

The legacy of diploid progenitors in allopolyploid gene expression patterns

Richard J. A. Buggs, Jonathan F. Wendel, Jeffrey J. Doyle, Douglas E. Soltis, Pamela S. Soltis and Jeremy E. Coate

Phil. Trans. R. Soc. B 2014 **369**, 20130354, published 23 June 2014

References

This article cites 102 articles, 30 of which can be accessed free

<http://rstb.royalsocietypublishing.org/content/369/1648/20130354.full.html#ref-list-1>

Subject collections

Articles on similar topics can be found in the following collections

[evolution](#) (693 articles)
[genetics](#) (98 articles)
[genomics](#) (76 articles)
[plant science](#) (90 articles)

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

rstb.royalsocietypublishing.org



CrossMark
click for updates

Review

Cite this article: Buggs RJA, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE. 2014 The legacy of diploid progenitors in allopolyploid gene expression patterns. *Phil. Trans. R. Soc. B* **369**: 20130354.

<http://dx.doi.org/10.1098/rstb.2013.0354>

One contribution of 14 to a Theme Issue 'Contemporary and future studies in plant speciation, morphological/floral evolution and polyploidy: honouring the scientific contributions of Leslie D. Gottlieb to plant evolutionary biology'.

Subject Areas:

evolution, genetics, genomics, plant science

Keywords:

polyploidy, transcriptomes, hybridization, whole-genome duplication

Author for correspondence:

Richard J. A. Buggs

e-mail: r.buggs@qmul.ac.uk

The legacy of diploid progenitors in allopolyploid gene expression patterns

Richard J. A. Buggs¹, Jonathan F. Wendel², Jeffrey J. Doyle³, Douglas E. Soltis^{4,5}, Pamela S. Soltis⁵ and Jeremy E. Coate⁶

¹School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK

²Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames IA 50011, USA

³L. H. Bailey Hortorium, Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA

⁴Department of Biology, and ⁵Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

⁶Department of Biology, Reed College, Portland, OR 97202, USA

Allopolyploidization (hybridization and whole-genome duplication) is a common phenomenon in plant evolution with immediate saltational effects on genome structure and gene expression. New technologies have allowed rapid progress over the past decade in our understanding of the consequences of allopolyploidy. A major question, raised by early pioneer of this field Leslie Gottlieb, concerned the extent to which gene expression differences among duplicate genes present in an allopolyploid are a legacy of expression differences that were already present in the progenitor diploid species. Addressing this question necessitates phylogenetically well-understood natural study systems, appropriate technology, availability of genomic resources and a suitable analytical framework, including a sufficiently detailed and generally accepted terminology. Here, we review these requirements and illustrate their application to a natural study system that Gottlieb worked on and recommended for this purpose: recent allopolyploids of *Tragopogon* (Asteraceae). We reanalyse recent data from this system within the conceptual framework of parental legacies on duplicate gene expression in allopolyploids. On a broader level, we highlight the intellectual connection between Gottlieb's phrasing of this issue and the more contemporary framework of *cis-* versus *trans-*regulation of duplicate gene expression in allopolyploid plants.

1. Introduction

'Little is known about the level of enzyme expression in polyploids or whether differences in the regulation of their diploid genomes are maintained when they are present together in a common tetraploid nucleus'

Gottlieb ([1], p. 378)

'... it may be that some differences in organ-specific transcript levels reflect a legacy from the true progenitors.'

Gottlieb ([2], p. 91)

A common phenomenon in plant evolution is the hybridization of two diploid species, accompanied by whole-genome duplication, to produce an allotetraploid species with two homoeologous sub-genomes (see contributions by Soltis *et al.* [3], Ramsey & Ramsey [4], Vanneste *et al.* [5], Jiao & Paterson [6]). New allotetraploids contain duplicate copies of every gene that was present in both their parents. Tracing the fate of these duplicated genes over time, with respect to both DNA sequence evolution and gene expression patterns, is key to understanding the short- and long-term consequences of allopolyploidization. A particularly interesting question is the extent to which allopolyploid formation causes novel patterns of gene expression, which may affect fitness. Given that allopolyploidization has given rise to many crop species, and evidence suggests that all extant seed plant lineages have experienced whole-genome doubling events, this is an important research area. It is more important, in fact, than any scientist realized when Leslie Gottlieb began to draw attention to this subject in the 1970s. As noted repeatedly in this issue, Gottlieb was an early pioneer in gel electrophoresis of proteins as a tool for the study of plant evolution [7]. After his initial application of this technique to diploid speciation and phylogenetics [8,9], he

turned his attention to protein expression in allopolyploids, first in *Stephanomeria elata* [10], then in *Tragopogon miscellus* [11,12] and later in *Clarkia gracilis* [13–15].

2. Leslie Gottlieb's question about parental legacies of gene expression

In his work on allopolyploid gene expression, Gottlieb emphasized the importance of considering the legacies of the diploid parental species on their descendent allopolyploids. When he saw a difference between the two homoeologous genomes of an allopolyploid, whether in the genomes themselves or in their expression, Gottlieb wanted to be sure that this was not due to retention of a previously evolved difference between the parental diploids, before he would consider that it was due to novel changes upon, or subsequent to, allopolyploidization. A brief review of Gottlieb's own experimental work on this topic, and his interpretation of his results and others', will show why he considered this to be important, and how he placed emphasis on it.

Gottlieb's opening contribution to the field of gene expression in allopolyploids was a study of glutamate oxaloacetate transaminase in *S. elata*, a species which he discovered during his graduate research to be an allotetraploid derived from the diploids *Stephanomeria exigua* and *Stephanomeria virgata* [16]. He showed that allotetraploid individuals possessed multiple fixed enzyme variants that were identical in their electrophoretic mobility to those variants in the diploid parents, resulting in much higher levels of population heterozygosity in the tetraploid than in the diploids [10]. This was the first study in a wild plant species to show this 'additive' pattern of homoeologue expression (fixed heterozygosity), albeit non-quantitatively, where all gene copies expressed in diploid parents are expressed in their allopolyploid descendants [10]. Gottlieb's first discovery in this field was thus an intact legacy from parental diploids in an allopolyploid.

Gottlieb and his graduate student Mikeal L. Roose began to work on the recently formed, natural allopolyploids *T. miscellus* and *Tragopogon mirus* in the mid-1970s. They analysed 13 enzyme systems in 23 populations of the two tetraploid species and their three parental diploid species [11]. They found fixed heterozygosity at many loci in the tetraploids, but contrary to their initial expectations, a few tetraploid individuals were homozygous at these loci. This homozygosity could have been a parental legacy if the parental diploids were harbouring undetected polymorphisms at these loci. After rigorous checks of diploid populations, Roose and Gottlieb became convinced that these results were not a parental legacy, but instead reflected loss of homoeologous loci, possibly due to recombination among the parental sub-genomes within the tetraploids.

Roose & Gottlieb [12] followed up this study on presence/absence of homoeologue expression with an elegant series of experiments to quantify the protein expression of alcohol dehydrogenase (ADH3) in the seeds of allopolyploid *T. miscellus* and its diploid progenitors *Tragopogon dubius* and *Tragopogon pratensis*. Using densitometry, they showed that the diploid parents differed in their levels of ADH3 expression: there was twice as much ADH3 protein in *T. pratensis* as in *T. dubius*. They also showed that *T. miscellus* had a total ADH3 abundance intermediate to that of the two parental diploids. Furthermore, because the *ADH3* genes of the two parental

diploids encode isozymes of different mobility, the protein expression levels of the two parental forms of ADH3 within the allotetraploid could be measured; the expression levels of the two forms were unequal and were proportional to their expression levels in their respective diploid progenitors.

In discussing their results, Roose & Gottlieb were careful to point out that 'differential expression of ADH3 genes results from the inheritance of differentially expressed genes from its diploid progenitors and is not to be interpreted as an example of an adaptive difference selected since its origin' ([12], p. 1083). As Gottlieb [1] later summarized in *Science*: 'the relative expression of the ADH gene in each of the diploid genomes is not influenced by the presence of the other genome' (p. 378). Today, we might interpret this as evidence that *trans*-regulatory effects were not equilibrating the expression of both forms of ADH between the sub-genomes, but rather that *cis*-factors predominated (see further discussion of *cis*- and *trans*- effects below). Roose and Gottlieb noted: 'To our knowledge this is the first study of isozyme expression in an allopolyploid and its diploid progenitors in which the diploid genomes specify different amounts of enzyme. It thus initiates studies of the extent to which divergent diploid genomes interact in determining the molecular characteristics of an allopolyploid' ([12], p. 1081).

In *Clarkia*, Holsinger & Gottlieb [13] used isozyme studies to determine the parentage of tetraploid *C. gracilis*, concluding that one parental species was *Clarkia amoena* subsp. *huntiana* and the other parental species was extinct. Ford & Gottlieb [14] subsequently studied the expression of cytosolic phosphoglucose isomerase (*PgiC*) in this allotetraploid using reverse-transcription polymerase chain reactions and sequencing and found that two copies of *PgiC1* were expressed, but only one copy of the paralogue *PgiC2* was expressed. While it might have been tempting to assume silencing of one homoeologue of *PgiC2* subsequent to allotetraploidization, Ford and Gottlieb found that this gene was commonly silenced in the diploid relatives of the allotetraploid and concluded: '*C. gracilis* expresses all the *PgiC* loci inherited from its diploid progenitors and there is no evidence of subsequent gene silencing'.

