Genome duplication and reorganization in *Aquilegia*

Gökçe Aköz[1,2] and Magnus Nordborg[1,*]
[1]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna Biocenter, Vienna, Austria
[2]Vienna Graduate School of Population Genetics, Vienna, Austria
[*]magnus.nordborg@gmi.oeaw.ac.at

Abstract

Background: Whole-genome duplications (WGD) have dominated the evolutionary history of plants. One consequence of WGD is a dramatic restructuring of the genome as it undergoes diploidization, a process under which rearrangements of various sizes scramble the genetic material, leading to a repacking of the genome. Here, we investigate the WGD history in the columbine genus *Aquilegia*, a basal eudicot, and trace the origins of its chromosomes, in particular the unique "nucleolus" chromosome 4 [1].

Results: Our results support the notion that basal eudicots share an ancient tetraploidy with core eudicots. Comparison to the grape genome reveals that the columbine genome has diverged relatively little from the ancestral tetraploid. As few as seven fusions and a single fission can explain the present-day karyotype of columbines. Unlike other columbine chromosomes of similar size, all of which are products of ancient fusion events, chromosome 4 appears to have reached its size due to an expansion of repetitive DNA, partly due to numerous segmental duplications.

Conclusions: The columbine genome provides an important insight into the karyotype evolution of basal eudicots, and also presents an unusual example of a chromosome that appears to decay almost as if it were a sex chromosome. More genomes from basal eudicots will be needed to understand eudicot karyotype evolution and the origins of this chromosome.

Background

Whole-genome duplication (WGD) is common in the evolutionary history of plants [reviewed in 2,3]. All flowering plants are descended from a polyploid ancestor, which in turn shows evidence of an even older WGD event shared by all seed plants [4]. These repeated cycles of polyploidy dramatically restructure plant genomes. Presumably driven by the "diploidization" process, whereby genomes are returned to an effectively diploid state, chromosomes are scrambled via fusions and fissions, lose both repetitive and genic sequences, or are lost entirely [5–12]. Intriguingly, gene loss after WGD is non-random: not only is there a bias against the retention of certain genes [13,14], but also against the retention of one of the WGD-derived paralog chromosomes [7,10,15–17].

We investigated the history of WGDs in the columbine genus *Aquilegia* for two reasons. The first is related to its phylogenetic position: as a basal eudicot, columbines are members of the very earliest diverging branch of the eudicots [18,19]. This matters because our understanding of eudicot karyotype evolution is limited to the heavily sampled core eudicots. Using the recently published *Aquilegia coerulea* genome [1], we addressed the question of whether all eudicots share an ancient tetraploidy. We also traced the origins of the columbine chromosomes with a particular focus on chromosome 4, which, compared to the rest of the genome: harbors more genetic polymorphism, has a higher transposable element density, has a lower gene density and reduced gene expression, shows less population structure worldwide, appears more permeable to gene flow, and carries the rDNA clusters [1]. Our findings uncovered additional features unique to chromosome 4, consistent with this chromosome having a distinct evolutionary history.

Results

Within genome synteny confirms columbine paleotetraploidy

Ancient WGDs have been commonly inferred from the distribution of divergences between gene duplicates. The simultaneous generation of gene duplicates via WGD is expected to produce a peak in the age distribution relative to the background age distribution of single gene duplicates [20–22]. Such a spike of ancient gene birth was the first evidence of paleotetraploidy in columbines [23], and was later supported by gene count-based modelling [24].

Given an assembled genome, a more direct method to infer ancient polyploidy events is to look for regions with conserved gene order [25,26]. Such conservation (a.k.a., synteny) decreases over time due to gene loss and rearrangements, but will still be detectable unless the extent of change is extreme. We detected a total of 121 synteny block pairs harboring at least five paralogous gene pairs within the columbine genome. The distribution of these blocks across the seven columbine chromosomes indicates pairwise synteny between large genomic regions (Fig. 1). This 1:1 relationship suggests a single round of WGD in columbine, and is further supported by similar levels of divergence between synteny pairs (Figs. S1 and S2).

