the plant journal

SEB Society for Experimental Biology

*The Plant Journal* (2017) **91,** 3–21 doi: 10.1111/tpj.13553

# Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention

**Terezie Mandáková[1], Zheng Li[2], Michael S. Barker[2] and Martin A. Lysak[1,*]**

[1]*Plant Cytogenomics Research Group, CEITEC–Central European Institute of Technology, Masaryk University, Brno 625 00, Czech Republic, and*

[2]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA*

Received 30 December 2016; revised 17 March 2017; accepted 23 March 2017; published online 31 March 2017.
*For correspondence (e-mail martin.lysak@ceitec.muni.cz).

## SUMMARY

**Hybridization and polyploidy followed by genome-wide diploidization had a significant impact on the diversification of land plants. The ancient At-α whole-genome duplication (WGD) preceded the diversification of crucifers (Brassicaceae). Some genera and tribes also experienced younger, mesopolyploid WGDs concealed by subsequent genome diploidization. Here we tested if multiple base chromosome numbers originated due to genome diploidization after independent mesopolyploid WGDs and how diploidization affected post-polyploid gene retention. Sixteen species representing 10 Brassicaceae tribes were analyzed by comparative chromosome painting and/or whole-transcriptome analysis of gene age distributions and phylogenetic analyses of gene duplications. Overall, we found evidence for at least 13 independent mesopolyploidies followed by different degrees of diploidization across the Brassicaceae. New mesotetraploid events were uncovered for the tribes Anastaticeae, Iberideae and Schizopetaleae, and mesohexaploid WGDs for Cochlearieae and Physarieae. In contrast, we found convergent patterns of gene retention and loss among these independent WGDs. Our combined analyses of genomic data for Brassicaceae indicate that extant chromosome number variation in many plant groups, and especially monophyletic taxa with multiple base chromosome numbers, can result from clade-specific genome duplications followed by diploidization. Our observation of parallel gene retention and loss across multiple independent WGDs provides one of the first multi-species tests of the predictability of patterns of post-polyploid genome evolution.**

**Keywords: whole-genome duplication, mesopolyploidy, post-polyploid diploidization, biased gene retention/loss, chromosome number variation, descending dysploidy, chromosomal rearrangement, Brassicaceae.**

## INTRODUCTION

Polyploidy or whole-genome duplication (WGD) is ubiquitous across the land plants (Wendel, 2015). At least 15% of flowering plant speciation events are the result of polyploidy (Wood *et al.*, 2009), and the populations of nearly 13% of diploid species probably harbor unnamed polyploid cytotypes (Barker *et al.*, 2016a). Ancient polyploidy has been identified throughout the evolutionary history of vascular plants (Jiao *et al.* 2011; Li *et al.*, 2015) and many angiosperm lineages have a polyploid ancestry (Cui *et al.*, 2006; Shi *et al.*, 2010; Arrigo and Barker, 2012; Jiao *et al.* 2014; Cannon *et al.*, 2015; McKain *et al.*, 2016; Barker *et al.*, 2016b). Many analyses suggest that polyploidy has played a significant role in the diversification of flowering plants (e.g. Soltis *et al.*, 2009, 2015; Vanneste *et al.*, 2014; Edger *et al.*, 2015; Tank *et al.*, 2015; Soltis and Soltis, 2016). Closely linked with polyploidy is the process of diploidization, changing the duplicated genome back into a diploid(-like) genome and concealing WGD events from detection (Wolfe, 2001; Barker *et al.*, 2012; Wendel, 2015; Dodsworth *et al.*, 2016; Mandáková *et al.*, 2016; Soltis *et al.*, 2016). During diploidization duplicated genes or larger chromosomal regions are lost, some paralogs gain a modified or new function (e.g., Adams and Wendel, 2005; Conant *et al.*, 2014; Douglas *et al.*, 2015; Edger *et al.*, 2015) and genome size usually decreases due to recombination-driven removal of repetitive sequences (e.g. Woodhouse

*et al.*, 2014; Soltis *et al.*, 2015) or as a consequence of chromosomal rearrangements, including descending dysploidy (e.g. Mandáková *et al.*, 2010a,b).

Genome evolution following polyploidy has many predictable aspects. Among descendants of even a single polyploidy event, some lineages may share nearly ancestral synteny whereas other taxa experience relatively more rearrangements and losses (Jaillon *et al.*, 2007; Tang *et al.*, 2008; Schnable *et al.*, 2012; de Miguel *et al.*, 2015; Murat *et al.*, 2014, 2015). One subgenome of a polyploid genome is often retained with more genes and synteny relative to the other subgenome(s) (Freeling *et al.*, 2012; Sankoff and Zheng, 2012; Schnable *et al.*, 2012; Garsmeur *et al.*, 2014; Murat *et al.*, 2014; Renny-Byfield *et al.*, 2015). In addition to biases resulting from differences in parental genomes, gene function may also influence fractionation. Some functional classes of genes are under-represented, whereas others are over-retained following WGD (e.g., Maere *et al.*, 2005; Rensing *et al.*, 2007; Barker *et al.*, 2008; Bekaert *et al.*, 2011; Geiser *et al.*, 2016; Li *et al.*, 2016). The patterns of convergent retention of functional groups of genes following WGDs are generally consistent with the dosage balance hypothesis (Birchler and Veitia, 2007, 2010, 2012; Edger and Pires, 2009; Freeling, 2009). According to this hypothesis, the drive to maintain dosage balance causes dosage-sensitive genes to be consistently over-retained following WGDs but lost following small-scale duplications. Numerous studies have found, from analyses of the descendants of single WGDs, that the biased signatures of gene retention and loss are consistent with this hypothesis (Maere *et al.*, 2005; Rensing *et al.*, 2007; Barker *et al.*, 2008; Shi *et al.*, 2010; Bekaert *et al.*, 2011; Geiser *et al.*, 2016). However, no study has yet evaluated whether convergent gene retention and loss, consistent with the dosage balance hypothesis, occurs among multiple related species that have independent WGDs.

Since the pioneering studies that revealed the paleotetraploid nature of Arabidopsis (Arabidopsis Genome Initiative 2000; Vision *et al.*, 2000), later described as the Alpha ($\alpha$) or At-$\alpha$ WGD (Bowers *et al.*, 2003; Barker *et al.*, 2009), several other WGD events post-dating At-$\alpha$ have been discovered. These WGDs can be divided broadly as mesopolyploid and neopolyploid events based upon their age (Mandáková *et al.*, 2010a; Kagale *et al.*, 2014a; Hohmann *et al.*, 2015). Neopolyploids may be identified by their elevated chromosome number, genome size and largely intact (sub)genomes, contributed by (often extant) parental species. Mesopolyploids, on the other hand, are cytologically more difficult to distinguish from 'true diploids', i.e. diploidized paleopolyploids (e.g. *Arabidopsis thaliana*), as partial diploidization blurs the primary polyploid character of these genomes (e.g. Lysak *et al.*, 2005; Mandáková *et al.*, 2010a; Wang *et al.*, 2011; Geiser *et al.*, 2016). Hybridization between diploidized mesopolyploid populations and species may lead to the origin of neoauto- or neoallopolyploid genomes (e.g. *Brassica napus*).

In mesopolyploid clades, re-diploidization is associated with genome rearrangements and descending dysploidies that sometimes generate a broad spectrum of diploid-like chromosome numbers (e.g. Lysak *et al.*, 2005, 2007; Mandáková *et al.*, 2010a; Salse, 2012). Thus, diploidization following mesopolyploid events frequently explains extant chromosome number variation, including intra-genus or intra-tribal variation in base chromosome numbers (i.e. the origin of dibasic and polybasic taxa). For instance, three base chromosome numbers in *Brassica* ($x$ = 8, 9 and 10) and multiple base numbers in the tribe Brassiceae originated from a WGD followed by independent diploidization (Lysak *et al.*, 2005, 2007).

Here we aimed to investigate whether multiple base chromosome numbers of the polybasic Brassicaceae genera and tribes resulted from genome diploidization after independent mesopolyploid WGDs post-dating the family-specific At-$\alpha$ paleopolyploidization. We approached this question by employing comparative chromosome painting (CCP) and/or whole-transcriptome analysis in 16 species from 10 Brassicaceae tribes (Table 1). To corroborate the CCP inferences of mesopolyploidy we analyzed gene age distributions for evidence of large gene duplication events consistent with polyploidy (Barker *et al.*, 2008, 2010). We also used a combination of among-lineage ortholog divergences and a recently developed phylogenomic approach, Multi-tAxon Paleopolyploidy Search (MAPS; Li *et al.*, 2015), to confirm the phylogenetic placement of the inferred WGDs. We show that several diploid-like Brassicaceae genera and tribes have undergone independent mesotetra- or mesohexaploidizations followed by genetic and genomic diploidization concealing the polyploid history of these taxa and generating the karyological variation. Furthermore, we found convergent patterns of gene retention and loss among the independent mesopolyploid WGDs.

## RESULTS

### The feasibility of CCP in 10 Brassicaceae species

We found that CCP with *A. thaliana* bacterial artificial chromosome (BAC) contigs is feasible in all analyzed species, but with different efficacies. The interpretation of *in situ* hybridization patterns is feasible only if the intensity of hybridized painting probes is sufficiently strong. This is dependent on several factors, such as the phylogenetic distance between *A. thaliana* and each target species, repeat content or the degree of pachytene chromosome clumping. Moreover, in polyploid genomes the interpretation of CCP patterns is complicated by diploidization and the associated genomic reorganization. Among 10 analyzed species (Table 1), *Lobularia libyca* had the weakest probe intensity.

**Table 1** Brassicaceae species analyzed in this study by comparative chromosome painting (CCP) and/or transcriptome sequencing (Tr)

| Species | Lineage[a]/Clade[b] | Tribe (no. of genera: no. of species)[c] | 2*n* | CCP/Tr |
|---|---|---|---|---|
| *Physaria gracilis* | I/A | Physarieae (7: 133) | 12 | +/– |
| *Physaria fendleri* | I/A | Physarieae | 12 | –/+ |
| *Stenopetalum nutans* | I/A | Microlepidieae (16: 56) | 8 | –/+ |
| *Pachycladon exilis* | I/A | Microlepidieae | 20 | –/+ |
| *Pachycladon fastigiatum* | I/A | Microlepidieae | 20? | –/+[d] |
| *Leavenworthia alabamica* | I/A | Cardamineae (12: 337) | 22 | +/– |
| *Schizopetalon walkeri* | II/? | Schizopetaleae (3: 16) | 18 | +/– |
| *Streptanthus farnthworthianus* | II/? | Thelypodieae (26: 244) | 28 | +/+ |
| *Cochlearia pyrenaica* | E II/C | Cochlearieae (2: 29) | 12 | +/+ |
| *Cochlearia officinalis* | E II/C | Cochlearieae | 24 | –/+ |
| *Iberis umbellata* | E II/C | Iberideae (2: 30) | 18 | +/+ |
| *Heliophila africana* | E II/? | Heliophileae (1: 90) | 20 | +/– |
| *Heliophila longifolia* | E II/? | Heliophileae | 22 | –/+ |
| *Biscutella baetica* | ?/C | Biscutelleae (2: 46) | 16 | +/+ |
| *Biscutella lyrata* | ?/C | Biscutelleae | 12 | +/+ |
| *Anastatica hierochuntica* | III/C | Anastaticeae (14: 72) | 22 | –/+ |
| *Lobularia libyca* | III/C | Anastaticeae | 22 | +/+ |

[a]Franzke *et al.* (2011) (E II = expanded lineage II).
[b]Huang *et al.* (2016).
[c]Al-Shehbaz (2012), Al-Shehbaz *et al.* (2014), Salariato *et al.* (2016).
[d]Gruenheit *et al.* (2012).

We used a double concentration of painting probes in this species. The lower fluorescence intensity may reflect the phylogenetic distance between Arabidopsis (clade A in Huang *et al.*, 2016) and *Lobularia* (clade C). All but three species (*Heliophila africana*, *Iberis umbellata* and *Physaria gracilis*) demonstrated extensive regions of pericentromeric heterochromatin. This caused clustering of non-homologous centromeres at pachytene. Consequently, interpretation of CCP patterns at proximal chromosomal regions was difficult for most species (Figure 1).

