Journal of Heredity Advance Access published May 23, 2016

Morphological convergence between an allopolyploid and one of its parental species correlates with biased gene expression and DNA loss

Douglas Alexander-Webber [1], Richard J Abbott [2] and Mark A Chapman [1,*]

[1] Centre for Biological Sciences, University of Southampton, Southampton, SO17 1BJ

[2] School of Biology, Sir Harold Mitchell Building, University of St Andrews, St Andrews, FIFE, KY16 9TH

[*] Author for correspondence

Email: M.Chapman@soton.ac.uk

Phone: +44 (0)2380 594396

Abstract

The contribution of gene expression modulation to phenotypic evolution is of major importance to an understanding of the origin of divergent or convergent phenotypes during and following polyploid speciation. Here we analysed genome-wide gene expression in two subspecies of the allotetraploid species, *Senecio mohavensis* A. Gray, and its diploid parents *S. flavus* (Decne.) Sch. Bip. and *S. glaucus* L.. The tetraploid is morphologically much more similar to *S. flavus*, leading to earlier confusion over its taxonomic status. By means of an analysis of transcriptomes of all three species, we show that gene expression divergence between the parent species is relatively low (ca. 14% of loci), whereas there is significant unequal expression between ca. 20-25% of the parental homoeologues (gene copies) in the tetraploid. The majority of the expression bias in the tetraploid is in favour of *S. flavus* homoeologues (ca. 65% of the differentially expressed loci), and overall expression of this parental species sub-genome is higher than that of the *S. glaucus* sub-genome. To determine whether absence of expression of a particular *S. glaucus* homoeologue in the allotetraploid could be due to loss of DNA, we carried out a PCR-based assay and confirmed that in three out of 10 loci the *S. glaucus* homoeologue appeared absent. Our results suggest that biased gene expression is one cause of the allotetraploid *S. mohavensis* being more similar in morphology to one of its parent, *S. flavus*, and that such bias could result, in part, from loss of *S. glaucus* homoeologues at some loci in the allotetraploid.

Introduction

Polyploidy is widespread throughout the flowering plants, clearly playing an important role in speciation and the generation of biodiversity (Adams and Wendel, 2005b; Jaillon*, et al.*, 2007; Soltis, Visger and Soltis, 2014). The reasons for the evolutionary success of polyploids have been much debated (Madlung, 2013). It has been suggested, for example, that the presence of multiple copies of each locus buffers against the effect of deleterious recessive mutations (e.g. Gu*, et al.*, 2003). Alternatively, increased fitness because of heterosis could arise due to an increase in heterozygosity of the genome (e.g. Birchler*, et al.*, 2010). Third, because each locus is present in duplicate, one copy is free to evolve a new or variant function (neofunctionalisation or subfunctionalisation) without causing a reduction in fitness because the duplicate copy can maintain the original function (Adams and Wendel, 2005a; Moore and Purugganan, 2005).

Following the formation of an allopolyploid, the genome of the resultant offspring contains the entire complement of DNA from the parental species (Soltis and Soltis, 2000). However, when natural polyploid species have been compared to their diploid progenitors, they very often show some loss of parental DNA, and there can be unequal loss from the different parents (Buggs*, et al.*, 2014; Hufton and Panopoulou, 2009; Tate*, et al.*, 2006). Similarly, gene expression in allopolyploids often shows differences in the overall contribution of the two parental genomes, as well as tissue-specific differences (Rapp, Udall and Wendel, 2009; Roulin*, et al.*, 2013; Wang, Wang and Paterson, 2012).

It has become clear that the elimination of parental DNA and changes in gene expression can arise very early in the evolution of an allopolyploid (Adams and Wendel, 2005a; Buggs*, et al.*, 2012; Chen, 2007). By studying artificial allopolyploids, the loss of parental DNA (Khasdan*, et al.*, 2010; Lukens*, et al.*, 2006; Skalická*, et al.*, 2005) and changes in parental gene expression (Gaeta*, et al.*, 2007; Hegarty*, et al.*, 2006) have been documented within the first few generations. These changes appear to 'stabilise' such that established polyploid species evolve a 'diploidised' genome, however large portions of the genome remain duplicated.

In an allopolyploid, the subgenomes are exposed to a novel transcriptional, translational and epigenetic environment; therefore alteration of gene expression is not unexpected (Adams*, et al.*, 2003). However, the link between gene expression variation and phenotypic evolution in polyploids is not well known. To this end, we chose to investigate gene expression variation in an allopolyploid and its diploid progenitor species where the allopolyploid is much more similar in morphology to one of its parental species. The allotetraploid, *Senecio mohavensis* A. Gray (Asteraceae) (2*n*=40),

comprises a subspecies restricted to the Mojave Desert area in North America, ssp. *mohavensis*, and a North African subspecies, ssp. *breviflorus* (Kadereit) M. Coleman (Coleman, Forbes and Abbott, 2001; Coleman*, et al.*, 2003; Liston, Rieseberg and Elias, 1989). Its allopolyploid status was established from molecular work (see below) with *S. flavus* (Dechne.) Sch. Bip. (2*n*=20) and *S. glaucus* L. ssp. *coronopifolius* (Maire) Alexander (2*n*=20) designated as its diploid parents. Both parent species occur in North Africa and are absent from North America (Kadereit*, et al.*, 2006), and are estimated to have diverged ca. 10 million years (MY) ago with the allopolyploid forming very recently, most likely within the last 1 MY (Coleman*, et al.*, 2003). The presence of *S. mohavensis* in North America is attributed to long distance dispersal from North Africa (Coleman *et al*., 2003).

Senecio mohavensis was initially thought to be a subspecies of *S. flavus* (i.e. *S. flavus* (Decne.) Schultz Bip. ssp. *breviflorus* Kadereit [Kadereit, 1984]) until it was established to be tetraploid (Coleman*, et al.*, 2001). This previous designation was primarily due to the morphological similarity between *S. mohavensis* ssp. *breviflorus* and *S. flavus* which both occur in North Africa. *Senecio flavus,* however, has non-radiate capitula, whereas *S. mohavensis* is typically radiate. Molecular analysis further revealed that *S. mohavensis* resembled *S. glaucus* with regard to both its plastid DNA and ITS sequences (Coleman*, et al.*, 2003; Comes and Abbott, 2001), but exhibited the additive genotype of its two parent species for sequences of two nuclear loci, *PgiC* and *Scyc2* (Chapman, 2004) and for a range of randomly amplified nuclear markers (Comes and Abbott, 2001; Kadereit*, et al.*, 2006).

By sequencing pooled transcriptomes of the four taxa we were able to investigate the gross divergence in expression between the parental homoeologues in *S. mohavensis* as well as identify the number of homoeologues in the tetraploid that are differentially expressed, and the direction of expression bias for these homoeologues. Several homoeologues showed differential expression, prompting a follow-up PCR-based assay to determine if the parental DNA had been lost, or was present but transcriptionally silenced.

