

Research review

Gene duplication and evolutionary novelty in plants

Author for correspondence:
Jonathan F. Wendel
Tel: +1 515 294 7172
Email: jfw@iastate.edu

Lex E. Fligel and Jonathan F. Wendel

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

Received: 25 February 2009
Accepted: 12 May 2009

Summary

New Phytologist (2009) **183**: 557–564
doi: 10.1111/j.1469-8137.2009.02923.x

Key words: evolution of novelty, gene duplication, land plants, neofunctionalization, polyploidy, subfunctionalization.

Duplication is a prominent feature of plant genomic architecture. This has led many researchers to speculate that gene duplication may have played an important role in the evolution of phenotypic novelty within plants. Until recently, however, it was difficult to make this connection. We are now beginning to understand how duplication has contributed to adaptive evolution in plants. In this review we introduce the sources of gene duplication and predictions of the various fates of duplicates. We also highlight several recent and pertinent examples from the literature. These examples demonstrate the importance of the functional characteristics of genes and the source of duplication in influencing evolutionary outcome.

Introduction

One of the realizations to emerge from comparative plant genomic studies is that plant gene families, as entities, are largely conserved, even over evolutionary timescales that encompass the diversification of all angiosperms and non-flowering plants (Rensing *et al.*, 2008). This central property of plant genomes indicates that major clades of plants have largely not invented new gene families, but that plant species have been endowed with a basic genetic toolkit of ancient origin. Yet, despite this broad evolutionary conservation of gene families, lineage-specific fluctuations in gene family size are frequent among taxa (Velasco *et al.*, 2007; Ming *et al.*, 2008; Rensing *et al.*, 2008), suggesting that the amazing diversity and lineage-specific phenotypic variation found among land plants may not be underlain by an equally diverse set of wholly novel genes. Instead, much of plant diversity may have arisen largely following the duplication and adaptive specialization of pre-existing genes.

This relatively recent perspective assigns gene duplication a central role in plant diversification, as a key process that

generates the raw material necessary for adaptive evolution. This notion has captivated plant biologists in particular, as no other group of organisms has a greater incidence of recent and historical polyploidy and hence duplicate gene content. Further, the once elusive connections between duplication-generating processes and subsequent adaptive evolution are now becoming clearer, bolstering the long-held view, based largely on evidence from comparative cytogenetics and more recently comparative genomics, that duplication is truly the 'stuff of evolution'. Here we attempt to provide perspective on these connections, highlighting empirical examples and insights that have furthered our understanding of gene duplication in adaptive evolution. We first provide a synopsis of the primary genetic and genomic mechanisms that produce duplicate genes. This is followed by a brief discussion of the theoretical framework that describes the spectrum of subsequent evolutionary possibilities for duplicated genes, and empirical examples of gene and genome duplications that are thought to have led to adaptive outcomes. Finally, we will explore several recent changes in perspective that are important to consider and which might suggest a future research agenda.

Duplications both great and small

Duplications of genomic content occur on various scales by independent mechanisms, including tandem and segmental duplications that often arise during DNA replication and recombination, and whole-genome duplications (polyploidy) that form by various means (Ramsey & Schemske, 1998). The genomes involved in the origin of polyploids range from near-identical (classic autopolyploids) to rather divergent (classic allopolyploids), with important implications for the initial conditions of the now co-resident duplicated genomes (the latter with more divergent regulatory elements and coding sequences) (Wendel & Doyle, 2005). In addition, transposable elements can create duplications by capturing and transporting gene copies (transduplication), or by stimulating intrachromosomal recombination events, and through reverse-transcriptase-mediated generation of cDNAs capable of genomic reintegration (retropositioning) (Hurles, 2004; Freeling *et al.*, 2008). A considerable body of literature exists for each of these categories, but to our knowledge the relative rates of gene duplication within any single plant lineage remain little investigated. Following the model used in *Drosophila* (Zhou *et al.*, 2008), the generation of this information should soon be possible, concomitant with the completion of several plant genome sequencing projects. A thorough enumeration of the various modes of gene duplication should prove useful, because, as we will point out later, the mode of duplication can influence evolutionary outcomes.