**New independent mesopolyploid events in Brassicaceae**

Previous experimental data (e.g. Lysak *et al.*, 2006; Mandáková and Lysak, 2008) showed that chromosome-specific BAC contigs of *A. thaliana* hybridize to unique chromosomal regions in Brassicaceae genomes with only the At-α WGD. In 10 crucifer species analyzed here, *A. thaliana* BAC contigs representing four ancestral chromosomes (AK1, AK2, AK6 and AK8) with 12 genomic blocks (GB) and about 48.5 Mb of the *A. thaliana* genome sequence, revealed two or three regions of chromosomal homoeology within the haploid chromosome complements (Figure 1). The two or three genomic copies of ancestral chromosomes and corresponding GBs were interpreted as resulting from mesotetra- or mesohexaploidization events that occurred in the ancestry of each species after the At-α WGD.

**Mesotetraploid genomes**

CCP in representatives of Biscutelleae (*Biscutella baetica*, 2*n* = 16 and *Biscutella lyrata*, 2*n* = 12, Figure 1a, b), Iberideae (*Iberis umbellata*, 2*n* = 18, Figure 1c), Anastaticeae (*L. libyca*, 2*n* = 22, Figure 1d), Schizopetaleae (*Schizopetalon wallkeri*, 2*n* = 18, Figure 1e) and Thelypodieae (*Streptanthus farnthworthianus*, 2*n* = 28, Figure 1f), revealed two genomic copies of the tested GB associations and unambiguously showed that these species have descended from mesotetraploid ancestors. In two *Biscutella* species and *I. umbellata*, the painted homoeologous GBs differed in length and, in some instances, fluorescence intensity. These differences were more pronounced for longer GBs, whereby a longer and brighter genomic copy was labeled as no. 1. In *L. libyca*, *S. wallkeri* and *S. farnthworthianus*, the two genomic copies could not be reliably differentiated. The CCP patterns in individual species are presented in Figure 1(a–f) and detailed in Data S1 in the Supporting Information.

**Mesohexaploid genomes**

In representatives of Cardamineae (*Leavenworthia alabamica*, 2*n* = 22, Figure 1i), Cochlearieae (*Cochlearia pyrenaica*, 2*n* = 12, Figure 1g), Heliophileae (*Heliophila africana*, 2*n* = 20, Figure 1h) and Physarieae (*Physaria gracilis,* 2*n* = 12, Figure 1j), CCP uncovered three copies of the GBs that comprise the four ancestral chromosomes. This level of duplication suggests two rounds of WGD in short succession in these species. For simplicity, these are referred to as whole-genome triplication (WGT) events. In all species, triplicated GBs differed significantly in length and, in some instances, fluorescence intensity. This difference was more pronounced in the case of longer GBs (the longest and usually brightest copy is always labeled as no. 1, whereas the shortest and usually the weakest copy is

 

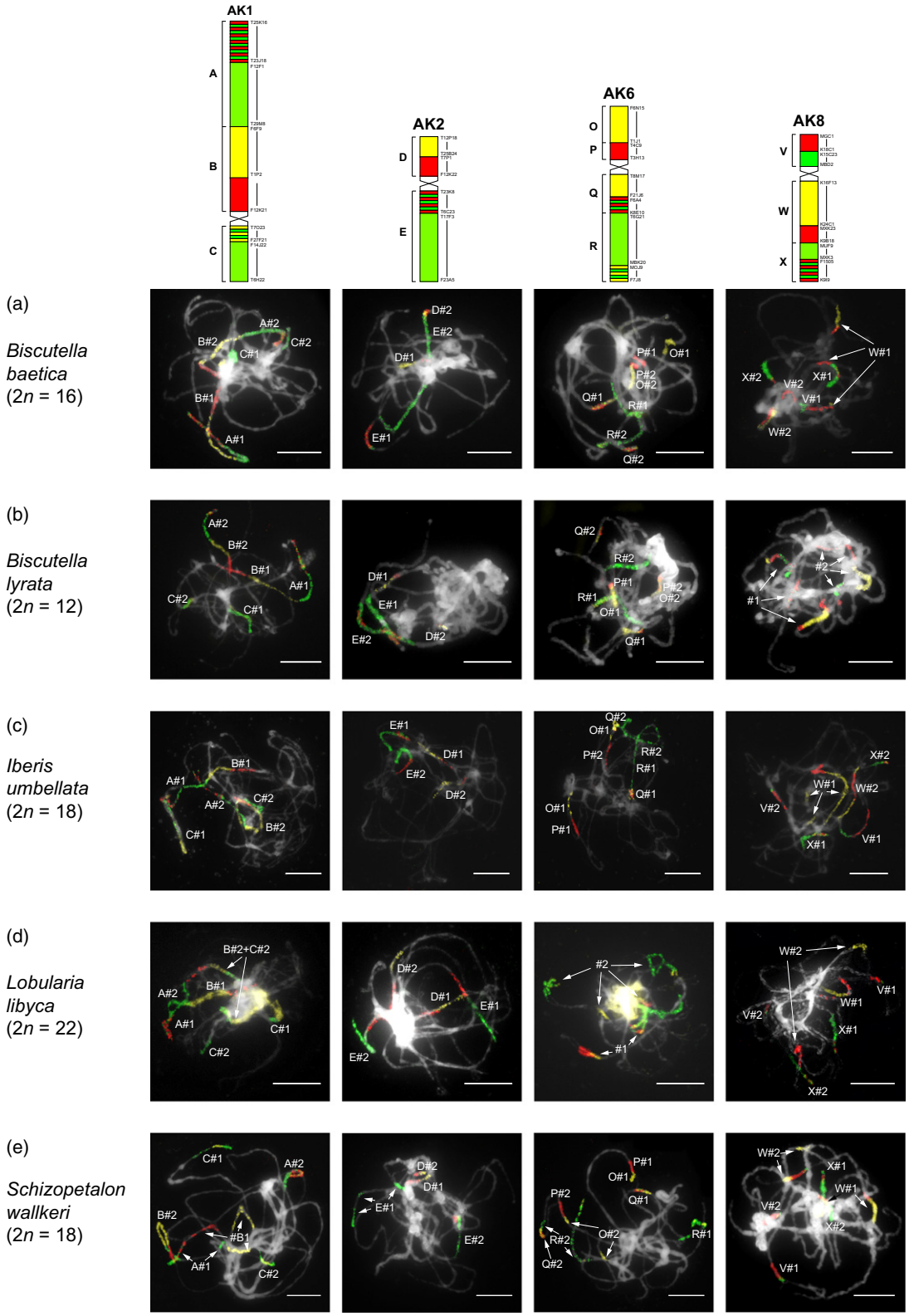


**Figure 1.** Results of comparative chromosome painting analyses in pachytene (haploid) chromosome complements of 10 Brassicaceae species. (a)–(f) Mesotetraploid species, (g)–(j) mesohexaploid species. The 12 genomic blocks (GBs) of four ancestral chromosomes were labeled by three different haptens/fluorochromes. Two (a–f) or three (g–j) copies of GBs were distinguished and assigned based on differences in their length and fluorescence intensity. All bars = 10 μm.

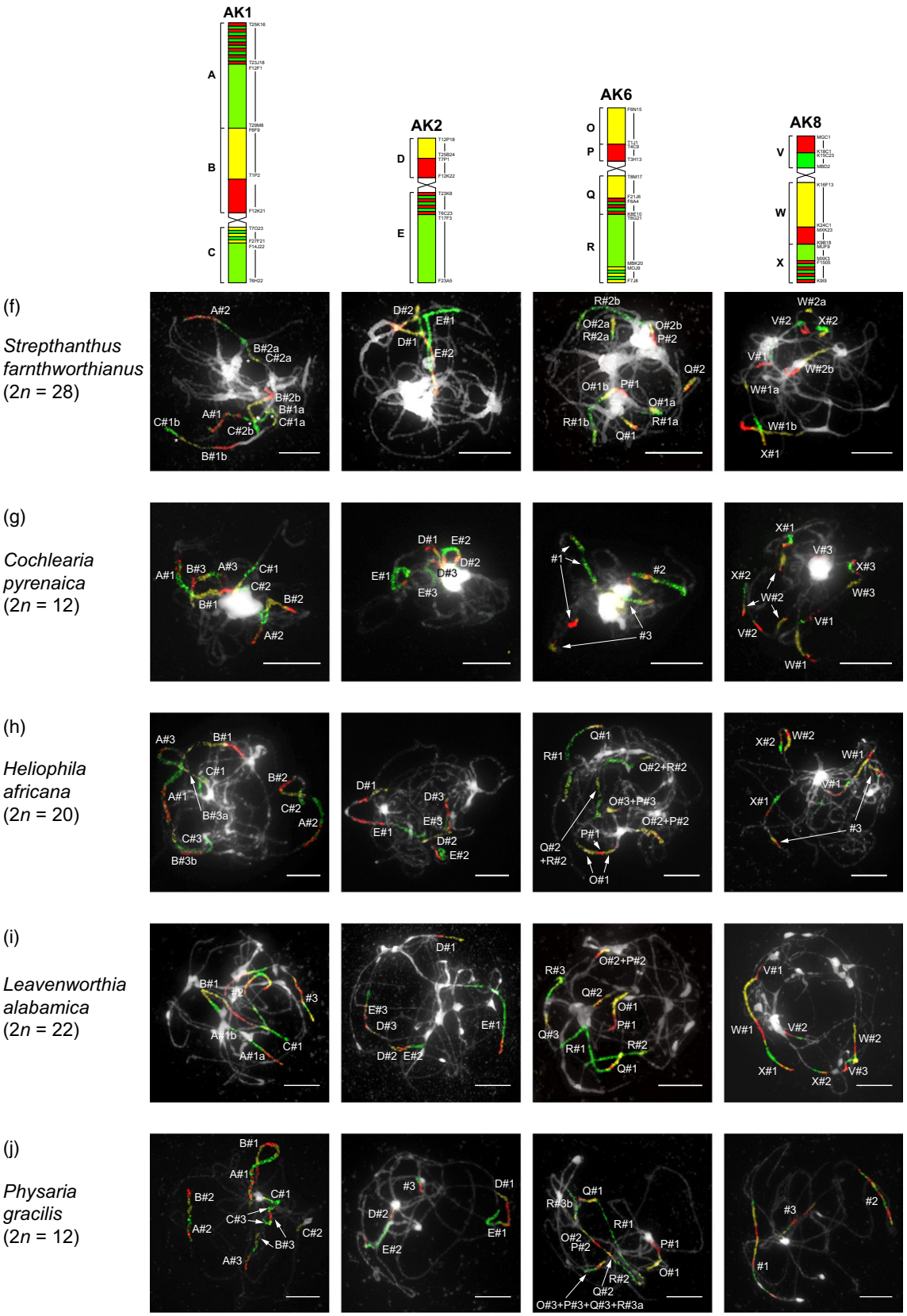


**Figure 1.** Continued

 

always labeled as no. 3). CCP patterns in individual species are presented in Figure 1g–j and detailed in Data S1.

**Transcriptome analyses of mesopolyploidies**

Our analyses of transcriptomes from 13 representative Brassicaceae species (Table 1) identified numerous mesopolyploidies and corroborated our cytogenetic results. For the 12 newly sequenced and one previously available (Gruenheit *et al.*, 2012) Brassicaceae transcriptomes, we assembled a mean of 40 427 unigenes with an average total assembly size of 29.5 Mb (Table S2). Analyses of gene age distributions found peaks of gene duplication consistent with WGDs in most taxa. Mixture models identified two or three significant normal distributions within the $K_s$ plot of gene duplications for each species (Table 2). These included visually well distinguished peaks in the gene age distributions of *Anastatica*, *Biscutella*, *Cochlearia*, *Heliophila*, *Pachycladon*, *Physaria* and *Stenopetalum*. The gene age distribution in *Lobularia* was relatively noisy, possibly an artifact of its relatively small sample size or potentially indicative of a more complex history of duplication and hybridization (Gaut and Doebley, 1997; Barker *et al.*, 2008; Doyle and Egan, 2010). Younger inferred WGD peaks in *Iberis* and *Streptanthus* were less distinguished from the initial peak of gene duplications (Figure 2). However, the relatively large number of duplications in the early age distribution is consistent with a young mesopolyploidy in their histories. Each species also contained an older, more fractionated peak located at a $K_s$ of about 0.7–1.2. These older peaks likely correspond to the shared At-α WGD in the ancestry of all Brassicaceae, whereas the younger and more prominent peaks represent recent mesopolyploidy. The ΔBIC values (BIC being the Bayesian information criterion) for models with and without inferred polyploid peaks were large and supported our inference of WGDs across all species (Table 2).