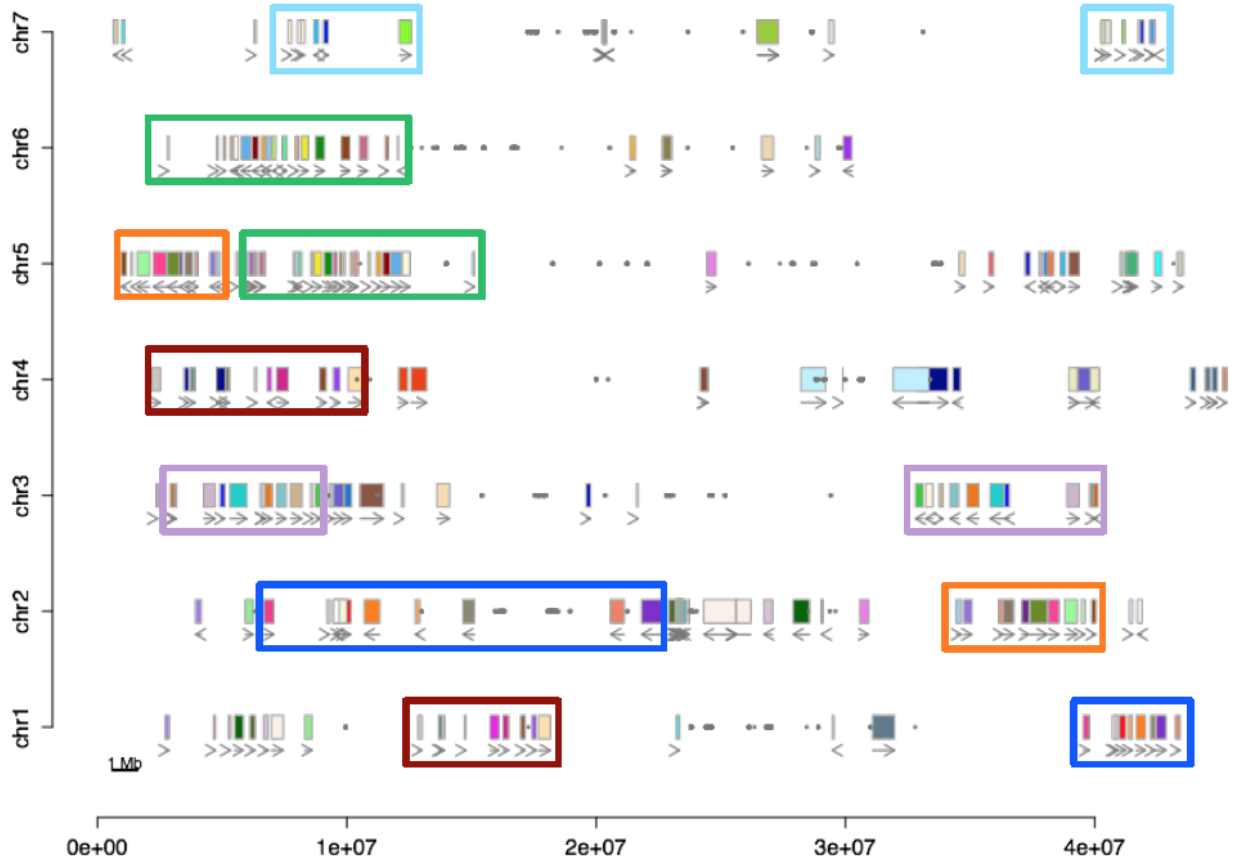

Fig. 1: Intragenomic synteny blocks in the columbine genome. Pairs of synteny blocks are denoted as uniquely colored small rectangles. Larger rectangles of the same color outline large regions of synteny. Arrows under the synteny blocks show the orientation of the alignment between collinear genes. Grey dots highlight BLAST hits of a 329 bp centromeric repeat monomer [1,27].

Columbines share ancient tetraploidy with core eudicots

All sequenced core eudicots appear to share a triplicate genome structure due to paleohexaploidization postdating the separation of monocots and eudicots [10,28–32, and Supplementary Note 5 in 33]. The tetraploidy in columbines, a basal eudicot, might be independent of this ancient "*gamma*" hexaploidy (Scenarios 1 and 2 in Fig. 2) or might be a remnant of a WGD at the base of all eudicots, which formed the first step of the *gamma*-hexaploidy in core eudicots (Scenario 3 in Fig. 2).