While it is obvious that polyploidy generates duplicated genes, each of the mechanisms listed above has been shown to play a considerable role in plant genomes (Rizzon *et al.*, 2006; W. Wang *et al.*, 2006; Freeling *et al.*, 2008; Ming *et al.*, 2008). Because of its prevalence in plants and obvious importance to the topic at hand, wherein the entire genomic complement of genes becomes instantly doubled, polyploidy deserves special mention. Among contemporary plants it has long been known that a high percentage of species are polyploid, with classical estimates ranging from *c.* 30 to 70% (Wendel, 2000; Tate *et al.*, 2005; Wendel & Doyle, 2005). This figure alone is considerably higher than that for other eukaryotic lineages. In addition to these recent polyploids, however, it is now understood that most, if not all, modern land plant genomes are built on the remnants of older polyploidy events (Soltis *et al.*, 2009). If we compound the high incidence of contemporary polyploidy with its cyclical recurrence throughout angiosperm evolution, it becomes evident that essentially all but the most recently formed plant gene families have experienced expansion through polyploidy.

Although polyploidy may be the largest contributor of duplicate genes, approx. 15–20% of the genic content of both *Arabidopsis thaliana* and rice (*Oryza sativa*) is comprised of tandemly arrayed gene clusters (Rizzon *et al.*, 2006). Within these clusters tandem duplications add duplicate genes to the array whereas segmental duplications copy and disperse fragments or entire arrays. With regard to evolution by duplica-

tion, tandem arrays have two relevant features. First, because of their mode of origin, tandem arrays often share regulatory elements and tend to be expressed in a coordinated manner (Schmid *et al.*, 2005). Secondly, linked tandem duplicates may 'homogenize' one another via unequal crossing over and/or gene conversion, and this homogenization tends to accelerate divergence among nonrecombining tandem array clusters (Baumgarten *et al.*, 2003). These twin processes are thought to be adaptively relevant to the evolution and function of disease resistance and abiotic stress response genes, which are over-represented among tandemly arrayed genes in *A. thaliana* and rice (Rizzon *et al.*, 2006). Perhaps the best known example is the nucleotide-binding site-leucine-rich repeat gene family, a group of *c.* 150 disease resistance proteins found in clusters throughout the *A. thaliana* genome (Meyers *et al.*, 1999). Apparently, tandem duplication has provided a means of amplifying adaptively important resistance genes while duplication via segmental isolation has permitted gene family diversification and long-term evolutionary plasticity (Baumgarten *et al.*, 2003; Meyers *et al.*, 2005).

In addition to small-scale genomic duplications arising from unequal crossing over and chromosomal anomalies, we are beginning to appreciate the impact of a category of mechanisms involving transposons and other reverse-transcriptase-mediated duplication. A striking example of the potential of transposon-mediated duplication is described by Freeling *et al.* (2008). By examining orthologous genomic alignments of sequences from *A. thaliana* with those from two outgroup species (papaya (*Carica papaya*) and grape (*Vitis vinifera*)), these authors found that ~11% of *A. thaliana* genes have 'transposed' either into or out of syntenic regions shared with its common ancestor with papaya. Although not all of these events need be transposon-mediated, it is likely many are because transposons contain the necessary replication and integration machinery. Additionally, genes and gene fragments are frequently found within the borders of some transposable elements, such as mutator-like transposable elements (MULEs) (Jiang *et al.*, 2004) and helitrons (Lai *et al.*, 2005). Once a transposable element has captured a gene, its amplification can generate gene duplications. Integration of the new copies may be near an existing gene, potentially altering gene expression patterns or leading to the formation of a new chimeric gene (the *jingwei* gene found in *Drosophila yakuba* and *Drosophila teisseiri* provides an example of these phenomena (Wang *et al.*, 2000)). Recent estimates suggest that we have underestimated the role of transposon proliferation as a force for generating new genes (W. Wang *et al.*, 2006; Freeling *et al.*, 2008) and modifying gene expression (Feschotte, 2008), and this remains a promising field of study.