To place these inferred WGDs on the Brassicaceae phylogeny and determine the number of independent WGDs, we used a combination of ortholog divergence (Barker *et al.*, 2008, 2009, 2010) and MAPS analyses (Li *et al.*, 2015). A comparison of the synonymous divergence of WGD paralogs relative to the divergence between orthologs of species may be used to assess whether a WGD occurred before or after lineage divergence. If the WGD paralog divergence is older than the ortholog divergence, then the WGD is likely to have occurred in a common ancestor of the taxa. Similarly, if the WGD paralog divergence is younger than the ortholog divergence, then the WGD(s) occurred after lineage divergence. In our analyses, the ortholog divergences between genera were generally older than the mesopolyploid $K_s$ divergences and support independent WGDs in *Anastatica*, *Heliophila*, *Iberis*,

**Table 2** Summary statistics of mixture model distributions inferred to represent the At-α whole-genome duplication and mesopolyploidies ('meso') across the Brassicaceae taxa analyzed

| Species | Mixture median ($K_s$) | % of data | ΔBIC (w/o At-α) | ΔBIC (w/o At-α and meso) |
|---|---|---|---|---|
| *Anastatica hierochuntica* | 0.41199 | 47.8 | 420.33 | 1861.5 |
| | 0.80368 | 40.5 | | |
| *Biscutella baetica* | 0.24466 | 46.3 | 6957.85 | 19538 |
| | 0.87186 | 21.9 | | |
| *Biscutella lyrata* | 0.3127 | 53.9 | 4406.47 | 4406.47 |
| | 0.73875 | 32.3 | | |
| *Cochlearia pyrenaica* | 0.52492 | 34.7 | 908.09 | 650.61 |
| | 1.1909 | 17.5 | | |
| *Cochlearia officinalis* | 0.49115 | 56.6 | 650.61 | 22306.2 |
| | 1.0583 | 29.2 | | |
| *Heliophila longifolia* | 0.42718 | 36.8 | 2906.4 | 16078.1 |
| | 0.98813 | 17.6 | | |
| *Iberis umbellata* | 0.12685 | 43.0 | 7297.67 | 23380.8 |
| | 0.62127 | 33.6 | | |
| *Physaria fendlerii* | 0.5491 | 73.6 | 851.22 | 851.22 |
| | 1.1817 | 23.1 | | |
| *Lobularia libyca* | 0.17719 | 24.6 | 670.25 | 1202.9 |
| | 0.68083 | 57.1 | | |
| *Pachycladon exilis* | 0.197 | 70.4 | 1643.55 | 30590.0 |
| | 0.73588 | 10.9 | | |
| *Streptanthus farnsworthianus* | 0.11151 | 33.9 | 1150.36 | 18974.1 |
| | 0.69468 | 38.8 | | |
| *Stenopetalum nutans* | 0.2452 | 42.3 | 6479.67 | 9378.44 |
| | 0.8432 | 22.8 | | |

BIC, Bayesian information criterion; w/o, without.

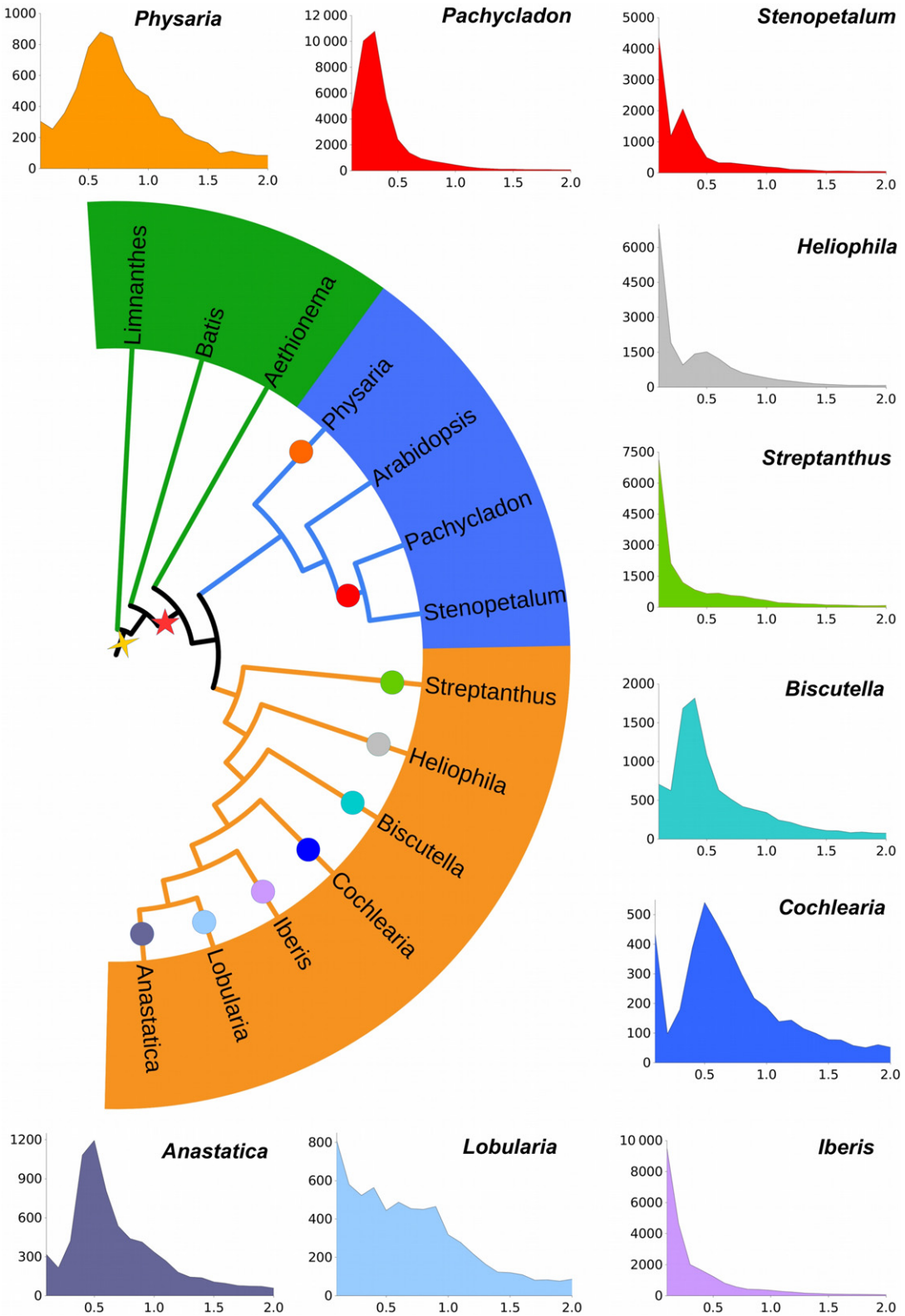


**Figure 2.** Phylogenetic placement of the nine mesopolyploidies in the Brassicaceae identified by transcriptome analyses.
Circles correspond to the inferred location of mesopolyploid events, colors match with 10 $K_s$ plots. Red and yellow stars indicate the inferred Brassicaceae-specific At-α and core Brassicales-specific At-β paleopolyploidies, respectively. Colors in the simplified phylogeny correspond to broadly defined lineage I (blue), expanded lineage II (orange) and Aethionemeae/Brassicales (green).

 

**Table 3** Summary of ortholog age distribution analyses. Median ortholog divergence is the median $K_s$ value of ortholog divergences of the selected species pair in each analysis

| Taxa 1 | Taxa 2 | Median ortholog divergence ($K_s$) | No. of orthologs |
|---|---|---|---|
| *Anastatica hierochuntica* | *Lobularia libyca* | 0.436 | 13 364 |
| *Iberis umbellata* | *Lobularia libyca* | 0.489 | 13 360 |
| *Iberis umbellata* | *Cochlearia officinalis* | 0.599 | 12 696 |
| *Lobularia libyca* | *Cochlearia pyrenaica* | 0.541 | 10 434 |
| *Cochlearia pyrenaica* | *Biscutella baetica* | 0.546 | 11 354 |
| *Iberis umbellata* | *Biscutella baetica* | 0.494 | 14 864 |
| *Biscutella baetica* | *Biscutella lyrata* | 0.169 | 17 178 |
| *Cochlearia pyrenaica* | *Cochlearia officinalis* | 0.017 | 9909 |
| *Streptanthus farnsworthianus* | *Biscutella lyrata* | 0.398 | 13 718 |
| *Heliophila longifolia* | *Streptanthus farnsworthianus* | 0.435 | 13 894 |
| *Heliophila longifolia* | *Lobularia libyca* | 0.472 | 13 048 |
| *Pachycladon fastigiatum* | *Stenopetalum nutans* | 0.083 | 6942 |
| *Pachycladon exilis* | *Physaria fendlerii* | 0.450 | 14 602 |

*Lobularia*, *Physaria* and *Streptanthus* (Table 3). However, we found that the ortholog divergence between *Pachycladon* and *Stenopetalum* was more recent than the paralog divergence of their inferred mesopolyploidy, consistent with a WGD in their common ancestor. We also used a recently developed phylogenomic approach, MAPS, to test the phylogenetic placement of these mesopolyploidies. Multi-species gene tree analyses with MAPS supported the ortholog divergence results and found no evidence that the inferred mesopolyploidies in *Anastatica*, *Heliophila*, *Iberis*, *Lobularia*, *Physaria* and *Streptanthus* were shared with other taxa. Consistent with independent WGDs within each lineage rather than a shared WGD(s) among genera, the fraction of shared gene duplications in each of these MAPS analyses was low across the analyses for nodes within the Brassicaceae (Figure 3; Table S3). Notably we recovered significant increases in shared gene families for the At-α WGD. Overall, our divergence and phylogenomic analyses consistently supported independent mesopolyploidies in the ancestry of *Anastatica*, *Heliophila*, *Iberis*, *Lobularia*, *Physaria* and *Streptanthus*.

We also found evidence that some of the inferred WGDs occurred in the common ancestor of pairs of genera. Our analyses of *Pachycladon* and *Stenopetalum* indicated that they share a mesopolyploidy. Although the median $K_s$ values of their WGD peaks are significantly different ($K_s$ = 0.197 versus 0.245), the ortholog divergence between *Pachycladon* and *Stenopetalum* was lower and more recent than their mesopolyploid paralog divergence (Table 3). Analyses of substitution rates across 3403 nuclear orthologs indicated that *Stenopetalum* is evolving about 21% faster than *Pachycladon* relative to Arabidopsis. Accounting for this heterogeneity in substitution rate aligned the median $K_s$ divergence estimates for the putative shared WGD. The median paralog divergence for the *Stenopetalum* WGD was $K_s$ = 0.245, whereas the median rate corrected paralog divergence for the *Pachycladon* WGD was $K_s$ = 0.238. Consistent with our ortholog divergence and rate analyses, more than 47% of the gene subtrees in the MAPS analyses supported a shared WGD in the ancestry of *Pachycladon* and *Stenopetalum* (Figure 3). Similarly, ortholog divergences were much younger than the paralog divergence of inferred mesopolyploidies in the two within-genera comparisons in *Biscutella* and *Cochlearia*. Overall, we found evidence for nine independent mesopolyploidies in the ancestry of the analyzed taxa.

Each species also contained a mixture with a median $K_s$ of about 0.7–1.2 that was consistently significant by large ΔBIC values. Previous analyses of the Brassicaceae found that paralogs from the At-α WGD diverge in this range of synonymous substitution (Barker *et al.*, 2009; Edger *et al.*, 2015). Analyses of ortholog divergence indicated that all Brassicaceae lineages diverged after these $K_s$ ~0.7–1.2 WGD peaks (Table 3), consistent with these peaks representing the At-α paleopolyploid event. Finally, our MAPS analyses, as expected, indicated that all Brassicaceae share the At-α WGD (Figure 3a, f). More than 35% of gene subtrees consistent with the species tree supported a shared gene duplication event in the ancestry of all Brassicaceae, consistent with At-α. The other MAPS analyses did not include a non-Brassicaceae outgroup (Figure 3b–e, g) and did not directly test the node where At-α is established to occur (Barker *et al.*, 2009; Edger *et al.*, 2015).