Materials and Methods

Morphological analysis

Achenes (i.e. single seeded fruits) of the four taxa were collected from the wild in the following countries: *S. glaucus* - Morocco, Tunisia and Israel; *S. flavus* - Morocco, Spain (the Canary Islands), and Egypt; *S. mohavensis* ssp. *mohavensis* - USA (Arizona, California and Nevada); and *S. mohavensis* ssp. *breviflorus* – Egypt, Israel and Saudi Arabia. Achenes were germinated on filter paper in a

growth room at 20°C and transplanted into a compost:gravel (50:50) mix in pots at the University of St Andrews greenhouse in 2003. In total, 15 *S. glaucus*, 12 *S. flavus*, eight *S. mohavensis* ssp. *mohavensis*, and 17 *S. mohavensis* ssp. *breviflorus* plants (Supplementary Table 1) were grown in a fully randomized block until flowering. Supplementary lighting provided a photoperiod of 16 h and water was supplied as necessary. Morphological characters recorded on each plant at the same flowering stage were plant height (mm), capitulum height and width (µm), number of calyculus bracts, number of ray florets (if present), ray floret length and width (µm, mean of three, if present), middle leaf length and perimeter (mm), and leaf area (mm^2). Two ratios were calculated from the leaf measurements: leaf dissection (perimeter/√area), and standardised leaf perimeter (perimeter/length).

Traits were compared between the four taxa using 1-way ANOVA in Minitab (ver. 17.1.0) with *post hoc* Tukey tests to determine significant differences. Principal components analysis (PCA; also in Minitab) to display the multi-dimensional variation in 2-dimensional space was conducted after excluding ray floret length and width from the data set due to absence of ray florets in non-radiate plants.

Plant material and preparation for RNA-seq

Achenes of *S. glaucus*, *S. flavus* and both subspecies of *S. mohavensis* (Table 1) were placed on damp filter paper in Petri dishes, incubated at 4°C for one week and transferred to a growth chamber (23°C, 16 h photoperiod) for germination. Upon germination, seedlings were transferred to a 1:1 mixture of Levington's M2+S compost (Scotts Miracle Gro, Surrey, UK) and Vermiculite (Sinclair Horticulture Ltd., Lincoln, UK) in 10 cm diameter pots in a greenhouse at the University of Southampton. Watering was carried out twice a day by flooding the benches for 15 minutes and lighting was supplemented to achieve a 16 h daylength.

After 3 weeks of growth, RNA extraction from 3-4 individuals of each taxon was carried out, but plants continued to grow until flowering to ensure a typical morphology. One organ type was analysed, a single fully expanded leaf, which was removed from each plant, placed individually in 1.5ml microcentrifuge tubes and immediately frozen in liquid nitrogen. All samples were collected at the same time on the same day. RNA extraction proceeded using a Qiagen RNeasy Plant kit (Qiagen, Crawley, UK) with on-column DNase digestion. RNA quantification was carried out with a NanoDrop 1000 (NanoDrop Products, Wilmington, DE, USA) and the individual samples from each taxon were combined in equimolar amounts to produce one bulk RNA sample for each taxon.

Each bulk was then prepared for sequencing using the KAPA Stranded mRNA-Seq Kit (Kapa Biosystems, London, UK). Three µg of RNA was converted to RNA-seq libraries following the manufacturer's protocol. The adapter was made by annealing two partially complementary oligonucleotides /P/GATCGGAAGAGCACACGTCTGAACTCCAGTCAC and ACACTCTTTCCCTACACGACGCTCTTCCGATC*T, where /P/ designates 5' phosphorylation and * is a phosphorothioate bond. Equal amounts of each oligonucleotide were combined and heated to 95°C followed by cooling slowly to 20°C. For library amplification (7 cycles) the primers CAAGCAGAAGACGGCATACGAGAT[X]GTGACTGGAGTTCA*G (where [X] represents an 8-bp sequence 'barcode' which differed between the four samples) and AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA*C were used. Following library quantification using a Bioanalyzer (Agilent Technologies, Santa Clara, CA) and Real-Time PCR quantification (KAPA Library Quantification kit) the four samples were combined in equimolar ratios and run for 300 cycles (paired-end) on a single lane of MiSeq (Illumina, UK) at the National Oceanography Centre, University of Southampton. Following the run, reads with ambiguous barcodes were removed and the remaining reads were de-convoluted by barcode.

RNA-seq de novo *assembly and annotation*

RNA-seq reads have been submitted to the NCBI SRA (http://www.ncbi.nlm.nih.gov/sra) under project number PRJNA322115. Prior to assembly, reads for each sample were trimmed of adapters and poor quality bases using Trimmomatic v.0.32 (Bolger, Lohse and Usadel, 2014) with settings LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:36.

To compare the expression of the parental (*S. flavus* [F] and *S. glaucus* [G]) orthologues in *S. mohavensis* (ssp. *mohavensis* [M] and ssp. *breviflorus* [B]) we first assembled the parental transcriptomes *de novo*. For this, reads from F and G were individually normalised and assembled using the Trinity (v 2.0.6; Grabherr*, et al.*, 2011) pipeline. Assembly utilised the orientation of the PE reads. Trinity allows for multiple transcripts to be assembled per gene, but because these were assembled *de novo*, in reality two such genes might represent different portions of the same gene. The assembled transcriptomes have been submitted to Data Dryad (http://dx.doi.org/10.5061/dryad.sj5t2).

Identification of orthologous loci between the F and G transcriptomes was done by identifying reciprocal best blast hits. This was carried out in Bioedit (Hall, 1999) with a minimum e-value cut-off of e^{-30}.

Mapping and Expression Analysis

We then tested to see how accurate the mapping of F and G reads to the 'correct' and 'incorrect' species was. The F and G *de novo* transcriptome sequences were combined into one fasta file to give a 'pseudo-tetraploid' and this was used as the reference for mapping (separately) the F and G reads. Mapping was carried out using Trinity which utilises bowtie (Langmead*, et al.*, 2009) and RSEM (Li and Dewey, 2011) for read mapping and abundance estimation, respectively. The F and G outputs were then compared to ascertain the proportion of F and G reads that mapped to the F and G sub-transcriptome. Given that only 1.8% of F and 5.6% of G reads mapped to the 'incorrect' sub-transcriptome (see results) we reasoned that mapping of reads from the F and G orthologues of ssp. *mohavensis* (M) and ssp. *breviflorus* (B) to the alternate sub-transcriptome would have minimal effect on the overall results (see also below).

We also mapped the F and G reads against a pseudo-tetraploid transcriptome which only contained the loci with reciprocal best blast hits ('reduced transcriptome'). Again, the proportion of reads mapping to the 'incorrect' sub-transcriptome was low (ca. 5%) plus the number of reads mapping to each transcript in the reduced transcriptome was strongly correlated with the number of reads that mapped to the same contig in the full transcriptome (see results). We present overall results for both analyses; however, for brevity, we present only the results of the reduced transcriptome for the expression analysis.

We then mapped the M and B reads, separately, to the reduced transcriptome using Trinity and the same settings as for the F and G reads. Again we used the output to ascertain the proportion of reads that mapped to each contig in each sub-transcriptome. Read counts (for all four samples) were converted to TPM (transcripts per million) to allow for differences in the number of reads in each library.

