped prior to analysis. One can envision that such analyses might reveal several phenomena, including: (i) an overall bias towards one parental genome in a reconstructed gene network; (ii) subfunctionalization of networks, such that the alternate homoeologous networks are used in different tissues/stages, in full or in part; and (iii) network neofunctionalization (all, or some) whereby one set of homoeologs from the same progenitor genome participates in a separate, novel, pathway. Phenomena such as these have been described in *Saccharomyces cerevisiae*, where co-expression networks appear to have partitioned for ancient sets of paralogs (Conant & Wolfe 2006), and in the neoallopolyploid plant *Arabidopsis suecica*, where genes with more highly correlated expression were more often derived from the same parent of origin (Chang *et al.* 2010).

Yet another mechanism that could modulate polyploid gene expression is epigenetic regulation (Madlung & Wendel 2013). Several high-throughput methods (e.g. bisulfite sequencing, ChIP-seq, DNase I hypersensitivity assays) extend the convenience of next-generation sequencing to the epigenetic arena and permit genome-wide surveys of DNA methylation, histone modifications and chromatin state, all of which have the potential to affect gene expression (Deal & Henikoff 2011; Meyer 2011; Furey 2012; Tsompana & Buck 2014; Zhang *et al.* 2014). Similarly, small RNAs (e.g. siRNAs and miRNAs), which also function in expression regulation, are becoming targets of next-generation sequencing as well (Abrouk *et al.* 2012; Gong *et al.* 2013; Li *et al.* 2014). When these features differ between parental genomes, expression of each homoeolog in the polyploid can diverge from that of their parents (Chelaifa *et al.* 2010; Buggs *et al.* 2011; Xu *et al.* 2014). Gene co-expression networks that show altered expression patterns between parental species and polyploids may be caused by these epigenetic effects.

Each of the foregoing aspects of duplicate gene expression feed into the short- and long-term patterns of gene loss and retention following the origin of a polyploid species or clade. A somewhat mysterious observation in this respect is the phenomenon of 'biased fractionation' (Woodhouse *et al.* 2010), whereby a polyploid lineage exhibits biased loss of genes from one progenitor genome as it returns to a diploid state. First demonstrated in *Arabidopsis* (Thomas *et al.* 2006), biased fractionation has been observed for a number of other taxa (Schnable *et al.* 2011; Tang *et al.* 2012; Renny-

Byfield *et al.* 2015), suggesting that this may be a common outcome of polyploidy. The causes underlying biased fractionation are not well-understood, but one can imagine how the interplay of phenomena such as gene dosage, maladaptive interactions, biased or novel expression patterns, and others introduced above may affect long-term gene retention bias. One may envision how an enhanced understanding of 'genes in context', that is, of regulatory interactions and of biological pathways and networks, may help elucidate the neutral and selective forces that govern loss and retention, and thereby affect phenotypes of relevance to the ecology and evolution of polyploids.

Evolutionary and ecological changes following polyploidy

While the genetic changes listed above are interesting in their own right, the ultimate goal of these studies is to discover the underlying changes that lead to ecologically or evolutionarily important phenotypic properties. A number of traits have been associated with polyploidy in plants, such as enlarged cell size, larger overall organism size and delayed development (i.e. the *gigas* effect); exploitation of new ecological niches; and physiological or biochemical novelty (Levin 1983; Fligel & Wendel 2009; Ramsey & Ramsey 2014; Soltis *et al.* 2014). For example, McIntyre (2012) showed that, for a set of *Claytonia* species of different ploidy levels, polyploids occupied a different niche compared with diploids. In *Achillea borealis*, Ramsey (2011) performed field transplant experiments showing that hexaploids had a distinct advantage over tetraploids in the xeric dune habitats in which they occur. Coate *et al.* (2013) showed an increased photoprotective capacity in *Glycine dolichocarpa* compared with its diploid progenitors, as well as identifying the genetic components that made that increased photoprotection possible. Importantly, these polyploidy-derived traits that allow for changes in habitat are the emergent consequences of the induced changes to different 'omes' and how they interact with the environment. These traits are not likely due to changes in the sequence or expression of a single gene or pair of homoeologs, but due to changes in a suite of genes and their interactions. The real goal of applying network analysis is to discover how all of these pieces interact to produce the interesting and unique phenotypes we see in polyploid plants.

Adding power to polyploid networks with prior knowledge and other data types

Polyploidization is accompanied by substantial rewiring of biological networks (De Smet & Van de Peer 2012). While gene co-expression networks can illuminate these

changes, the quality of the reconstructed network and the resulting inferences depends both upon the data used to generate the network and the specifics of the biological questions being asked. When considering the underlying transcriptional correlates of a particular phenotypic difference (e.g. fruit size, drought tolerance) between two genotypes, ecotypes or species, it is a common practice to conduct RNA-seq expression profiling experiments, and then subject the transcriptional data to differential expression analysis. Extending these analyses to co-expression networks can allow for further inference (see above, Fig. 1); however, more extensive and thoughtful sampling and the integration of prior knowledge or other data types can increase the inferential power of these networks even further.

The nuances behind network analysis design have been extensively discussed (Albert 2005; Horvath 2009; Krouk *et al.* 2013; Mitra *et al.* 2013). In general terms, the power of co-expression network analysis to provide insight depends largely on the question, on the extent of sampling, and on the amount of prior knowledge available (Krouk *et al.* 2013). A recent example from *Helianthus* exemplifies what can be learned from appropriately tailoring the experimental data to the biological question. Marchand *et al.* (2014) utilized gene expression data in *H. annuus* under nine hormonal treatments from seven time points to build a gene regulatory network (GRN) for drought stress in sunflowers, with a focus on an informed set of candidate genes. From this analysis, they: (i) uncovered hub genes for the drought stress GRN; (ii) discovered a role for nitrate transporters in regulating transpiration; and (iii) connected the abscisic acid-dependent and abscisic acid-independent pathways. Through this targeted approach, Marchand *et al.* were able to infer causal relationships involved in transcriptional regulation, in addition to co-expression patterns. This perspective sets the stage for further analyses, such as evaluating gene differentiation among *Helianthus* species and cultivars with various adaptations to drought stress. This approach can be applied to a wide range of experiments utilizing gene expression networks.

Gene co-expression networks are relatively straightforward to generate, and with the addition of prior knowledge and proper sampling, have successfully identified gene-to-gene connections related to phenotype and function. These co-expression relationships, however, reflect the results of a series of direct molecular interactions, for example protein-protein interaction, protein-DNA interaction, membership in metabolic networks and small RNA regulatory interactions. For example, consider a hypothetical protein-DNA interaction that results in joint upregulation of a set of genes. This coregulation may be due to the involvement of these genes in a multimeric complex, and hence direct

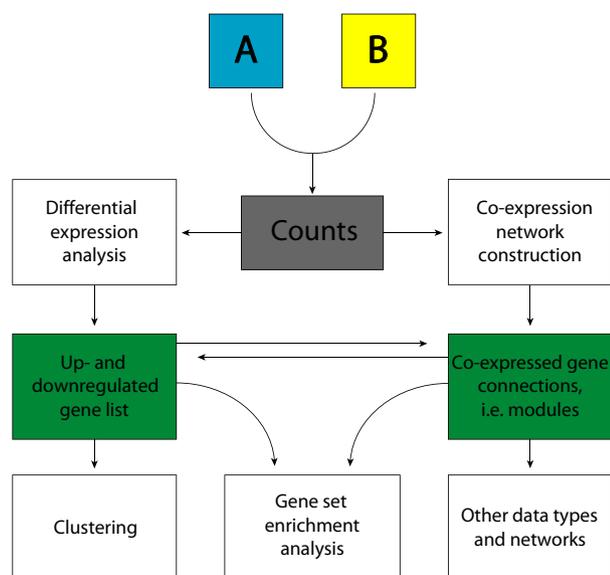


Fig. 1 Differences between traditional gene expression analysis and a network-focused approach. Blue box A and yellow box B represent organisms with divergent phenotypes. From these two organisms, high-throughput gene expression data are generated, processed, mapped and converted into counts per gene model (grey box). These data typically are subjected to differential expression analyses (left) and/or co-expression network construction (right). These two forms of analysis lead to complementary but different results: a list of up- and downregulated genes or a set of co-expression connections between genes, respectively (green boxes). Importantly, part of co-expression network construction statistically analyses these connections and groups them into modules of significantly connected genes. From these two sets of results, further analyses often are performed. Differentially expressed genes often are subjected to cluster analysis, similar to co-expression analysis, or are used for gene set enrichment, such as analysing the list for Gene Ontology term and KEGG pathway enrichment. The gene co-expression connection results typically are queried for gene set enrichment, specifically looking at enrichment within modules. In addition, other layers of connection information, such as protein–protein interaction data, may be stacked on top of the gene co-expression networks, allowing for further inferences of genes important for divergence of phenotypes. Of note, the differential expression data set may be one of these layers, providing information on expression modulation within the connected network.