In a later study, Ford & Gottlieb [15] examined the allotetraploids *Clarkia delicata* and *Clarkia similis*, whose diploid parental species had previously been identified conclusively and which were extant. Previous isozyme work by Smith-Huerta [17] suggested that silencing of half of the *PgiC* copies had occurred in both allotetraploids. Ford & Gottlieb [15] showed that one of the four gene copies of *PgiC* had been disrupted in both of the allotetraploids, with different mutational events occurring in each one. However, they argued that this silencing could not be proved to be due to tetraploidy *per se* and 'may have more to do with the peculiar properties and history' ([15], p. 706) of the *PgiC* loci: one paralogue had also been silenced in some diploids and retained in duplicate in allotetraploid *C. gracilis* (see above [14]). Though the true sequence of events remains unknown, this emphasizes the importance that Gottlieb placed on parental legacies.

3. Challenges in evaluating parental legacies in allopolyploids

The concept of parental legacies in allopolyploid gene expression is simple, but obviously parental legacies need to be

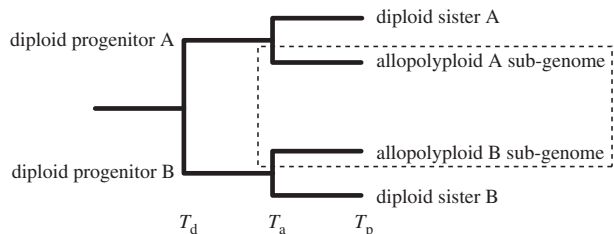


Figure 1. Lineages in the evolution of an allopolyploid. Two diploid progenitors speciate at time T_d and hybridize to form an allopolyploid at time T_a . The two genomes within the allopolyploid then have independent evolutionary trajectories from the diploid genomes of the parental species. Experiments on gene expression in natural allopolyploids take place at the present, time T_p , and patterns of gene expression in the diploid sister genomes are used as surrogates for the patterns of gene expression in the parental diploid species at time T_a .

assessed before we can make inferences about the effects of allopolyploidization on gene expression patterns. As Gottlieb noted, however, there are several considerations and challenges that merit attention when assessing the extent of parental legacies. Given that we still face some of the same challenges that affected Gottlieb's work despite huge methodological advances in transcriptomics, we outline the key considerations below.

(a) Available natural study systems

The biggest challenge in studying allopolyploid gene expression is finding individuals that are truly representative of the original diploid parents of an allopolyploid. This may be impossible due to evolution of the diploid lineages subsequent to the allopolyploidization event (figure 1). As Stebbins had pointed out in 1971:

One cannot assume that the diploid ancestor or ancestors of a modern polyploid species still exist in their original form, unless good evidence for their existence has been obtained. Extinction or cytogenetic modification of diploid ancestors since they participated in the origin of a polyploid are likely possibilities that must always be taken into account. ([18], p. 140)

Gottlieb was acutely aware of this issue in his own research, applying this insight to gene expression patterns. Of his study systems, *S. elata*, *C. delicata* and *C. similis* were of unknown age, and in the case of *C. gracilis*, one diploid progenitor species appeared to be extinct. Gottlieb's early discoveries of continuity in gene expression between diploid progenitors in *Stephanomeria* and *Tragopogon* led him to emphasize and thereby draw attention to the legacy of diploid progenitors when studying allopolyploid gene expression. He considered that in many allopolyploid study systems, it is not possible to directly study the parents even if the progenitor species remain extant: the diploids that we treat as 'parents' in our studies are not the actual parents, but closely related diploid lineages (figure 1). Gottlieb was very rigorous about this in his own research and also in commenting on the research of others. When Adams *et al.* [19] showed for the first time organ-specific reciprocal silencing in an allopolyploid (in cotton), which they suggested was a consequence of polyploidy, Gottlieb commented in *Heredity* ([2], p. 91):

An alternate hypothesis that the difference in expression patterns was a legacy from the diploid progenitors was rejected by the finding that transcripts were present in all of the tested organs of plants representing both diploid progenitors. However, the tested plants are more than a million years away from the

progenitors of cotton and it may be that some differences in organ-specific transcript levels reflect a legacy from the true progenitors. An analysis similar to the one carried out by Adams *et al.* should now be carried out on very recent allotetraploids with extant and identified diploid parents. ([11,12,15,20])

With this comment, Gottlieb was showing a very high level of stringency. He was suggesting that reciprocal organ-specific gene silencing could have occurred in the two diploid parents of the natural allotetraploids and then been convergently lost in the two diploid lineages since allotetraploidization, while remaining unchanged in the allotetraploid over the same time-frame. This is a less parsimonious explanation for the data than that proposed by Adams *et al.* [19], especially when we consider that loss of expression of a gene in a diploid involves zero expression of that gene in a tissue, whereas silencing of a homoeologue in an allopolyploid still allows expression of one copy of that gene in a tissue. Such was Gottlieb's emphasis on the legacy of parental diploids that he wanted this null expectation to be rigorously tested before he was convinced that differential expression among homoeologues in an allopolyploid was due to changes upon or subsequent to allopolyploidization.

Gottlieb sought to solve this problem, as he implied by the choice of papers he cited at the end of the quote from [2] above (i.e. [11,12,15,20]), by working on a natural allopolyploid of very recent origin. When he and Roose began to work on the *Tragopogon* system, they argued that it provided an exceptional opportunity in this respect, writing:

The only unambiguous examples of the recent natural origin of allotetraploid plant species are the two tetraploid species of *Tragopogon*, *T. mirus* and *T. miscellus*, which were discovered and elegantly described by Ownbey [20]. The original populations of both species described by Ownbey are nearly all still extant, and *T. miscellus* has become one of the most common weeds in vacant lots in and around Spokane, Washington, and to the east. These species provide crucial evidence about the initial genetic and biochemical consequences of allotetraploidy because they originated during the present century and it is likely that their genomes have undergone little if any modification since their origin. ([11], p. 819)

Today, several additional recent natural allopolyploids have been found that can be used for such studies: *Mimulus peregrinus* [21,22], *Spartina anglica* [23], *Cardamine schulzii* [24,25] and *Senecio cambrensis* [26]. However, both *M. peregrinus* and *S. cambrensis* have formed via crosses between diploids and tetraploids. Moreover, even in a very recent natural polyploid, we cannot be sure of the exact individuals that acted as parents, and populations of diploids will contain polymorphisms and differences in gene expression patterns that can be hard to take into account [27].

In the absence of a very recent origin, the ideal model system for assessing parental legacies in allopolyploids would need to possess a well-understood phylogenetic framework and have available sufficient genomic and transcriptomic data to permit detailed gene expression analysis. Many genera meet the latter requirement, including *Arabidopsis*, *Gossypium*, *Triticum* and *Glycine*, genera that have proved to be useful models for the study of allopolyploidy. These genera, and others, also offer the advantages of being experimentally facile and provide the opportunity to study artificially produced F_1 hybrids and allopolyploids. In addition, some genera, such as *Glycine*, *Oryza* and *Nicotiana*, contain multiple allopolyploids of different parentage and/or greatly different ages, thus providing a window

into the long-term effects of allopolyploidization on gene expression. Notwithstanding the many important insights derived from these model systems, the major solution to the essential difficulty of accurately assessing diploid legacies continues to be Gottlieb's approach of studying very recent allopolyploids, especially if new high-throughput methods allow broad surveying of polymorphism and gene expression differences within diploid populations.

(b) Technological issues

In the last 10 years, progress in technology has greatly enhanced our ability to study allopolyploid gene expression, permitting, in some cases, genome-wide comparisons of homoeologous gene expression within allopolyploids to those of their parental genes. Ideally, experiments like Roose & Gottlieb's [12] on ADH in *T. miscellus*, where they made absolute measurements of protein expression of both homoeologues and their progenitor genes, are needed on a genome-wide scale.

Initial progression from Gottlieb's studies based on single proteins came in the form of examining RNA transcripts from single genes, for example by cDNA-single-stranded conformation polymorphism in *Gossypium* [19] and by cDNA cleaved amplified polymorphic sequences (CAPSs) in *Tragopogon* [28,29]. Both of these methods can distinguish between homoeologues, but are gel-based and therefore not fully quantitative in the measurement of expression levels. At the same time, anonymous surveys of many genes were carried out using cDNA amplified fragment length polymorphism gels in, for example, *Brassica* [30,31], *Tragopogon* [29], *Triticum* [32,33] and *Arabidopsis* [34].

The first quantitative genome-wide studies of gene expression in allopolyploids compared to their parents were undertaken using microarrays that were unable to distinguish between homoeologous variants of genes. This meant that within allopolyploids, only the overall expression of a particular gene could be measured, but without knowledge of how each homoeologous copy was being regulated. Thus, the only measure of parental legacies was whether or not overall expression was equal to or different from the 'mid-parent value' (MPV), the mean expression of the two parental species (often represented by sister genomes of the allopolyploid sub-genomes; figure 1) at that locus. Such studies were conducted on several allopolyploids, including *Arabidopsis* [35], *Gossypium* [36,37], *Senecio* [26,38], *Brassica* [39], *Triticum* [40–42] and *Spartina* [23,43].

For a better understanding of parental legacies in allopolyploid gene expression, genome-wide methods that distinguish between homoeologues are needed. One method is homoeologue-specific microarrays, as used in *Gossypium* [37,44,45]. Another is quantitative SNP assays, such as the Sequenom MassARRAY, as used in *Tragopogon* [46,47] and *Gossypium* [48]. A third is high-throughput RNA sequencing, as used in *Arabidopsis* [49], *Gossypium* [50], *Nicotiana* [51], *Glycine* [52], *Brassica* [53,54] and *Coffea* [55] allopolyploids. These methods provide the most comprehensive data that we have so far on allopolyploid gene expression, though in general they measure gene expression relative to other genes rather than in an absolute sense (but see [56]), and post-transcriptional regulation can result in levels of protein expression that may not reflect RNA levels in the transcriptome [57,58].