An interesting dimension to the foregoing discussion is that the various mechanisms that lead to gene duplication need not be mutually exclusive. For example, transposon release has been shown to coincide with polyploidy (Hanson *et al.*, 1999; Kashkush *et al.*, 2003; Madlung *et al.*, 2005), potentially leading to an episodic expansion of transposon-mediated

duplication shortly after polyploidy. Thus, to borrow a metaphor from Wessler & Carrington (2005), polyploidy doubles the number of cards in the deck, and, through transposon release, could initiate the process of shuffling as well. In addition, this increased transposon activity has been found to alter the expression levels of nearby genes (Kashkush *et al.*, 2003), which may further induce phenotypic alteration following polyploidy (Chen & Ni, 2006). Although there are few if any convincing connections between adaptation *per se* and polyploidy-induced transposon proliferation, the scale of the phenomenon suggests that these processes have played an important role in evolutionary adaptation of nascent polyploids.

It is also important to recognize that there are varying degrees of the functional extent of 'duplication' for each of the above mechanisms. For example, a gene duplication arising from retropositioning typically creates an intron-free copy that is removed from its original regulatory context, whereas whole-genome duplication involves the duplication of regulatory sequences (typically more similar in autopolyploidy than allopolyploidy) and intervening (nongenic) sequences, in the process doubling higher order features such as genetic interaction networks. These distinctions are important; not all duplicates are created equally, nor do they stand the same chance of retention (Paterson *et al.*, 2006). With regard to these distinctions, it is clear that the creation of duplicate genes needs to be considered in light of their subsequent fates; this topic is addressed in the following section.

Theoretical models of evolution following gene duplication, with examples

A notable feature of duplication, when compared with other forms of mutation, is that it creates genetic redundancy. This redundancy has long been thought to foster evolutionary innovation, as the constraints of purifying selection are expected to be relaxed on duplicate loci thereby creating the opportunity for duplicates to explore new evolutionary terrain. Although this concept did not originate with Ohno (for a history see Taylor & Raes, 2004), it was broadly popularized in his book '*Evolution by gene duplication*' (1970). In Ohno's classic formulation, if given sufficient time, one copy of a duplicate pair can acquire a beneficial mutation ('neofunctionalization') resulting in retention of both divergent copies. Alternatively, one copy can accumulate a mutation(s) that renders it nonfunctional and leads to mutational obliteration ('nonfunctionalization' in Ohno's words; 'pseudogenization' in modern terms), consequently maintaining the other copy through purifying selection. In recent years, with the recognition of the highly duplicate nature of eukaryotic genomes, these concepts of evolution by duplication have been the source of great interest, leading to the development of a significant body of theory. Much of this material has been reviewed elsewhere; see, for example, Conant & Wolfe (2008). Here we provide a brief overview relevant to adaptive evolution.

One noteworthy contribution to the theory of evolution by duplication has been the realization that mutations may accumulate among duplicates in such a way that they *partition* aggregate ancestral functions such that both gene copies must be preserved to carry out their complementary ancestral roles. This process, which can arise in the absence of natural selection, has been termed 'subfunctionalization' under the Duplication-Degeneration-Complementation (DDC) model of Force *et al.* (1999); it posits a mechanism that creates a stable safe-haven for preservation of both members of a duplicate pair (Lynch & Force, 2000). Importantly, for DDC-subfunctionalization to occur, it is necessary that the ancestral gene had at least two necessary functions (broadly defined here to include multiple expression domains, for example). If expression in multiple cell lines, tissues, or organs is necessary for a given protein product, then a duplicate gene pair encoding this protein may experience expression DDC-subfunctionalization by regulatory rather than coding mutations. In support of this view, recent empirical results have exposed duplicate gene expression patterns consistent with DDC-subfunctionalization on broad scales in both *A. thaliana* (Duarte *et al.*, 2006; Ha *et al.*, 2009) and cotton (*Gossypium hirsutum*) (Flagel *et al.*, 2008).