coregulation of these components by the protein–DNA interaction complex; alternatively, this interaction may only be made possible when the right multimeric complex is present, which itself promotes co-upregulation of the whole pathway. Supplementing gene co-expression networks with this kind of information leads to more informative network reconstructions better capable of discerning the underlying biological interactions (Li & Jackson 2015; however, see Bloom & Adami 2003 for an example of the caveats of integrating data sets).

Protein–protein interaction networks have been used to consider the consequences of gene and genome duplications, most notably in *Arabidopsis thaliana*, where the effects of both small-scale and whole-genome duplications were evaluated during network construction (Arabidopsis Interactome Mapping Consortium 2011). The Arabidopsis Interactome Mapping Consortium found that shared protein–protein interactions between paralogs decreased rapidly for duplicates derived from both small-scale and genome-wide duplications, followed by a period of much slower decrease. This trend was mirrored by the divergence in protein sequence, suggesting that, following duplication, the rate of protein evolution and the maintenance of interacting partners are connected. Similarly, specific protein–DNA binding interactions may be assessed (e.g. ChIP-chip or ChIP-seq; Heyndrickx *et al.* 2014) and layered on gene co-expression networks (Angelini & Costa 2014). Other, less specific assays (e.g. DNase I hypersensitivity assay) can also provide information with respect to the protein-bound regions of the genome (e.g. Zhang *et al.* 2012), which can subsequently be used to inform more specific protein binding assays (Zhu *et al.* 2015). Biologically, these protein–DNA interactions may include different forms of transcriptional regulation, such as enrichment or depletion of histones or transcription factor binding sites; in the context of allopolyploidy, differences in protein–DNA interactions between duplicate genomes could contribute to novel phenotypes. Ha *et al.* (2011), for example, used ChIP-seq to show that homoeologous gene expression patterns were altered via histone modification differences in *Arabidopsis*. Similar experiments in other polyploid species could help determine the underpinnings of their unique phenotypes or adaptations.

Most sophisticated network analyses have been conducted in model organisms, such as yeast and human, where multiple, layered data types (e.g. RNA-seq plus epigenetic surveys, phenotypes, transcription factor binding sites, etc.) have led to increased understanding of gene regulatory network analyses (Madan Babu & Teichmann 2003; Gao *et al.* 2004; Zhu *et al.* 2008; Cookson *et al.* 2009; Mason *et al.* 2009; Bocklandt *et al.* 2011; Langfelder *et al.* 2012). In yeast, for example, integration of genotypic, expression, protein–protein interaction and transcription factor binding data led to the development of an extremely well-refined and useful gene regulatory network (Zhu *et al.* 2008). Through integration of well-constructed gene co-expression networks with prior knowledge and these other forms of interaction data, we may best address questions central to understanding the ecological success of polyploid lineages and their specific adaptations.

Empirical example from *Gossypium*

Network analyses have clear advantages in model systems because of the availability of genomic and other 'omic' resources and established experimental methods. In plants, the best developed models for most molecular biological experiments are *Arabidopsis*, rice and several other species with well-developed transformation technologies. These species also have the distinct, bioinformatic advantage of diploidy; as with many other types of analyses, direct application of established protocols and pipelines to polyploid species presents complications arising from redundant genomes, particularly from a bioinformatics perspective. As these challenges become addressed (Mithani *et al.* 2013; Page *et al.* 2013a, 2014; Duchemin *et al.* 2015; Page & Udall 2015), we can now envision experiments that will inform us about network processes in polyploid species.

As an example, we present here an application relevant to the question of the evolution of fibre development and morphology in polyploid *Gossypium hirsutum*, the species that provides most of the world's cotton crop. Phenotypically, fibres from modern cultivated cotton are vastly improved relative to those of their wild progenitors (Fig. 2), having longer, stronger and finer fibre as a consequence of several millennia of strong directional selection under domestication. Morphologically, 'fibres' are single-celled epidermal trichomes arising from the surface of the ovule, already evident on the day the flower opens and thereafter growing rapidly over a period of a couple of months through the stages of primary and secondary wall synthesis, followed by maturation and programmed cell death (Haigler *et al.* 2012). In our earlier work, we showed that the transcriptome of these cells is highly dynamic and that domestication has led to a massive transcriptional and proteomic rewiring (Chaudhary *et al.* 2008; Hovav *et al.* 2008a,b; Rapp *et al.* 2009, 2010; Bao *et al.* 2011; Hu *et al.* 2013, 2014). One insight emerging from this body of work is that the profilin gene family contains six members (PRF1–PRF6), five of which are upregulated in fibres early in development in modern cotton species relative to its wild progenitor (Bao *et al.* 2011). Moreover, this same upregulation was shown to have been repeatedly and unknowingly selected under domestication in three independently domesticated cottons. Subsequent differential gene expression analysis of two fibre developmental stages in multiple wild and domesticated accessions suggested that PRF1, PRF3 and PRF4 may be key players in the difference between wild and domestication cotton fibre development (Yoo & Wendel 2014). As profilin plays an important role in actin polymerization, it is reasonable to propose that proteins relevant to cytoskeletal behaviour were

targeted by humans during domestication and crop improvement (Bao *et al.* 2011; Yoo & Wendel 2014). These observations raise many questions; for example, have the profilin genes themselves been the targets of selection, or have their upstream regulators? Are all three important in specifying the domesticated fibre phenotype, or is it one gene that encodes the key player, with others passively co-upregulated? What other genes are up- or downregulated, or coexpressed, during fibre development that may be related to PRF1, PRF3 and PRF4?

Both differential gene expression (DGE) and co-expression network analyses have the potential to increase our understanding of the role of profilin genes in fibre development, although from different angles. To illustrate this, we used RNA-seq data from an ongoing analysis of fibre development for both DGE and network reconstruction (J. Gallagher *et al.*, unpublished). These data consist of four developmental stages (5, 10, 15 and 20 days post-anthesis, or flower opening, hereafter 'dpa') from three accessions each of wild and domesticated *G. hirsutum*. Standard DGE analysis (see Methods) of the data revealed 3811 genes upregulated under domestication during fibre development (~domestication + dpa, domesticated>wild, adjusted *P*-value < 0.05), including 167 transcription factors and 42 genes with known or suspected involvement in cytoskeleton and cell wall organization (Table 1). The numbers of upregulated and downregulated genes between developmental stages and between wild and domesticated cotton demonstrate several notable results, such as the observation of incredible transcriptional stasis from 10 to 15 dpa in wild cotton and massive change during the transition to secondary wall synthesis in domesticated *G. hirsutum* between 15 and 20 dpa (Fig. 3; Bao *et al.* 2011; Rapp *et al.* 2010; Yoo & Wendel 2014). These results illustrate and confirm the massive scope of transcriptional rewiring that has accompanied the morphological transformations engineered by human selection on this single-celled structure. Of the profilins, we find PRF1 (Gorai.009G028500) is significantly upregulated between domesticated and wild cottons at 10, 15 and 20 dpa (Fig. 4A), in contrast to the earlier quantitative PCR (QPCR) results where PRF1 to PRF5 were all upregulated by domestication (Bao *et al.* 2011); however, this new result was similar to a previous RNA-seq study where only PRF1 was found to be significantly upregulated by domestication at 10 and 20 dpa (Yoo & Wendel 2014), possibly indicative of the more sensitive nature of QPCR. Previous microarray results also indicate upregulation of a profilin from 7 to 20 dpa (Rapp *et al.* 2010), which coincides with the upregulation of one of the profilins (Gorai.003G061200) from 15 to 20 dpa in domesticated