In addition to these RNA-based methods, studies of allopolyploid gene expression have gone full circle and started to

return to the protein level. Gel electrophoresis of proteins has been carried out in studies of *Brassica* [59], *Citrus* [60] and *Musa* [61] allopolyploids, though these methods are not fully quantitative and can only distinguish between homoeologues if they differ in amino acid sequence in a manner that affects gel migration. Mass spectroscopy studies have been carried out in *Arabidopsis* [62], *Brassica* [63,64], *Gossypium* [65–67] and *Tragopogon* [68] allopolyploids; these studies are quantitative when isobaric tags for relative and absolute quantification are used [62,66–68] and can distinguish between homoeologues if they differ in mass.

(c) Multiple comparisons and their terminology

New data from these experimental approaches have yielded rapid progress in our knowledge of the patterns of gene expression in allopolyploids, but have also generated terminological complexity because different methods have measured slightly different aspects of gene expression. These various phenomena have sometimes been referred to by similar terms. Thus, terms essential to discussion of parental legacies, such as 'genome dominance', 'additivity' and 'MPV' have been used in several ways. Much of the resulting confusion relates to whether we are measuring the expression of the two homoeologues relative to each other, or their combined expression levels relative to those of their diploid parents.

Many studies have detected the 'dominance' of one parental genome within an allopolyploid over the other parental genome. In studies that measure the relative expression of each homoeologue, dominance refers to the homoeologue from one parent being expressed at a higher level than the homoeologue from the other parent [45,69,70]. In studies that measure overall levels of expression of genes without distinguishing between homoeologues, dominance refers to the overall expression level of the gene mimicking the expression level of one diploid parent and not the other [71]. Only recently have RNA-seq methods come into use that can measure both of these forms of dominance simultaneously and hence relate them to one another biologically [50]. A suggested terminology to distinguish these two forms of genome dominance has been developed [50,72,73]. 'Expression-level dominance' refers to patterns of overall gene expression (termed 'genome dominance' in [71]), and 'homoeologue expression bias' refers to patterns in the ratio of homoeologous expression (referred to as 'genome dominance' in [70]). A summary of how these terms can be used to describe various patterns of gene expression is shown in figure 2. Expression-level dominance could in principle co-occur with any value of homoeologue expression bias (though Yoo *et al.* [50] showed that in *Gossypium* it most commonly occurs by alteration of the expression level of the homoeologue from the parent not being mimicked) and in some cases, expression-level dominance could co-occur with additive homoeologue expression.

In the context of this review, we emphasize that expression-level dominance cannot be detected without measurements of expression levels in the progenitors. On the other hand, homoeologue expression bias is not defined in terms of parental expression levels and can be detected in the absence of parental expression data, as long as the parental alleles can be identified. In some cases, homoeologue expression bias is caused by a parental legacy, but in many cases it is not. Hence, as illustrated in figure 2, extra adjectives are needed to describe whether or not homoeologue

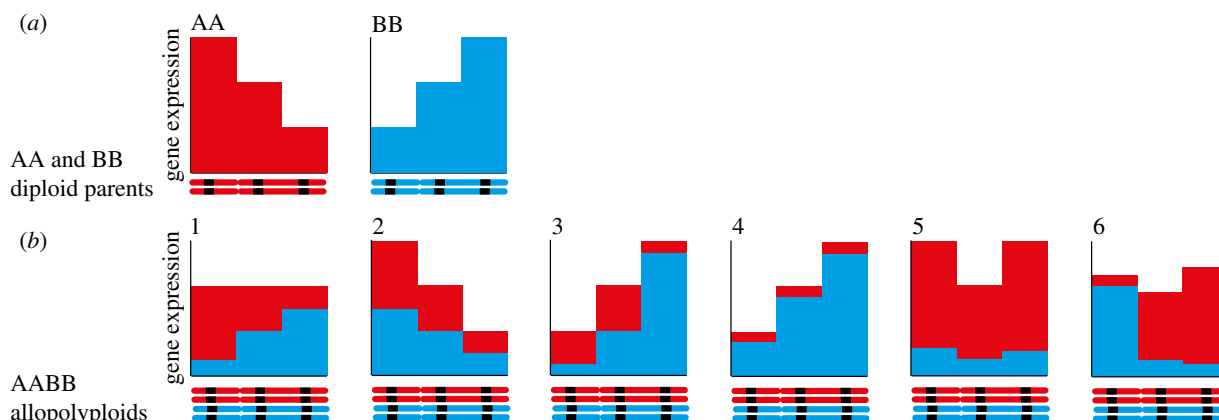


Figure 2. Patterns of gene expression in allopolyploids and their diploid parents. In row (a), example levels of gene expression for three sets of homologous genes in two diploids are shown. Row (b) shows six example patterns of gene expression in an allotetraploid formed from the two diploids, with the contribution of the two homoeologues shown in red and blue: (1) Additive patterns of gene expression, both in terms of absolute level of expression and in terms of their expression ratio of homoeologues towards one another, (2) expression-level dominance by the A sub-genome at all loci, with equal expression of homoeologues at each locus, (3) expression-level dominance by the B sub-genome at all loci, with additive expression of homoeologues at each locus in terms of their expression ratio towards one another, (4) expression-level dominance by the B sub-genome at all loci, with homoeologous expression bias in favour of the B sub-genome, (5) expression-level dominance by different sub-genomes at different loci, and homoeologous expression bias in favour of the A sub-genome at all loci and (6) no expression-level dominance, and homoeologous expression bias differing among loci. (Adapted from [72], pp. 97–98, with permission from Nature Publishing Group.)

expression bias is a parental legacy: if it is, it is often described as ‘additive’. In this context, additive means that relative levels of expression of each homoeologue in an allopolyploid are the same as the relative levels of expression in the two diploid progenitors.

However, this is not the only way in which the word ‘additivity’ is used in the context of allopolyploid gene expression. It should not be used to describe just any case that has equal expression of two homoeologues, because when parental diploids differ in levels of expression of a gene, unbiased (equal) expression of its two homoeologues in an allopolyploid is a non-additive pattern of gene expression. A common, legitimate but different use of the word ‘additive’ is to describe a situation where the combined expression of both homoeologues is equal to the mean of the levels of expression of that locus in both diploid parents [35,38,40–42,69,74]. In Roose & Gottlieb’s [12] quantitative experiment on ADH3 in *T. miscellus* (see above), we see a pattern of gene expression that was additive both in terms of absolute expression levels and also in terms of relative expression of each homoeologue. However, a pattern of homoeologue expression can be additive in one sense but not another. For example, in figure 2b, example 3, levels of homoeologue gene expression are additive in terms of being proportional to the relative expression levels of each gene in the parental diploids, but non-additive in terms of absolute levels of gene expression, because the absolute level of expression of both genes is lower than those of the diploid parents, owing to expression-level dominance. If we think in terms of absolute levels of expression of each homoeologue, additivity and expression-level dominance cannot co-occur at one locus.

To further complicate matters, Gianinetti [75] expresses the opinion that additivity must be measured in terms of absolute levels of gene expression per cell. In most studies published thus far, absolute levels of gene expression per cell or per unit of tissue mass have not been measured ([56] but see, for example, [68]). Microarray studies, while they measure overall expression levels at a locus, usually start with exactly the same amount of total RNA from each

sample being tested, and all measurements are relative to the expression levels of other genes or to total transcription per cell (termed ‘transcriptome-normalized expression’ in [56]). If total transcription per cell (i.e. transcriptome size) differs between a polyploid and its diploid progenitors, then a gene that is ‘additive’ for transcriptome-normalized expression may demonstrate expression-level dominance in terms of expression per cell or vice versa (figure 3). Transcriptome size has been shown to vary between a recently formed allotetraploid and its diploid progenitors [56], as well as between growth conditions or disease states within ploidy levels [76,77], yet few studies have considered additivity in such an absolute sense.

How we think about additivity affects another term frequently used when investigating parental legacies of gene expression: ‘MPV’. This term originates from heritability studies on quantitative phenotypic traits, referring to a trait in an offspring whose value is the mean of the value in the two parents. In studies of allopolyploid gene expression, the MPV is usually (e.g. [35,38,40–42,69,74]) used to refer to an additive pattern of gene expression, when additivity is measured in terms of overall expression of both homoeologues (i.e. the MPV is calculated by summing expression of the two parents at a given locus before dividing by two). As well as defining additivity in terms of absolute expression levels, as opposed to relative measures of expression of homoeologues used in most allopolyploid gene expression studies [56], Gianinetti’s recent critique [75] of the use of the MPV in allopolyploid gene expression studies also argues that because allopolyploids contain two copies of each parental genome, absolute additivity should be a sum, not a mean; he calls this the ‘summed parent value’ (SPV). He bases this on the assumption that absolute levels of gene expression correlate with absolute copy numbers of genes in cells; thus, the null hypothesis is twice as much gene expression in a tetraploid cell as in a diploid cell [78]. This might seem a mistaken assumption if availability of energy and metabolites, rather than gene copy number, is the limiting factor determining overall gene expression levels in a cell. Differences in gene expression due to copy number are most