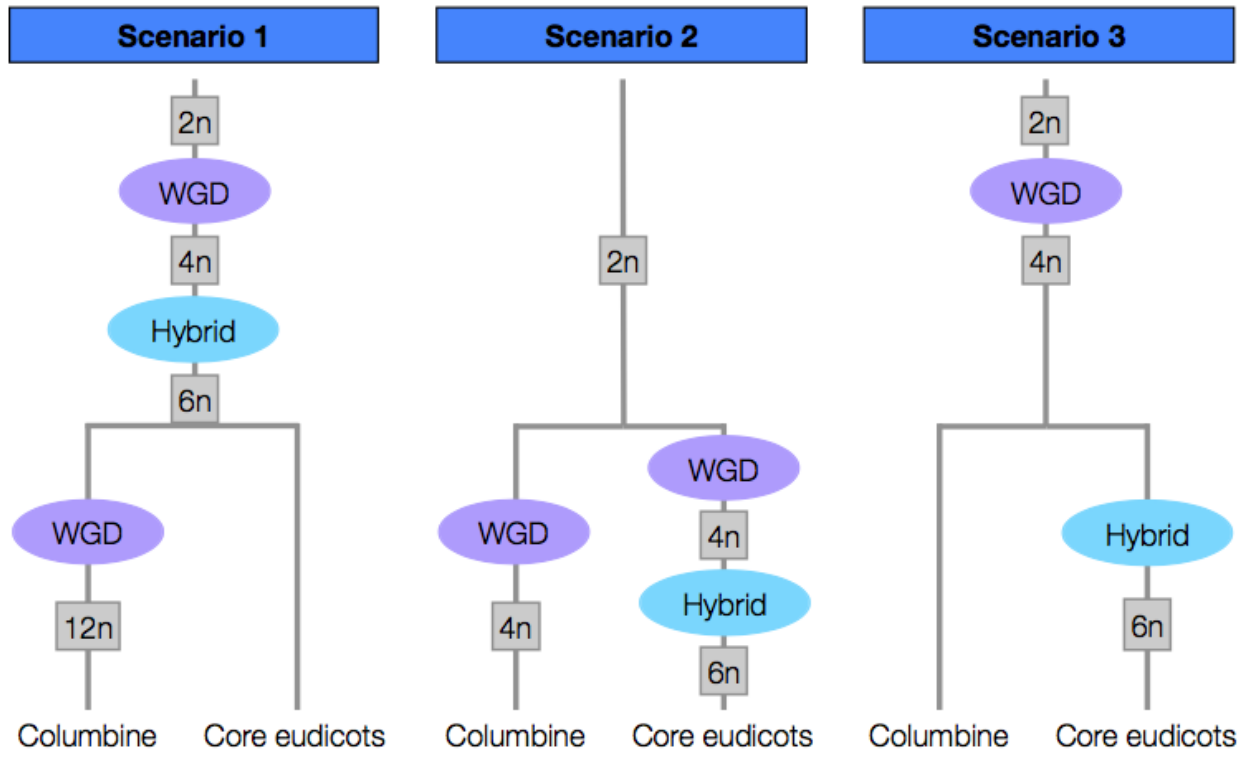

Fig. 2: Three scenarios for the relationship between columbine tetraploidy and core eudicot "*gamma*" hexaploidy. The *gamma* hexaploidy is a two-step process: a single round of WGD creates tetraploids (4n) which then cross with diploids (Hybrid). **Scenario 1**: *gamma* hexaploidization precedes the split between columbine and core eudicots, with the former undergoing an additional tetraploidization. **Scenario 2**: Both *gamma* hexaploidization and columbine tetraploidization occur after the split between columbines and core eudicots. **Scenario 3**: Columbine tetraploidy is derived from the ancient tetraploidization that was the first step of the process leading to *gamma* hexaploidy (presumably a hybridization with diploid species in the lineage ancestral to all core eudicots).

We used the grape (*Vitis vinifera*) genome as a representative of the core eudicots to distinguish between the three scenarios in Fig. 2. Grape has experienced the least number of chromosomal rearrangements post-*gamma* and thus strongly resembles the ancestral pre-hexaploid genome [34]. Given the ploidy level of columbine under each scenario, we can predict the synteny relationship between the homologous chromosomes of grape and columbine, which is simply the ratio of haploid chromosome set in grape to that of in columbine. If tetraploidy in columbines is lineage-specific and superimposed on the *gamma*-hexaploidy (Scenario 1), we would expect to find a 3:6 ratio of grape and columbine synteny blocks. Instead, we observe a 3:2 relationship (Figs. 3 and S3) as expected under Scenarios 2 or 3. A similar 3:2 pattern is found in

comparisons between grape and sacred lotus [35]. This strongly suggests that basal eudicots do not share the triplicate genome structure of core eudicots, ruling out Scenario 1.

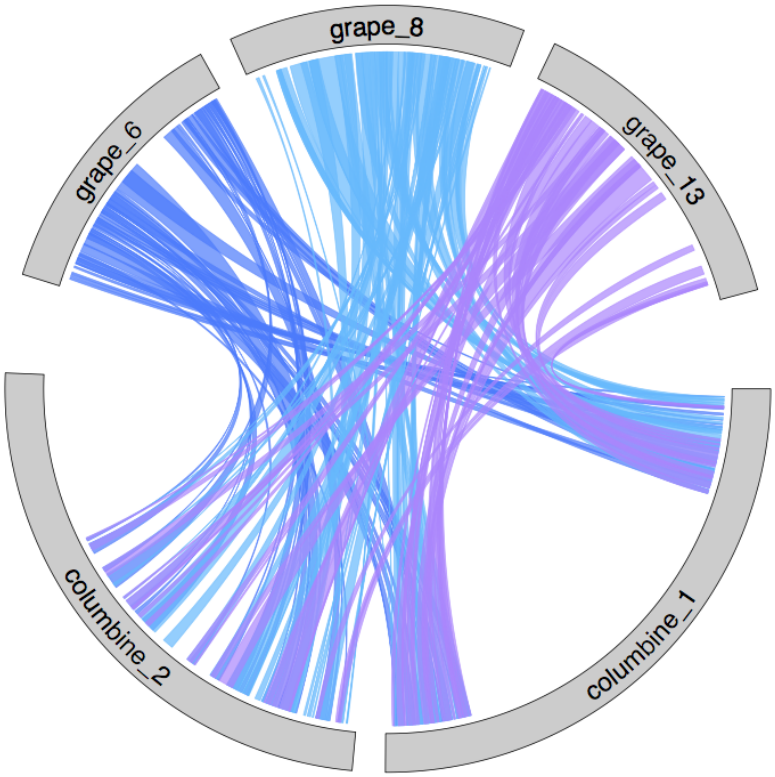

Fig. 3: Synteny between the homologous regions of columbine and grape. The results are shown here only for columbine chromosomes 1, 2 and grape chromosomes 6, 8 and 13 but reflect the overall synteny relationship of 3:2 between grape:columbine chromosomes (see Fig. S3 for the genome-wide synteny). This pattern argues against Scenario 1, but is consistent with either Scenario 2 or Scenario 3.

To distinguish between the two remaining scenarios, we compared the divergence at synonymous sites (Ks) between columbine paralogs, grape paralogs and columbine-grape homologs. In agreement with the analysis of Jiao et al. [36], the Ks distribution for grape paralogs shows two major peaks, as expected under the two-step model for *gamma* hexaploidy (Fig. 4). However, columbine paralogs and columbine-grape homologs each show a single peak of divergence — and the peaks overlap each other and the "older" divergence peak of grape paralogs. This suggests that columbine tetraploidy is derived from the tetraploidization that eventually led to *gamma* hexaploidy in core eudicots (Scenario 3).