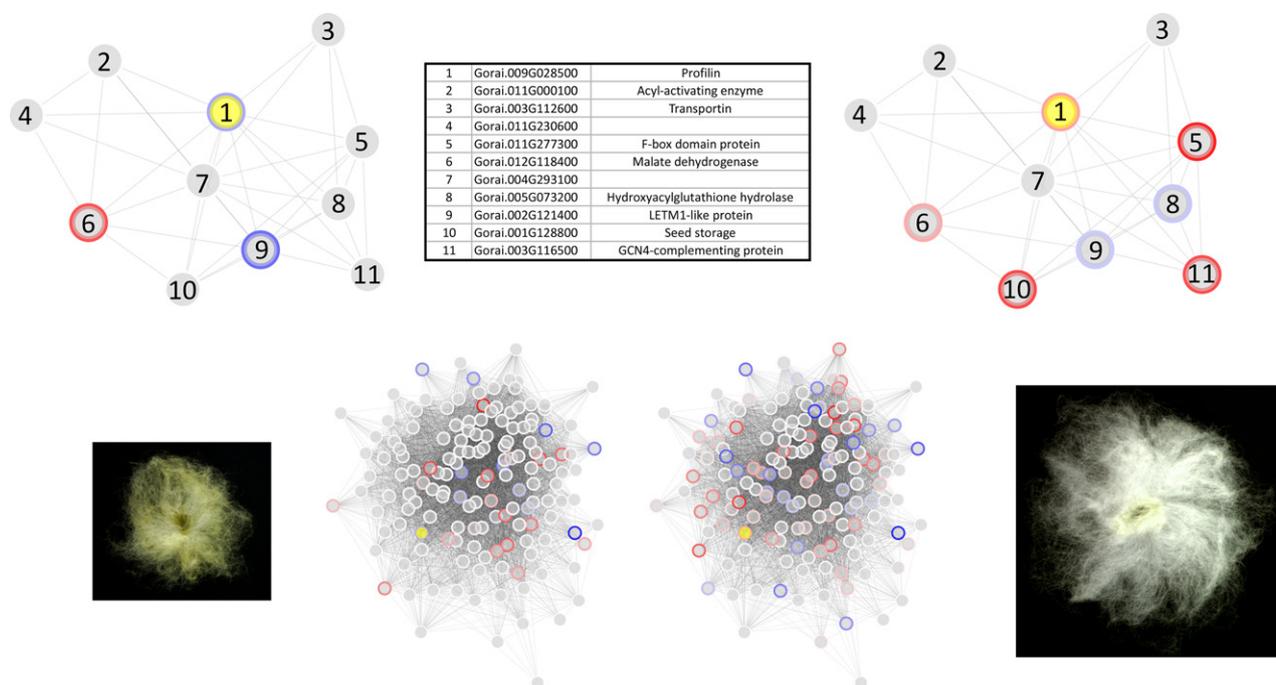


Fig. 2 Homoeolog expression bias for genes possibly coregulated with profilin 1 in fibre from wild (left) and domesticated (right) cotton, *Gossypium hirsutum*. Unsupervised network reconstruction via WGCNA of fibre RNA from four time points (5, 10, 15, and 20 dpa) produces several modules, including one ('module 5') that demonstrates upregulation in domesticated *G. hirsutum* (vs. wild) for all four stages. This module contains a total of 1508 genes, including two profilin genes, one of which (PRF1) has been implicated in fibre morphology differences between multiple wild and domesticated cotton species/accessions (Bao *et al.* 2011; Yoo & Wendel 2014). Homoeolog expression biases at 5 dpa were mapped (as an example) to a reduction in module 5 which contains only the top 10% (top) or top 25% (bottom) nearest neighbours to profilin 1 (shown as a yellow circle) for wild (bottom left) and domesticated (bottom right) *G. hirsutum*. Blue circle borders indicate bias towards A-derived homoeologs, red circle borders indicate bias towards D-homoeologs, and the intensity of the colour indicates the intensity of homoeolog expression bias for that gene. Identities of the genes in the top 10% of nearest neighbours and their putative annotations (Paterson *et al.* 2012) are shown. Representative images of cotton fibre are shown for the wild (far left) and domesticated (far right) cottons.

cotton (Fig. 4A). The differences in results of these several studies may be due to the greater variance in RNA-seq (vs. QPCR or microarrays), as direct inspection of individual gene expression profiles based on normalized read counts suggests that at least half of the profilins are differentially expressed between wild and domesticated cotton while exhibiting diverse developmental programmes (Fig. 4A). These expression patterns are highly consistent with those of Bao *et al.* (2011) and further our understanding of variation among profilins, both among gene family members and between wild and domesticated cotton fibres. Without prior interest in the profilin family, however, it is likely the differences in expression of these profilins, and their potential relevance to the cotton fibre developmental programme, would be lost in the morass of differentially expressed gene lists.

The same data were subjected to unsupervised network reconstruction via WGCNA, using default parameters (Langfelder & Horvath 2008). From this analysis, the whole fibre transcriptome was clustered into 48

co-expression modules, with the profilin genes located in five different modules. While similar results could be obtained through more complicated DGE analyses than employed above, network reconstruction (as used here) provides a relatively simple and readily accessible method for distilling complex information into partitioned sets of putatively functionally related genes. Although no expression changes of PRF2 were suggested by DGE results, PRF2 was clustered into the same module (ME) with PRF1 (ME5), which represents genes continuously upregulated during fibre development in domesticated cotton (Fig. 4B). While other modules containing profilin genes (ME1 with PRF5 and Gorai.001G025300, ME2 with PRF4, ME3 with PRF6 and ME12 with PRF3 and Gorai.003G061200; Fig 4B) may indicate coordinated expression changes by domestication limited to a certain developmental stage (such as ME1 and ME2 at 10 dpa), we chose to focus on the module containing PRF1 and PRF2 to exemplify what can be gained through network analysis. The genes comprising this module number 1508, less than half the

Table 1 Gene families related to cytoskeleton and cell wall organization

Gene family	Exemplary references
Actin	Bouget <i>et al.</i> (1996), Cleary & Smith (1998), Gallagher & Smith (1997) and Vidali & Hepler (2001)
Actin-related proteins (ARPs)	Schafer & Schroer (1999)
Actin capping	Cooper & Schafer (2000) and Hart <i>et al.</i> (2000)
Annexin	Morgan & Fernandez (1997) and Morgan & Pilar Fernandez (1997)
Callose synthase	Irshad <i>et al.</i> (2008)
CESA	Arioli <i>et al.</i> (1998) and Betancur <i>et al.</i> (2010)
CESA-like	Cutler & Somerville (1997)
COBRA-like	Roudier <i>et al.</i> (2002)
Cofilin/ADF	Carlier <i>et al.</i> (1997)
eEF1 α	Sun <i>et al.</i> (1997)
EMP24/GP25L	Wang <i>et al.</i> (2008) and Zhou <i>et al.</i> (2010)
Fimbrin/Plastin	Kovar <i>et al.</i> (2000) and McCurdy & Kim (1998)
Formin	Deeks <i>et al.</i> (2010)
Fragmin	Furuhashi & Hatano (1989) and Huang <i>et al.</i> (2011)
Gelsolin	Huang <i>et al.</i> (2011) and Sun <i>et al.</i> (1999)
Glycosyl hydrolase	Irshad <i>et al.</i> (2008)
Kinesin	Oppenheimer <i>et al.</i> (1997)
Methylesterase inhibitor/pectinesterase	Irshad <i>et al.</i> (2008)
MS4A/Peroxidase	Irshad <i>et al.</i> (2008) and Shigeto <i>et al.</i> (2013)
Myosin	Kinkema & Schiefelbein (1994), Kinkema <i>et al.</i> (1994) and Yamamoto <i>et al.</i> (1995)
Profilin	Christensen <i>et al.</i> (1996) and Huang <i>et al.</i> (1996)
Severin	Brown <i>et al.</i> (1982) and Huang <i>et al.</i> (2011)
Tubulin	Amos (2000), Schroer (2001) and Wick (2000)
Villin	Khurana <i>et al.</i> (2010) and Tang <i>et al.</i> (2012)
Xyloglucan transferase	Irshad <i>et al.</i> (2008)