**Biased gene retention and loss**

Given the numerous mesopolyploidies and karyotypic variation revealed by our analyses, we evaluated the consistency of post-polyploid genome evolution across the diverse Brassicaceae species sampled. To explore the patterns of gene retention and loss following each mesopolyploidy and the At-α WGD, we annotated each

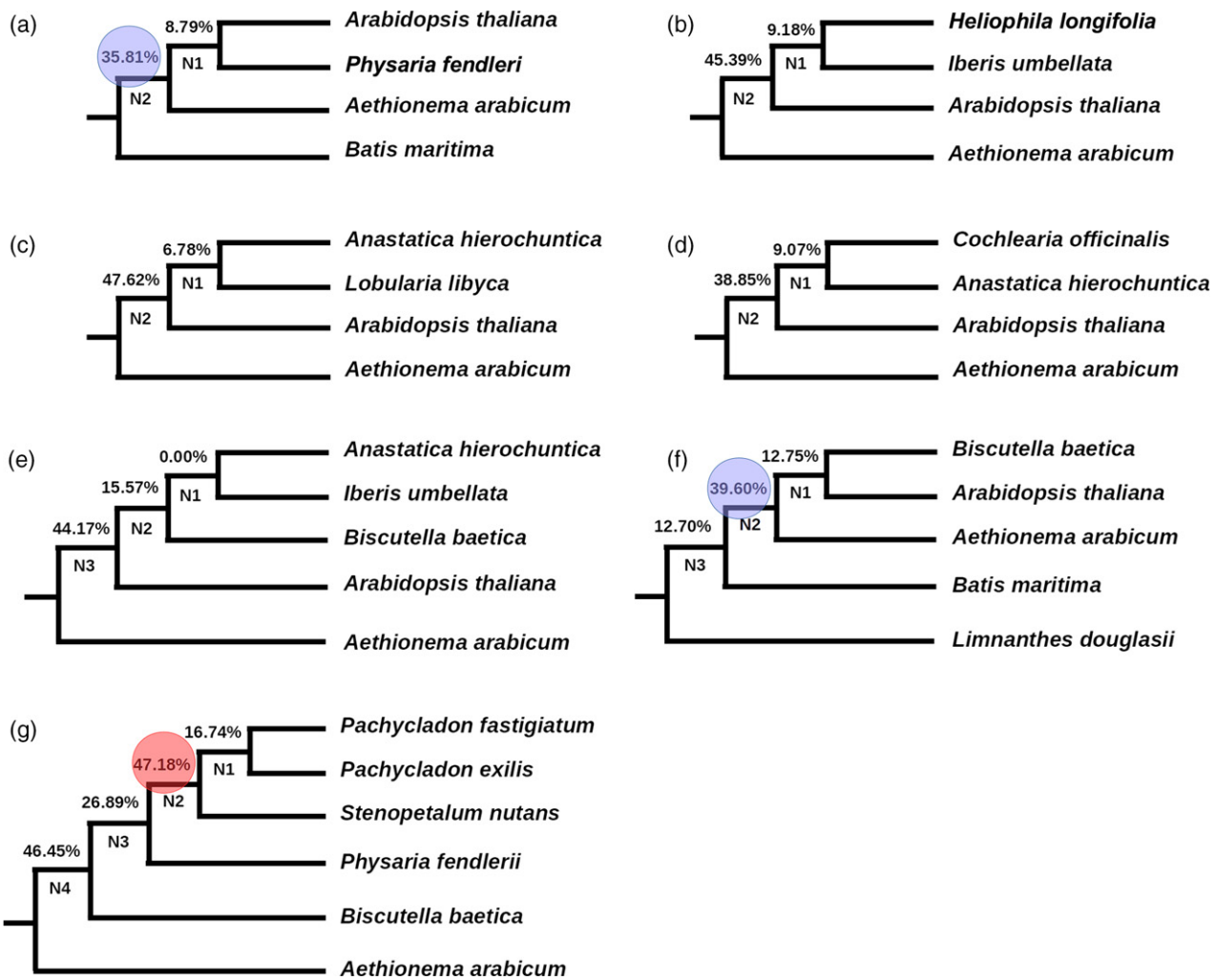


**Figure 3.** Species trees with gene duplication at each node as inferred by MAPS.
The percentage of gene subtrees consistent with the species tree that support a shared duplication is indicated at each node. (a), (f) Blue circles highlight inferred Brassicaceae-specific At-α paleopolyploid events at the associated nodes. In (a)–(f) the low percentage of gene subtrees at other nodes support independent mesopolyploid events in *Anastatica*, *Biscutella*, *Cochlearia*, *Heliophila*, *Iberis*, *Lobularia*, *Physaria* and *Streptanthus*. (g) A red circle highlights the inferred mesopolyploidy shared by *Pachycladon* and *Stenopetalum*.

transcriptome using the Arabidopsis Gene Ontologies (GOs). For each species, we used the mixture model $K_s$ ranges for each WGD to identify paralogs from the mesopolyploidies and the At-α WGD. The GO categories for these mesopolyploidy and At-α WGD paralogs were then analyzed using a simulated chi-square test to identify significantly over- and under-retained GO categories among these WGD paralogs. We also used a hierarchical clustering approach to group WGD GO category profiles by similarity. Consistent with our expectations, we found evidence that the patterns of paralog retention and loss from all nine mesopolyploidies and the At-α WGD were significantly biased with respect to GO categories and largely convergent across the independent mesopolyploidies (Figure 4). Overall, the GO categories retained and lost following all the WGDs were generally similar in profile, but there were some consistent differences of GO category enrichment based on WGD age. Hierarchical clustering resolved two major clusters of gene retention and loss among the mesopolyploid and At-α paralogs. One cluster was largely composed of genes retained or lost from independent mesopolyploidies, whereas the other cluster consisted mostly of genes retained and lost following the shared At-α WGD (Figure 4a). Chi-square analyses of GO category retention and loss identified many significantly over- and under-retained functional categories of paralogs (Figure 4b). Paralogs retained from the At-α event were significantly enriched across our sampled species for genes associated with other cellular processes, other metabolic processes, other binding, transferase activity, protein metabolism, kinase activity, transcription factor activity and signal transduction. Many of these GO categories were

significantly under-retained among the paralogs in mesopolyploidies. In contrast, the independent mesopolyploidies consistently demonstrated significant over-retention of genes associated with other intracellular components, other cytoplasmic components, other membranes, chloroplast, plastid, extracellular and electron transport. Similarly, many of these GO categories were significantly under-retained among At-α paralogs across our sampled species (Figure 4).

Although these contrasting patterns of gene retention and loss between mesopolyploidies and the At-α WGD were largely consistent across taxa, there were a few

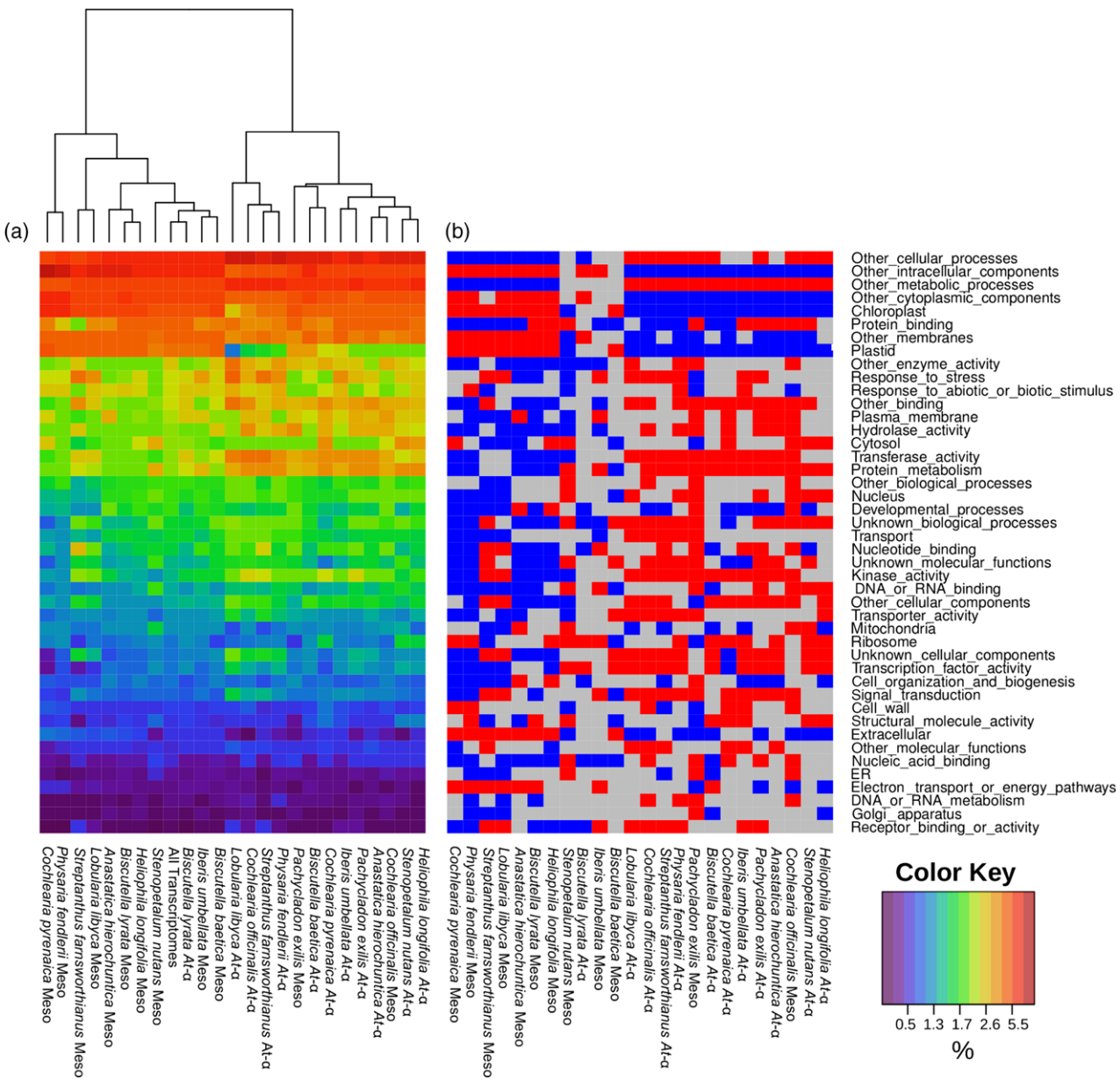


**Figure 4.** Gene Ontology (GO) annotations of Brassicaceae whole transcriptomes and paleologs.
Each column represents the annotated GO categories retained in duplicate following independent mesopolyploidies or the At-α whole-genome duplication from each analyzed species, whereas the rows represent a particular GO category.
In (a) the colors of the heat map represent the percentage of the transcriptome represented by a particular GO category, with red being highest and purple lowest. The overall ranking of GO category rows was determined by the ranking of GO annotations among the total pooled transcriptomes. Hierarchical clustering was used to organize the heatmap columns.
In (b) GO categories are identified that are significantly over- (red) or under-retained (blue) among paleologs relative to the total pooled transcriptomes, as determined by residuals from chi-square tests. GO categories with gray boxes in panel (b) are not represented in significantly different numbers from those expected by chance.

exceptions to this pattern (Figure 4). The paralogs from the At-α event of *B. lyrata* clustered with the mesopolyploid paralogs of most species. However, relatively few GO categories of *B. lyrata*'s At-α paralogs were significantly over- or under-retained, and its mesopolyploid paralogs had a similar pattern of GO categories to most other mesopolyploidies. It appears that *B. lyrata*'s At-α paralogs clustered with the mesopolyploidies because it is an outlier rather than sharing great similarity. Low-quality tissue or other variation in sample preparation for *B. lyrata* may have caused poor sampling of the genes. This is supported by the fact that the differential gene retention and loss in *B. lyrata*'s congener, *B. baetica*, follows the general pattern. In contrast, the retained paralogs from the mesopolyploidies of *Cochlearia officinalis* and *Pachycladon exilis* had a pattern of GO categories similar to the At-α paralogs. These mesopolyploid paralogs are not clearly outliers and seem to have a similar pattern of retention to At-α paralogs from their own and other species' genomes. The pattern of gene retention following the mesopolyploidy in *Pachycladon* was significantly different than gene retention and loss from the same WGD in *Stenopetalum*. Although *Pachycladon*'s nuclear substitution rate is about 21% slower than *Stenopetalum*, the pattern of gene retention did not appear to reflect a simply slower rate of loss to the mesopolyploid pattern. It may be that a significantly different substitution rate affects the pattern of genome fractionation, especially if genes are lost by accumulating substitutions. As *Pachycladon* is relatively slowly evolving and *C. officinalis* is a neomesopolyploid species ($2n = 4x = 24$), duplicated At-α paralogs may have not yet undergone gene fractionation comparable with the remaining mesopolyploid species. For *C. officinalis*, this is also supported by gene retention and loss patterns in the mesopolyploid *C. pyrenaica*, which follows the general pattern.