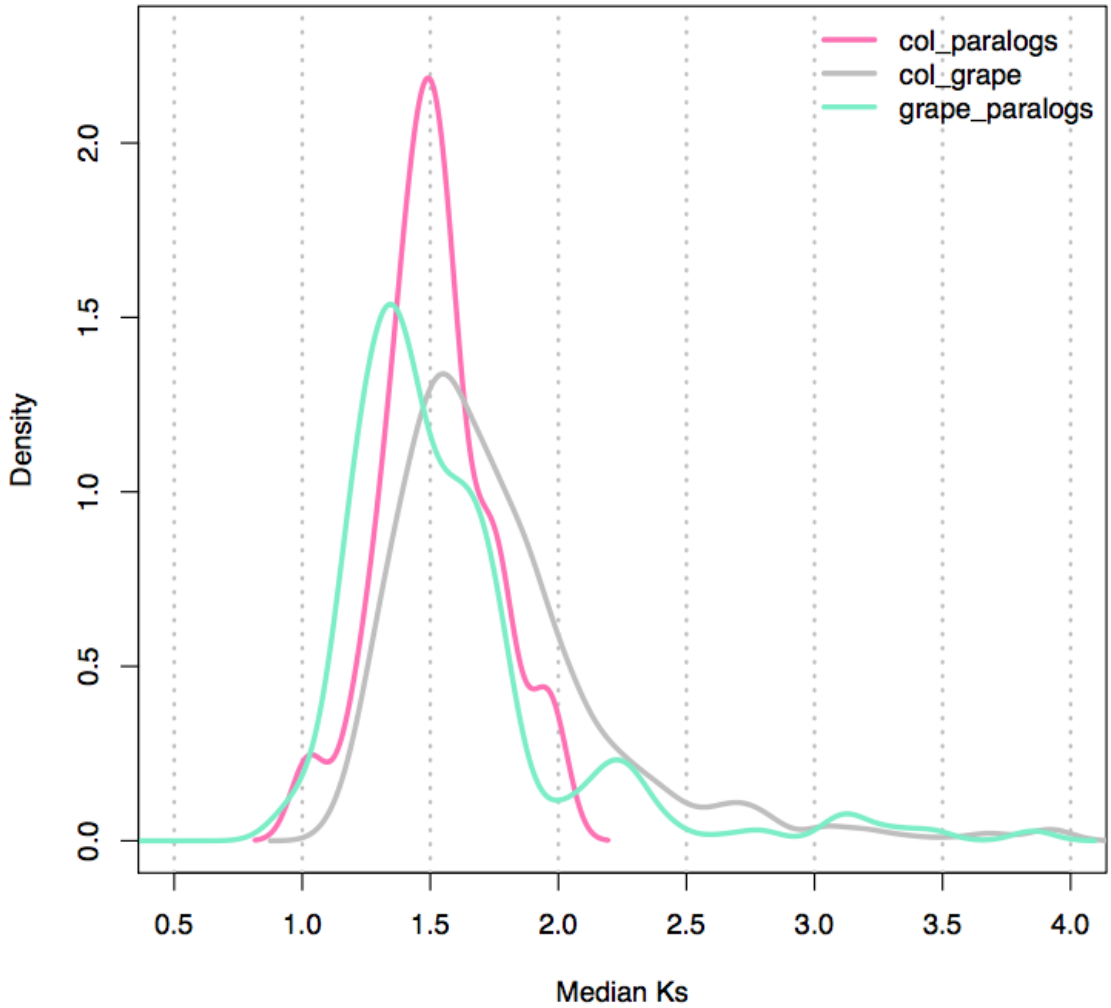

Fig. 4: The distribution of the median Ks across syntenic regions. Synteny blocks are identified within columbine (col_paralogs), between columbine and grape (col_grape) and within grape (grape_paralogs). Note that only the putative WGD-derived blocks (median Ks=1-2) are kept in columbine (Fig. S2).

To further explore the hypothesis of a shared WGD by all eudicots, we focused on the gene order similarity between the homologous regions of columbine and grape. If columbine and grape have descended from a common tetraploid ancestor, they should have inherited diploidization-driven differential gene order on the paralogous chromosomes of the ancestor (Fig. S4). As a result, we expect to see the alternative paralogous gene orders to be uniquely shared between two different pairs of columbine and grape chromosomes. To detect this, we first searched for at least three consecutive genes aligning between a pair of columbine and grape chromosomes and then looked at the distribution of these genes on all the columbine and grape chromosomes. This way of reconstructing chromosomes clearly shows that each of the paralogous chromosome pairs in columbine has a match to a single grape chromosome, with respect to its gene

order (Fig. S5). This result was corroborated by a second approach where we quantify the similarity between columbine and grape chromosomes. When we performed a pairwise alignment between each sliding window of genes on a columbine chromosome and all the genes on a grape homolog, we again see that each member of columbine paralogs gets the best hit to a single grape chromosome (Figs. 5 and S6-8). Reshuffling genes on grape chromosomes further indicates that this pattern of clustering is highly unlikely to be produced by chance alone (p=0-0.05).