Having identified homoeologous F and G loci through reciprocal best blast (above), the proportion of M and B reads mapping to the F and G copies could be determined for these loci. As we were working with bulked samples we were not able to analyse replicate samples per taxon, however edgeR can compare single samples per taxon for differential expression (Robinson, McCarthy and Smyth, 2010). Although this is likely to result in some false-positives and false-negatives (Auer and Doerge, 2010) we have no reason to think this would be biased in favour of any subset of the loci, and hence our analysis of overall bias in homoeologue expression is not because of a lack of replication. Given that 5% of reads might be mapping to the 'wrong' sub-transcriptome, we also present results after having removed loci where one or more parental (F or G) read mapped to the other sub-transcriptome.

Using Bioedit the G orthologue of each pair was used to blast against the TAIR *Arabidopsis* cDNA library (https://www.arabidopsis.org/) with a cut-off of e^{-20}, and the top hit was used in the analysis of Gene Ontology. For this, the *Arabidopsis* blast hit for the subset of differentially expressed homoeologues was compared to the *Arabidopsis* blast hit for the remainder of the *Senecio* transcripts in agriGO (Du*, et al.*, 2010) to determine if any GO categories were significantly (Hypergeometric test with FDR 0.05) over-represented.

PCR assay for homoeologue deletion

In several instances, apparent silencing of the *S. glaucus* homoeologue in both M and B was revealed (see results), however an alternative explanation is that this parental copy of the locus has been deleted from the *S. mohavensis* genome. For 16 loci which were apparently expressed in both parents (F and G), but where expression was only detected from the F sub-transcriptome in both subspecies of *S. mohavensis* (M and B) we designed polymerase chain reaction (PCR) primers which would amplify only the G copy (Supplementary Table 2). To do this we aligned homoeologous parental loci and identified ~150-400 bp regions of the transcript where species-specific primers could be designed.

DNA was extracted using a CTAB-based protocol and PCR amplification followed standard conditions (see Chapman and Burke, 2012 for details). Amplification of these primer pairs was first tested on DNA from four (*S. flavus*) and six (*S. glaucus*) individuals (Supplementary Table 3). After resolving PCR products on agarose gels we focussed on 10 loci which successfully amplified in just *S. glaucus*, hence determining the PCR primers would be homoeologue-specific in M and B (maximum one individual failed to amplify). These homoeologue-specific primer pairs were then PCR amplified from four individuals of each subspecies of *S. mohavensis*.

Results

Morphological analysis

Of the 12 traits examined, both subspecies of *S. mohavensis* were significantly different from *S. glaucus* for nine, and significantly different from *S. flavus* for two traits (leaf length and leaf area) (Figure 1). For another three traits *S. mohavensis* could not be compared to *S. flavus* because ray florets were absent in *S. flavus*. The difference between *S. mohavensis* and *S. glaucus* is clearly

evident in the PCA plot, whereas some individuals of *S. mohavensis* overlap with *S. flavus,* emphasising the convergent phenotypes of these two species (Figure 2).

RNA-seq analysis - Sequencing and assembly

Each library comprised 4.5 – 7.3 M 2 x 300b PE reads (Table 2). After trimming, approximately 2% of reads were removed either because they were too short or contained poor quality bases (Table 2). Normalisation of the F and G libraries prior to assembly reduced the number of reads to be assembled to 1.69M and 1.46M, respectively (the normalised libraries were only used for the assembly and not for the mapping). The *de novo* assembly of the F and G samples resulted in transcriptomes comprising 74,723 and 87,363 transcripts from 65,174 and 75,276 genes, respectively (Table 3). Contig N50 and median length were 1112 and 434 bp for F and 746 and 368 bp for G; the F transcriptome comprised 54.0 MB of sequence and the G transcriptome 50.2 MB (Table 3).

Of the F transcripts, 41,211 (55.2%) had a blast hit in the G transcriptome, and of the G transcripts, 54,814 (62.7%) had a hit in the F transcriptome. Of these, 19,151 were reciprocal best blast hits and were used for comparison of orthologue expression.

Mapping and Expression Analysis

Mapping of F and G to the pseudo-tetraploid transcriptome resulted in 98.2% of F reads mapping to the F transcripts and 94.4% of G reads mapping to the G transcripts. When just the transcripts with reciprocal best blast hits were used as the reference ('reduced transcriptome'; 19,151 transcripts from each species) the results were very similar (95.5% and 94.7%, respectively). We reason therefore that in the analysis of the M (ssp. *mohavensis*) and B (ssp. *breviflorus*) transcriptomes of *S. mohavensis*, approximately 95% of reads will map to the 'correct' sub-transcriptome. We present just the results using the reduced transcriptome as the reference, because the number of reads that mapped to each locus in the full transcriptome was so similar (Spearman's rank correlations for expression in the full transcriptome vs. expression in the reduced transcriptome were 0.98 for F and 0.97 for G).

Following mapping of the F and G reads to the reduced transcriptome (above), differentially expressed (DE) loci were identified using edgeR. Expression (TPM) of the F reads mapping to the F portion of the reduced transcriptome was compared to the expression of the G reads mapping to the G portion. This revealed that 14.2 % of loci were differentially expressed (2,713/19,139) at 5%

FDR, with around half of DE loci showing greater expression in F (1,298) and half in G (1,415) (Figure 3).

We then mapped the M and B reads to the reduced transcriptome and carried out differential expression analysis for each locus. The expression of M reads mapping to the F portion of the reduced transcriptome was compared to M reads mapping to the G portion. The same was then carried out for the B reads. For both M and B, ca. 65% of the mapped reads mapped to the F homoeologues and 35% to the G homoeologues (both summed TPM and summed FPKM were also skewed ca. 65%:35%). After excluding loci with no expression from either homoeologue, the differential expression analysis revealed 25.0% (4,566/18,272) and 20.2% (3,600/17,857) of loci to exhibit biased homoeologue expression in M and B, respectively, at 5% FDR (Figure 3). For both subspecies the majority of DE loci were biased in favour of expression of the *S. flavus* homoeologue. In M and B, 68.4% and 69.4% of DE loci, respectively, showed greater expression of the F homoeologue than the G homoeologue. A large number of these loci exhibited no expression from the *S. glaucus* copy (see 'G silent' in Figure 3).

To analyse overlap in the DE loci between the different comparisons we first reduced the list of loci to compare with only those expressed in one or both diploids and expressed in both M and B. This resulted in 17,598 loci for comparison. There was considerable overlap between the loci differentially expressed in all three comparisons (i.e. between F and G, and between orthologues within M and within B; Figure 4). Of 6,365 loci showing differential expression in at least one comparison, 3,046 (47.9%) were differentially expressed in at least two of the three comparisons and 1,090 (17.1%) were differentially expressed in all comparisons. Loci that were differentially expressed in the F vs. G comparison were significantly more likely to be differentially expressed within M and/or B than by chance (2 x 2 contingency table, χ^2 test $P < 10^{-5}$). Of these DE loci the majority were DE in the same direction in the F-G comparison as they were in the comparisons of orthologue expression within each subspecies of *S. mohavensis*. Of the 1090 loci DE in all three comparisons, 1050 (96.3%) were consistently DE expressed and only 31 (2.8%) were DE in a different direction between F-G and both subspecies of *S. mohavensis*. The remaining 9 (0.8%) loci were DE in different directions between subspecies of *S. mohavensis*. Of the 687 loci DE between F and G and DE in one, but not both, subspecies of *S. mohavensis*, 640 (93.2%) were DE in the same direction.