## DISCUSSION

### Cytogenetic and transcriptomic evidence of clade-specific mesopolyploid events across Brassicaceae

Here for the first time we combined large-scale comparative chromosome painting and transcriptome analyses to identify and characterize WGD events across the Brassicaceae. Using painting probes for half of the ancestral $n = 8$ chromosome complement we uncovered new mesotetraploidy events putatively specific for Anastaticeae (*Anastatica*- and *Lobularia*-specific WGD), Iberideae and Schizopetaleae and two new mesohexaploidy events in the ancestry of *Cochlearia* (Cochlearieae) and *Physaria* (Physarieae). Furthermore, we confirmed a mesotetraploidy in *Biscutella* (Biscutelleae; Geiser *et al.*, 2016) and Thelypodieae (Burrell *et al.*, 2011; Kagale *et al.*, 2014a), another event in Microlepidieae (WGD shared by *Pachycladon* and *Stenopetalum*; Mandáková *et al.*, 2010a,b), as well as the mesohexaploidy events in *Leavenworthia* (Cardamineae, Haudry *et al.*, 2013) and Heliophileae (Mandáková *et al.*, 2012).

Although cytogenetic analyses can reveal the existence and nature of mesopolyploidies, it may be difficult to discern if they represent shared or independent ancestral WGDs. This is particularly true if post-polyploidization genome reshuffling obscures either ancestrally shared or genome-specific cytogenetic signatures. Analyses of newly sequenced transcriptomes from 10 Brassicaceae genera corroborated our cytogenetic analyses. Our combined single species and phylogenomic approaches found that nearly all of the analyzed genera experienced an independent mesopolyploidy. Overall, our cytogenetic and transcriptomic analyses revealed six new independent mesopolyploidies in the ancestry of polybasic Brassicaceae genera and tribes (Figure 2).

### Mesotetraploidy events

CCP and transcriptome analysis unveiled a previously unknown WGD in *I. umbellata* ($n = 9$) belonging to 27 *Iberis* species from the rather small tribe Iberideae (two genera, 30 species; Al-Shehbaz, 2012). The WGD is most probably shared by the whole genus and various base numbers ($n = 7$, 9 and 11; Warwick and Al-Shehbaz, 2006) reflect the different extent of post-mesopolyploid diploidization in *Iberis*.

Another newly described mesotetraploidy occurred in the ancestry of the *S. walkeri* genome ($n = 9$). *Schizopetalon* (10 species) together with *Atacama* (1 species) and *Mathewsia* (6 species) constitute the small tribe Schizopetaleae confined exclusively to central Chile and parts of Peru and Argentina (Salariato *et al.*, 2016). More data are needed to determine if the WGD is shared by all of Schizopetaleae and the CES clade (Salariato *et al.*, 2016).

By cytogenetic and transcriptomic analysis of *S. farnthworthianus* ($n = 14$) we demonstrated that this species, and most likely the whole Thelypodieae (26 genera, 244 species), dominated by species with $n = 14$ and 13, have a neo/mesotetraploid origin based on $n = 7$. CCP analysis suggested that the *S. farnthworthianus* genome originated by duplication of the proto-Calepineae karyotype (PCK; $n = 7$; Mandáková and Lysak, 2008). This is in accordance with the inclusion of Thelypodieae into Lineage II (clade B *sensu* Huang *et al.*, 2016). The Thelypodieae-specific WGD was already suspected by Burrell *et al.* (2011) from genetic mapping in *Caulanthus amplexicaulis* ($n = 14$) and further corroborated by transcriptome sequencing in *Pringlea antiscorbutica* ($n = 12$) and *Stanleya pinnata* ($n = 12$ or 14) (Kagale *et al.*, 2014a). Chromosome number variation across Thelypodieae ($n = 7$, 9, 10, 11, 12, 13 and 14) and our data suggest that the primary neo/mesotetraploid genome had 14 chromosomes and that the lower chromosome counts represent descending dysploidies.

 

Despite the prevalent low, diploid-like chromosome numbers ($n$ = 4–7, 10) of the endemic Australian/New Zealand crucifers classified as members of Microlepidieae (16 genera, about 56 species; Heenan *et al.*, 2012), the group has descended from allotetraploid ancestors with approximately 16 chromosomes (Mandáková *et al.*, 2010a, b). Besides the 15 endemic Australian genera, Microlepidieae contain 11 species of the genus *Pachycladon* endemic to New Zealand (10 species) and Tasmania (1 species). Unlike the Australian species, *Pachycladon* species show a remarkable stasis of chromosome numbers (all $n$ = 10) and ancestral genome structures (Mandáková *et al.*, 2010b). The Australian genera and *Pachycladon* species differ in the extent of genome diploidization, whereby the Australian taxa have undergone more extensive post-polyploid genome repatterning relative to 10 *Pachycladon* chromosomes built by largely preserved 16 (2 × 8) ancestral chromosomes. Surprisingly, we found that *Pachycladon* and *Stenopetalum nutans*, a representative of the Australian Microlepidieae taxa, shared the mesotetraploid WGD. Given that they began with the same parental genomes, the extent of diploidization differentiating the two Microlepidieae lineages is probably explained by a much slower tempo of diploidization in *Pachycladon*. Consistent with this hypothesis, our analyses of nuclear substitution rates indicate that the Australian taxa are evolving more than 20% faster than the New Zealand *Pachycladon*. These results indicate that rates of diploidization are correlated with differences in substitution rates. Previous research has found that paleopolyploid genomes of relatively slowly evolving taxa, such as *Vitis vinifera* (Cenci *et al.*, 2013; Murat *et al.*, 2015), are less rearranged than the genomes of faster-evolving taxa. These results are the first to demonstrate this pattern in taxa that descend from the same mesopolyploidy that is recent enough to reconstruct their karyotypes.

**Mesohexaploidy events are not rare in the Brassicaceae**

Until recently, the only mesohexaploidy known in the Brassicaceae was the WGT detected in *Brassica* and probably shared by all Brassiceae (Lysak *et al.*, 2005, 2007; Parkin *et al.*, 2005). Later, a tribe-specific WGT was reported to occur in the ancestry of the Heliophileae (Mandáková *et al.*, 2012) and in the North American *Leavenworthia* (Haudry *et al.*, 2013). Here, by uncovering WGT events in Cochlearieae and Physarieae, we show that mesohexaploidy events might be more frequent than initially thought.

*Leavenworthia* (eight species) is confined to New World and the available chromosome counts mapped on a phylogenetic tree (Urban and Bailey, 2013) suggest that $n$ = 11 is a descending dysploidy from $n$ = 15, and most likely both chromosome numbers were derived from the primary hexaploid genome with $n$ = 24 chromosomes ($2n = 6x = 48$). Considering that *Leavenworthia* and *Selenia* (five species) form a well-defined monophyletic group, with some species of *Leavenworthia* and *Selenia* co-occurring, it can be assumed that the WGT pre-dates the divergence of two sister genera and that it contributed to diversification and species radiation within this North American Cardamineae subclade.

Physarieae is one of the most species-rich Brassicaceae clades with seven genera harboring about 133 species. They are primarily distributed in North America, but several species are disjunctly distributed in Argentina and Bolivia (Al-Shehbaz, 2012). Physarieae was repeatedly confirmed to be a monophyletic group with two large subclades: DDNLS (five genera) and PP (*Paysonia* and *Physaria*) (Fuentes-Soriano and Al-Shehbaz, 2013). *Paysonia* has only eight species with diploid-like chromosome numbers of $n$ = 7, 8 and 9, whereas *Physaria* is a large genus of 106 species with more variable chromosome numbers ($n$ = 4, 5, 6, 7, 8, 9, 10, 12 and higher counts; Warwick and Al-Shehbaz, 2006). Both transcriptome and cytogenetic data suggest that at least *Physaria* has undergone an ancestral WGT event followed by species radiations and genome diploidization towards the extant spectrum of diploid-like chromosome numbers.

Cochlearieae contains 29 species in only two genera–*Cochlearia* (20 species) and *Ionopsidium* (9 species) (Al-Shehbaz, 2012). We analyzed a species with the lowest chromosome number known in the tribe ($n$ = 6) and obtained convincing data indicating that the genus *Cochlearia* has descended from a mesohexaploid ancestor. Subsequent diploidizing reduction of chromosome numbers resulted in the origin of diploid-like chromosome complements, and some of the mesohexaploids formed new allopolyploid species (Koch *et al.*, 1998). Interestingly, Kagale *et al.* (2014a) missed the *Cochlearia*-specific WGT by analyzing the neomesohexaploid genome of *C. officinalis* ($n$ = 12).

Mandáková *et al.* (2012), who analyzed seven different *Heliophila* species, concluded that probably the whole Heliophileae containing more than 90 species, endemic to South Africa and Namibia, experienced a shared WGT. Here, by analyzing two more species by CCP (*H. africana*, $n$ = 10) and transcriptome sequencing (*Heliophila longifolia*, $n$ = 11), we further corroborated our previous conclusion on incidence of this tribe-specific WGT event.

**How outstanding is the number of mesopolyploidies observed in the Brassicaceae?**

In all there are 13 mesopolyploid WGDs, including the WGT and WGD specific for Brassiceae and *Orychophragmus* (Lysak *et al.*, 2005, 2007), which occurred in the Brassicaceae crown group after the At-α paleotetraploidization and after the split of Aethionemeae and the crown group (Figure 5). Of the 49 recognized Brassicaceae tribes (Al-Shehbaz, 2012), at least 11 (22%) have a mesopolyploid

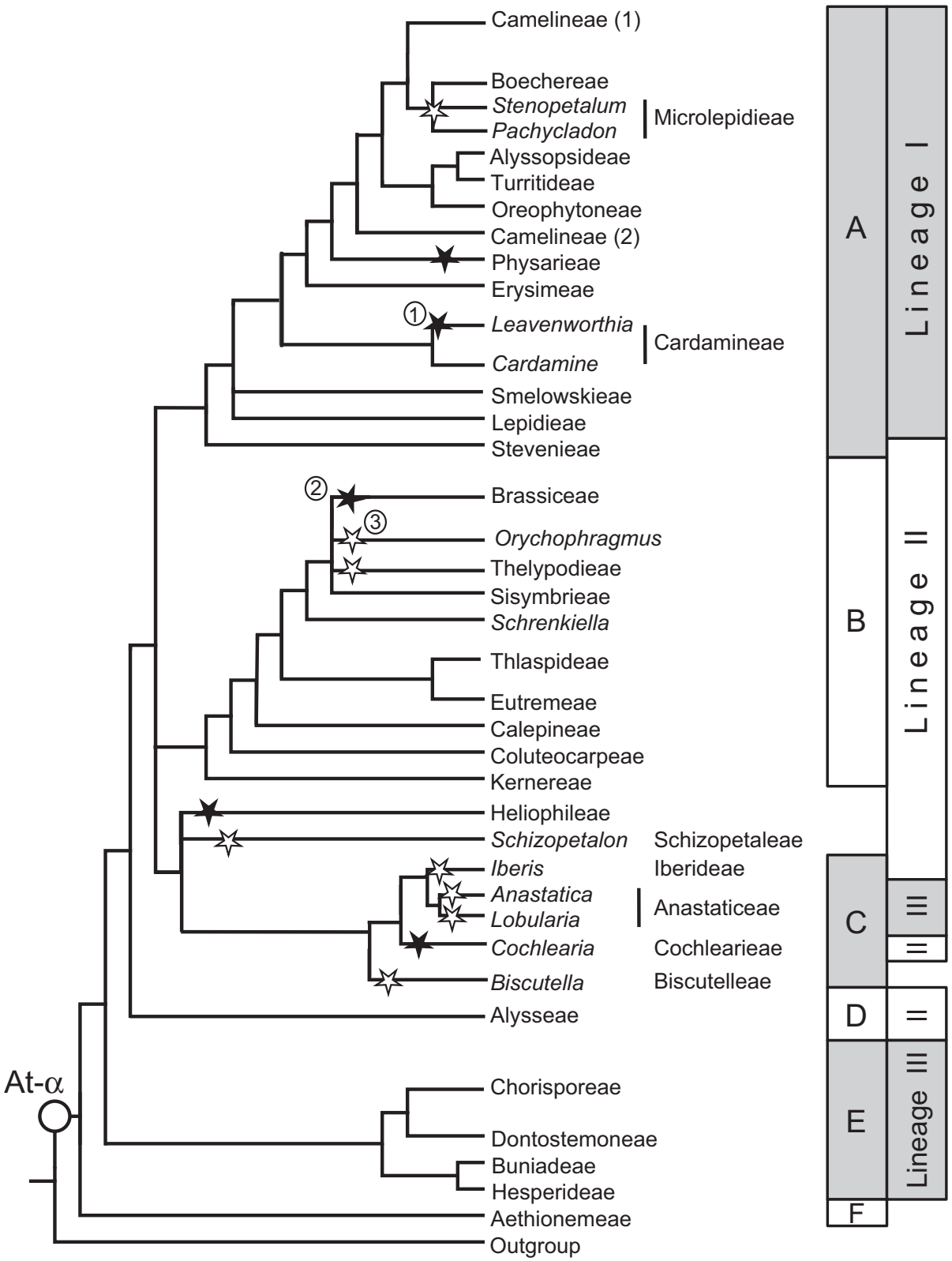


**Figure 5.** Phylogenetic scheme of the family Brassicaceae with the placement of 13 known mesotetraploid (white stars) and mesohexaploid (black stars) events. ① Whole-genome duplication (WGD) in *Leavenworthia* (Haudry *et al.*, 2013), ② WGD in Brassiceae (Lysak *et al.*, 2005) and ③ WGD in *Orychophragmus* (Lysak *et al.*, 2007). The schematic tree topology is based on published phylogenetic reconstructions by Franzke *et al.* (2011) and Huang *et al.* (2016); clade/lineage classification follows Huang *et al.* (2016), Franzke *et al.* (2011) and Al-Shehbaz (2012).