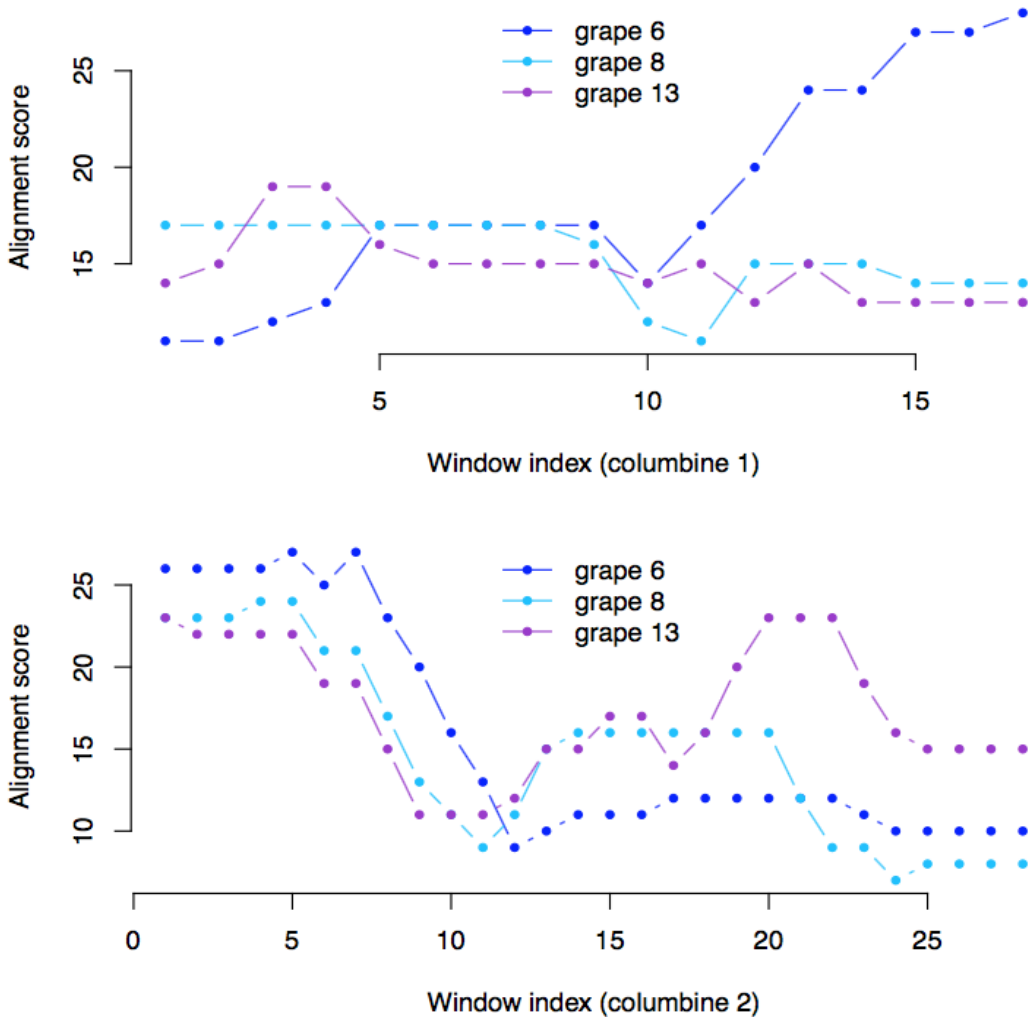

Fig. 5: Examples of gene order similarity between the homologous regions of columbine and grape. For successive windows of genes within a given columbine chromosomal region, the best alignment score with respect to each of the three grape chromosomes harboring homologous regions, is given. For example, columbine chromosomes 1 and 2 share a paralogous region homologous to grape chromosomes 6, 8, and 13 (Figs. S3 and S9). The chromosome 1 region **(top panel)** appears to be most closely related to grape chromosome 6, whereas its paralogous counterpart on chromosome 2 **(bottom panel)** appears to be most closely related to grape

chromosome 13. See Fig. S5 for the correspondence between gene orders and scores, which peak towards the end of each columbine region. Note that the results presented here are shown for a window size of 12 genes but remain significant for all the window sizes tested (p=0-0.05).

A eudicot-wide WGD is further supported by the observation that a chromosomal fusion, presumably experienced by the common tetraploid ancestor, is still detectable in the genomes of columbine and grape despite their separation of around 125 million years [37]. The first hint comes from the composition of the chromosomes: columbine chromosome 5 and grape chromosome 7 share the two chromosomal origins (Fig. S9). If these fused chromosomes were created by a single fusion event in the common tetraploid ancestor of eudicots, they should match each other with respect to gene order on each of the two homologous portions ("orange" and "green" portions in Fig. 6). This is what we see: columbine chromosome 5 and grape chromosome 7 cluster together with respect to their gene order on the "orange" portion (Fig. S6). For the "green" portion, columbine chromosome 5 matches grape chromosome 4 (Fig. S7), which used to be fused to grape chromosome 7 [38]. Additional support for this ancestral fusion event comes from cacao chromosome 1, which a similar pattern of chromosomal ancestry [39].