Another important contribution to the theory of evolution by duplication is the recognition that a single protein can perform multiple catalytic or structural functions. This has been famously demonstrated for structural eye crystallin proteins, which also have enzymatic functions when expressed outside the eye (Piatigorsky & Wistow, 1991). For these 'shared genes' (shared in the sense of a single gene being employed by unrelated cellular processes), the selective optimization of one function may lead to a decline in another function, creating an adaptive conflict. Under this scenario, gene duplication and subsequent functional specialization between duplicates can provide a solution to the optimization problem. This has been termed 'escape from adaptive conflict' (EAC). EAC can generate observed patterns that may be misconstrued as either neofunctionalization or DDC-subfunctionalization (Des Marais & Rausher, 2008), the important distinction being that DDC-subfunctionalization may occur purely as a result of neutral mutations, whereas EAC requires positive natural selection on both copies of a duplicate gene pair (Conant & Wolfe, 2008; Des Marais & Rausher, 2008). In a similar vein, both neofunctionalization and EAC require positive natural selection, although for neofunctionalization this selection need only influence one of the two duplicate genes. Consequently, the EAC model has the greatest number of conditions that must be met, and because of this it may occur less frequently than neofunctionalization or DDC-subfunctionalization. However, from an experimental perspective, it is difficult to distinguish between neutral processes and selection, and thus it is difficult to empirically tease apart neofunctionalization and DDC-subfunctionalization from EAC. For these reasons we lack good estimates of the relative roles of adaptive and neutral processes in shaping gene

evolution following duplication (Des Marais & Rausher, 2008).

Finally, an important contribution to the evolution by duplication theory is the observation that duplications must maintain proper dosage balance among dosage-sensitive genes (Veitia *et al.*, 2008). If a duplication event produces a dosage imbalance in a finely tuned gene network or protein complex it can lead to a reduction in the functional efficiency of these interactions. Under this scenario, selection against dosage imbalance would favor the return to a single-copy state. This process is one of several factors that might account for the differential retention of different classes of genes following whole-genome duplication events (Paterson *et al.*, 2006). As an example, Thomas *et al.* (2006) proposed that, following an ancient polyploidization event in *A. thaliana*, chromosomal clusters of interacting dosage-sensitive genes were preferentially preserved, whilst their homoeologous clusters were shed. The dosage-balance hypothesis (Veitia *et al.*, 2008) has resulted in an appreciation of the necessity of considering gene duplication in a more interdependent context: for any single locus, duplication may relax selection; however, this relaxed selection may be counterbalanced by other genomic dependences, such as dosage-sensitive interactions with the products of other genes.

Collectively, the neofunctionalization, DDC-subfunctionalization, EAC, and dosage-balance models form a theoretical framework for understanding evolution following gene duplication. It should be noted, however, that these models, although useful, are neither mutually exclusive nor likely to capture many evolutionary intricacies and complexities. For example, there have been two duplications of the B class floral identity gene in *Aquilegia vulgaris* (Kramer *et al.*, 2007), leading to complex and overlapping expression patterns for three paralogs that are not readily explained by any single model of duplicate gene evolution but instead require, minimally, a combination of DDC-subfunctionalization and neofunctionalization. Notwithstanding these shortcomings, the models provide a useful framework for interpreting patterns and mechanisms that underlie duplicate gene retention.

Empirical studies are beginning to reveal additional factors that influence duplicate gene retention and by extension the potential for adaptive evolution via duplication. As noted above, mode of duplication (e.g. whole-genome vs tandem) is important because genes and gene products do not exist in isolation. In addition, it has become evident that the function of a gene can alter its probability of retention. Although the primary fate of duplicate genes is a return to single copy (Lynch & Conery, 2000), among those duplicates that have been retained in *A. thaliana*, transcription factors and kinases tend to be preferentially retained after polyploidy, whereas various classes of structural and metabolic genes preferentially return to a single-copy state following whole-genome duplication (Paterson *et al.*, 2006). An additional factor influencing the probability of

duplicate gene retention is its connectivity (i.e. the number of interacting partners in a molecular network). In yeast, for example, genes with high connectivity tend to have more pleiotropic effects than do genes with low connectivity (He & Zhang, 2006), the latter being preferentially retained, suggesting that duplication of highly connected genes with pleiotropic activities is largely harmful (Li *et al.*, 2006). The preceding example highlights the link between network connectivity and duplicate gene retention; however, the concept of connectivity can also be extended to the maintenance of optimal stoichiometries among gene products in multiprotein complexes (i.e. the dosage-balance hypothesis). For this reason the term 'connectivity' has doubly important meanings in the context of duplicate gene retention.