 

ancestry. All but one (as there is not a sufficient number of chromosome counts for Schizopetaleae) mesopolyploid tribes is polybasic, and thus multiple base chromosome numbers in a genus or tribe may indicate an ancient WGD followed by independent diploidizing descending dysploidies. Given our observations in the Brassicaceae, it may be that most polybasic genera of plants have descended from mesopolyploid ancestors. However, most analyses of genomic data from families of plants have not revealed evidence for WGDs in the history of each analyzed genus. For example, genomic analyses that included data from many genera of the Asteraceae (Barker *et al.*, 2008, 2016b) support numerous ancestral polyploidies, but there is no evidence that every genus experienced an independent ancestral WGD. This is particularly surprising in light of our present findings and the enormous taxonomic (about 1623 genera with 24 700 species; Funk *et al.*, 2009) and karyological (Jones, 1985) diversity of the Asteraceae. Similar results have been observed in studies of the Poaceae (Jiao *et al.* 2014; McKain *et al.*, 2016), Fabaceae (Cannon *et al.*, 2015) and Orchidaceae, one of the largest families of vascular plants with about 880 genera and up to 27 000 species where only one paleopolyploid WGD has been revealed so far (Cai *et al.*, 2015). In light of these comparisons, it seems that the frequency of mesopolyploidy in the Brassicaceae may be a relative outlier. It should, however, be noted that this conclusion is premature as many analyses of plant genomes have not included such karyotypically variable lineages and more mesopolyploidies may be found as these lineages are sequenced.

Recent transcriptomic analyses found neo/mesopolyploid events in eight species from five Brassicaceae tribes (Kagale *et al.*, 2014a). However, five of these WGDs are probably best characterized as neopolyploidies based on their elevated euploid chromosome numbers. These neopolyploids include *Capsella bursa-pastoris* ($2n = 4x = 32$; Douglas *et al.*, 2015), *Armoracia rusticana* ($2n = 4x = 32$), *Draba lactea* ($2n = 4x$, $6x = 32$, 48; Grundt *et al.*, 2005), *Lepidium densiflorum* ($2n = 4x = 32$; http://brassibase.cos.uni-heidelberg.de/) and *Lepidium meyenii* ($2n = 8x = 64$; Quiros *et al.*, 1996; Zhang *et al.*, 2016). The puzzling chromosome number of *Lepidium sativum* ($2n = 24$) most likely represents diploidization of a neo/mesotetraploid genome ($2n = 4x = 32$), as suggested by Graeber *et al.* (2014) and observed in Camelina sativa (Kagale *et al.*, 2014b) and *Cardamine cordifolia* (Mandáková *et al.*, 2016).

### Ancestral mesopolyploid genomes and structural diploidization

Polyploid genomes of the Brassicaceae originate through duplication of either an ancestral crucifer karyotype (ACK) with $n = 8$ or a PCK with $n = 7$ (Lysak *et al.*, 2016). Mesotetraploidy thus yields genomes with 16 or 14 chromosomes, whereas mesohexaploidy results in genomes with 24 or 21 chromosomes (for the sake of simplicity hybridization between $n = 8$ and $n = 7$ genomes is not considered). The degree of diploidization, i.e. the reduction of ancestral chromosome numbers ($n = 14$, 16, 21 or 24) to extant, diploid-like chromosome numbers, appears to differ among the groups of mesopolyploid origin (Figure 6). The lowest chromosome numbers, and thus the highest degree of structural diploidization, are characteristic of the mesohexaploid taxa (46–73% genome diploidization) and the mesotetraploid Australian Microlepidieae (66%). This pattern may be explained by the age of mesohexaploidies, whereby a two-step hybridization event resulting in a hexaploid genome theoretically needs more time than a one-step tetraploidization, and/or by higher levels of genome duplication ($6x$ versus $4x$) increasing the frequency of

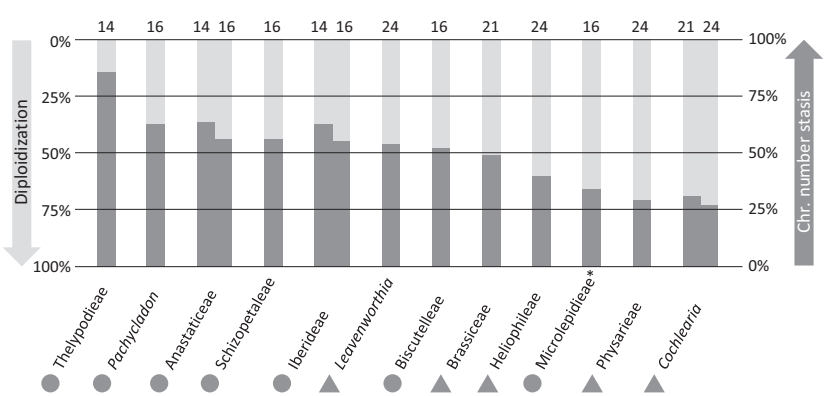


**Figure 6.** Relative percentage of post-polyploid diploidization and ancestral chromosome number stasis in 12 mesopolyploid crucifer taxa.
The percentage of diploidization and chromosome number stasis, respectively, corresponds to the ratio between an average extant chromosome number and the inferred ancestral chromosome number (e.g. if the average extant chromosome number equals 6.5 and the ancestral number is 24, the level of diploidization is equal 73%). Numbers above the graph denote the inferred ancestral mesopolyploid chromosome number for a given taxon; circular and triangular symbols refer to mesotetraploid and mesohexaploid events, respectively. Two possible ancestral chromosome numbers are given for Anastaticeae, Cochlearieae and Iberideae. As the diploidization tempo is strikingly different for *Pachycladon* and the remaining Microlepidieae genera (*), these groups are shown separately.

inter-genome recombination and, at the same time, buffering potentially deleterious consequences of chromosomal rearrangements. The least diploidized genomes were found among Thelypodieae and *Pachycladon* species (14 and 37%, respectively), i.e. two mesopolyploid groups which originated either relatively recently (Thelypodieae) or with a slow genome diploidization (*Pachycladon*).

**Gene retention and loss following multiple, independent mesopolyploidies**

Despite substantial cytological variation across the nine independent mesopolyploidies studied in our analyses, we observed a convergent pattern among the GO categories of genes retained and lost. Eight of the nine independent mesopolyploidies had similar patterns of GO category retention among their paralogs following WGD. The pattern of the other mesopolyploidy was similar to the GO patterns of genes retained in duplicate following the At-α event. The overall pattern of convergent, biased GO category enrichment following WGD across the nine mesopolyploidies was consistent with the predictions of the dosage balance hypothesis (DBH). The DBH predicts that genes with many connections or in dosage-sensitive pathways will be more highly retained following polyploidy (Birchler and Veitia, 2007, 2010, 2012; Edger and Pires, 2009; Freeling, 2009). The higher than expected retention is proposed to maintain the stoichiometry and 'balance' of these biological pathways. Similarly, the same genes will be lost at a higher rate from small-scale duplications to minimize dosage problems. Among the many hypotheses for duplicate gene retention (Kondrashov and Kondrashov, 2006; Freeling, 2009), the DBH is the only hypothesis that explicitly predicts the parallel retention and loss of functionally related genes across species following WGD (Freeling and Thomas, 2006; Freeling, 2009; Conant *et al.*, 2014). Previous research has found evidence consistent with the DBH from analyses of GO enrichment among the genes duplicated by paleopolyploidy, including the At-α WGD in the Brassicaceae (Maere *et al.*, 2005; Thomas *et al.*, 2006; Bekaert *et al.*, 2011; Conant *et al.*, 2014) and other plant WGDs (Rensing *et al.*, 2007; Barker *et al.*, 2008; Shi *et al.*, 2010; Geiser *et al.*, 2016).

Consistent with the DBH, we also found parallel patterns of gene retention and loss among the GO categories of At-α paralogs. Notably, the patterns of GO category retention across the mesopolyploid and At-α paralogs exhibit some differences. Previous analyses have found differences in the patterns of gene retention and loss across WGDs of different ages in the Brassicaceae (Maere *et al.*, 2005; Bekaert *et al.*, 2011) and the Actinidiaceae (Shi *et al.*, 2010). Although we may expect little difference in the pattern of gene retention and loss from any WGD under the DBH, as observed in some analyses (e.g. Barker *et al.*, 2008), dosage sensitivity appears to be fluid and may change over time (Conant *et al.*, 2014). Thus, we may observe variation in the functional patterns of gene retention and loss across WGDs of different ages.

Our results extend past analyses by demonstrating convergent biased gene retention and loss following numerous independent polyploidies. Previous analyses of gene retention and loss for multiple species found parallel patterns of gene retention and loss across multiple WGDs in the Asteraceae (Barker *et al.*, 2008), including rapid emergence of this pattern in *Tragopogon* polyploids only 40 generations old (Buggs *et al.*, 2012). Thus, our observation of convergent gene retention and loss across multiple mesopolyploidies of the Brassicaceae is a new but expected result. The DBH is the most likely explanation for the parallel retention, as it is difficult to conceive that sub- or neofunctionalization would consistently yield the same pattern of paralog retention across nine independent polyploidies in 13 different species. Further, our chromosome painting analyses uncovered significant differences in the degree of chromosomal and genome-wide reorganization and diploidization. Despite these differences, we still found that these genomes retain similar GO categories of genes across their independent WGDs as expected by the DBH.

**Mesopolyploidy and diversification**

Polyploidization is frequently linked with adaptive radiation and diversification (Schranz *et al.*, 2012; Vanneste *et al.*, 2014; Tank *et al.*, 2015; Shimizu-Inatsugi *et al.*, 2017; Soltis and Soltis, 2016). However, establishing a causal link between mesopolyploidies and diversification is not straightforward, as several intrinsic and extrinsic factors have to be considered (e.g. age of a WGD, geographical and paleoclimatic constraints, frequency of neopolyploid formation, extinction rate). In Brassicaceae, independent WGDs appear to pre-date the origins of some species-rich groups, such as Brassiceae (47 genera: 229 species), Heliophileae (2: 100), Physarieae (7: 135) and Thelypodieae (28: 249), and moderately large groups such as Anastaticeae (13: 74) and the Microlepidieae (16: 56) (see also Data S2). This pattern is, however, not universal as mesopolyploidy also occurs in the ancestry of relatively small genera and tribes, such as Biscutelleae (2 genera: 45 species), Cochlearieae (2: 29), Iberideae (2: 30), Schizopetaleae (3: 16) and *Leavenworthia* (8 species). Among 11 Brassicaceae tribes harboring more than 100 species each (Al-Shehbaz, 2012), most likely only three (Brassiceae, Physarieae and Thelypodieae) have a mesopolyploid origin, whereas the species richness of the remaining seven tribes (Alysseae, Arabideae, Boechereae, Cardamineae, Erysimeae, Euclidieae and Lepidieae) is explained by neopolyploidy (Hohmann *et al.*, 2015). In the light of our results, we analyzed the percentage of neopolyploids within mesopolyploid clades and concluded that, with the exception of *Cochlearia* (82%), neopolyploidy maximally explains 25% of the species

richness of these clades (Figure S1). Thus, a preliminary conclusion can be drawn that Brassicaceae mesopolyploids diversify by independent diploidizations presumably associated with reproductive isolation and speciation instead of subsequent rounds of (neo)polyploidization.