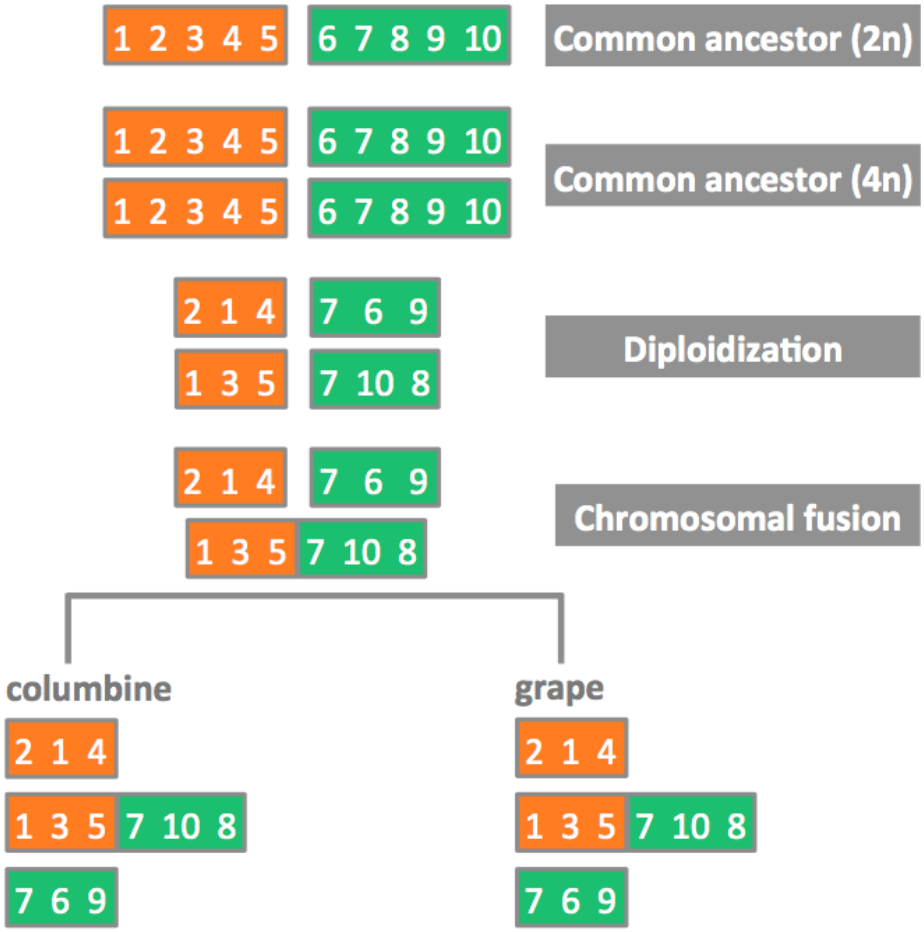

Fig. 6: Schematic of predicted synteny patterns in the case of shared ancestral fusion. Two ancestral chromosomes (orange and green rectangles, with genes depicted as numbers) undergo WGD. Paralogous chromosome pairs diverge as a part of the diploidization process. A fusion joins one version of the "orange" chromosome ('1, 3, 5') with one version of the "green" chromosome ('7, 10, 8'). If this event took place in the common tetraploid ancestor of eudicots, the fused chromosomes in columbine and grape should also carry these versions on their "orange" and "green" portions. In the hypothetical example here, diploidization precedes the fusion event but may well happen afterwards with no effect on the predicted synteny patterns.

Current columbine chromosomes have mostly been generated via fusions

It is widely accepted that genome shuffling post-WGD has shaped the present-day karyotypes of all plant genomes [34]. Nevertheless, the extent of genome shuffling as a part of the "re-diploidization" process seems to vary widely: only 3 chromosomal rearrangements post-*gamma* are enough to explain the current structure of the grape genome (Fig. 7) while almost 150 chromosomal rearrangements were necessary for the sunflower genome to reach its current karyotype after several rounds of WGD [12]. To

check where columbine falls in this spectrum, we identified chromosomal rearrangements likely to have happened after the tetraploidy shared by all eudicots: if the pre-WGD ancestral eudicot karyotype had a haploid number of 7 chromosomes [28], only seven columbine-specific fusions and a single fission are enough to explain the reduction in columbine chromosome number from n=13 to n=7 after the ancestral fusion event (Fig. 7). These rearrangements involve all the chromosomes in columbine apart from chromosomes 4 and 6, the former of which paradoxically shows the greatest erosion of synteny with grape chromosomes (Fig. S9). Given all the evidence suggesting a “decaying” nature of columbine chromosome 4 [1], we repeated the analysis of grape-columbine synteny detection with relaxed parameter settings. We did this by decreasing the minimum number of aligned gene pairs within a block (from 5 to 3) and increasing the maximum genic distance between matches (from 20 to 30). This allowed us to extend the synteny blocks towards more proximal regions (Fig. S10). Further zooming into the synteny relationship between grape chromosomes that are homologous to columbine chromosome 4 confirmed that there is no evidence of a fusion event (Fig. 8).