An additional consideration is that different types of proteins vary substantially in their functional plasticity and/or resistance to mutational diminution of function. One might envision that some proteins may require several key substitutions before acquiring a new function, while others may be more mutationally labile or fewer steps away from adopting a new function. Given this variation, it is easy to imagine that these differences might alter the likelihood of neofunctionalization. An excellent case in point is the terpene synthase gene family in Norway spruce (*Picea abies*). These genes modify secondary metabolites and appear to have undergone repeated rounds of neofunctionalization (Keeling *et al.*, 2008). Within this gene family a small number of key amino acid substitutions among paralogs has radically altered substrate specificity and terpenoid product profiles. These small changes are suggested to have facilitated the genesis, via neofunctionalization of paralogs, of a broader diversity of secondary metabolites in conifers, compounds hypothesized to play a crucial role in warding off pathogens and herbivores.

As a counter example to the terpene synthases, the *LEAFY* transcription factor appears to lack functional plasticity. *LEAFY* is found as a single copy in many plant genomes, and when duplicated there is strong evidence for neutral mutational drift and little evidence for positive selection (Baum *et al.*, 2005), thus precluding neofunctionalization (insofar as it has been studied). Interestingly, this pattern is quite different from that found in other plant transcription factor families, such as the sizable MADS-box family, which has undergone multiple duplications followed by diversifying positive selection (Martínez-Castilla & Alvarez-Buylla, 2003). Based on these data, we might speculate that *LEAFY* is inherently more constrained than are members of the terpene synthase or MADS-box families, thereby making it less likely to evolve new functions. These differences in functional plasticity between gene families may, in part, explain biased patterns of duplicate loss and retention (Paterson *et al.*, 2006).

Finally, these empirical examples impel a reconsideration of the basic theoretical framework for duplicate gene preservation. Specifically, the theoretical models are largely agnostic with regard to the functional properties of duplicate genes and

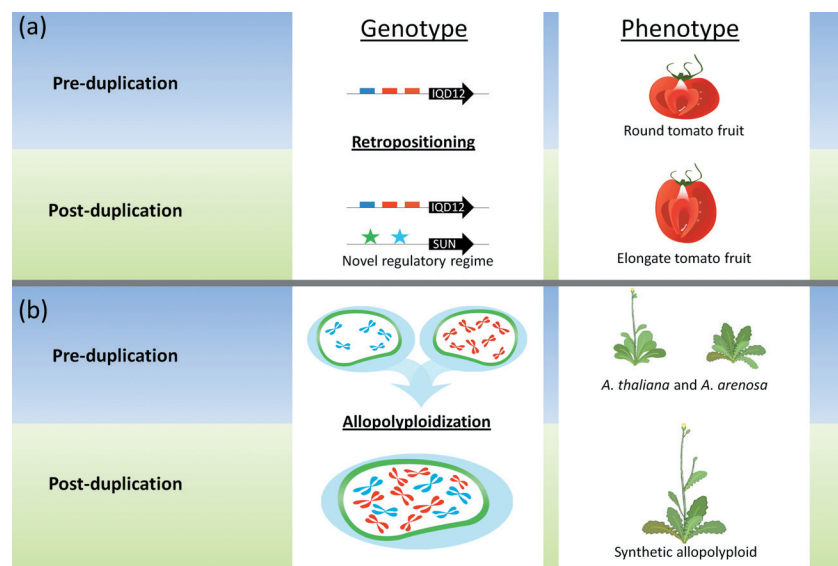


Fig. 1 Evolution following duplication in tomato (*Solanum lycopersicum*) and *Arabidopsis thaliana*. Two duplication events of vastly different sizes have contributed to adaptive evolution in tomato and *A. thaliana*. In both species the genotypic change (a duplication event) and associated phenotypic response are depicted pre- and post-duplication. (a) In the study by Xiao *et al.* (2008), the retropositioning of the *IQD12* gene leads to the novel up-regulation of its descendant locus, *SUN*, during tomato fruit development. This alteration has a profound phenotypic effect, producing the elongate fruit shape found in popular tomato cultivars. (b) Ni *et al.* (2009) show that a synthetic allopolyploid, derived from the diploid species *A. thaliana* and *Arabidopsis arenosa*, displays expression alteration among genes that exert control over circadian rhythms. These expression changes stem from epigenetic perturbations associated with allopolyploidy, and lead to improved starch and sugar accumulation and greater plant biomass within the synthetic allopolyploid.

their protein products. For example, the probability of neofunctionalization for a particular gene may be largely contingent on its intrinsic properties, such as functional category, plasticity, or physical and network connectivity, rather than extrinsic factors such as the effective population size and mutation rate of the organism. Thus there is a need for additional development of gene duplication theory such that it captures more of the inherent biological complexity.