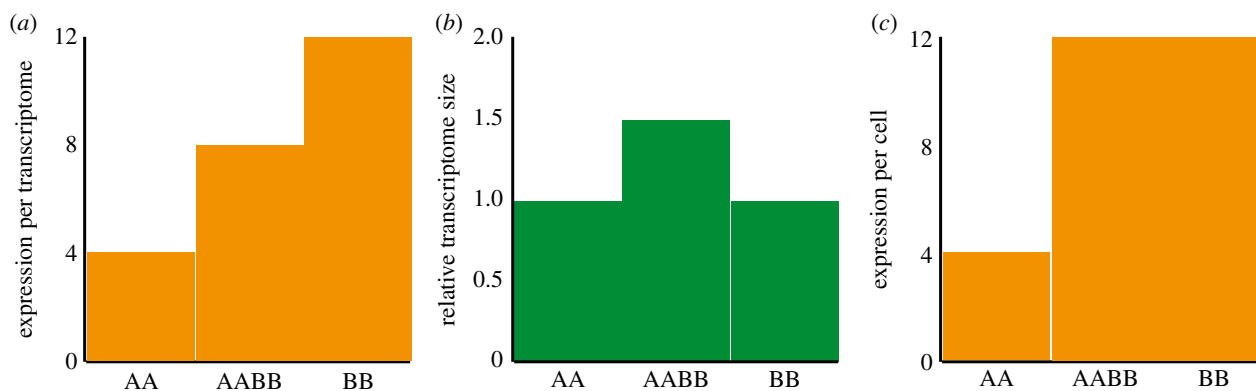


Figure 3. The effect of differences in total transcription (transcriptome size) on the interpretation of additivity and expression-level dominance. (a) Expression levels in an AABB tetraploid and its diploid progenitors (AA and BB) for a gene exhibiting additive overall expression (combined expression of homoeologues in the tetraploid is equal to the mid-parent value of the diploids). As in all standard transcript-profiling experiments, expression levels for individual genes are normalized to total transcription or transcriptome size (e.g. in an RNA-Seq experiment, expression is reported as reads per kilobase per million reads (RPKM)). (b) Relative transcriptome sizes, with the AABB tetraploid having 1.5-fold more transcripts per cell than in either diploid. (c) After adjusting for differences in transcriptome size (multiplying expression per transcriptome by transcriptome size), the gene shown in panel (a) exhibits expression-level dominance by the B sub-genome on a per-cell basis. Expression levels of the individual homoeologues are not shown and could show homoeologue expression bias in either direction, or equal expression.

likely when segmental duplications of genes lead to a different partitioning of resources among particular genes [79,80]. Indeed, it was shown that fewer than 20% of genes exhibited a doubling of expression with a doubling of gene copy number in a *Glycine* allotetraploid, and total transcription per cell was only 1.4-fold higher in the tetraploid than in models of its diploid progenitors [56]. Gianinetti admits that this hypothesis is not generally upheld by experiments, which show MPV to be common and SPV rare (though there are exceptions, particularly when cell size increases [78,81]). Despite this, his hypothesis is an interesting thought-experiment that emphasizes that each gene copy in an allopolyploid tends to be less expressed than its homologue in a diploid parental species, and we do need to think about why this is the case.

It is also worth noting that in practical terms, MPVs have been estimated in at least two ways in the polyploidy literature. They can be calculated either by measuring gene expression levels in the parents separately, and taking an average [35,38,42,69,74], or by physically mixing total RNA from both parents and measuring gene expression levels in the mix [40,41,46]. Both methods have shortcomings: the former relies on accurate normalizations and may fail to factor in the effects of different affinities of different gene variants for microarray baits or PCR primers. The latter method relies on accurate measurements of total RNA concentrations and volumes.

(d) Conclusion regarding challenges

The study of allopolyploid gene expression has greatly benefited from the rigour and stringency of Leslie Gottlieb as an early pioneer. Owing to new technologies, we can now generate huge volumes of data, but data volumes alone cannot solve every difficulty associated with thorough understanding of the evolution of gene expression patterns in allopolyploids. We still need careful experimental designs and good biological understanding of the histories of the study systems we use. Communication and synthesis of results from a variety of study systems and experimental methods will be facilitated by a generally accepted set of terms to describe patterns of gene expression evolution, which to some extent relies

upon a shared theoretical understanding of the underlying biological processes involved.

4. The case study of *Tragopogon* allopolyploids

As noted above, Gottlieb argued [2,11] that recent allopolyploids such as *T. miscellus* and *T. mirus* could help shed light on diploid parental legacies in allopolyploid gene expression. Since Gottlieb's experiments on these species, experiments have been done using CAPS analyses on DNA reverse-transcribed from RNA; this method assays for single nucleotide differences between homoeologous gene copies at restriction enzyme cut sites [46,82]. Subsequent experiments used Sequenom MassARRAY technology, which can give relative measures of the expression of each homoeologue in an allopolyploid [46]. Though these studies mention the influence of parental gene expression patterns on allopolyploid gene expression, this was not a major focus of the analyses. Below, we re-examine the data from these studies in an effort to quantify more precisely the legacy of parental patterns of gene expression, to seek one answer to Gottlieb's question. As noted above, this is not the only system from which such answers can be, or have been, sought, but we use it here as a good case study, being the system that Gottlieb advocated and was most familiar with.

(a) *Tragopogon mirus*

Expression of 13 genes was examined using CAPS, in up to seven tissues, in 10 plants of *T. mirus* and two plants of each of its diploid parental species, *T. porrifolius* and *T. dubius*, randomly sampled from local natural populations [82]. For 11 of these genes, expression was found in all tissues of the parental diploid species. In the allopolyploid *T. mirus* plants, 1518 plant \times gene \times tissue \times homoeologue combinations were assayed. Of these, 85% showed the parental pattern of gene expression (i.e. both homoeologues expressed), and 15% of assays showed silencing of a homoeologue that could have arisen since hybridization. However, 12% showed silencing of a homoeologue in all tissues of a plant, which is most likely to indicate loss of that homoeologue from the genome

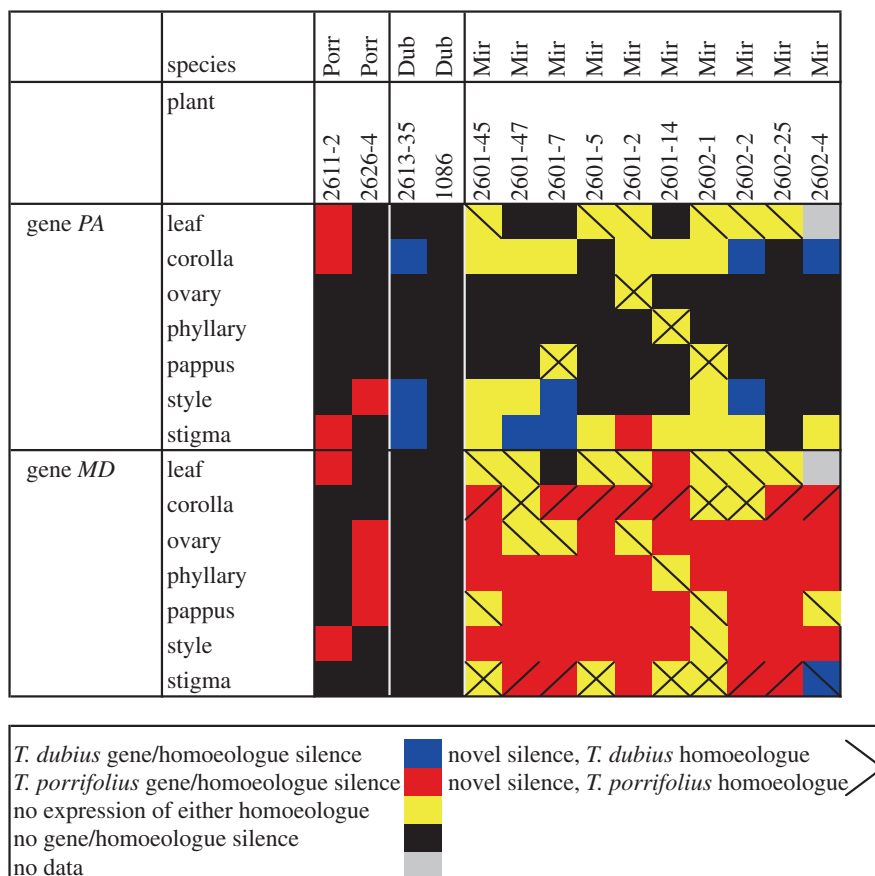


Figure 4. The legacy of diploid parents in *T. mirus*. Tissue-specific expression of homoeologues in *T. mirus* (Mir) for two genes, peroxisomal *MD* and *PA*, that show tissue-specific expression in diploid parental species *T. dubius* (Dub) and *T. porrifolius* (Porr), as reported in [82].

(as shown in other studies in *Tragopogon* allopolyploids [28,29,83–85]). This leaves 3% of assays that showed novel changes in the control of homoeologue expression subsequent to allopolyploidization.

For two of the 13 genes in the *T. mirus* study [82], peroxisomal NAD-malate dehydrogenase (*MD*) and peroxidase (*PA*), tissue-specific silencing was observed in the diploid parental species. It is these cases that would have been of particular interest to Gottlieb as they could give a legacy of tissue-specific homoeologue silencing in allopolyploids that could be spuriously attributed to gene expression changes since allopolyploidization if the parental gene expression patterns were not known. Figure 4 shows the expression of these two genes in parents and allopolyploids. Working out which patterns of gene expression are a legacy from parental diploids is complicated by the fact that gene expression patterns are polymorphic within the diploid species, and none of the cases of gene silencing is found in both individuals of each species. In total, there are 276 plant \times gene \times tissue \times homoeologue combinations in the study for these two genes. If we make the assumption that homoeologue silencing is a legacy from a parental diploid if the same tissue was silent for that gene in either one of the two individual plants assayed within each parental diploid species, then of the 276 assays, 221 (80%) show patterns of homoeologue gene expression that could have been inherited from parental species, and 55 (20%) show novel silencing. In eight *T. mirus* individuals, the *T. porrifolius* homoeologue of gene *MD* is silent in all tissues, which could be due to homoeologue loss from the genome. If it is, then 25 cases of novel silencing could be due to homoeologue loss, leaving only 20 cases

(7.2% of all assays) of novel silencing due to lack of expression of a gene that is present in the genome. These figures may overestimate the diploid parental legacy, because they assume that a silencing event is parental if it is found in any one individual assayed within each parental species. However, as the authors noted, one conclusion, which Gottlieb would have heartily endorsed, is clear:

These studies caution us against attributing all patterns of tissue-specific expression in allopolyploids to processes arising after cross-fertilization between the parental diploids, if the diploids are not known or not examined: genes that show tissue-specific expression in diploids may be likely to do the same in polyploids. ([82], p. 182)

Taking all 13 of the genes in this study of *T. mirus* together, there are 72 cases of tissue-specific gene expression in the allopolyploids, where one homoeologue is expressed and one is not, of which 18% might be due to a legacy of tissue-specific gene silencing from diploid parents (table 1).