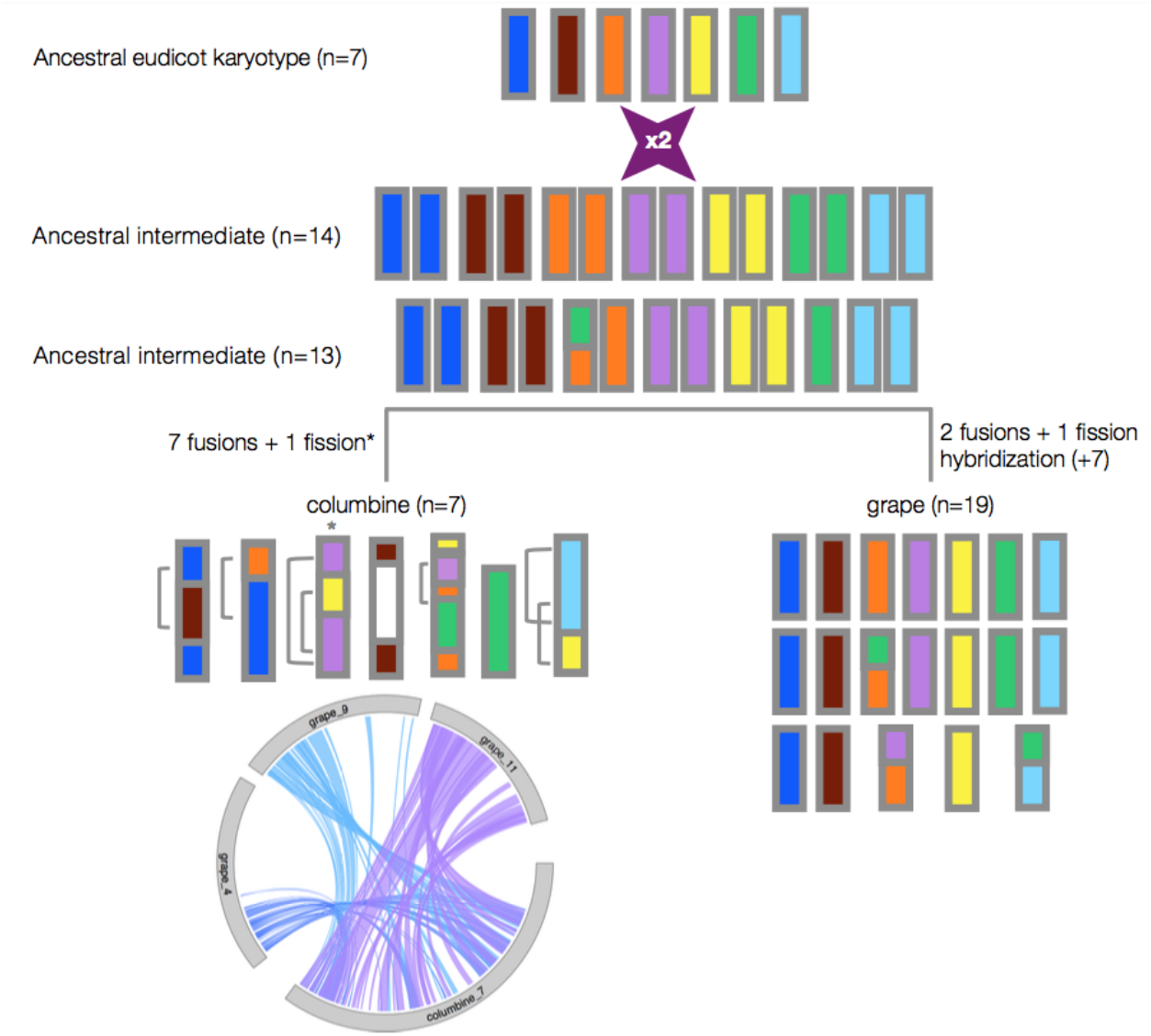

Fig. 7: Tracing the genome reshuffling events in the columbine genome following tetraploidization. If all eudicots share the same ancestral karyotype (n=7), only 7 fusions (brackets) and a single fission (*) are enough to explain the current structure of columbine genome. Of these 7 fusions, 5 of them are between different chromosomes while 2 of them are between WGD-derived paralogous chromosomes. Columbine chromosomes 3 and 7 are examples of the latter case. The circos plot zooms into the synteny between columbine chromosome 7 and paralogous grape chromosomes 4, 9 and 11 to show that both ends of columbine chromosome 7 align to the same region of a given grape chromosome as expected from the fusion of paralogous chromosomes.

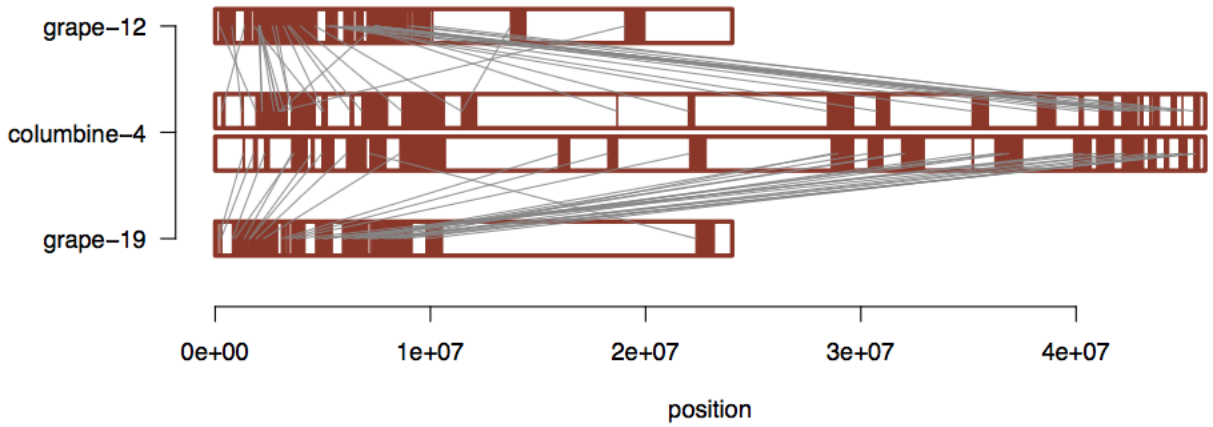

Fig. 8: Synteny between columbine chromosome 4 and grape chromosomes 12 and 19. Much smaller grape chromosomes look like the compact versions of columbine chromosome 4. Note that this result is generated with the most relaxed parameter combination in Fig. S10, but holds true for a less relaxed combination of parameters as well (Fig. S11).

The lack of a fusion event on columbine chromosome 6 might explain the fact that it is the smallest chromosome of columbine (Figs. 7 and S9). However, chromosome 4 is comparable in size to the remaining chromosomes, all of which are products of ancient fusion events. The observations that chromosome 4 has a higher proportion of genes in tandem duplicates (0.37 versus genome-wide mean of 0.22) and a greater extent of intra-chromosomal synteny (indicative of segmental duplications) (Fig. S12) suggest that chromosome 4 has reached a comparable size partly due to numerous tandem and segmental duplications and partly due to an expansion of repetitive DNA [1]. These results reinforce the idea that chromosome 4 has followed a distinct evolutionary path from the rest of genome.

Dosage balance hypothesis explains biased gene retention in columbine

The loss of duplicated genes after polyploidy is reported to be nonrandom for several organisms, including: yeast [40]; *Arabidopsis* [13]; rice [41]; paramecium [42]; maize [43] and poplar [44]. Genes involved in connected molecular functions like kinases, transcription factors and ribosomal proteins are dosage-sensitive [45] and losing some, but not all of them following WGD might upset the stoichiometric relationship between these proteins [46–48]. The “dosage balance hypothesis” predicts that these “connected” genes will be retained post-WGD while the very same genes will be removed after tandem or any small scale duplications by dosage-related purifying selection.