As mentioned, around 5% of reads from F and G mapped back to the opposite sub-transcriptome, hence for these loci we might be incorrectly estimating the contribution of the two homoeologues in M and B. We therefore repeated the above after removing any loci that had 1 or more reads from one diploid (F or G) mapping to the transcriptome of the other. This reduced the number of loci to

compare to 14,204, but had a marginal effect on the overall results, with a similar proportion of loci differentially expressed (at 5% FDR) between F and G (14.2% in the first analysis, 14.0% in the second analysis) as well as within M (25.0% in the first and 24.1% in the second) and B (20.2% in the first and 18.4% in the second). For the F vs. G comparison, again approximately equal proportions of loci were expressed at a higher level in one species than the other (48.2% G > F, 51.8% F > G), whereas within M and B more loci were biased towards the F orthologue than the G orthologue (70.5 % and 72.8 % F > G in M and B, respectively).

The list of differentially expressed loci was tested for over-representation of GO terms using the top *Arabidopsis* blast hit in agriGO. Of the loci differentially expressed between F and G, 382 had an *Arabidopsis* hit (e^{-20}; n = 330 after excluding duplicate hits), and amongst these loci the GO terms 'response to far red light', 'tetrapyrrole binding ', and 'structural constituent of cytoskeleton' were significantly over-represented (Table 4A). The equivalent analysis of loci DE in both subspecies of *S. mohavensis* resulted in 406 loci with an *Arabidopsis* hit (353 after excluding duplicate hits), and over-representation of two terms identified in the first comparison ('tetrapyrrole binding ', and 'structural constituent of cytoskeleton') as well as more general terms related to chlorophyll and the chloroplast (Table 4B). Finally, in the analysis of homoeologues differentially expressed in one or both subspecies of *S. mohavensis*, 957 had a significant hit to *Arabidopsis* (789 after excluding duplicate hits) and two GO terms relating to the cytoskeleton were significantly over-represented (Table 4C).

PCR assay for homoeologue deletion

From the list of loci which exhibited no expression from the *S. glaucus* homoeologue in both *S. mohavensis* subspecies, but were expressed in both parental taxa, *S. glaucus*-specific primers were designed from a comparison of the *S. glaucus* and *S. flavus* orthologues. Of these 16, ten proved species-specific based on PCR amplification and hence could be used to determine if the *S. glaucus* copy was indeed present in the genome of *S. mohavensis*.

When PCR was carried out using these ten primer pairs, three showed no amplification (hence are potentially deleted from *S. mohavensis*) and four amplified in all eight *S. mohavensis* individuals tested (hence are present in the genome of *S. mohavensis*). For the remaining loci the interpretation is less clear, and the amplification in seven of eight individuals (two loci) or three of eight (one locus) could be explained by primer site divergence in some individuals, or that the deletion is segregating in the species.

Discussion

The pattern of gene expression evolution following whole genome duplication is of major interest to an understanding of how polyploids evolve phenotypically following their origin. Moreover, because most plants have a history of polyploidy in their ancestry (Masterson 1994; Wood et al. 2009), changes in gene expression following polyploidy are of relevance to how the majority of plant taxa evolve. In allopolyploids, it has generally been found that the expression of parental homoeologues is highly variable – some loci are expressed as in the parents, others show transgressive expression, while others are silenced (Adams*, et al.*, 2003; Buggs*, et al.*, 2014; Hegarty*, et al.*, 2008; Roulin*, et al.*, 2013; Wang*, et al.*, 2012).

For the allotetraploid *Senecio mohavensis* we were primarily interested in determining if any biased homoeologue expression was in favour of the *S. flavus* copy at the expense of the *S. glaucus* copy. *Senecio flavus* and *S. glaucus* differ in a number of morphological characters, and for most traits *S. mohavensis* is more similar to *S. flavus* than to *S. glaucus* as confirmed by our morphometric analysis (Figures 1 and 2). This morphological similarity is so pronounced that it was assumed that *S. mohavensis* and *S. flavus* were conspecific (Kadereit, 1984), until chromosome counts distinguished their different ploidies (Coleman*, et al.*, 2001), and later molecular work showed that *S. mohavensis* possessed cpDNA and ITS sequences strongly divergent from *S. flavus* and more similar to Mediterranean diploid *Senecio* species (Coleman*, et al.*, 2003; Comes and Abbott, 2001), which includes *S. glaucus*, the presumed second parent (Kadereit*, et al.*, 2006).

In terms of overall gene expression within the two subspecies of *S. mohavensis* (M and B) we did indeed find a bias in favour of the *S. flavus* sub-transcriptome. This bias was quite strong with approximately a 2:1 ratio of expression from the *S. flavus* and *S. glaucus* homoeologues. On a locus-by-locus basis, significant differences in expression between the diploid parents *S. flavus* and *S. glaucus* occurred for about 14% of loci, whereas within the two subspecies of tetraploid *S. mohavensis* differential expression between the gene copies of *S. flavus* and *S. glaucus* origin occurred at 20-25% of loci (Figure 3). When the direction of biased expression was examined more closely it was evident that the bias between orthologues in *S. flavus* and *S. glaucus* was approximately 50:50, i.e., approximately half of orthologues showed greater expression in *S. flavus* with the other half showing greater expression in *S. glaucus*. However, in both subspecies of *S. mohavensis*, the homoeologues showing greatest expression were more often those inherited from the *S. flavus* parent. Thus, about two thirds of differentially expressed loci exhibited greater expression of the *S. flavus* homoeologue and in many of these instances the *S. glaucus* homoeologue was not expressed (Figure 3). As mentioned, *S. mohavensis* was formed from a cross between

maternal *S. glaucus* and paternal *S. flavus* (presuming that cpDNA in *Senecio* is maternally transmitted, as is the case for most angiosperms) therefore the nuclear genome exhibits biased expression from the paternal sub-genome. This biased parental expression has been reported in other studies (e.g. Buggs *et al*., 2010; Flagel *et al*., 2008), however there is no general pattern for whether maternal or paternal homoeologues are expressed at the greater level.

Although very few GO terms were revealed to be over-represented in the differentially expressed loci (Table 4), the presence of related GO terms referring to the cytoskeleton in all comparisons could suggest selection on cell structure. Considering the GO terms identified in the comparison of *S. flavus* and *S. glaucus*, it is interesting that the leaves of *S. flavus* are conspicuously purple when compared to other *Senecio* species, including *S. glaucus*, and this could indicate that divergence in the light/shade response has occurred between these species.