## EXPERIMENTAL PROCEDURES

### Plant material

Seeds of each species were donated by botanical gardens and individuals, or collected in the field (Table S1). All plants were grown from seed in a greenhouse or growth chamber until flowering.

### Chromosome preparation and chromosome counts

In preparation for CCP, entire inflorescences were fixed in the ethanol:acetic acid fixative (3:1) overnight and stored in 70% ethanol at −20°C. Selected inflorescences or individual flower buds were rinsed and digested as described by Lysak and Mandáková (2013); the same protocol was applied for the preparation of chromosome spreads from individual flower buds. Suitable slides staged with a phase contrast microscope were used for CCP experiments as well as to establish chromosome numbers from mitoses of tapetal tissues. Prior to CCP, slides were pre-treated by pepsin and RNase (Lysak and Mandáková, 2013). The large amount of cytoplasm covering pachytene chromosomes in *H. africana*, *S. farnthworthianus* and *S. walkeri* required a pepsin treatment twice as long as that needed for the other species. Chromosome numbers were recorded from microscope photographs after counterstaining with DAPI (4′,6-diamidino-2-phenylindole; 2 mg $ml^{-1}$) in Vectashield.

### Comparative chromosome painting

Preparation of painting probes and their localization on chromosomes followed established protocols (Lysak and Mandáková, 2013). Briefly, chromosome-specific BAC contigs of *A. thaliana*, corresponding to 12 GBs that comprise four ancestral crucifer chromosomes (see Figure 1 and Table 2 in Lysak *et al.*, 2016), were used as painting probes. DNA of individual BAC clones was isolated by a standard alkaline lysis protocol and labeled with biotin-, digoxigenin- and Cy3-dUTP by nick translation. Individually labeled probes were pooled according to the 12 GBs, ethanol precipitated, dissolved in hybridization buffer and hybridized to pepsin- and RNase-treated chromosome preparations for about 60 h (for details see Lysak and Mandáková, 2013). The concentration of painting probes was 200 ng/BAC/slide for mesotetraploid species (except for a double concentration of 400 ng/BAC/slide in *Lobularia libyca*) and 300 ng/BAC/slide for mesohexaploid species. After the hybridization, biotin- and digoxigenin-labeled painting probes were immunodetected using fluorescently labeled antibodies, dehydrated and stained with DAPI. Painted chromosomes were photographed using an Olympus BX-61 epifluorescence microscope equipped with a Zeiss CoolCube camera. Monochromatic images were pseudocolored and merged using Adobe Photoshop CS5 software.

### Transcriptome assembly

RNA was isolated from fresh young leaves using the RNeasy plant mini kit (Qiagen, http://www.qiagen.com/). Raw reads for each species were demultiplexed, cleaned with SnoWhite (Dlugosch *et al.*, 2013) and assembled into contigs (Table S2). Two different assembly strategies were used for our two different types of sequence data. SOAPdenovo-Trans (Xie *et al.*, 2014) was used to assemble the paired end Illumina reads using a *k*-mer of $k = 57$. All other parameters were set to default. For our single Ion Torrent transcriptome, we treated the data similarly to 454 reads and assembled them with MIRA version 3.2.1 (Chevreux *et al.*, 2004) using the 'accurate.est.denovo.454' assembly mode. As MIRA may split up high-coverage contigs into multiple contigs, we used CAP3 at 94% identity to further assemble the MIRA contigs and singletons (Huang and Madan, 1999).

### DupPipe analyses of WGDs from transcriptomes of single species

For each transcriptome, we used our DupPipe pipeline to construct gene families and estimate the age of gene duplications (Barker *et al.*, 2008, 2010). We translated DNA sequences and identified reading frames by comparing the Genewise alignment to the best hit protein from a collection of proteins from 25 plant genomes from Phytozome (Goodstein *et al.*, 2012). For all DupPipe runs, we used protein-guided DNA alignments to align our nucleic acids while maintaining the reading frame. We estimated synonymous divergence ($K_s$) using PAML with the F3X4 model (Yang, 2007) for each node in our gene family phylogenies. We identified peaks of gene duplication as evidence of ancient WGDs in histograms of the age distribution of gene duplications ($K_s$ plots). We used a mixture model, EMMIX (McLachlan *et al.*, 1999), to identify significant peaks and estimate their median $K_s$ values. Significant peaks were identified by comparing the ΔBIC values for models with and without inferred polyploid peaks. We used the BIC instead of the Akaike information criterion (AIC) because the BIC has more severe penalties for increasing parameters (peaks, in this application).

### Estimating orthologous divergence

To estimate the average ortholog divergence of conifer taxa and compare it with observed paleopolyploid peaks, we used our previously described RBH Ortholog pipeline (Barker *et al.*, 2010). Briefly, we identified orthologs as reciprocal best blast hits in pairs of transcriptomes. Using protein-guided DNA alignments, we estimated the pairwise synonymous ($K_s$) divergence for each pair of orthologs using PAML with the F3X4 model (Yang, 2007). We plotted the distribution of ortholog divergences and calculated the median divergence to compare against the synonymous divergence of paralogs from inferred WGDs in related lineages. For *Pachycladon* and *Stenopetalum* we also used the orthologs to assess substitution rate heterogeneity (see Methods S1).

### MAPS analyses of WGDs from transcriptomes of multiple species

To further infer and locate paleopolyploidy in our data sets, we used our recently developed gene tree sorting and counting algorithm, the MAPS tool (Li *et al.*, 2015). The species trees for MAPS analyses were based on previously published phylogenies (Huang *et al.*, 2016). We circumscribed and constructed nuclear gene family phylogenies from multiple species for the MAPS analyses. For further information on the MAPS analyses see Methods S2.

### Gene Ontology annotations and paleolog retention and lost patterns

Gene Ontology annotations of all Brassicaceae transcriptomes and paleologs were obtained through discontiguous MegaBlast searches against annotated *A. thaliana* transcripts from TAIR

(Swarbreck *et al.*, 2008) to find the best hit with length of at least 100 bp and an e-value of at least 0.01. For each species, we assigned paralogs to the independent mesopolyploidies and At-$\alpha$ WGD based on the $K_s$ ranges identified in mixture model analyses (McLachlan *et al.*, 1999; Barker *et al.*, 2008). Further information on the GO analysis can be found in Methods S3.

## ACKNOWLEDGEMENTS

The work was supported by the Czech Science Foundation (grant 13-10159S) and the CEITEC 2020 (LQ1601) project to TM and MAL, and by NSF-IOS-1339156 and NSF-EF-1550838 to MSB and ZL. We thank J. Busch, A. Doust, P. B. Heenan, Ch. König and the Millenium Seed Bank, Royal Botanic Gardens, Kew for providing seeds of species used in this study. The authors declare no conflict of interest.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Relationship between species richness and percentage of neopolyploid species for 11 mesopolyploid crucifer taxa.

**Table S1.** Origin and collection data for species used in the present study.

**Table S2.** Assembly statistics for 13 transcriptomes used in this study.

**Table S3.** Sampling and summary statistics for Multi-tAxon Paleopolyploidy Search (MAPS) analyses.

**Data S1.** Patterns of chromosomal homoeology revealed by comparative chromosome painting in 10 mesopolyploid Brassicaceae species.

**Data S2.** Mesopolyploid Brassicaceae clades, diversification and colonization.

**Methods S1.** Estimating substitution rate heterogeneity.

**Methods S2.** Multi-tAxon Paleopolyploidy Search (MAPS) analysis.

**Methods S3.** Gene Ontology annotations.

## REFERENCES

**Adams, K.L. and Wendel, J.F.** (2005) Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **8**, 135–141.

**Al-Shehbaz, I.A.** (2012) A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon*, **61**, 931–954.

**Al-Shehbaz, I.A., German, D.A., Mummenhoff, K. and Moazzeni, H.** (2014) Systematics, tribal placements, and synopses of the *Malcolmia* s.l. segregates (Brassicaceae). *Harv. Pap. Bot.* **19**, 53–71.

**Arrigo, N. and Barker, M.S.** (2012) Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* **15**, 140–146.

**Barker, M.S., Kane, N.C., Matvienko, M., Kozik, A., Michelmore, R.W., Knapp, S.J. and Rieseberg, L.H.** (2008) Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol. Biol. Evol.* **25**, 2445–2455.

**Barker, M.S., Vogel, H. and Schranz, M.E.** (2009) Paleopolyploidy in the Brassicales: analyses of the Cleome transcriptome elucidate the history of genome duplications in Arabidopsis and other Brassicales. *Genome Biol Evol.* **1**, 391–399.

**Barker, M.S., Dlugosch, K.M., Dinh, L., Challa, R.S., Kane, N.C., King, M.G. and Rieseberg, L.H.** (2010) EvoPipes.net: bioinformatic tools for ecological and evolutionary genomics. *Evol Bioinform Online*, **6**, 143–149.

**Barker, MS, Baute, GJ and Liu, SL.** (2012) Duplications and turnover in plant genomes. In *Plant Genome Diversity, Volume 1* (Wendel, J.F., Greilhuber, J,, Dolezel, J. and Leitch, I.J., eds). Vienna: Springer, pp. 155–169.

**Barker, M.S., Arrigo, N., Baniaga, A.E., Li, Z. and Levin, D.A.** (2016a) On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* **210**, 391–398.

**Barker, M.S., Li, Z., Kidder, T.I., Reardon, C.R., Lai, Z., Oliveira, L., Scascitelli, M. and Rieseberg, L.H.** (2016b) Most compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. *Am. J. Bot.* **103**, 1203–1211.

**Bekaert, M., Edger, P.P., Pires, J.C. and Conant, G.C.** (2011) Two-phase resolution of polyploidy in the Arabidopsis metabolic network gives rise to relative and absolute dosage constraints. *Plant Cell*, **23**, 1719–1728.

**Birchler, J.A. and Veitia, R.A.** (2007) The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell*, **19**, 395–402.

**Birchler, J.A. and Veitia, R.A.** (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* **186**, 54–62.

**Birchler, J.A. and Veitia, R.A.** (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl Acad. Sci. USA*, **109**, 14746–14753.

**Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A.H.** (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**, 433–438.

**Buggs, R.J., Chamala, S., Wu, W., Tate, J.A., Schnable, P.S., Soltis, D.E., Soltis, P.S. and Barbazuk, W.B.** (2012) Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr. Biol.* **22**, 248–252.

**Burrell, A.M., Taylor, K.G., Williams, R.J., Cantrell, R.T., Menz, M.A. and Pepper, A.E.** (2011) A comparative genomic map for *Caulanthus amplexicaulis* and related species (Brassicaceae). *Mol. Ecol.* **20**, 784–798.

**Cai, J., Liu, X., Vanneste, K. *et al.*** (2015) The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* **47**, 65–72.

**Cannon, S.B., McKain, M.R., Harkess, A. *et al.*** (2015) Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol. Biol. Evol.* **32**, 193–210.

**Cenci, A., Combes, M.C. and Lashermes, P.** (2013) Differences in evolution rates among eudicotyledon species observed by analysis of protein divergence. *J. Hered.* **104**, 459–464.

**Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A.J., Müller, W.E.G., Wetter, T. and Suhai, S.** (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**, 1147–1159.

**Conant, G.C., Birchler, J.A. and Pires, J.C.** (2014) Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* **19**, 91–98.

**Cui, L., Wall, P.K., Leebens-Mack, J.H. *et al.*** (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res.* **16**, 738–749.

**Dlugosch, K.M., Lai, Z., Bonin, A., Hierro, J. and Rieseberg, L.H.** (2013) Allele identification for transcriptome-based population genomics in the invasive plant Centaurea solstitialis. *G3*, **3**, 359–367 http://paperpile.com/b/bWo3Lp/8ihcT.

**Dodsworth, S., Chase, M.W. and Leitch, A.R.** (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* **180**, 1–5.

**Douglas, G.M., Gos, G., Steige, K.A. *et al.*** (2015) Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc. Natl Acad. Sci. USA*, **112**, 2806–2811.

**Doyle, J.J. and Egan, A.N.** (2010) Dating the origins of polyploidy events. *New Phytol.*, **186**, 73–85.

**Edger, P.P. and Pires, J.C.** (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* **17**, 699–717.

**Edger, P.P., Heidel-Fischer, H.M., Bekaert, M. *et al.*** (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl Acad. Sci. USA*, **112**, 8362–8366.

**Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A. and Mummenhoff, K.** (2011) Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* **16**, 108–116.

**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.

**Freeling, M. and Thomas, B.C.** (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* **16**, 805–814.

**Freeling, M., Woodhouse, M.R., Subramaniam, S., Turco, G., Lisch, D. and Schnable, J.C.** (2012) Fractionation mutagenesis and similar

consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr. Opin. Plant Biol.* **15**, 131–139.

**Fuentes-Soriano, S. and Al-Shehbaz, I.** (2013) Phylogenetic relationships of mustards with multiaperturate pollen (Physarieae, Brassicaceae) based on the plastid *ndhF* gene: implications for morphological diversification. *Systematic Bot.* **38**, 178–191.

**Funk, V.A., Susanna, A., Stuessy, T.F. and Bayer, R.J., eds.** (2009) *Systematics, Evolution, and Biogeography of Compositae*. Vienna: International Association for Plant Taxonomy.

**Garsmeur, O., Schnable, J.C., Almeida, A., Jourda, C., D'Hont, A. and Freeling, M.** (2014) Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol.* **31**, 448–454.

**Gaut, B.S. and Doebley, J.F.** (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl Acad. Sci. USA*, **94**, 6809–6814.

**Geiser, C., Mandáková, T., Arrigo, N., Lysak, M.A. and Parisod, C.** (2016) Repeated whole-genome duplication, karyotype reshuffling and biased retention of stress-responding genes in Buckler Mustard. *Plant Cell*, **28**, 17–27.

**Goodstein, D.M., Shu, S., Howson, R. *et al.*** (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186.

**Graeber, K., Linkies, A., Steinbrecher, T. *et al.*** (2014) *DELAY OF GERMINATION 1* mediates a conserved coat-dormancy mechanism for the temperature- and gibberellin-dependent control of seed germination. *Proc. Natl Acad. Sci. U S A*, **111**, E3571–E3580.

**Gruenheit, N., Deusch, O., Esser, C., Becker, M., Voelckel, C. and Lockhart, P.** (2012) Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants. *BMC Genom.* **13**, 92.

**Grundt, H.H., Obermayer, R. and Borgen, L.** (2005) Ploidal levels in the arctic-alpine polyploid *Draba lactea* (Brassicaceae) and its low-ploid relatives. *Bot. J. Linn. Soc.* **147**, 333–347.

**Haudry, A., Platts, A.E., Vello, E. *et al.*** (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat. Genet.* **45**, 891–898.

**Heenan, P.B., Goeke, D.F., Houliston, G.J. and Lysak, M.A.** (2012) Phylogenetic analyses of ITS and *rbcL* DNA sequences for sixteen genera of Australian and New Zealand Brassicaceae result in the expansion of the tribe Microlepidieae. *Taxon*, **61**, 970–979.

**Hohmann, N., Wolf, E.M., Lysak, M.A. and Koch, M.A.** (2015) A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell*, **27**, 2770–2784.

**Huang, X. and Madan, A.** (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877.

**Huang, C.H., Sun, R., Hu, Y. *et al.*** (2016) Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol. Biol. Evol.* **33**, 394–412.

**Jaillon, O., Aury, J.M., Noel, B. *et al.*** (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

**Jiao, Y., Wickett, N.J., Ayyampalayam, S.**, et al. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**, 97–100.

**Jiao, Y., Li, J., Tang, H. and Paterson, A.H.** (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell*, **26**, 2792–2802.

**Jones, A.G.** (1985) Chromosomal features as generic criteria in the *Asteraceae*. *Taxon*, **34**, 44–54.

**Kagale, S., Robinson, S.J., Nixon, J. *et al.*** (2014a) Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell*, **26**, 2777–2791.

**Kagale, S., Koh, C., Nixon, J. *et al.*** (2014b) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat. Commun.* **5**, 3706.

**Koch, M., Huthmann, M. and Hurka, H.** (1998) Isozymes, speciation and evolution in the polyploid complex *Cochlearia* L. (Brassicaceae). *Botanica Acta*, **111**, 411–425.

**Kondrashov, F.A. and Kondrashov, A.S.** (2006) Role of selection in fixation of gene duplications. *J. Theor. Biol.* **239**, 141–151.

**Li, Z., Baniaga, A.E., Sessa, E.B., Scascitelli, M., Graham, S.W., Rieseberg, L.H. and Barker, M.S.** (2015) Early genome duplications in conifers and other seed plants. *Sci. Adv.* **1**, e1501084.

**Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y. and De Smet, R.** (2016) Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell*, **28**, 326–344.

**Lysak, M.A. and Mandáková, T.** (2013) Analysis of plant meiotic chromosomes by chromosome painting. *Methods Mol. Biol.* **990**, 13–24.

**Lysak, M.A., Koch, M.A., Pecinka, A. and Schubert, I.** (2005) Chromosome triplication found across the tribe Brassicaceae. *Genome Res.* **15**, 516–525.

**Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K. and Schubert, I.** (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl Acad. Sci. USA*, **103**, 5224–5229.

**Lysak, M.A., Cheung, K., Kitschke, M. and Bures, P.** (2007) Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiol.* **145**, 402–410.

**Lysak, M.A., Mandáková, T. and Schranz, M.E.** (2016) Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Curr. Opin. Plant Biol.* **30**, 108–115.

**Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. and Van de Peer, Y.** (2005) Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA*, **102**, 5454–5459.

**Mandáková, T. and Lysak, M.A.** (2008) Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (Brassicaceae). *Plant Cell*, **20**, 2559–2570.

**Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K. and Lysak, M.A.** (2010a) Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell*, **22**, 2277–2290.

**Mandáková, T., Heenan, P.B. and Lysak, M.A.** (2010b) Island species radiation and karyotypic stasis in *Pachycladon* allopolyploids. *BMC Evol. Biol.* **10**, 367.

**Mandáková, T., Mummenhoff, K., Al-Shehbaz, I.A., Mucina, L., Mühlhausen, A. and Lysak, M.A.** (2012) Whole-genome triplication and species radiation in the southern African tribe Heliophileae (Brassicaceae). *Taxon*, **61**, 989–1000.

**Mandáková, T., Gloss, A.D., Whiteman, N.K. and Lysak, M.A.** (2016) How diploidization turned a tetraploid into a pseudotriploid. *Am. J. Bot.* **103**, 1187–1196.

**McKain, M.R., Tang, H., McNeal, J.R., Ayyampalayam, S., Davis, J.I., dePamphilis, C.W., Givnish, T.J., Pires, J.C., Stevenson, D.W. and Leebens-Mack, J.H.** (2016) A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biol. Evol.* **8**, 1150–1164.

**McLachlan, G.J., Peel, D., Basford, K.E. and Adams, P.** (1999) The EMMIX software for the fitting of mixtures of normal and t-components. *J. Stat. Softw.* **4**, 1–14.

**deMiguel, M., Bartholomé, J., Ehrenmann, F. *et al.*** (2015) Evidence of intense chromosomal shuffling during conifer evolution. *Genome Biol. Evol.* **pii**: evv185.

**Murat, F., Zhang, R., Guizard, S., Flores, R., Armero, A., Pont, C., Steinbach, D., Quesneville, H., Cooke, R. and Salse, J.** (2014) Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16K protogenes. *Genome Biol Evol.* **6**, 12–33.

**Murat, F., Zhang, R., Guizard, S., Gavranovic′, H., Flores, R., Steinbach, D., Quesneville, H., Tannier, E. and Salse, J.** (2015) Karyotype and gene order evolution from reconstructed extinct ancestors highlight contrasts in genome plasticity of modern rosid crops. *Genome Biol. Evol.* **7**, 735–749.

**Parkin, I.A.P., Gulden, S.M., Sharpe, A.G., Lukens, L., Trick, M., Osborn, T.C. and Lydiate, D.J.** (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics*, **171**, 765–781.

**Quiros, C.F., Epperson, A., Hu, J. and Holle, M.** (1996) Physiological studies and determination of chromosome number in maca, *Lepidium meyenii* (Brassicaceae). *Econ. Bot.* **50**, 216–223.

**Renny-Byfield, S., Gong, L., Gallagher, J.P. and Wendel, J.F.** (2015) Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Mol. Biol. Evol.* **32**, 1063–1071.

**Rensing, S.A., Ick, J., Fawcett, J.A., Lang, D., Zimmer, A., Van de Peer, Y. and Reski, R.** (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol. Biol.* **7**, 130.

**Salariato, D.L., Zuloaga, F.O., Franzke, A., Mummenhoff, K. and Al-Shehbaz, I.A.** (2016) Diversification patterns in the CES clade (Brassicaceae tribes

Cremolobeae, Eudemeae, Schizopetaleae) in Andean South America. *Bot. J. Linn. Soc.* **181**, 543–566.

**Salse, J.** (2012) *In silico* archeogenomics unveils modern plant genome organisation, regulation and evolution. *Curr. Opin. Plant Biol.* **15**, 122–130.

**Sankoff, D. and Zheng, C.** (2012) Fractionation, rearrangement and subgenome dominance. *Bioinformatics*, **15**, i402–i408.

**Schnable, J.C., Wang, X., Pires, J.C. and Freeling, M.** (2012) Escape from preferential retention following repeated whole genome duplications in plants. *Front Plant Sci.* **3**, 94.

**Schranz, M.E., Mohammadin, S. and Edger, P.P.** (2012) Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin. Plant Biol.*, **15**, 147–153.

**Shi, T., Huang, H. and Barker, M.S.** (2010) Ancient genome duplications during the evolution of kiwifruit (*Actinidia*) and related Ericales. *Ann. Bot.* **106**, 497–504.

**Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J. and Shimizu, KK.** (2017) Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Mol. Ecol.* **26**, 193–207.

**Soltis, P.S. and Soltis, D.E.** (2016) Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **30**, 159–165.

**Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Depamphilis, C.W., Wall, P.K. and Soltis, P.S.** (2009) Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**, 336–348.

**Soltis, P.S., Marchant, D.B., Van de Peer, Y. and Soltis, D.E.** (2015) Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* **35**, 119–125.

**Soltis, D.E., Visger, C.J., Marchant, D.B. and Soltis, P.S.** (2016) Polyploidy: pitfalls and paths to a paradigm. *Am. J. Bot.* **103**, 1146–1166.

**Swarbreck, D., Wilks, C., Lamesch, P. *et al.*** (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* **36**, D1009–D1014.

**Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M. and Paterson, A.H.** (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954.

**Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B. and Harmon, L.J.** (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* **207**, 454–467.

**The Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

**Thomas, B.C., Pedersen, B. and Freeling, M.** (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* **16**, 934–946.

**Urban, I. and Bailey, C.D.** (2013) Phylogeny of *Leavenworthia* and *Selenia* (Brassicaceae). *Syst. Bot.* **38**, 723–736.

**Vanneste, K., Baele, G., Maere, S. and Van de Peer, Y.** (2014) Analysis of 41 plant genomes supports a wave of successful genome duplications at the Cretaceous-Tertiary boundary. *Genome Res.* **24**, 1334–1347.

**Vision, T.J., Brown, D.G. and Tanksley, S.D.** (2000) The origins of genomic duplications in *Arabidopsis*. *Science*, **290**, 2114–2117.

**Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I. and Cheng, F.** (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**, 1035–1039.

**Warwick, S.I. and Al-Shehbaz, I.A.** (2006) Brassicaceae: chromosome number index and database on CD-Rom. *Plant Syst. Evol.* **259**, 237–248.

**Wendel, J.F.** (2015) The wondrous cycles of polyploidy in plants. *Am. J. Bot.* **102**, 1753–1756.

**Wolfe, K.H.** (2001) Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* **2**, 333–341.

**Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B. and Rieseberg, L.H.** (2009) The frequency of polyploid speciation in vascular plants. *Proc. Natl Acad. Sci. USA*, **106**, 13875–13879.

**Woodhouse, M.R., Cheng, F., Pires, J.C., Lisch, D., Freeling, M. and Wang, X.** (2014) Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc. Natl Acad. Sci. USA*, **111**, 5283–5288.

**Xie, Y., Wu, G., Tang, J. *et al.*** (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, **30**, 1660–1666.

**Yang, Z.** (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.

**Zhang, J., Tian, Y., Yan, L. *et al.*** (2016) Genome of plant maca (*Lepidium meyenii*) illuminates genomic basis for high altitude adaptation in the central Andes. *Mol Plant*, **9**, pii: S1674-2052(16)30053-3.

 