(b) *Tragopogon miscellus*

A larger gene expression study was carried out on *T. miscellus* [46] and its diploid parental species, using both CAPS and Sequenom methods. The main emphasis of the paper presenting these data was on the high level of tissue-specific gene expression in diploid progenitors and in natural approximately 40-generation-old allopolyploids, and the apparent relaxation of the tissue-specific patterns observed in the diploid parents in F₁ hybrids and synthetic (S₁) allopolyploids. The discovery of high levels of tissue-specific gene expression in the parental diploids is of course in agreement with Gottlieb's emphasis on the legacy of diploid parents in allopolyploid gene expression.

Table 1. Summary of the frequency of tissue-specific silence (TSS) of homoeologues in the allopolyploids *T. mirus* and *T. miscellus* and its attribution to diploid parental legacies versus novel changes, based on CAPS and Sequenom data in [82] and [46]. See main text for further details.

	<i>T. mirus</i> CAPS	<i>T. miscellus</i> long- liguled CAPS	<i>T. miscellus</i> short- liguled CAPS	<i>T. miscellus</i> long- liguled Sequenom	<i>T. miscellus</i> short- liguled Sequenom
total assays	1794	2016	1906	13 196	9516
silences due to gene loss	186	223	246	849	888
TSS of one homoeologue	72	55	55	357	515
attributable to parental legacy	13 (18%)	29 (53%)	30 (55%)	145 (41%)	146 (28%)
apparently novel	59 (82%)	26 (47%)	25 (45%)	212 (59%)	369 (72%)

Here, we seek to use the data from this earlier study [46] to quantify the possible contribution of the diploid legacy in the natural allopolyploids.

The CAPS study [46] examined 18 genes in up to seven tissues of three individuals of the two diploid parental species (*T. dubius* and *T. pratensis*), five first-generation diploid (F_1) hybrids, three first-generation synthetic allopolyploids (S_1), 10 natural long-liguled allopolyploids (formed from a cross with *T. dubius* as the maternal parent) and 10 natural short-liguled allopolyploids (formed from a cross with *T. dubius* as the paternal parent). Only one gene, *TDF72* (a putative adenine-DNA glycosylase), displayed tissue-specific expression in the diploid progenitors, with expression only in the leaves of *T. dubius*. Unlike genes *MD* and *PA* in the *T. mirus* study, the pattern of expression of *TDF72* was the same in all three individuals of *T. dubius* studied. Gene *TDF72* showed no expression of the *T. dubius* allele in any of the tissues of any of the F_1 and S_1 plants, but in the natural allopolyploids, 15 of the 20 individuals assayed showed expression of the *T. dubius* homoeologue in leaf tissue (we present a new diagram of this in figure 5), and occasional tissue-specific expression in other tissues. Analysing these results, we find that of the 234 assays of this gene in natural *T. miscellus*, 209 (89%) showed parental patterns of gene expression, and 25 (11%) showed non-parental patterns. Of the 25 assays showing non-additive parental patterns, four may have been due to homoeologue loss from the genome. Results for all 18 of the genes analysed in this study of *T. miscellus* are given in table 1, showing that 53% of the assays showing tissue-specific expression in long-liguled *T. miscellus* might be attributed to parental legacy, and 55% in the short-liguled form.

An experiment using quantitative Sequenom MassARRAY allelotyping assays [46] followed a similar design, but examined more genes and more F_1 and S_1 individuals. It assayed the diploids as *in vitro* 'hybrids' (i.e. equimolar mixed transcriptomes from two diploid individuals of different species). Heat map visualizations of the 126 genes assayed showed some genes to have patterns that could be discerned by eye as legacies of diploid parental gene expression patterns. The more striking of these are shown in figure 5. Some of these show parental patterns maintained through the F_1 and S_1 generations as well as in natural populations (such as 08428_719 and 31924_322). Other genes show a gradual diminishing in the occurrence of parental patterns, either by a reduced frequency of occurrence in natural populations (28476_597) or a reduced difference between expressions of homoeologues within the tissues (28117_519). Others showed a loss of parental patterns of tissue-specific expression in the F_1 hybrids,

but a recovery of these patterns in the S_1 allopolyploids and natural allopolyploids (15567_808).

In analysing the entire Sequenom dataset, the authors found that those genes that lack expression of parental homoeologues in specific tissues (termed tissue-specific silence (TSS) by the authors) in the natural allopolyploids tended to fit patterns in the diploid parents more than those of F_1 hybrids and S_1 allopolyploids. They describe this pattern as follows:

We then asked whether the same genes showed TSS in the diploid *in vitro* 'hybrids', F_1 hybrids, S_1 allopolyploids, and natural allopolyploids. There was a significant correlation between the percentage of TSS shown by individual genes in the diploid *in vitro* 'hybrids' and in the natural allopolyploids ($R^2 = 0.307$, $F = 48.23$, $p < 0.0001$), for 111 genes assayed using Sequenom, which were expressed in at least one tissue in every diploid *in vitro* 'hybrid'. There was a weaker correlation between F_1 s and natural allopolyploids ($R^2 = 0.097$, $F = 11.73$, $p < 0.0009$) and between S_1 s and natural allopolyploids ($R^2 = 0.106$, $F = 12.92$, $p < 0.0005$). Therefore, the same genes tended to show TSS in the diploid *in vitro* 'hybrids' and natural allopolyploids despite loss of TSS upon hybridization. It must be emphasized that in the diploids TSS involves total non-expression of that gene in a tissue, whereas in allopolyploids exhibiting TSS, the expression of one homoeologous gene copy is retained. ([46], p. 553)

For the purposes of this review, we have reanalysed the data from this experiment to assess the proportion of TSS in natural populations of *T. miscellus* that could be attributed to inactivity in the parents. For 111 genes assayed using Sequenom, which were expressed in at least one tissue in every diploid *in vitro* 'hybrid', there were 13 196 working assays (i.e. plant \times gene \times tissue \times homoeologue combinations) for long-liguled natural *T. miscellus*, and 9516 for short-liguled *T. miscellus*. Of these working assays, there were 357 (2.7%) cases of TSS in which one homoeologue was not expressed in long-liguled *T. miscellus* that could not be attributed to homoeologue loss from the genome, and 515 cases in short-liguled *T. miscellus* (5.4%). Of these cases of TSS, 41% in long-liguled and 28% in short-liguled *T. miscellus* represented TSS of a homoeologue corresponding to a gene that was also not expressed in the same tissue in at least one diploid individual (for example, the *T. dubius*-derived homoeologue in a *T. miscellus* plant was silent in pappus tissue, and at least one *T. dubius* diploid individual showed lack of expression of that gene in pappus tissue); thus, the lack of expression of a homoeologue in *T. miscellus* potentially represents a legacy from the diploid parents (table 1).

(c) Conclusions from work on *Tragopogon* allopolyploids

The results outlined above, and in table 1, suggest that a significant fraction (18–55%) of cases of tissue-specific homoeologue silencing in natural populations of both