Our gene expression findings therefore suggest that (1) overall there are more genes showing differential homoeologue expression within the allotetraploid subspecies than between the equivalent orthologues in the parental species *S. flavus* and *S. glaucus*, and (2) *S. flavus* homoeologues are expressed at a greater level, on average, than *S. glaucus* homoeologues in the allotetraploid, *S. mohavensis*. In polyploids, however, there is considerable evidence that genomes become restructured over time, with the loss of some parental DNA. When comparing genome size of polyploids and diploid parents there is a trend for the polypoid genome to be smaller in size than the sum of the parents (Leitch and Bennett, 2004); although this is not a universal finding (e.g. Verma and Rees, 1974). In addition, studies show that parental DNA segments are often eliminated from polyploid offspring, sometimes within just one or a few generations (Khasdan*, et al.*, 2010; Lukens*, et al.*, 2006; Skalická*, et al.*, 2005; Vallejo-Marin *et al*., 2015). If this loss was biased in favour of loss from the *S. glaucus* sub-genome in *S. mohavensis* (and this loss was from coding DNA or from non-coding DNA which affected gene expression) then we might find reduced expression of *S. glaucus* homoeologues in *S. mohavensis*, simply because they are not present in the genome of the allotetraploid.

We investigated this by identifying loci from *S. glaucus* that were apparently not expressed from the genomes of the two subspecies of *S. mohavensis*. Primers were designed to specifically amplify the *S. glaucus* orthologue, and in 40% of cases (4/10) the locus was confirmed to be present in all individuals of *S. mohavensis*. For a further 30% (3/10) we found no amplification across eight individuals of *S. mohavensis* which likely suggests the locus has been deleted (or mutated sufficiently that the primers fail to anneal) (Supplementary Table 2). It should be noted that with such small sample size this estimate of homoeologue loss has very wide confidence intervals (1.6 – 58.4%). The

other three loci showed amplification in only a subset of the *S. mohavensis* individuals which could result from polymorphism for presence/absence of the deletion (or of allelic variation which precluded PCR amplification). It should be noted that the species-specificity of the primers was tested on *S. glaucus* individuals from a variety of geographic sources (Morocco, Tunisia, Israel), reducing the chances that allelic differences in primer annealing sites between *S. glaucus* and the *S. mohavensis* homoeologue gave rise to non-successful amplification.

One caveat of our RNA-seq data is that we relied on bulk transcriptome RNA-seq data from the four taxa under investigation; we are therefore unable to rule out the possibility that variation in gene expression and homoeologue silencing exists among individuals of the parent species. This variation could contribute to the expression differences found here between the allopolyploid homoeologues simply if a silenced copy had been inherited during the origin of *S. mohavensis* (see Buggs*, et al.*, 2014; Gottlieb, 1982). Furthermore, in any study of an ancient polyploid, the degree to which gene expression divergence between the parental taxa we sampled, and the actual parental taxa that were involved in the initial establishment of the polyploid, is unknown (Buggs*, et al.*, 2014). Whilst our bulks contained RNA from only 3 or 4 individuals, this small sample size is unlikely to strongly influence the overall pattern of gene expression bias we uncovered, even though accurate identification of specific genes that do and do not show significant differences in expression between taxa might be impacted (Auer and Doerge, 2010).

Another caveat concerns the fact that only one organ type at one particular stage of development was analysed. However, we focussed on seedling leaf tissue because at this growth stage plants of each taxon are phenotypically more similar to each other than at later stages. This reduced the likelihood that gross differences in phenotype result in differences in the cell types used in the RNA extraction (see also Koenig*, et al.*, 2013; Swanson-Wagner*, et al.*, 2012). Other studies, in soybean and cotton (Chaudhary*, et al.*, 2009; Roulin*, et al.*, 2013), have shown differences among tissue types in terms of gene expression divergence, although there was often a significant overlap (ca. 50% in soybean) in the genes identified as differentially expressed in two different tissues.

The observation of significant overlap between loci differentially expressed between the parental species *S. flavus* and *S. glaucus* and those differentially expressed (as homoeologues) within *S. mohavensis* (significantly more than expected by chance alone) suggests that a large portion of the loci have maintained parental-like expression (whether differentially expressed or not) when combined in a polyploid cell. This would require that *cis*-acting factors are, on average, dominant over any *trans* factors in terms of affecting expression in the allotetraploid. This predominance of *cis*-acting regulation was also observed in cotton and *Tragopogon* tetraploids (Buggs*, et al.*, 2014;

Chaudhary*, et al.*, 2009). In other instances, however, a mixture of both *cis*- and *trans*-acting factors contribute to the overall expression patterns of hybrids (e.g. Yoo, Szadkowski and Wendel, 2009). It is worth noting here that recently Buggs *et al.*, (2014) put forward the argument that *cis*-acting factors should predominate in a polyploid where the parental species are diverged to such a degree that transcription factors (TFs) from one sub-genome are incompatible with the TF binding sites on the other. This model could well explain the situation in *S. mohavensis* where the parental genomes shared a common ancestor ca. 10 MY ago (Coleman*, et al.*, 2003), and have diverged to such a degree that artificial crosses between the parents fail.

The cause of the differential homoeologue expression we uncovered within *S. mohavensis* appears to be a combination of loss of parental DNA and homoeologue-specific silencing. Overall, a larger proportion of homoeologues were differentially expressed in *S. mohavensis* than between the same loci in the parental species, and in particular we noted a large number of loci inherited from the *S. glaucus* parent that were not sequenced in the RNA-seq data (yet expressed in both parents; Figure 3). Using a PCR assay for 10 such silenced homoeologues, we determined that the parental copy was present in the genome of all or most *S. mohavensis* individuals for six, whereas for three we were unable to amplify the *S. glaucus* copy, potentially suggesting this homoeologue is absent. For the final locus some individuals appeared to have lost the homoeologue, whereas others had retained it. Thus, the evolution of *S. mohavensis* has involved divergence in expression as well as loss of parental DNA, in a biased manner, and some of this DNA loss is apparently segregating.

Conclusions

Our finding, that the genome of the allotetraploid *S. mohavensis* appears to show deletion and loss of expression of some of the parental DNA inherited from *S. glaucus*, provides preliminary evidence that since the initial allopolyploid event the genome has become restructured, and more "*flavus*-like". It is worth noting that an analysis of random nuclear markers (Random Amplified Polymorphic DNAs; RAPDs) in a sample of Mediterranean *Senecio*s, including one individual of each of *S. flavus* and *S. mohavensis* subsp. *breviflorus*, noted that "…the nuclear genome of subspecies *breviflorus* is predominantly ''*flavus*-like''" (Comes and Abbott, 2001), p. 1955). Both the biased loss of parental DNA and specific silencing of *S. glaucus* homoeologues in *S. mohavensis* correlate with the allotetraploid being morphologically much more similar to *S. flavus* than to its other parent *S. glaucus.* Whether overall genomic and transcriptomic divergence is the cause of this morphological

convergence, or whether a few loci correspond to this convergence and the transcriptomic divergence is a (partly *trans*-acting) result of this remains unknown.

Acknowledgements

We would like to thank the Genetics Society (UK) for a Genes and Development summer studentship for DAW to carry out work in the lab of MAC in 2015, Christopher Soper for helping with the morphological analysis in the lab of RJA, and Alex Harkess (University of Georgia) for advice on the library prep for RNA-seq. We also thank the reviewers for their helpful comments which improved the initial version of our manuscript.