Examples of adaptive evolution following gene duplication

To further explore the adaptive consequences of gene duplication, we highlight two recent studies in some detail. The first entails the evolution of a novel fruit shape in tomato (*Solanum lycopersicum*) (Xiao *et al.*, 2008). Phenotypically this transition involved the evolution of an elongated fruit from a round ancestor, a novelty that was probably valued by early domesticators (Fig. 1a). The molecular evolution that underlies this event appears to have been created by the chance duplication and transposition of a gene (*SUN*) into a new regulatory context. *SUN* and its progenitor (*IQD12*) belong to a gene family that contains a plant-specific, 67 amino acid motif (called IQ67) that is involved in calmodulin signaling. It appears that *IQD12* was linked to a *cop*ia-like retrotransposon. During a transpositioning event, this element failed to recognize its natural 3' long terminal repeat (LTR) border and continued read-through

transcription, stopping only after picking up a ~25-kb genomic fragment, including the entire *IQD12* coding domain. This fragment was then integrated into a different chromosome, creating a new copy (*SUN*), which is translationally identical to *IQD12*. In the transposition process the sequence just upstream from the 5' start site of the *IQD12* gene was left behind, thereby dissociating the gene from its former regulatory context. By chance this gene was inserted near the 5' end of another gene called *DEFL1*, thus placing the duplicated *SUN* gene in a new regulatory environment. Expression assays show that, in its new location, *SUN* is expressed at much higher levels during the early stages of fruit development, and that this up-regulation is clearly correlated with an elongated fruit shape. Additionally, over-expression of the progenitor locus, *IQD12*, in transgenic plants is sufficient to confer the elongated fruit phenotype, indicating that this regulatory change alone can explain the transition to a novel fruit shape.

This study shows that an important shift in tomato fruit shape originated via the duplication of a small genomic fragment. This evolutionary transition reflects a fortuitous insertion of a pre-existing gene into a new regulatory context. Additionally, this example highlights a point made earlier, that the mode of duplication may have a significant effect on the evolutionary outcome. Had the *SUN* locus been created by a large-scale genomic duplication, it would probably have remained associated with its original regulatory regime. An analysis in *A. thaliana* that demonstrates that duplicate pairs which arose

from small-scale duplication events tend to have a greater level of expression divergence than do pairs from larger events corroborates this assertion (Casneuf *et al.*, 2006). Thus, small-scale duplications may have profound phenotypic consequences; in this case the transposon-mediated duplication of *SUN* has separated it from its former regulatory regime, placing it in a new environment, one which by chance had a profound impact on fruit shape.

A second and very recent example, reported by Ni *et al.* (2009), offers an interesting contrast in that it involves polyploidy. *Arabidopsis thaliana* itself is not only a paleopolyploid, exemplifying the recurrent, episodic, and cyclical nature of polyploidy in angiosperms, but it has been involved in a relatively recent neopolyploidization event leading to the origin of the natural allopolyploid *Arabidopsis suecica*. This natural allopolyploid is readily resynthesized in the laboratory from its model progenitors *A. thaliana* and *Arabidopsis arenosa*. An interesting feature of this allopolyploid, and many others, is that it grows to a larger stature and produces more biomass than either of its parents (Fig. 1b). In this study the authors investigated the cause of this growth increase using comparative gene expression profiling. Among 128 genes up-regulated in the allotetraploid relative to its parents, ~67% were found to have either *circadian clock associated 1* (*CCA1*) or evening-element binding sites in their upstream regulatory regions. These binding sites are the targets of *CCA1* and *late elongated hypocotyl* (*LHY*), both important circadian regulators responsible for suppressing carbon fixation during the night. This finding led the authors to suspect that the alteration in circadian rhythms might be the key to the vigorous growth of the synthetic allopolyploid. Further analyses showed that *CCA1* and *LHY* were epigenetically suppressed in the allopolyploid and that this suppression strongly correlates with increased starch synthesis and chlorophyll content, ultimately leading to greater plant biomass.