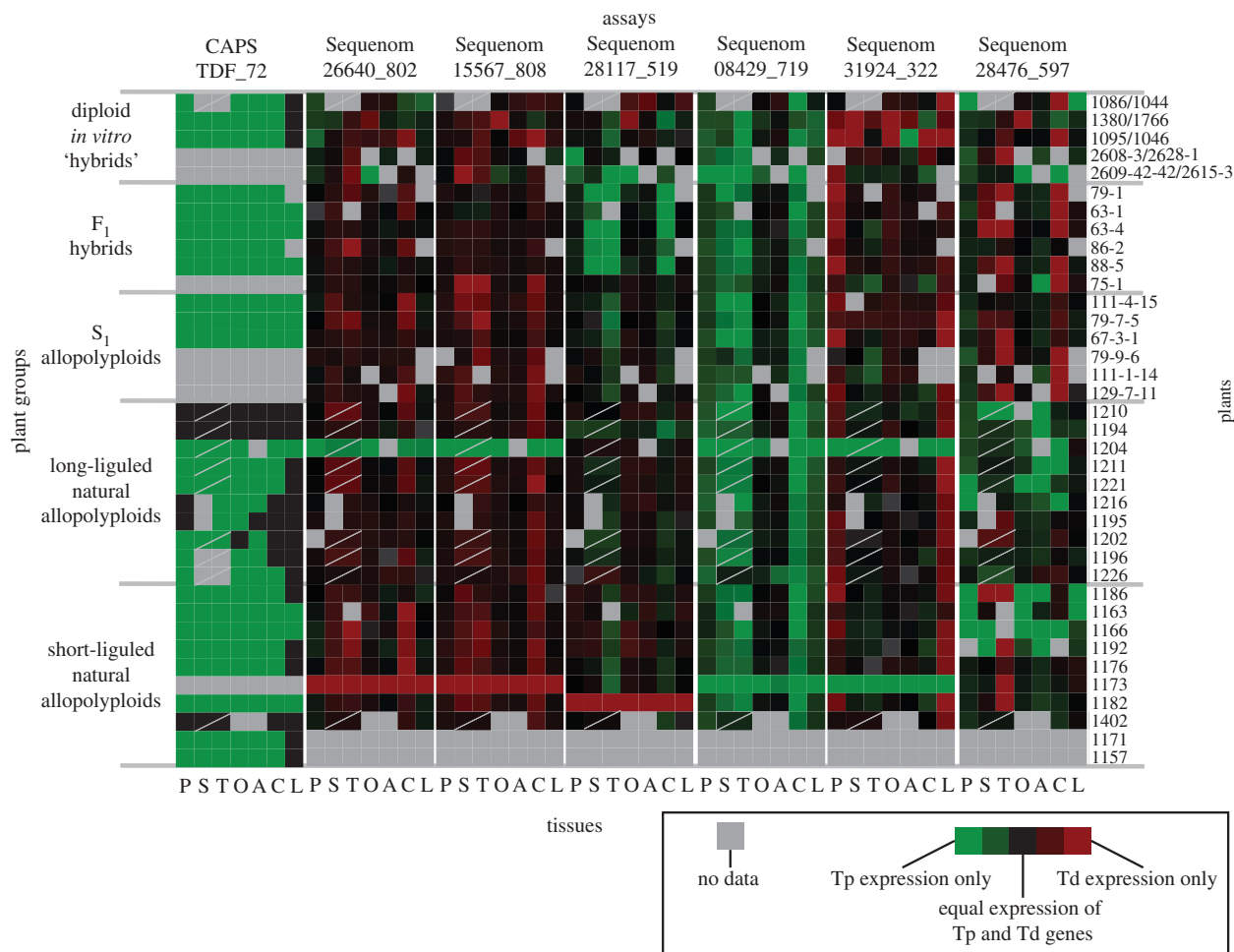


Figure 5. The legacy of diploid parents in *T. miscellus*. Tissue-specific relative expression of gene copies derived from *T. dubius* (Td) and *T. pratensis* (Tp) in diploid *in vitro* 'hybrids', F_1 hybrids, S_1 synthetic allopolyploids and natural populations of *T. miscellus*. Results for seven gene pairs (ID numbers given correspond to those in the original paper) are shown, measured using CAPS or Sequenom analysis. Columns represent tissues, and lines represent plants. Columns are grouped by gene, and rows are grouped by plant types. Tissue abbreviations are as follows: P, phyllary; S, style; T, stigma; O, ovary; A, pappus; C, corolla; L, leaf. Colours show relative tissue-specific expression of Td and Tp gene copies (see legend). Cells joined by diagonal grey lines represent groups of tissues that were assayed together. This figure is derived from figs. 2 and S2 in [46], plus a new representation of *TDF72* results.

T. mirus and *T. miscellus* may be attributed to the legacy of parental diploid patterns of tissue-specific gene expression (noting, however, that these results are based on only a very small fraction of the genes in the genome). These are maximum estimates of the parental legacy, as they assume that a pattern seen in an allopolyploid is derived from a parental diploid even if that pattern was seen in only one of several diploid individuals. In only a few cases, where a tissue-specific pattern for a particular gene was fixed in all diploids sampled, can we be certain that a pattern was inherited from diploid ancestors. These results therefore show that some patterns of tissue-specific homoeologue expression are vertically inherited from diploid parents and show with greater certainty that many tissue-specific patterns of homoeologue expression in allopolyploids are actually novel. Parental legacies do not explain the majority of cases of tissue-specific gene expression in allopolyploids in these experiments: saltational changes in gene expression do occur either upon allopolyploidization, or in the first 40 generations of allopolyploidy. One limitation of the experiments reviewed above is that some patterns of gene expression that are present in some of the allopolyploids might be found in unsampled diploid plants, though it should be noted that the diploids are not as abundant as the allopolyploids in the areas where the allopolyploid have formed.

5. Parental legacies and *cis-* versus *trans-*regulation

A more contemporary way of looking at the issue of parental legacies of gene expression in allopolyploids may invoke the conceptual framework of *cis-* versus *trans-*regulation of gene expression. As mentioned above, Roose & Gottlieb's discovery [12] that differential expression of *ADH3* homoeologues in *T. miscellus* is a legacy of differentially expressed genes in its diploid progenitors may be interpreted as the maintenance of *cis-*regulation of the two gene copies, without interference from *trans-*factors. This topic is of current interest both with respect to the relative prevalence of *cis-* versus *trans-* controls of duplicate gene expression in allopolyploids, and in the related topic of evolutionary divergence among closely related species. The latter topic is an active area of interest in both plants and animals [48,49,74,86–89], and both *cis-* [86] and *trans-* effects [87] have been reported to predominate in the causes of divergence among diploids.

This conceptual framework is important to the interpretation of duplicate gene expression at the allopolyploid level. If homoeologue expression levels quantitatively mimic those of the parental diploids, this may be interpreted as *cis-*regulation in the allopolyploid derivative, due to *cis-* differences that evolved during diploid divergence. If, on the other hand,

duplicate gene expression in the allopolyploid diverges from those of the parental diploids, *trans*-regulatory evolution may be invoked. Chaudhary *et al.* [48] studied homoeologue expression levels for 63 gene pairs in 24 tissues in naturally occurring allopolyploid *Gossypium*, a synthetic allopolyploid of the same genomic composition, and models of the diploid progenitor species. They reported that most gene expression alterations are caused by *cis*-regulatory divergence between the diploid progenitors. Similarly, Yoo *et al.* [50] used RNA-seq to explore patterns of duplicate gene expression for about 30 000 gene pairs in diploid and allopolyploid species of cotton, as well as in synthetic F₁ hybrids and allopolyploids, showing a complex mix of *cis*- and *trans*-regulation characterizing allele-specific expression in diploid hybrids and homoeologue expression in allopolyploids. Shi *et al.* [49] used Illumina sequencing to compare expression levels in *Arabidopsis thaliana* autotetraploids ($2n = 20$), and *Arabidopsis arenosa* autotetraploids ($2n = 32$). They reported that 19% of the expression differences were associated with *cis*- effects, 8% exhibited *trans*- effects and a further 5% were associated with both *cis*- and *trans*- effects.

One potential cause of 'expression-level dominance' could be that *trans*-regulatory phenomena overwhelm any *cis*- differences that may have evolved between the diploid parents; that is, once the two divergent promoter regions are united by allopolyploidy and placed in a common regulatory environment, *trans*-acting factors generate equivalent levels of transcription of both homoeologues (a contrasting possible cause of expression-level dominance is shown in figure 3). It may also be, for example, that de-repression of genes in new allopolyploids, as in *Arabidopsis* [74] and *Tragopogon* [46] allopolyploids, may represent examples of *trans*-regulatory control.

We might hypothesize that *cis*-regulation is most likely to be maintained, and *trans*-regulation avoided, in an allopolyploid if the two parental genomes are divergent at the nucleotide level. The more similar the genomes, the more likely it is that the transcription factors of one genome might be compatible with transcription factor binding sites on the other genome. Hence, it could be that legacies of divergent parental gene expression are more likely to be maintained in allopolyploids with parents that are less genomically similar, or 'wide' hybrids. In recent years, there has been growing interest in the effect of parental divergence on the establishment and success of polyploids [90–95], but as far as we are aware the possible effects of parental divergence on the maintenance of divergent patterns of gene expression among homoeologues

has not yet been explored. A possibly fruitful avenue for future research would be to explore the extent and relative rates of *cis*- and *trans*- divergence among diploids, and the relationships among these phenomena and levels of duplicate gene co-regulation in a broad spectrum of allopolyploids. It is also useful to remind ourselves that, because distinguishing *cis*- and *trans*- effects relies on identifying species-specific patterns of expression [86], inferences about the regulation of gene expression in allopolyploids are complicated by the same considerations concerning polymorphisms in progenitor species and divergence subsequent to polyploid formation that Gottlieb recognized as important constraints on evolutionary hypotheses based on isozyme data.

6. Conclusion

Most reviews of allopolyploid gene expression in the last 10 years have emphasized novel and variable patterns of gene expression [96–102], due to multiple recent discoveries of alteration of gene expression patterns [23,26,45,46,103] and variation among tissues and progenies of similar allopolyploids [50,104]. In this review, we have concentrated on the converse: continuity among allopolyploids and their diploid progenitors, inspired by the early insights of Leslie Gottlieb in this field. While accurate assessment of parental legacies is a prerequisite for the documentation of novel patterns of gene expression in allopolyploids, it is not straightforward to measure or to describe. Despite this, results from *Tragopogon* allopolyploids, a study system commended by Gottlieb, show us that, while making an important contribution, parental legacies of gene expression are often not the causes of expression differences among homoeologues in allopolyploids. These studies and others indicate that gene expression evolution at the allopolyploid level is far more complex than mere inheritance of a parental legacy, and that effects are nonlinear, in many cases being caused by interacting *cis*- and *trans*-factors whose interactions and stoichiometric responses to hybridization, genome doubling and environmental conditions are rapid and highly complex.

Funding statement. R.J.A.B. acknowledges support from Natural Environment Research Council Fellowship NE/G01504X/1; J.E.C. and J.J.D. acknowledge support from National Science Foundation (NSF) grant nos. IOS-0822258 and DEB-1257522; P.S.S. and D.S.S. acknowledge support from US NSF grant nos. DEB-0614421, MCB-0346437 and DEB-0919254; J.F.W. acknowledges support from NSF grant nos. IOS-0817707 and MCB-1118646.