Tables:

Table 1 – Collection information for the plants used in the RNA-seq analysis

Species	Origin	Accession
S. glaucus	Houmt Souk, Tunisia	HS9
S. glaucus	Sbeitla, Tunisia	SB3
S. glaucus	Sbeitla, Tunisia	SB32
S. flavus	Tata, Morocco	14454
S. flavus	Asni, Morocco	751
S. flavus	Ait-Baha, Morocco	14388
S. flavus	Canary Islands, Spain	26145
S. mohavensis ssp. *mohavensis*	Arizona, USA	10190
S. mohavensis ssp. *mohavensis*	Arizona, USA	44790
S. mohavensis ssp. *mohavensis*	California, USA	645-3
S. mohavensis ssp. *mohavensis*	Nevada, USA	5375
S. mohavensis ssp. *breviflorus*	Uema Figra, Saudi Arabia	7077
S. mohavensis ssp. *breviflorus*	Arava Valley, Israel	7HA3
S. mohavensis ssp. *breviflorus*	Nahal Paran, Israel	1219

Table 2 – Read statistics (taxa named by first letter as in text)

	F	G	M	B
Number of raw reads	7,050,319	4,513,220	7,360,535	5,290,211
Number of trimmed reads	6,928,009	4,417,812	7,175,799	5,194,478
% trimmed reads	98.27	97.89	97.49	98.19
Number of normalised reads	1,689,071	1,460,103	NA[1]	NA
% Normalised reads	32.52	21.08	NA	NA

[1]NA, not applicable

Table 3 – Assembly statistics (taxa named by first letter as in text)

	F	G
Total trinity 'genes'	65,174	75,276
Total trinity transcripts	74,723	87,363
Percent GC	39.71	40.27
Contig N50	1112	746
Median contig length	434	368
Mean contig length	723	576
Total assembled bases	54,037,457	50,293,940
Blast hits (%) to other species	41,211 (55.2%)	54,814 (62.7%)
Reciprocal best blast hits (%)	19,151 (25.6%)	19,151 (21.9%)

Table 4 – GO terms significantly over-represented (FDR < 0.05) in the differential expression comparison of (A) *S. flavus* and *S. glaucus* and (B) the parental homoeologues within *S. mohavensis*.

	Description	Number in input list	Number in reference	p-value	FDR
A	response to far red light	7	7	2.20E-07	0.00013
	response to red light	7	7	2.20E-07	0.00013
	structural constituent of cytoskeleton	8	14	3.90E-05	0.019
	tetrapyrrole binding	8	16	0.00014	0.034
B	tetrapyrrole binding	11	19	2.3e-06	0.0011
	structural constituent of cytoskeleton	10	18	1.1e-05	0.0025
	structural molecule activity	38	156	2.2e-05	0.0031
	chlorophyll binding	9	16	2.7e-05	0.0031
	cytoskeleton	20	47	1.8e-07	6.5e-05
	non-membrane-bounded organelle	53	230	4.7e-06	0.00057
	cytoskeletal part	15	40	3.5e-05	0.0032
	plastoglobule	11	27	0.00016	0.012
C	structural constituent of cytoskeleton	14	14	2.50E-08	2.10E-05
	cytoskeleton	25	38	2.40E-06	0.0013

Figure Legends:

Figure 1 - Morphometric comparison of the four taxa under investigation. Values of 0 indicate floral traits that could not be measured on the non-radiate plants. Bars with different letters above indicate that means are significantly different (*t*-test, $P < 0.05$ with Tukey test).

Figure 2 - Principal components analysis (PCA) of the morphometric data. Taxa are indicated with different symbols (diamond, *S. glaucus*; open square, *S. flavus*; triangle *S. mohavensis* ssp. *mohavensis*; cross, *S. mohavensis* ssp. *breviflorus*).

Figure 3 - Results of the differential expression analysis of the RNA-seq data. Segments of each pie chart denote the different classes of loci. Left, differential expression between *S. flavus* and *S. glaucus*; centre, differential expression between the *S. flavus* and *S. glaucus* homoeologues within *S. mohavensis* ssp. *mohavensis*; right, differential expression between the *S. flavus* and *S. glaucus* homoeologues within *S. mohavensis* ssp. *breviflorus*. 'F silent' and 'G silent' indicate no expression was detected from the *S. flavus* or *S. glaucus* copy, respectively. 'G > F' and 'F > G' indicate significantly biased expression (but not silencing of one copy). N.S. indicates no significant difference.

Figure 4 - Overlap of loci that were differentially expressed in the three different comparisons.

Literature Cited:

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, U.S.A. 100:4649-4654.

ADAMS KL, WENDEL JF. 2005a Novel patterns of gene expression in polyploid plants. Trends in Genetics. 21(10):539-543.

ADAMS KL, WENDEL JF. 2005b Polyploidy and genome evolution in plants. Current Opinion in Plant Biology. 8(2):135-141.

AUER P, DOERGE R. 2010 Statistical design and analysis of RNA sequencing data. Genetics. 185:405-416.

BIRCHLER J, YAO H, CHUDALAYANDI S, VAIMAN D, VEITIA R. 2010 Heterosis. The Plant Cell. 22:2105-2112.

BOLGER A, LOHSE M, USADEL B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114-2120.

BUGGS R, WENDEL J, DOYLE J, SOLTIS D, SOLTIS P, COATE J. 2014 The legacy of diploid progenitors in allopolyploid gene expression patterns. Philosophical Transactions of the Royal Society B-Biological Sciences. 369:20130354.

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. Current Biology. 22(3):248-252.

BUGGS RJA, ELLIOTT NM, ZHANG L, KOH J, VICCINI LF, SOLTIS DE, SOLTIS PS. 2010 Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. New Phytologist 186:175–183

CHAPMAN MA. 2004. The taxonomy of *Senecio* sect. *Senecio*: hybridisation and speciation. PhD thesis, St Andrews, UK: University of St Andrews.

CHAPMAN MA, BURKE JM. 2012 Evidence of selection on fatty acid biosynthetic genes during the evolution of cultivated sunflower. Theoretical and Applied Genetics. 125(5):897-907.

CHAUDHARY B, FLAGEL L, STUPAR R, UDALL J, VERMA N, SPRINGER NM, WENDEL JF. 2009 Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). Genetics. 182:503-517.

CHEN Z. 2007 Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annual Review of Plant Biology. 58:377-406.

COLEMAN M, FORBES DG, ABBOTT RJ. 2001 A new subspecies of *Senecio mohavensis* (Compositae) reveals Old-New World species disjunction. Edinburgh Journal of Botany. 58:389-403.

COLEMAN M, LISTON A, KADEREIT JW, ABBOTT RJ. 2003 Repeat intercontinental dispersal and Pleistocene speciation in disjunct Mediterranean and desert *Senecio* (Asteraceae). American Journal of Botany. 90(10):1446-1454.

COMES HP, ABBOTT RJ. 2001 Molecular phylogeography, reticulation and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). Evolution. 55(10):1943-1962.

DU Z, ZHOU X, LING Y, ZHANG Z, SU Z. 2010 agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Research. 38:W64–W70.

FLAGEL LE, UDALL JA, NETTLETON D, WENDEL JF. 2008 Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct dhases of expression evolution. BMC Biology 6: 16.

GAETA R, PIRES J, INIGUEZ-LUY F, LEON E, OSBORN T. 2007 Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. The Plant Cell. 19:3403-3417.