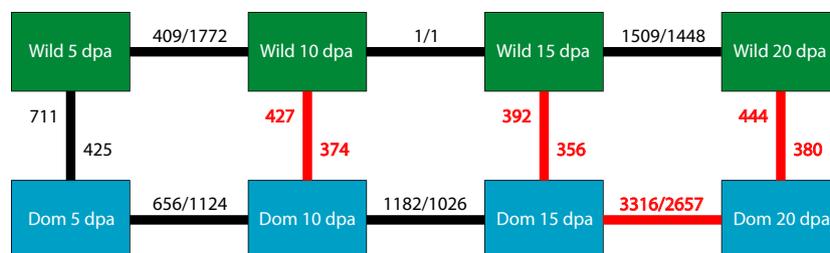


Fig. 3 Results of differential gene expression analysis for four developmental stages (in days post-anthesis, 'dpa') of cotton fibre in representatives of wild and domesticated *Gossypium hirsutum*. Green and blue boxes represent wild (top) and domesticated (bottom) cotton gene expression at a given developmental stage, respectively. Shown above horizontal lines are the numbers of statistically significantly upregulated genes in the developmental stages to their left and right, respectively. To the left and right of the vertical lines are shown, respectively, the number of statistically significantly upregulated genes in the comparisons between the two cotton phenotypes. Comparisons including at least one significantly upregulated profilin homolog are highlighted in red.

number of upregulated genes in domesticated vs. wild cotton from the DGE gene set. Furthermore, only 1062 are common between the two gene lists, winnowing the DGE candidate gene list by ~2800, but identifying an additional ~500 genes that may be related to PRF1 and its role in fibre development that are overlooked by standard DGE analysis. Included in these ~1500 genes are 71 transcription factors, as well as 16 genes with known or suspected involvement in fibre development

(J. Jareczek, unpublished), including 11 that also are present in the DGE gene list. In addition, we extracted putative protein-protein interactions of profilins from a cotton bioinformatics database GraP (Zhang *et al.* 2015a) and identified a phosphatidylinositol phosphate kinase 10 gene (Gorai.004G153600) that was both bioinformatically predicted to interact with PRF1 and also present in the DGE and co-expression module gene lists. Given that protein kinases are key players in intracellular

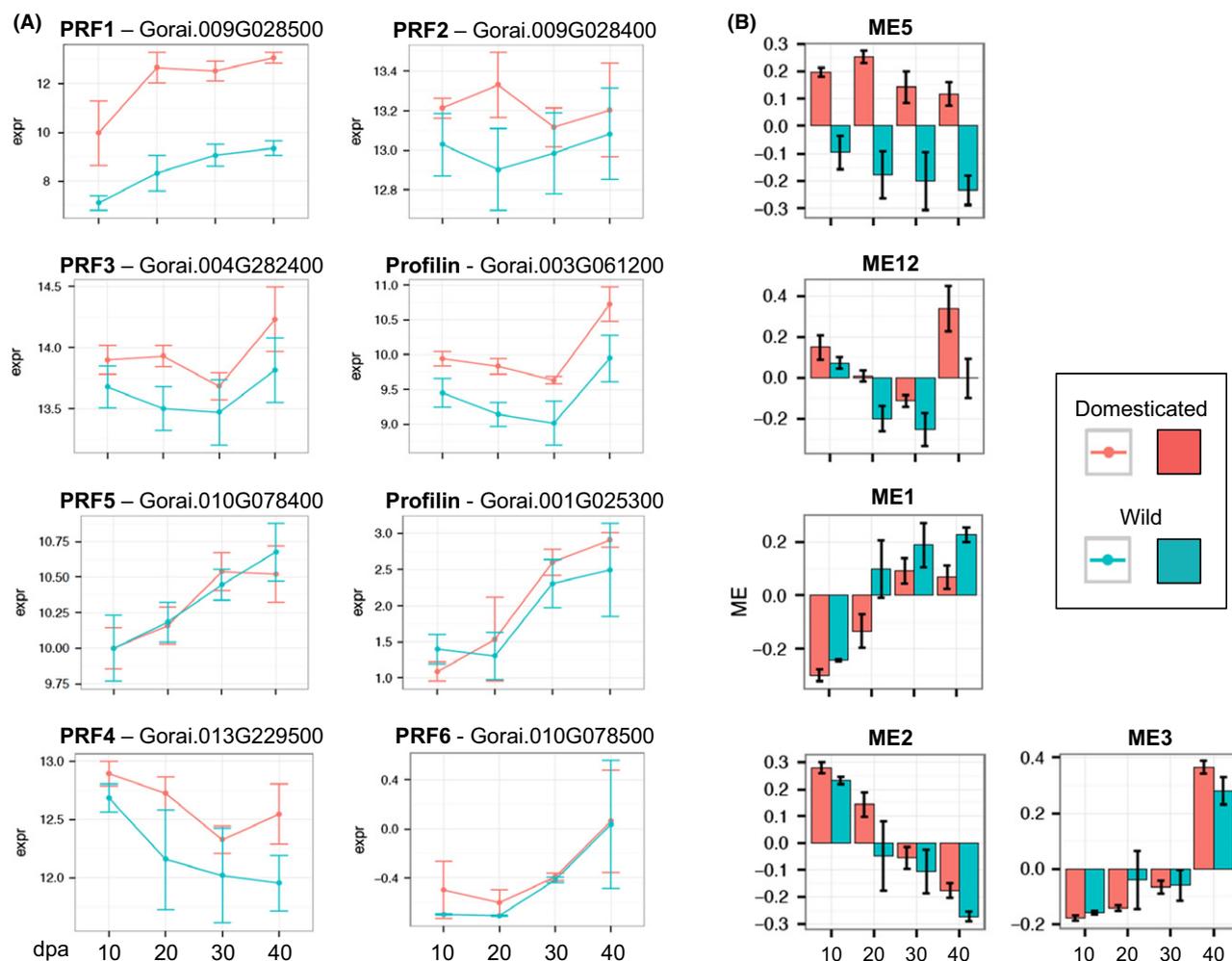


Fig. 4 Patterns of expression for (A) profilin family members and (B) profilin-containing co-expression modules during cotton fibre development. (A) Individual gene expression profiles for eight profilin family members between wild (blue) and domesticated (orange) *Gossypium hirsutum* during fibre development. Expression is displayed using *rlog*-transformed read counts for each profilin gene, with the standard error among replicates shown. Each profilin is associated with an annotated gene (i.e. Gorai) from the published D-genome reference (Paterson *et al.* 2012). (B) Expression patterns for network modules (ME) containing one or more profilin family members (PRF): PRF1 and PRF2 in module ME5; PRF3 and Gorai.003G061200 in ME12; PRF5 and Gorai.001G025300 in ME1; PRF4 in ME2; and PRF6 in ME3. Within-module relative expression, summarized and standardized as per Langfelder & Horvath (2007), is shown for each accession and fibre stage in all profilin-containing modules. For each module, the expression of all member genes was summarized into one representative expression profile, which is zero-centred and relative between accessions and among developmental stages.

signal transduction, further investigation into this gene of interest could reveal other functional genes or pathways that act in coordination with profilin at important time points in fibre development.