References

- Gottlieb LD. 1982 Conservation and duplication of isozymes in plants. *Science* **216**, 373–379. (doi:10.1126/Science.216.4544.373)
- Gottlieb LD. 2003 Plant polyploidy: gene expression and genetic redundancy. *Heredity* **91**, 91–92. (doi:10.1038/Sj.Hdy.6800317)
- Soltis PS, Liu X, Marchant DB, Visger CJ, Soltis DE. 2014 Polyploidy and novelty: Gottlieb's legacy. *Phil. Trans. R. Soc. B* **369**, 20130351. (doi:10.1098/rstb.2013.0351)
- Ramsey J, Ramsey TS. 2014 Ecological studies of polyploidy in the 100 years following its discovery. *Phil. Trans. R. Soc. B* **369**, 20130352. (doi:10.1098/rstb.2013.0352)
- Vanneste K, Maere S, Van de Peer Y. 2014 Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Phil. Trans. R. Soc. B* **369**, 20130353. (doi:10.1098/rstb.2013.0353)
- Jiao Y, Paterson AH. 2014 Polyploidy-associated genome modifications during land plant evolution. *Phil. Trans. R. Soc. B* **369**, 20130355. (doi:10.1098/rstb.2013.0355)
- Gottlieb LD. 1971 Gel electrophoresis: new approach to study of evolution. *Bioscience* **21**, 939–944. (doi:10.2307/1296112)
- Gottlieb LD. 1973 Genetic differentiation, sympatric speciation, and origin of a diploid species of *Stephanomeria*. *Am. J. Bot.* **60**, 545–553. (doi:10.2307/2441378)
- Gottlieb LD. 1973 Enzyme differentiation and phylogeny in *Clarkia franciscana* and *Clarkia rubicunda*. *Genetics* **74**, S99.
- Gottlieb LD. 1973 Genetic control of glutamate oxaloacetate transaminase isoenzymes in diploid

- plant *Stephanomeria exigua* and its allotetraploid derivative. *Biochem. Genet.* **9**, 97–107. (doi:10.1007/Bf00485595)
11. Roose ML, Gottlieb LD. 1976 Genetic and biochemical consequences of polyploidy in *Tragopogon*. *Evolution* **30**, 818–830. (doi:10.2307/2407821)
 12. Roose ML, Gottlieb LD. 1980 Biochemical properties and level of expression of alcohol dehydrogenases in the allotetraploid plant *Tragopogon miscellus* and its diploid progenitors. *Biochem. Genet.* **18**, 1065–1085. (doi:10.1007/Bf00484339)
 13. Holsinger KE, Gottlieb LD. 1988 Isozyme variability in the tetraploid *Clarkia gracilis* (Onagraceae) and its diploid relatives. *Syst. Bot.* **13**, 1–6. (doi:10.2307/2419235)
 14. Ford VS, Gottlieb LD. 1999 Molecular characterization of *PgiC* in a tetraploid plant and its diploid relatives. *Evolution* **53**, 1060–1067. (doi:10.2307/2640811)
 15. Ford VS, Gottlieb LD. 2002 Single mutations silence *PgiC2* genes in two very recent allotetraploid species of *Clarkia*. *Evolution* **56**, 699–707. (doi:10.1111/j.0014-3820.2002.tb01381.x)
 16. Gottlieb LD. 1969 *The role of hybridization in the evolution of the annual species of Stephanomeria (Compositae)*. Ann Arbor, MI: University of Michigan.
 17. Smith-Huerta NL. 1986 Isozymic diversity in three allotetraploid *Clarkia* species and their putative diploid progenitors. *J. Hered.* **77**, 349–354.
 18. Stebbins GL. 1971 *Chromosomal evolution in higher plants*. London, UK: Edward Arnold.
 19. Adams KL, Cronn R, Percifield R, Wendel JF. 2003 Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl Acad. Sci. USA* **100**, 4649–4654. (doi:10.1073/pnas.0630618100)
 20. Ownbey M. 1950 Natural hybridization and amphiploidy in the genus *Tragopogon*. *Am. J. Bot.* **37**, 487–499. (doi:10.2307/2438023)
 21. Vallejo-Marin M. 2012 *Mimulus peregrinus* (Phrymaceae): a new British allopolyploid species. *PhytoKeys* **14**, 1–14. (doi:10.3897/phytokeys.14.3305)
 22. Buggs RJA. 2012 Monkeying around with ploidy. *Mol. Ecol.* **21**, 5159–5161. (doi:10.1111/mec.12005)
 23. Chelaifa H, Monnier A, Ainouche M. 2010 Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina × townsendii* and *Spartina anglica* (Poaceae). *New Phytol.* **186**, 161–174. (doi:10.1111/j.1469-8137.2010.03179.x)
 24. Urbanska KM, Hurka H, Landolt E, Neuffer B, Mummenhoff K. 1997 Hybridization and evolution in *Cardamine* (Brassicaceae) at Urnerboden, Central Switzerland: biosystematic and molecular evidence. *Plant Syst. Evol.* **204**, 233–256. (doi:10.1007/BF00989208)
 25. Mandáková T, Kovařík A, Zozomová-Lihová J, Shimizu-Inatsugi R, Shimizu KK, Mummenhoff K, Marhold K, Lysak MA. 2013 The more the merrier: recent hybridization and polyploidy in *Cardamine*. *Plant Cell* **25**, 3280–3295. (doi:10.1105/tpc.113.114405)
 26. Hegarty M, Barker G, Wilson I, Abbott RJ, Edwards KJ, Hiscock SJ. 2006 Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Curr. Biol.* **16**, 1652–1659 (doi:10.1016/j.cub.2006.06.071)
 27. Doyle JJ, Egan AN. 2010 Dating the origins of polyploidy events. *New Phytol.* **186**, 73–85. (doi:10.1111/j.1469-8137.2009.03118.x)
 28. Tate J, Joshi P, Soltis K, Soltis P, Soltis D. 2009 On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). *BMC Plant Biol.* **9**, 80. (doi:10.1186/1471-2229-9-80)
 29. Tate JA, Ni ZF, Scheen AC, Koh J, Gilbert CA, Lefkowitz D, Chen ZJ, Soltis PS, Soltis DE. 2006 Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* **173**, 1599–1611. (doi:10.1534/genetics.106.057646)
 30. Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007 Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* **19**, 3403–3417. (doi:10.1105/tpc.107.054346)
 31. Xu Y, Zhong L, Wu X, Fang X, Wang J. 2009 Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta* **229**, 471–483. (doi:10.1007/s00425-008-0844-8)
 32. He P, Friebe BR, Gill BS, Zhou JM. 2003 Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant Mol. Biol.* **52**, 401–414. (doi:10.1023/A:1023965400532)
 33. Kashkush K, Feldman M, Levy AA. 2002 Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* **160**, 1651–1659.
 34. Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002 Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiol.* **129**, 733–746. (doi:10.1104/pp.003095)
 35. Wang J *et al.* 2006 Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* **172**, 507–517. (doi:10.1534/genetics.105.047894)
 36. Udall JA *et al.* 2007 Spotted cotton oligonucleotide microarrays for gene expression analysis. *BMC Genomics* **8**, 81. (doi:10.1186/1471-2164-8-81)
 37. Hovav R, Udall JA, Chaudhary B, Rapp R, Flagel L, Wendel JF. 2008 Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc. Natl Acad. Sci. USA* **105**, 6191–6195. (doi:10.1073/pnas.0711569105)
 38. Hegarty MJ *et al.* 2005 Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. *Mol. Ecol.* **14**, 2493–2510. (doi:10.1111/j.1365-294X.2005.02608.x)
 39. Gaeta RT, Yoo S-Y, Pires JC, Doerge RW, Chen ZJ, Osborn TC. 2009 Analysis of gene expression in resynthesized *Brassica napus* allopolyploids using *Arabidopsis* 70mer oligo microarrays. *PLoS ONE* **4**, e4760. (doi:10.1371/journal.pone.0004760)
 40. Chelaifa H *et al.* 2013 Prevalence of gene expression additivity in genetically stable wheat allohexaploids. *New Phytol.* **197**, 730–736. (doi:10.1111/nph.12108)
 41. Chagué V *et al.* 2010 Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytol.* **187**, 1181–1194. (doi:10.1111/j.1469-8137.2010.03339.x)
 42. Pumphrey MO, Bai J, Laudencia-Chinguanco D, Anderson O, Gill BS. 2008 Non-additive expression of homeologous genes is established upon polyploidization in hexaploid wheat. *Genetics* **181**, 1147–1157. (doi:10.1534/genetics.108.096941)
 43. Chelaifa H, Mahe F, Ainouche M. 2010 Transcriptome divergence between the hexaploid salt-marsh sister species *Spartina maritima* and *Spartina alterniflora* (Poaceae). *Mol. Ecol.* **19**, 2050–2063. (doi:10.1111/j.1365-294X.2010.04637.x)
 44. Udall JA, Swanson JM, Nettleton D, Percifield RJ, Wendel JF. 2006 A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* **173**, 1823–1827. (doi:10.1534/genetics.106.058271)
 45. Flagel LE, Udall J, Nettleton D, Wendel J. 2008 Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.* **6**, 16. (doi:10.1186/1741-7007-6-16)
 46. Buggs RJA *et al.* 2011 Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr. Biol.* **21**, 551–556. (doi:10.1016/j.cub.2011.02.016)
 47. Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2010 Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Mol. Ecol.* **19**, 132–146. (doi:10.1111/j.1365-294X.2009.04469.x)
 48. Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF. 2009 Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics* **182**, 503–517. (doi:10.1534/genetics.109.102608)
 49. Shi X, Ng DWK, Zhang C, Comai L, Ye W, Jeffrey CZ. 2012 *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat. Commun.* **3**, 950. (doi:10.1038/ncomms1954)
 50. Yoo MJ, Szadkowski E, Wendel JF. 2013 Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**, 171–180. (doi:10.1038/hdy.2012.94)
 51. Bombarely A, Edwards K, Sanchez-Tamburrino J, Mueller L. 2012 Deciphering the complex leaf transcriptome of the allotetraploid species *Nicotiana tabacum*: a phylogenomic perspective. *BMC Genomics* **13**, 406. (doi:10.1186/1471-2164-13-406)