GOTTLIEB LD. 1982 Conservation and duplication of isozymes in plants. Science. 216:373-380.

GRABHERR MG, HAAS BJ, YASSOUR M, LEVIN JZ, THOMPSON DA, AMIT I, ADICONIS X, FAN L, RAYCHOWDHURY R, ZENG QD, CHEN ZH, MAUCELI E, HACOHEN N, GNIRKE A, RHIND N, DI PALMA F, BIRREN BW, NUSBAUM C, LINDBLAD-TOH K, FRIEDMAN N, REGEV A. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 29(7):644-652.

GU Z, STEINMETZ L, GU X, SCHARFE C, DAVIS R, LI W-H. 2003 Role of duplicate genes in genetic robustness against null mutations. Nature. 421:63-66.

HALL TA. 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41:95-98.

HEGARTY MJ, BARKER GL, BRENNAN AC, EDWARDS KJ, ABBOTT RJ, HISCOCK SJ. 2008 Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in *Senecio*. Philosophical Transactions of the Royal Society B-Biological Sciences. 363:3055-3069.

HEGARTY MJ, BARKER GL, WILSON ID, ABBOTT RJ, EDWARDS KJ, HISCOCK SJ. 2006 Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Current Biology. 16(16):1652-1659.

HUFTON A, PANOPOULOU G. 2009 Polyploidy and genome restructuring: a variety of outcomes. Current Opinion in Genetics and Development. 19:600-606.

JAILLON O, AURY J-M, NOEL B, POLICRITI A, CLEPET C, CASAGRANDE A, CHOISNE N, AUBOURG S, VITULO N, JUBIN C, VEZZI A, LEGEAI F, HUGUENEY P, DASILVA C, HORNER D, MICA E, JUBLOT D, POULAIN J, BRUYÈRE C, BILLAULT A, SEGURENS B, GOUYVENOUX M, UGARTE E, CATTONARO F, ANTHOUARD V, VICO V, DEL F, C., ALAUX M, DI G, G., DUMAS V, FELICE N, PAILLARD S, JUMAN I, MOROLDO M, SCALABRIN S, CANAGUIER A, LE C, I., MALACRIDA G, DURAND E, PESOLE G, LAUCOU V, CHATELET P, MERDINOGLU D, DELLEDONNE M, PEZZOTTI M, LECHARNY A, SCARPELLI C, ARTIGUENAVE F, PÈ M, VALLE G, MORGANTE M, CABOCHE M, ADAM-BLONDON A, WEISSENBACH J, QUÉTIER F, WINCKER P, CHARACTERIZATION F-IPCFGG. 2007 The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 449(7161):463-467.

KADEREIT JW. 1984 Some notes on annual species of *Senecio* L. (Asteraceae) from northern Africa and the Canary Islands. Botanische Jahrbuch der Systematischen. 104:509-517.

KADEREIT JW, URIBE-CONVERS S, WESTBERG E, COMES HP. 2006 Reciprocal hybridization at different times between *Senecio flavus* and *Senecio glaucus* gave rise to two polyploid species in north Africa and south-west Asia. New Phytologist. 169(2):431-441.

KHASDAN V, YAAKOV B, KRAITSHTEIN Z, KASHKUSH K. 2010 Developmental timing of DNA elimination following allopolyploidization in wheat. Genetics. 185:387-390.

KOENIG D, JIMENEZ-GOMEZ JM, KIMURA S, FULOP D, CHITWOOD DH, HEADLAND LR, KUMAR R, COVINGTON MF, DEVISETTY UK, TAT AV, TOHGE T, BOLGER A, SCHNEEBERGER K, OSSOWSKI S, LANZ C, XIONG G, TAYLOR-TEEPLES M, BRADY SM, PAULY M, WEIGEL D, USADEL B, FERNIE AR, PENG J, SINHA NR, MALOOF JN. 2013 Comparative transcriptomics reveals patterns of selection in

domesticated and wild tomato. Proceedings of the National Academy of Sciences of the United States of America. 110(28):E2655-E2662.

LANGMEAD B, TRAPNELL C, POP M, SALZBERG S. 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 10:R25.

LEITCH I, BENNETT M. 2004 Genome downsizing in polyploid plants. Biology Journal of the Linnean Society. 82:651-663.

LI B, DEWEY C. 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323.

LISTON A, RIESEBERG LH, ELIAS TS. 1989 Genetic similarity is high between intercontinental disjunct species of *Senecio* (Asteraceae). American Journal of Botany. 76(3):383-388.

LUKENS L, PIRES J, LEON E, VOGELZANG R, OSLACH L, OSBORN T. 2006 Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. Plant Physiology. 140:336-348.

MADLUNG A. 2013 Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity. 110:99-104.

MOORE RC, PURUGGANAN MD. 2005 The evolutionary dynamics of plant duplicate genes. Current Opinion in Plant Biology. 8(2):122-128.

RAPP R, UDALL J, WENDEL JF. 2009 Genomic expression dominance in allopolyploids. BMC Biology. 7:18.

ROBINSON MD, MCCARTHY DJ, SMYTH GK. 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26:139-140.

ROULIN A, AUER P, LIBAULT M, SCHLUETER J, FARMER A, MAY G, STACEY G, DOERGE R, JACKSON S. 2013 The fate of duplicated genes in a polyploid plant genome. The Plant Journal. 73:143-153.

SKALICKÁ K, LIM K, MATYASEK R, MATZKE M, LEITCH A, KOVARIK A. 2005 Blackwell Publishing, Ltd. Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. New Phytologist. 166:291-303.

SOLTIS D, VISGER C, SOLTIS P. 2014 The polyploidy revolution then…and now: Stebbins revisited. American Journal of Botany. 101:1057-1078.

SOLTIS PS, SOLTIS DE. 2000 The role of genetic and genomic attributes in the success of polyploids. Proceedings of the National Academy of Sciences of the USA. 97(13):7051-7057.

SWANSON-WAGNER R, BRISKINE R, SCHAEFER R, HUFFORD MB, ROSS-IBARRA J, MYERS CL, TIFFIN P, SPRINGER NM. 2012 Reshaping of the maize transcriptome by domestication. Proceedings of the National Academy of Sciences of the United States of America. 109(29):11878-11883.

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(3):1599-1611.

VERMA S, REES H. 1974 Nuclear DNA and the evolution of allotetraploid *Brassiceae*. Heredity. 33:61-68.

WANG Y, WANG X, PATERSON AH. 2012 Genome and gene duplications and gene expression divergence: a view from plants. Annals of the New York Academy of Sciences. 1256:1-14.

YOO M, SZADKOWSKI E, WENDEL J. 2009 Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity. 110:171-180.