We note that in the foregoing we considered the aggregate expression of both duplicated copies (homoeologs) of each profilin gene, in that *G. hirsutum* is an allopolyploid species. In the context of the present study, we are especially interested in how network approaches can illuminate processes in *polyploids*, which requires analyses wherein each duplicated copy of each gene is bioinformatically distinguished and

transcriptionally measured. From the standpoint of understanding polyploidy, we might then ask whether instances of homoeolog expression bias are randomly distributed among modules, or whether instead they are coordinated at this level, which would be suggestive of selection for network coordination and optimization (Blanc & Wolfe 2004; Conant & Wolfe 2006; Chang *et al.* 2010). This phenomenon, termed ‘concerted divergence’, was seen by Blanc & Wolfe (2004) in *Arabidopsis thaliana*, where paralogs from an ancient genome duplication event were found to diverge in gene expression and form two parallel networks with other paralogs

from the same event. To the extent that concerted divergence of homoeologs exists, it may indicate ecologically or evolutionarily interesting sub- or neofunctionalization of gene regulatory networks following genome merger and doubling. Therefore, evaluating homoeolog expression bias in the context of networks can provide insight into the underlying architecture of relevant phenotypes. For example, by overlaying relative homoeolog expression on our profilin example (Fig. 2), we illustrate patterns of homoeolog expression bias for 5 dpa fibre from wild and domesticated cotton. A first observation is that there exists no detectable overall or module-wide bias in homoeolog usage for either wild or domesticated cotton in 5 dpa fibre. A second observation is that PRF1 homoeolog usage switches from A-homoeolog biased to D-homoeolog biased in wild vs. domesticated cottons, respectively. Third, a similar trend is observed for several genes when looking at the top 10% of nearest neighbours (most closely connected genes) to profilin; that is homoeolog usage bias that was previously near-equal becomes biased towards the D-homoeolog under domestication. This is not true for all nearest neighbours, as one gene coding for a putative LETM1-like protein (Gorai.002G121400) remained slightly A-biased and a second, a putative hydroxyacylglutathione hydrolase gene (Gorai.005G073200), went from near-equal homoeolog usage to being slightly A-biased. At a broader scale, for those genes in the top 25% of nearest neighbours to PRF1, the shift under domestication is more evenly spread between A-bias and D-bias. These observations are notable, as long, spinnable fibres are found only in the A-genome parent of the polyploid species, indicating possible recruitment of the alternative homoeolog for certain pathway segments during the phenotypic transition from the coarse brown fibre of wild cotton to the fine white fibre of domesticated cotton; however, further data are required to evaluate this veracity of this speculation.

The above example is intended to illustrate some possibilities for simple co-expression network analyses in polyploid species, how they can reveal genomic interactions within a polyploid nucleus that would not be evident from more traditional DGE analyses, and in a manner potentially connected to phenotypes of ecological or evolutionary relevance. Additional analytical possibilities will emerge in the future, as the resources and tools are developed. For example, in the above we constructed co-expression networks based on few samples and aggregate expression of homoeologs. As such, we are limited to what essentially are aggregate co-expression relationships of both homoeologs simultaneously, and for only a single cell type (fibres). Extension to multiple tissues may reveal novel network interactions among loci/homoeologs that may be indicative of sub-

or neofunctionalization. Furthermore, evidence from other types of analyses (e.g. ChIP-seq, methyl-seq, protein–protein interactions) can be integrated and layered in a network context (Hawkins *et al.* 2010; Gomez-Cabrero *et al.* 2014) to provide further understanding of the interactions among genes, their regulation and their evolution in a duplicated context. Such integrative network analysis holds promise for increasing our understanding of the complex foundations of the novel and/or adaptive phenotypes and ecological traits of polyploids (Madan Babu & Teichmann 2003; Gao *et al.* 2004; Zhu *et al.* 2008; Cookson *et al.* 2009; Mason *et al.* 2009; Bocklandt *et al.* 2011; Langfelder *et al.* 2012). The power of network-based analyses will become more fully realized as additional molecular relationships and interdependencies become layered on the expression data, such that a more complete understanding emerges into omics relationships and their connections to ecological and evolutionary traits of interest.

Conclusions

Despite substantial gains in understanding the immediate and long-term consequences of polyploidy, there remain many unanswered questions. Recent research in multiple polyploid systems has described patterns of genomic and transcriptomic change (as well as other measurable omics) on different timescales. These patterns have illuminated many of the phenomena associated with genomic merger (hybridization) and doubling, as well as subsequent cladogenesis and diversification. An exciting prospect is that we are able to reveal the underpinnings of complex phenotypes and ecological traits. As the generation of multiple data types becomes more accessible, we have the opportunity to reveal multiple subcellular connections that previously were hidden from view, thereby facilitating our understanding of how omics changes manifest into evolutionarily and ecologically important traits. While massive, integrated data sets (e.g. Zhu *et al.* 2008; Gerstein *et al.* 2012) are necessary to truly understand the nuances of interplay among polyploid 'omes', the integration of multiple data sets from multiple perspectives will ultimately increase our understanding of the formation of novel and adaptive traits in polyploids, and their evolutionary and ecological significance.

Acknowledgements

The authors thank Josef Jareczek for assistance with identifying cell wall and cytoskeleton genes. The authors are supported by grants from the National Science Foundation and Cotton Incorporated. J. P. Gallagher is supported by National Science Foundation Graduate Research Fellowship DGE1247194.

Methods

RNA was collected from fibres as described previously (Yoo & Wendel 2014). RNA-sequencing was performed on Illumina Hi-Seq2500 at Iowa State University DNA Facility (<http://www.dna.iastate.edu/>). Reads were mapped to the *Gossypium raimondii* genome (Paterson *et al.* 2012) using GSNAP with SNP-tolerant mapping (Wu & Nacu 2010; Page & Udall 2015). Differential gene expression analysis was conducted in R software v.3.2.0 (R Core Team 2015) with package DESEQ2 (Love *et al.* 2014). For weighted gene co-expression network analysis, raw read count data were *rlog*-transformed (regularized logarithm built into DESEQ2) and then analysed using the WGCNA package in R with default parameters (Langfelder & Horvath 2008, 2012). The resulting networks were visualized using CYTOSCAPE v3.2.0 (Shannon *et al.* 2003).