52. Ilut DC, Coate JE, Luciano AK, Owens TG, May GD, Farmer A, Doyle JJ. 2012 A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. *Am. J. Bot.* **99**, 383–396. (doi:10.3732/ajb.1100312).
53. Higgins J, Magusin A, Trick M, Fraser F, Bancroft I. 2012 Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. *BMC Genomics* **13**, 247. (doi:10.1186/1471-2164-13-247)
54. Jiang J, Shao Y, Du K, Ran L, Fang X, Wang Y. 2013 Use of digital gene expression to discriminate gene expression differences in early generations of resynthesized *Brassica napus* and its diploid progenitors. *BMC Genomics* **14**, 72. (doi:10.1186/1471-2164-14-72)
55. Combes M-C, Dereeper A, Severac D, Bertrand B, Lashermes P. 2013 Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytol.* **200**, 251–260. (doi:10.1111/nph.12371).
56. Coate JE, Doyle JJ. 2010 Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. *Genome Biol. Evol.* **2**, 534–546. (doi:10.1093/gbe/evq038)
57. Vogel C, Marcotte EM. 2012 Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232.
58. Coate JE, Bar H, Doyle JJ. 2014 Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). *Plant Cell* **26**, 136–150. (doi:10.1105/tpc.113.119966)
59. Kong F, Mao S, Jiang J, Wang J, Fang X, Wang Y. 2011 Proteomic changes in newly synthesized *Brassica napus* allotetraploids and their early generations. *Plant Mol. Biol. Rep.* **29**, 927–935. (doi:10.1007/s11105-011-0301-x)
60. Gancel A-L, Grimplet JRM, Sauvage FO-X, Ollitrault P, Brillouet J-M. 2006 Predominant expression of diploid mandarin leaf proteome in two citrus mandarin-derived somatic allotetraploid hybrids. *J. Agric. Food Chem.* **54**, 6212–6218. (doi:10.1021/jf060657p).
61. Carpentier SC, Panis B, Renaut J, Samyn B, Vertommen A, Vanhove A-C, Swennen R, Sergeant K. 2011 The use of 2D-electrophoresis and de novo sequencing to characterize inter- and intra-cultivar protein polymorphisms in an allopolyploid crop. *Phytochemistry* **72**, 1243–1250. (doi:10.1016/j.phytochem.2010.10.016)
62. Ng DWK, Zhang C, Miller M, Shen Z, Briggs SP, Chen ZJ. 2012 Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors. *Heredity* **108**, 419–430. (doi:10.1038/hdy.2011.92)
63. Albertin W, Alix K, Balliau T, Brabant P, Davanture M, Malosse C, Valot B, Thiellement H. 2007 Differential regulation of gene products in newly synthesized *Brassica napus* allotetraploids is not related to protein function nor subcellular localization. *BMC Genomics* **8**, 56. (doi:10.1186/1471-2164-8-56)
64. Albertin W, Balliau T, Brabant P, Chevre AM, Eber F, Malosse C, Thiellement H. 2006 Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. *Genetics* **173**, 1101–1113. (doi:10.1534/genetics.106.057554).
65. Hu G, Houston NL, Pathak D, Schmidt L, Thelen JJ, Wendel JF. 2011 Genomically biased accumulation of seed storage proteins in allopolyploid cotton. *Genetics* **189**, 1103–1115. (doi:10.1534/genetics.111.132407)
66. Bao Y, Hu G, Flagel LE, Salmon A, Bezanilla M, Paterson AH, Wang Z, Wendel JF. 2011 Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (*Gossypium*). *Proc. Natl Acad. Sci. USA* **108**, 21 152–21 157. (doi:10.1073/pnas.1115926109)
67. Hu G, Koh J, Yoo M-J, Grupp K, Chen S, Wendel JF. 2013 Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. *New Phytol.* **200**, 570–582. (doi:10.1111/nph.12381)
68. Koh J, Chen S, Zhu N, Yu F, Soltis PS, Soltis DE. 2012 Comparative proteomics of the recently and recurrently formed natural allopolyploid *Tragopogon mirus* (Asteraceae) and its parents. *New Phytol.* **196**, 292–305. (doi:10.1111/j.1469-8137.2012.04251.x)
69. Flagel LE, Wendel JF. 2010 Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* **186**, 184–193. (doi:10.1111/j.1469-8137.2009.03107.x)
70. Schnable JC, Springer NM, Freeling M. 2011 Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl Acad. Sci. USA* **108**, 4069–4074. (doi:10.1073/pnas.1101368108)
71. Rapp RA, Udall JA, Wendel JF. 2009 Genomic expression dominance in allopolyploids. *BMC Biol.* **7**, 18. (doi:10.1186/1741-7007-7-18)
72. Buggs RJA. 2013 Unravelling gene expression of complex crop genomes. *Heredity* **110**, 97–98. (doi:10.1038/hdy.2012.61)
73. Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012 Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* **196**, 966–971. (doi:10.1111/j.1469-8137.2012.04365.x)
74. Wang JL, Tian L, Lee HS, Chen ZJ. 2006 Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in *Arabidopsis* allopolyploids. *Genetics* **173**, 965–974. (doi:10.1534/genetics.106.056580)
75. Gianinetti A. 2013 A criticism of the value of midparent in polyploidization. *J. Exp. Bot.* **64**, 4119–4129. (doi:10.1093/jxb/ert263)
76. Lovén J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, Levens DL, Lee TI, Young RA. 2012 Revisiting global gene expression analysis. *Cell* **151**, 476–482. (doi:10.1016/j.cell.2012.10.012)
77. Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R, Bühler J. 2012 Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151**, 671–683. (doi:10.1016/j.cell.2012.09.019)
78. Guo M, Davis D, Birchler JA. 1996 Dosage effects on gene expression in a maize ploidy series. *Genetics* **142**, 1349–1355.
79. Freeling M. 2009 Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **60**, 433–453. (doi:10.1146/annurev.arplant.043008.092122).
80. Birchler JA, Veitia RA. 2012 Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl Acad. Sci. USA* **109**, 14 746–14 753. (doi:10.1073/pnas.1207726109)
81. Yao H, Kato A, Mooney B, Birchler J. 2011 Phenotypic and gene expression analyses of a ploidy series of maize inbred Oh43. *Plant Mol. Biol.* **75**, 237–251. (doi:10.1007/s11103-010-9722-4)
82. Buggs RJA, Elliott NM, Zhang LJ, Koh J, Viccini LF, Soltis DE, Soltis PS. 2010 Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytol.* **186**, 175–183. (doi:10.1111/j.1469-8137.2010.03205.x)
83. Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE. 2009 Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity* **103**, 73–81. (doi:10.1038/hdy.2009.24)
84. Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2012 Rapid, repeated and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr. Biol.* **22**, 248–252. (doi:10.1016/j.cub.2011.12.027)
85. Soltis PS, Plunkett GM, Novak SJ, Soltis DE. 1995 Genetic variation in *Tragopogon* species: additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae). *Am. J. Bot.* **82**, 1329–1341. (doi:10.2307/2446255)
86. Wittkopp PJ, Haerum BK, Clark AG. 2004 Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**, 85–88. (doi:10.1038/nature02698)
87. Bell GDM, Kane NC, Rieseberg LH, Adams KL. 2013 RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol. Evol.* **5**, 1309–1323. (doi:10.1093/gbe/evt072)
88. Meiklejohn CD, Coolon JD, Hartl DL, Wittkopp PJ. 2014 The roles of *cis*- and *trans*-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Res.* **24**, 84–95. (doi:10.1101/gr.156414.113)
89. Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, Brazma A, Odom DT, Marioni JC. 2012 Extensive compensatory *cis*–*trans* regulation in the evolution of mouse gene expression. *Genome Res.* **22**, 2376–2384. (doi:10.1101/gr.142281.112)

90. Buggs RJA, Soltis PS, Mavrodiev EV, Symonds VV, Soltis DE. 2008 Does phylogenetic distance between parental genomes govern the success of polyploids? *Castanea* **73**, 74–93. (doi:10.2179/0008-7475(2008)73[74:DPDBPG]2.0.CO;2)
91. Buggs RJA, Soltis PS, Soltis DE. 2009 Does hybridization between divergent progenitors drive whole-genome duplication? *Mol. Ecol.* **18**, 3334–3339. (doi:10.1111/j.1365-294X.2009.04285.x)
92. Buggs RJA, Soltis PS, Soltis DE. 2011 Biosystematic relationships and the formation of polyploids. *Taxon* **60**, 324–332.
93. Chapman MA, Burke JM. 2007 Genetic divergence and hybrid speciation. *Evolution* **61**, 1773–1780. (doi:10.1111/j.1558-5646.2007.00134.x)
94. Paun O, Forest F, Fay MF, Chase MW. 2009 Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytol.* **182**, 507–518. (doi:10.1111/j.1469-8137.2009.02767.x)
95. Paun O, Forest F, Fay MF, Chase MW. 2011 Parental divergence and hybrid speciation in angiosperms revisited. *Taxon* **60**, 1241–1244.
96. Adams KL. 2007 Evolution of duplicate gene expression in polyploid and hybrid plants. *J. Hered.* **98**, 136–141. (doi:10.1093/jhered/esl061)
97. Adams KL, Wendel JF. 2005 Novel patterns of gene expression in polyploid plants. *Trends Genet.* **21**, 539–543. (doi:10.1016/j.tig.2005.07.009)
98. Chen ZJ, Ni Z. 2006 Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* **28**, 240–252. (doi:10.1002/bies.20374)
99. Hegarty MJ, Hiscock SJ. 2008 Genomic clues to the evolutionary success of polyploid plants. *Curr. Biol.* **18**, R435–R444. (doi:10.1016/j.cub.2008.03.043)
100. Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008 Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* **42**, 443–461. (doi:10.1146/annurev.genet.42.110807.091524)
101. Soltis PS, Soltis DE. 2009 The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* **60**, 561–588. (doi:10.1146/annurev.arplant.043008.092039)
102. Comai L, Madlung A, Josefsson C, Tyagi A. 2003 Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? *Phil. Trans. R. Soc. Lond. B* **358**, 1149–1155. (doi:10.1098/rstb.2003.1305)
103. Ha M *et al.* 2009 Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc. Natl Acad. Sci. USA* **106**, 17 835–17 840. (doi:10.1073/pnas.0907003106)
104. Bardil A, de Almeida JD, Combes MC, Lashermes P, Bertrand B. 2011 Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. *New Phytol.* **192**, 760–774. (doi:10.1111/j.1469-8137.2011.03833.x)