With regard to phenotypic evolution following genome duplication, this study adds interesting facets to our understanding. First, the locus of evolution is not genetic, but rather epigenetic, and involved a temporal shift in gene expression among regulatory genes. Secondly, it appears that allopolyploidy (coupling genome merger with duplication) alone is responsible for the change, because the synthetic allopolyploid lines were formed only a few generations before the experiment. It remains to be seen how common these and other physiologically relevant alterations are in other allopolyploids, but similar genome-wide shifts in expression patterns are common in natural and synthetic *Brassica*, *Gossypium*, *Tragopogon*, and wheat (*Triticum*) allopolyploids (Doyle *et al.*, 2008). The study by Ni *et al.* (2009) illustrates the importance of the instantaneous shifts in genetic networks and their associated metabolism caused by allopolyploidy, which is likely to serve as an important source of evolutionary novelty. The principle that 'more is different' is often used to characterize complex systems, and this concept may aptly describe the emergence of new phenotypes following allopolyploidy.

Conclusions

Recent years have witnessed a breathtaking increase in the availability of genome sequence data, providing a vastly improved ability to document and study the dynamics of duplicate gene evolution. We now understand that all plant genomes harbor a history of ancient and recent gene and genome duplication, and that these duplicates have and continue to originate from several, potentially interconnected processes. A theoretical framework for describing the potential outcomes of gene duplication has been developed, but in addition, we now appreciate that gene-specific factors, such as functional properties, connectivity, and mutational plasticity, play important roles in defining the probability of duplicate gene preservation and adaptive evolution. We have highlighted examples in tomato and *A. thaliana*, in which expression alteration following duplication has led to phenotypic evolution at different scales and following different paths (Fig. 1). These studies demonstrate that expression patterns and associated phenotypic novelty can evolve very quickly. These findings are underscored by recent explorations of the transcriptomic response among synthetic allopolyploids (Adams *et al.*, 2003; J. Wang *et al.*, 2006; Chaudhary *et al.*, 2009; Ha *et al.*, 2009), which suggest that dynamic expression changes may occur immediately upon allopolyploid formation. Because of the genomic scale and potential phenotypic effects of gene expression change following polyploidy, we suspect that expression alteration following polyploidy will prove to be a significant source of evolutionary novelty among plants.

Acknowledgements

We thank Anna Krush for her expertise in executing the figure design, and gratefully acknowledge the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service and the National Science Foundation Plant Genome Research Program for their support of duplicate gene evolution research in the Wendel laboratory.

References

- 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, USA* **100**: 4649–4654.
- Baum DA, Yoon HS, Oldham RL. 2005. Molecular evolution of the transcription factor *LEAFY* in Brassicaceae. *Molecular Phylogenetics and Evolution* **37**: 1–14.
- Baumgarten A, Cannon S, Spangler R, May G. 2003. Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics* **165**: 309–319.
- Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y. 2006. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biology* **7**: R13.
- Chaudhary B, Flagel LE, Stupar RM, Udall JA, Verma N, Springer NM, Wendel J. 2009. Reciprocal silencing, transcriptional bias and functional