References

- Abrouk M, Zhang R, Murat F *et al.* (2012) Grass microRNA gene paleohistory unveils new insights into gene dosage balance in subgenome partitioning after whole-genome duplication. *Plant Cell*, **24**, 1776–1792.
- Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences of the USA*, **100**, 4649–4654.
- Albert R (2005) Scale-free networks in cell biology. *Journal of Cell Science*, **118**, 4947–4957.
- Albertin W, Marullo P (2012) Polyploidy in fungi: evolution after whole-genome duplication. *Proceedings of the Royal Society B*, **279**, 2497–2509.
- Amos LA (2000) Focusing-in on microtubules. *Current Opinion in Structural Biology*, **10**, 236–241.
- Angelini C, Costa V (2014) Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. *Frontiers in Cell and Developmental Biology*, **2**, Article 51.
- Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an *Arabidopsis* interactome map. *Science*, **333**, 601–607.
- Arioli T, Peng L, Betzner AS *et al.* (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science*, **279**, 717–720.
- Bao Y, Hu GJ, Flagel LE *et al.* (2011) Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (*Gossypium*). *Proceedings of the National Academy of Sciences of the USA*, **108**, 21152–21157.
- Barker MS, Kane NC, Matvienko M *et al.* (2008) Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution*, **25**, 2445–2455.
- Bekaert M, Edger PP, Pires JC, Conant GC (2011) Two-phase resolution of polyploidy in the *Arabidopsis* metabolic network gives rise to relative and absolute dosage constraints. *Plant Cell*, **23**, 1719–1728.
- Betancur L, Singh B, Rapp RA *et al.* (2010) Phylogenetically distinct cellulose synthase genes support secondary wall thickening in arabidopsis shoot trichomes and cotton fiber. *Journal of Integrative Plant Biology*, **52**, 205–220.
- Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell*, **19**, 395–402.
- Birchler JA, Veitia RA (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences of the USA*, **109**, 14746–14753.
- Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: biological implications. *Trends in Genetics*, **21**, 219–226.
- Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell*, **16**, 1679–1691.
- Bloom JD, Adami C (2003) Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets. *BMC Evolutionary Biology*, **3**, 21.
- Bocklandt S, Lin W, Sehl ME *et al.* (2011) Epigenetic predictor of age. *PLoS ONE*, **6**, e14821.
- Bouget FY, Gerttula S, Shaw SL, Quatrano RS (1996) Localization of actin mRNA during the establishment of cell polarity and early cell divisions in *Fucus* embryos. *Plant Cell*, **8**, 189–201.
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**, 433–438.
- Brown SS, Yamamoto K, Spudich JA (1982) A 40,000-dalton protein from *Dictyostelium discoideum* affects assembly properties of actin in a Ca²⁺-dependent manner. *Journal of Cell Biology*, **93**, 205–210.
- Buggs RJA, Doust AN, Tate JA *et al.* (2009) Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity*, **103**, 73–81.
- Buggs RJA, Elliott NM, Zhang LJ *et al.* (2010) Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytologist*, **186**, 175–183.
- Buggs RJA, Zhang LJ, Miles N *et al.* (2011) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Current Biology*, **21**, 551–556.
- Cañas RA, Canales J, Muñoz-Hernández C *et al.* (2015) Understanding developmental and adaptive cues in pine through metabolite profiling and co-expression network analysis. *Journal of Experimental Botany*, **66**, 3113–3127.
- Carlier MF, Laurent V, Santolini J *et al.* (1997) Actin depolymerizing factor (ADF/cofilin) enhances the rate of filament turnover: implication in actin-based motility. *Journal of Cell Biology*, **136**, 1307–1322.
- Chalhoub B, Denoeud F, Liu SY *et al.* (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, **345**, 950–953.
- Chang PL, Dilkes BP, McMahon M, Comai L, Nuzhdin SV (2010) Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. *Genome Biology*, **11**, R125.
- Chaudhary B, Hovav R, Rapp R *et al.* (2008) Global analysis of gene expression in cotton fibers from wild and domesticated *Gossypium barbadense*. *Evolution & Development*, **10**, 567–582.

- Chelaifa H, Monnier A, Ainouche M (2010) Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina x townsendii* and *Spartina anglica* (Poaceae). *New Phytologist*, **186**, 161–174.
- Chester M, Gallagher JP, Symonds VV *et al.* (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proceedings of the National Academy of Sciences of the USA*, **109**, 1176–1181.
- Christensen HE, Ramachandran S, Tan CT *et al.* (1996) *Arabidopsis* profilins are functionally similar to yeast profilins: identification of a vascular bundle-specific profilin and a pollen-specific profilin. *Plant Journal*, **10**, 269–279.
- Cleary AL, Smith LG (1998) The Tangled1 gene is required for spatial control of cytoskeletal arrays associated with cell division during maize leaf development. *Plant Cell*, **10**, 1875–1888.
- Coate JE, Powell AF, Owens TG, Doyle JJ (2013) Transgressive physiological and transcriptomic responses to light stress in allopolyploid *Glycine dolichocarpa* (Leguminosae). *Heredity*, **110**, 160–170.
- Coate JE, Bar H, Doyle JJ (2014) Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). *Plant Cell*, **26**, 136–150.
- Comai L, Tyagi AP, Winter K *et al.* (2000) Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell*, **12**, 1551–1568.
- Conant GC, Wolfe KH (2006) Functional partitioning of yeast co-expression networks after genome duplication. *Plos Biology*, **4**, e109.
- Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics*, **9**, 938–950.
- Conant GC, Birchler JA, Pires JC (2014) Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Current Opinion in Plant Biology*, **19**, 91–98.
- Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. *Nature Reviews Genetics*, **10**, 184–194.
- Cooper JA, Schafer DA (2000) Control of actin assembly and disassembly at filament ends. *Current Opinion in Cell Biology*, **12**, 97–103.
- Cutler S, Somerville C (1997) Cellulose synthesis: cloning in silico. *Current Biology*, **7**, R108–R111.
- De Smet R, Van de Peer Y (2012) Redundancy and rewiring of genetic networks following genome-wide duplication events. *Current Opinion in Plant Biology*, **15**, 168–176.
- De Smet R, Adams KL, Vandepoele K *et al.* (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences of the USA*, **110**, 2898–2903.
- Deal RB, Henikoff S (2011) Histone variants and modifications in plant gene regulation. *Current Opinion in Plant Biology*, **14**, 116–122.
- Deeks MJ, Fendrych M, Smertenko A *et al.* (2010) The plant formin AtFH4 interacts with both actin and microtubules, and contains a newly identified microtubule-binding domain. *Journal of Cell Science*, **123**, 1209–1215.
- Dong SW, Adams KL (2011) Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytologist*, **190**, 1045–1057.
- Duarte JM, Cui LY, Wall PK *et al.* (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Molecular Biology and Evolution*, **23**, 469–478.
- Duchemin W, Dupont PY, Campbell MA, Ganley AR, Cox MP (2015) HyLiTE: accurate and flexible analysis of gene expression in hybrid and allopolyploid species. *BMC Bioinformatics*, **16**, 8.
- Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. *New Phytologist*, **183**, 557–564.
- Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytologist*, **186**, 184–193.
- Flagel LE, Wendel JF, Udall JA (2012) Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. *BMC Genomics*, **13**, 302.
- Force A, Lynch M, Pickett FB *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**, 1531–1545.
- Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nature Reviews Genetics*, **13**, 840–852.
- Furuhashi K, Hatano S (1989) A fragmin-like protein from plasmodium of *Physarum polycephalum* that severs F-actin and caps the barbed end of F-actin in a Ca²⁺-sensitive way. *Journal of Biochemistry*, **106**, 311–318.
- Gallagher K, Smith LG (1997) Asymmetric cell division and cell fate in plants. *Current Opinion in Cell Biology*, **9**, 842–848.
- Gao F, Foat BC, Bussemaker HJ (2004) Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics*, **5**, 31.
- Gerstein MB, Kundaje A, Hariharan M *et al.* (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature*, **489**, 91–100.
- Gomez-Cabrero D, Abugessaisa I, Maier D *et al.* (2014) Data integration in the era of omics: current and future challenges. *BMC Systems Biology*, **8** (Suppl. 2), I1.
- Gong L, Kakrana A, Arikat S, Meyers BC, Wendel JF (2013) Composition and expression of conserved microRNA genes in diploid cotton (*Gossypium*) species. *Genome Biology and Evolution*, **5**, 2449–2459.
- Grover CE, Gallagher JP, Szadkowski EP *et al.* (2012) Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytologist*, **196**, 966–971.
- Guan XY, Song QX, Chen ZJ (2014) Polyploidy and small RNA regulation of cotton fiber development. *Trends in Plant Science*, **19**, 516–528.
- Guo H, Wang XY, Gundlach H *et al.* (2014) Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). *Genetics*, **197**, 1153–1163.
- Ha M, Ng DWK, Li WH, Chen ZJ (2011) Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Research*, **21**, 590–598.
- Haigler CH, Betancur L, Stiff MR, Tuttle JR (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Frontiers in Plant Science*, **3**, 104.