Figure 1

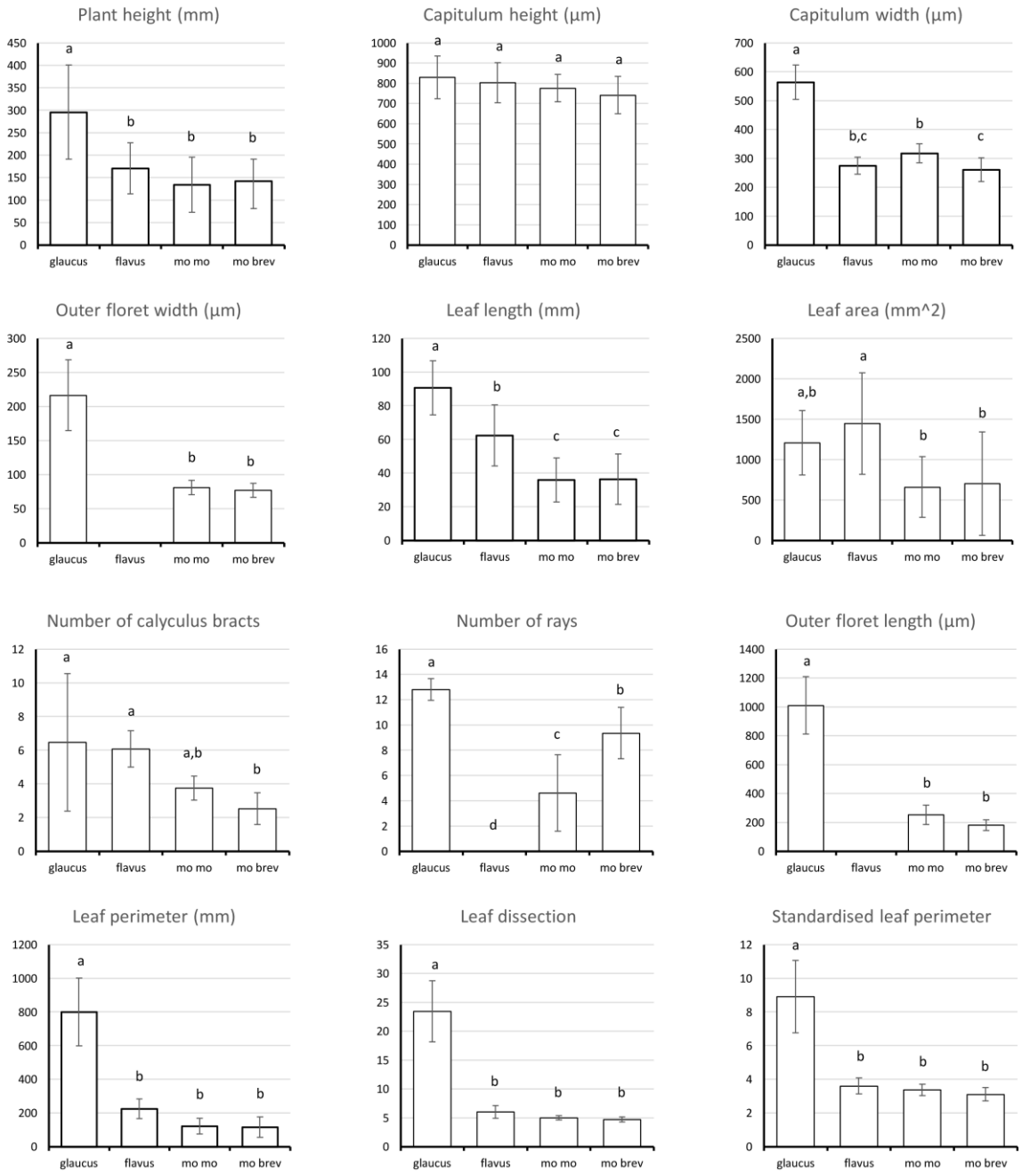

Figure 2

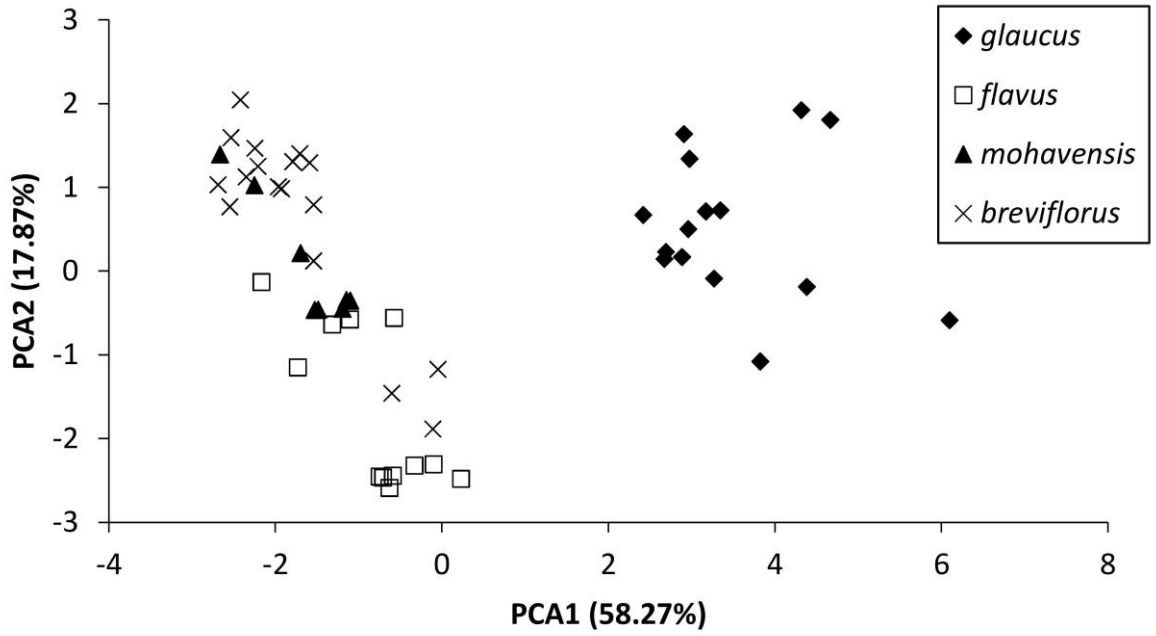

Figure 3

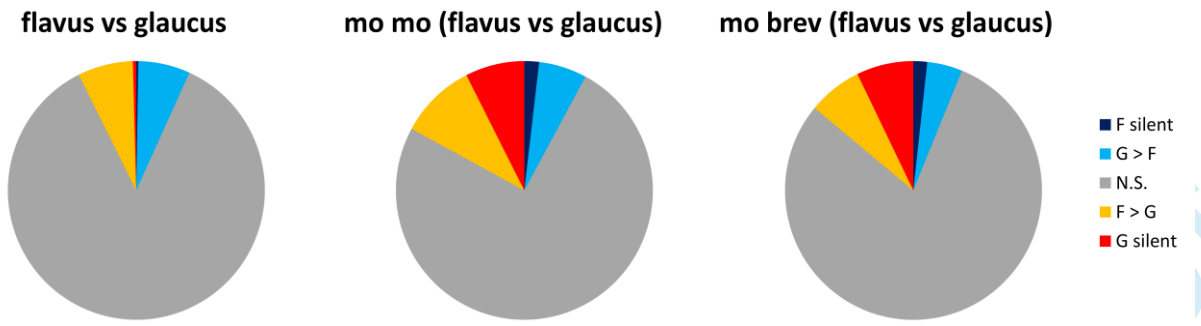

Figure 4