- divergence of homoeologs in polyploid cotton (*Gossypium*). *Genetics* 182: 503–517.
- Chen ZJ, Ni Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* 28: 240–252.
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics* 9: 938–950.
- Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454: 762–765.
- Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* 42: 443–461.
- Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Molecular Biology and Evolution* 23: 469–478.
- Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* 9: 397–405.
- Flagel L, Udall J, Nettleton D, Wendel J. 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biology* 6: 16.
- Force A, Lynch M, Pickett FB, Amores A, Yan Y-L, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531–1545.
- Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D. 2008. Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. *Genome Research* 18: 1924–1937.
- Ha M, Kim E-D, Chen ZJ. 2009. Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proceedings of the National Academy of Sciences, USA* 106: 2295–2300.
- Hanson RE, Islam-Faridi MN, Crane CF, Zwick MS, Czeschin DG, Wendel JF, McKnight TD, Price HJ, Stelly DM. 1999. Ty1-copia-retrotransposon behavior in a polyploid cotton. *Chromosome Research* 8: 73–76.
- He X, Zhang J. 2006. Toward a molecular understanding of pleiotropy. *Genetics* 173: 1885–1891.
- Hurles M. 2004. Gene duplication: the genomic trade in spare parts. *PLoS Biology* 2: e206.
- Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR. 2004. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431: 569–573.
- Kashkush K, Feldman M, Levy AA. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* 33: 102–106.
- Keeling CI, Weishaar S, Lin RPC, Bohlmann J. 2008. Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proceedings of the National Academy of Sciences, USA* 105: 1085–1090.
- Kramer EM, Holappa L, Gould B, Jaramillo MA, Setnikov D, Santiago PM. 2007. Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. *Plant Cell* 19: 750–766.
- Lai J, Li Y, Messing J, Dooner HK. 2005. Gene movement by helitron transposons contributes to the haplotype variability of maize. *Proceedings of the National Academy of Sciences, USA* 102: 9068–9073.
- Li L, Huang Y, Xia X, Sun Z. 2006. Preferential duplication in the sparse part of yeast protein interaction network. *Molecular Biology and Evolution* 23: 2467–2473.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154: 459–473.
- Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic *Arabidopsis* polyploids. *Plant Journal* 41: 221–230.
- Martínez-Castilla LP, Alvarez-Buylla ER. 2003. Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proceedings of the National Academy of Sciences, USA* 100: 13407–13412.
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant Journal* 20: 317–332.
- Meyers BC, Kaushik S, Nandety RS. 2005. Evolving disease resistance genes. *Current Opinion in Plant Biology* 8: 129–134.
- Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KLT *et al.* 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452: 991–996.
- Ni Z, Kim E-D, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ. 2009. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457: 327–331.
- Ohno S. 1970. *Evolution by gene duplication*. New York, NY, USA: Springer-Verlag.
- Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC. 2006. Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends in Genetics* 22: 597–602.
- Piatigorsky J, Wistow G. 1991. The recruitment of crystallins: new functions precede gene duplication. *Science* 252: 1078–1079.
- Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29: 467–501.
- Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y *et al.* 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64–69.
- Rizzon C, Ponger L, Gaut BS. 2006. Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Computational Biology* 2: e115.
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics* 37: 501–506.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.
- Tate JA, Soltis DE, Soltis PS. 2005. Polyploidy in plants. In: Gregory TR, ed. *The evolution of the genome*. San Diego, CA, USA: Elsevier, 371–426.
- Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annual Review of Genetics* 38: 615–643.
- 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.
- Veitia RA, Bottani S, Birchler JA. 2008. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* 24: 390–397.
- Velasco R, Zharkikh A, Troggo M, Cartwright DA, Cestaro A, Pruss D, Pindo M, FitzGerald LM, Vezzulli S, Reid J *et al.* 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* 2: e1326.
- Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L *et al.* 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172: 507–517.
- Wang W, Zhang J, Alvarez C, Llopart A, Long M. 2000. The origin of the *jingwei* gene and the complex modular structure of its parental gene, *yellow emperor*, in *Drosophila melanogaster*. *Molecular Biology and Evolution* 17: 1294–1301.
- Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, Liu D, Zhang J,

- Vang S *et al.* 2006. High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* 18: 1791–1802.
- Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.
- Wendel JF, Doyle JJ. 2005. Polyploidy and evolution in plants. In: Henry RJ, ed. *Plant diversity and evolution*. Wallingford, UK: CABI Publishing, 97–117.
- Wessler SR, Carrington JC. 2005. Genome studies and molecular genetics: the consequences of gene and genome duplication in plants. *Current Opinion in Plant Biology* 8: 119–121.
- Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319: 1527–1530.
- Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, Li X, Ding Y, Yang S, Wang W. 2008. On the origin of new genes in *Drosophila*. *Genome Research* 18: 1446–1455.



About *New Phytologist*

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at www.newphytologist.org.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *Early View* – our average submission to decision time is just 29 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £139 in Europe/\$259 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office (newphytol@lancaster.ac.uk; tel +44 1524 594691) or, for a local contact in North America, the US Office (newphytol@ornl.gov; tel +1 865 576 5261).