- Hart MC, Korshunova YO, Cooper JA (2000) Mapping of the mouse actin capping protein beta subunit gene. *BMC Genomics*, **1**, 1.
- Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. *Nature Reviews Genetics*, **11**, 476–486.
- Heyndrickx KS, Van de Velde J, Wang CM, Weigel D, Vandepoele K (2014) A functional and evolutionary perspective on transcription factor binding in *Arabidopsis thaliana*. *Plant Cell*, **26**, 3894–3910.
- van Hoek MJ, Hogeweg P (2009) Metabolic adaptation after whole genome duplication. *Molecular Biology and Evolution*, **26**, 2441–2453.
- Horvath S (2009) *Weighted Gene Co-Expression Network Page*. <http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/>
- Hovav R, Chaudhary B, Udall JA, Flagel L, Wendel JF (2008a) Parallel domestication, convergent evolution and duplicated gene recruitment in allopolyploid cotton. *Genetics*, **179**, 1725–1733.
- Hovav R, Udall JA, Chaudhary B *et al.* (2008b) The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. *Plos Genetics*, **4**, e25.
- Hu GJ, Koh J, Yoo MJ *et al.* (2013) Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. *New Phytologist*, **200**, 570–582.
- Hu GJ, Koh J, Yoo MJ *et al.* (2014) Proteomics profiling of fiber development and domestication in Upland cotton (*Gossypium hirsutum* L.). *Planta*, **240**, 1237–1251.
- Hu G, Koh J, Yoo MJ, Chen S, Wendel JF (2015) Gene-expression novelty in allopolyploid cotton: a proteomic perspective. *Genetics*, **200**, 91–104.
- Huang S, McDowell JM, Weise MJ, Meagher RB (1996) The *Arabidopsis* profilin gene family (Evidence for an ancient split between constitutive and pollen-specific profilin genes). *Plant Physiology*, **111**, 115–126.
- Huang S, Xiang Y, Ren H (2011) Actin-binding proteins and actin dynamics in plant cells. In: *The Plant Cytoskeleton* (ed Liu B), pp. 57–80. Springer, New York.
- Hughes TE, Langdale JA, Kelly S (2014) The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Research*, **24**, 1348–1355.
- Ichihashi Y, Aguilar-Martinez JA, Farhi M *et al.* (2014) Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proceedings of the National Academy of Sciences of the USA*, **111**, E2616–E2621.
- International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, **345**, 1251788.
- Irshad M, Canut H, Borderies G, Pont-Lezica R, Jamet E (2008) A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers. *Bmc Plant Biology*, **8**, 94.
- Jiao YN, Wickett NJ, Ayyampalayam S *et al.* (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**, 97–100.
- Khurana P, Henty JL, Huang SJ *et al.* (2010) *Arabidopsis* VIL-LIN1 and VILLIN3 have overlapping and distinct activities in actin bundle formation and turnover. *Plant Cell*, **22**, 2727–2748.
- Kinkema M, Schiefelbein J (1994) A myosin from a higher plant has structural similarities to class V myosins. *Journal of Molecular Biology*, **239**, 591–597.
- Kinkema M, Wang H, Schiefelbein J (1994) Molecular analysis of the myosin gene family in *Arabidopsis thaliana*. *Plant Molecular Biology*, **26**, 1139–1153.
- Koh J, Chen SX, Zhu N *et al.* (2012) Comparative proteomics of the recently and recurrently formed natural allopolyploid *Tragopogon mirus* (Asteraceae) and its parents. *New Phytologist*, **196**, 292–305.
- Kovar DR, Staiger CJ, Weaver EA, McCurdy DW (2000) AtFim1 is an actin filament crosslinking protein from *Arabidopsis thaliana*. *Plant Journal*, **24**, 625–636.
- Krouk G, Lingeman J, Colon AM, Coruzzi G, Shasha D (2013) Gene regulatory networks in plants: learning causality from time and perturbation. *Genome Biology*, **14**, 123.
- Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. *Bmc Systems Biology*, **1**, 54.
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
- Langfelder P, Horvath S (2012) Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software*, **46**, i11.
- Langfelder P, Castellani LW, Zhou Z *et al.* (2012) A systems genetic analysis of high density lipoprotein metabolism and network preservation across mouse models. *Biochimica et Biophysica Acta*, **1821**, 435–447.
- Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. *Science*, **320**, 481–483.
- Levin DA (1983) Polyploidy and novelty in flowering plants. *American Naturalist*, **122**, 1–25.
- Li Y, Jackson SA (2015) Gene network reconstruction by integration of prior biological knowledge. *G3*, **5**, 1075–1079.
- Li AL, Liu DC, Wu J *et al.* (2014) mRNA and small RNA transcriptomes reveal insights into dynamic homeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell*, **26**, 1878–1900.
- Li FG, Fan GY, Lu CR *et al.* (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology*, **33**, 524–530.
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.
- Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics*, **154**, 459–473.
- Madan Babu M, Teichmann SA (2003) Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Research*, **31**, 1234–1244.
- Madlung A, Wendel JF (2013) Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenetic and Genome Research*, **140**, 270–285.
- Maere S, De Bodt S, Raes J *et al.* (2005) Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences of the USA*, **102**, 5454–5459.
- Makino T, McLysaght A (2012) Positionally biased gene loss after whole genome duplication: evidence from human, yeast, and plant. *Genome Research*, **22**, 2427–2435.

- Marchand G, Huynh-Thu VA, Kane NC *et al.* (2014) Bridging physiological and evolutionary time-scales in a gene regulatory network. *New Phytologist*, **203**, 685–696.
- Mason MJ, Fan G, Plath K, Zhou Q, Horvath S (2009) Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics*, **10**, 327.
- Matsushita SC, Tyagi AP, Thornton GM, Pires JC, Madlung A (2012) Allopolyploidization lays the foundation for evolution of distinct populations: evidence from analysis of synthetic *Arabidopsis* allohexaploids. *Genetics*, **191**, 535–547.
- McAllister C, Blaine R, Kron P *et al.* (2015) Environmental correlates of cytotype distribution in *Andropogon gerardii* (Poaceae). *American Journal of Botany*, **102**, 92–102.
- McCurdy DW, Kim M (1998) Molecular cloning of a novel fibrin-like cDNA from *Arabidopsis thaliana*. *Plant Molecular Biology*, **36**, 23–31.
- McIntyre PJ (2012) Polyploidy associated with altered and broader ecological niches in the *Claytonia perfoliata* (Portulacaceae) species complex. *American Journal of Botany*, **99**, 655–662.
- Meyer P (2011) DNA methylation systems and targets in plants. *FEBS Letters*, **585**, 2008–2015.
- Mithani A, Belfield EJ, Brown C *et al.* (2013) HANDS: a tool for genome-wide discovery of subgenome-specific base-identity in polyploids. *BMC Genomics*, **14**, 653.
- Mitra K, Carvunis AR, Ramesh SK, Ideker T (2013) Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics*, **14**, 719–732.
- Morgan RO, Fernandez MP (1997) Annexin gene structures and molecular evolutionary genetics. *Cellular and Molecular Life Sciences*, **53**, 508–515.
- Morgan RO, Pilar Fernandez M (1997) Distinct annexin subfamilies in plants and protists diverged prior to animal annexins and from a common ancestor. *Journal of Molecular Evolution*, **44**, 178–188.
- Muller HJ (1925) Why polyploidy is rarer in animals than in plants. *American Naturalist*, **59**, 346–353.
- Ng DWK, Zhang C, Miller M *et al.* (2012) Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors. *Heredity*, **108**, 419–430.
- Ni Z, Kim ED, Ha M *et al.* (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature*, **457**, 327–331.
- van Noort V, Snel B, Huynen MA (2003) Predicting gene function by conserved co-expression. *Trends in Genetics*, **19**, 238–242.
- Ohno S (1970) *Evolution by Gene Duplication*. Springer-Verlag, Berlin.
- Oppenheimer DG, Pollock MA, Vacik J *et al.* (1997) Essential role of a kinesin-like protein in *Arabidopsis* trichome morphogenesis. *Proceedings of the National Academy of Sciences of the USA*, **94**, 6261–6266.
- Orr HA (1990) Why polyploidy is rarer in animals than in plants revisited. *American Naturalist*, **136**, 759–770.
- Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell*, **13**, 1735–1747.
- Page JT, Udall JA (2015) Methods for mapping and categorization of DNA sequence reads from allopolyploid organisms. *Bmc Genetics*, **16**, S4.
- Page JT, Gingle AR, Udall JA (2013a) PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. *G3*, **3**, 517–525.
- Page JT, Huynh MD, Liechty ZS *et al.* (2013b) Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *G3*, **3**, 1809–1818.
- Page JT, Liechty ZS, Huynh MD, Udall JA (2014) BamBam: genome sequence analysis tools for biologists. *BMC Research Notes*, **7**, 829.
- Pang MX, Woodward AW, Agarwal V *et al.* (2009) Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). *Genome Biology*, **10**, R122.
- Papp B, Pál C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature*, **424**, 194–197.
- Paterson AH, Wendel JF, Gundlach H *et al.* (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*, **492**, 423–427.
- Pfeifer M, Kugler KG, Sandve SR *et al.* (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science*, **345**, 1250091.
- Qian W, He X, Chan E, Xu H, Zhang J (2011) Measuring the evolutionary rate of protein-protein interaction. *Proceedings of the National Academy of Sciences of the USA*, **108**, 8725–8730.
- R Core Team (2015) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramsey J (2011) Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences of the USA*, **108**, 7096–7101.
- Ramsey J, Ramsey TS (2014) Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B*, **369**, 20130352.
- Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, **29**, 467–501.
- Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics*, **33**, 589–639.
- Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. *Bmc Biology*, **7**, 18.
- Rapp RA, Haigler CH, Flagel L *et al.* (2010) Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. *Bmc Biology*, **8**, 139.
- Renny-Byfield S, Gallagher JP, Grover CE *et al.* (2014) Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence. *Genome Biology and Evolution*, **6**, 559–571.
- Renny-Byfield S, Gong L, Gallagher JP, Wendel JF (2015) Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Molecular Biology and Evolution*, **32**, 1063–1071.
- Roudier F, Schindelman G, DeSalle R, Benfey PN (2002) The COBRA family of putative GPI-anchored proteins in *Arabidopsis*. A new fellowship in expansion. *Plant Physiology*, **130**, 538–548.
- Roulin A, Auer PL, Libault M *et al.* (2013) The fate of duplicated genes in a polyploid plant genome. *Plant Journal*, **73**, 143–153.
- Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytologist*, **186**, 123–134.

- Schafer DA, Schroer TA (1999) Actin-related proteins. *Annual Review of Cell and Developmental Biology*, **15**, 341–363.
- Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences of the USA*, **108**, 4069–4074.
- Schroer TA (2001) Microtubules don and doff their caps: dynamic attachments at plus and minus ends. *Current Opinion in Cell Biology*, **13**, 92–96.
- Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA (2014) Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids. *BMC Genomics*, **15**, 701.
- Seoighe C, Gehring C (2004) Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends in Genetics*, **20**, 461–464.
- Shannon P, Markiel A, Ozier O *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**, 2498–2504.
- Shi XL, Ng DWK, Zhang CQ *et al.* (2012) Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nature Communications*, **3**, 950.
- Shigeto J, Kiyonaga Y, Fujita K, Kondo R, Tsutsumi Y (2013) Putative cationic cell-wall-bound peroxidase homologues in *Arabidopsis*, AtPrx2, AtPrx25, and AtPrx71, are involved in lignification. *Journal of Agricultural and Food Chemistry*, **61**, 3781–3788.
- Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution*, **14**, 348–352.
- Soltis PS, Soltis DE (2012) *Polyploidy and Genome Evolution*, p. 420. Springer-Verlag, Berlin Heidelberg.
- Soltis PS, Liu XX, Marchant DB, Visger CJ, Soltis DE (2014) Polyploidy and novelty: Gottlieb's legacy. *Philosophical Transactions of the Royal Society B*, **369**, 20130351.
- Stebbins GL (1940) The significance of polyploidy in plant evolution. *American Naturalist*, **74**, 54–66.
- Stephens SG (1951) Possible significance of duplication in evolution. *Advances in Genetics*, **4**, 247–265.
- Sun Y, Carneiro N, Clore AM *et al.* (1997) Characterization of maize elongation factor 1A and its relationship to protein quality in the endosperm. *Plant Physiology*, **115**, 1101–1107.
- Sun HQ, Yamamoto M, Mejillano M, Yin HL (1999) Gelsolin, a multifunctional actin regulatory protein. *Journal of Biological Chemistry*, **274**, 33179–33182.
- Swanson-Wagner R, Briskine R, Schaefer R *et al.* (2012) Reshaping of the maize transcriptome by domestication. *Proceedings of the National Academy of Sciences of the USA*, **109**, 11878–11883.
- Tang HB, Woodhouse MR, Cheng F *et al.* (2012) Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics*, **190**, 1563–1574.
- Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Research*, **16**, 934–946.
- Tsompana M, Buck MJ (2014) Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin*, **7**, 33.
- Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, **10**, 725–732.
- Vidali L, Hepler PK (2001) Actin and pollen tube growth. *Protoplasma*, **215**, 64–76.
- Wagner A (2001) The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Molecular Biology and Evolution*, **18**, 1283–1292.
- Wang XY, Paterson AH (2011) Gene conversion in angiosperm genomes with an emphasis on genes duplicated by polyploidization. *Genes*, **2**, 1–20.
- Wang JL, Tian L, Madlung A *et al.* (2004) Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics*, **167**, 1961–1973.
- Wang Y, Zhang WZ, Song LF *et al.* (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiology*, **148**, 1201–1211.
- Wendel JF (2015) The wondrous cycles of polyploidy in plants. *American Journal of Botany*, **102**, 1753–1756.
- Wick S (2000) Plant microtubules meet their MAPs and mimics. *Nature Cell Biology*, **2**, E204–E206.
- Wood TE, Takebayashi N, Barker MS *et al.* (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the USA*, **106**, 13875–13879.
- Woodhouse MR, Schnable JC, Pedersen BS *et al.* (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *Plos Biology*, **8**, e1000409.
- Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.
- Xiong ZY, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences of the USA*, **108**, 7908–7913.
- Xu CM, Bai Y, Lin XY *et al.* (2014) Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Molecular Biology and Evolution*, **31**, 1066–1076.
- Yamamoto K, Kikuyama M, Sutoh-Yamamoto N, Kamitsubo E, Katayama E (1995) Myosin from alga Chara: unique structure revealed by electron microscopy. *Journal of Molecular Biology*, **254**, 109–112.
- Yoo MJ, Wendel JF (2014) Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *Plos Genetics*, **10**, e1004073.
- Yoo MJ, Szadkowski E, Wendel JF (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity*, **110**, 171–180.
- Yoo MJ, Liu XX, Pires JC, Soltis PS, Soltis DE (2014) Nonadditive gene expression in polyploids. *Annual Review of Genetics*, **48**, 485–517.
- Zhang WL, Zhang T, Wu YF, Jiang JM (2012) Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell*, **24**, 2719–2731.
- Zhang WL, Zhang T, Wu YF, Jiang JM (2014) Open chromatin in plant genomes. *Cytogenetic and Genome Research*, **143**, 18–27.
- Zhang L, Guo J, You Q *et al.* (2015a) GraP: platform for functional genomics analysis of *Gossypium raimondii*. *Database*, **2015**, bav047.

- Zhang TZ, Hu Y, Jiang WK *et al.* (2015b) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology*, **33**, 531–537.
- Zhou C, Yin YB, Dam P, Xu Y (2010) Identification of novel proteins involved in plant cell-wall synthesis based on protein-protein interaction data. *Journal of Proteome Research*, **9**, 5025–5037.
- Zhu J, Zhang B, Smith EN *et al.* (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics*, **40**, 854–861.
- Zhu B, Zhang W, Zhang T, Liu B, Jiang J (2015) Genome-wide prediction and validation of intergenic enhancers in *Arabidopsis* using open chromatin signatures. *Plant Cell*, **27**, 2415–2426.

J.P.G., C.E.G., G.H. and J.F.W. conceived of the concept of the study. G.H. and C.E.G. performed example analysis. J.P.G., C.E.G., G.H. and J.F.W. wrote the study.

Data accessibility

RNA-seq data are archived at Dryad: doi:<http://dx.doi.org/10.5061/dryad.256hn> and will be made publicly available upon final analysis. Scripts for performing the example analysis are available at <https://github.com/Wendellab/MolEcol2016>.