To test this hypothesis, we focused on a total number of 1302 genes across 76 syntenic regions (with median Ks values between 1-2) and applied Fisher's exact test to check whether genes potentially retained post-WGD (Supplementary Data 1) are enriched for any Gene Ontology (GO) category. Consistent with the dosage balance hypothesis, we found that these genes are enriched for the GO categories "structural constituent of ribosome", "transcription factor activity", "translation" ($p < 0.001$) and "protein tyrosine kinase activity" ($p < 0.01$). Tandemly duplicated genes (n=6972), on the other hand, are depleted for the GO categories "structural constituent of ribosome", and "translation" ($p=10^{-17}$), consistent with the dosage balance hypothesis.

Discussion

The evolutionary history of plants abounds with shared and independent WGD events and columbine is no exception. The alignment between chromosomal regions in a 1:1 ratio suggests a single round of WGD in columbine (Fig. 1), confirming the previous findings from Vanneste et al. [23] and Tiley et al. [24]. This tetraploidy signal, however, seems to have arisen from a WGD event shared by all eudicots, which contradicts the argument of a lineage-specific polyploidy in columbine [23], and supports prior studies hinting at a eudicot-wide WGD [26,35,36,49].

Free from recent WGD events that erase the signal of more ancient ones [26], columbines have a relatively well-preserved genome structure. Its karyotype is derived from a tetraploid eudicot genome (n=14) ~ 135 million years ago (mya) [36], which experienced both intra- and interchromosomal reshuffling, potentially driven by diploidization. The tetraploid genome was scrambled even further with fusion and fission events from about 125 mya onwards [37], generating the present-day karyotype of columbine (Fig. 7). While it is not clear to what extent these additional rearrangements reflect the "decay" of the tetraploid genome as it becomes diploid [50], it is clear that diploidization has acted nonrandomly at the gene level, favoring the retention of dosage-sensitive genes.

Falling in between the grape and the cacao genomes with respect to the spectrum of karyotype shuffling post-WGD, the columbine genome has experienced substantially more chromosomal fusions than chromosomal fissions (Figs. 7 and S9). One of these fusion events shows an interesting overlap with *Populus* and *Eucalyptus* [51]. The pre-polyploid ancestral chromosome fused with its WGD-derived paralog, independently generating chromosome 7 of *Aquilegia*, chromosome 18 of *Populus* and chromosome 3 of *Eucalyptus*. Another interesting pattern appears on a *fusion-resistant* chromosome of columbine: chromosome 4 has a similar assembly size as other columbine chromosomes, but this is due to duplications and repetitive elements. These findings raise several interesting questions: How random are chromosomal fusions? Can we learn more about genome evolution by looking at chromosomal rearrangements shared

between distantly related species? Do any other plant species carry a chromosome like that of columbine chromosome 4? Could chromosome 4 be an ancient sex chromosome? Can we trace back a more recent history of chromosome 4 by looking at more closely related species? All these questions call for additional genomes, especially from the basal eudicots. This will extend the scale of comparative genomics for a deeper understanding of genome evolution in eudicots, particularly in columbine.

Conclusions

The lack of independent "wash-rinse-repeat" cycles of polyploidization [2] in columbine and grape helped us trace the history of the *gamma* hexaploidy. This event happened in two steps: a single round of WGD shared by all eudicots was followed by hybridization in the lineage ancestral to all core eudicots. Having diverged relatively little from the tetraploid ancestor, columbine provides an initial insight into the karyotype evolution of basal eudicots. The peculiar chromosome 4 makes this genus all the more fascinating; it has the features of a sex chromosome in a system with no sex!

Materials and Methods

Synteny detection

We performed all genes (CDS)-against-all genes (CDS) BLAST for the latest version of *Aquilegia coerulea* reference genome (v3.1) using SynMap tool [29] in the online CoGe portal [52]. We also looked at the synteny within *Vitis vinifera* (v12) and between *V. vinifera* and *A. coerulea* using both default and more relaxed parameter combinations in DAGChainer. We filtered the raw output files for both within grape and grape-to-columbine synteny. For the former, we only kept the blocks that are syntenic between the polyploidy-derived paralogous chromosomes of grape as identified by Jaillon et al. [28]. For the latter, we required that a given columbine chromosome is overall syntenic to all the three paralogous chromosomes of grape. So, for a given pair of columbine and grape chromosomes, we only kept the blocks if the columbine chromosome also matches to the other members of paralogous grape chromosomes.

The raw output files can be regenerated from the following links with parameter combinations (D:A) specifying the maximum genic distance between two matches (D) and the minimum number of aligned gene pairs (A) to form a collinear syntenic block:

- **D:A=20:5 (default combination)**

A. coerulea - A. coerulea: https://genomevolution.org/r/s3cb.
V. vinifera - V. vinifera: https://genomevolution.org/r/k80c.
V. vinifera -A. coerulea: https://genomevolution.org/r/pxyq.

342 - **D:A=20:3**
343 *A. coerulea - A. coerulea*: https://genomevolution.org/r/s5aw.
344 *V. vinifera - V. vinifera*: https://genomevolution.org/r/s7hq.
345 *V. vinifera -A. coerulea*: https://genomevolution.org/r/s2g3.
346 - **D:A=30:3**
347 *A. coerulea - A. coerulea*: https://genomevolution.org/r/slco.
348 *V. vinifera - V. vinifera*:https://genomevolution.org/r/s7hs.
349 *V. vinifera -A. coerulea*: https://genomevolution.org/r/s2gc.
350 - **D:A=0:1 (gene-to-gene blast)**
351 *A. coerulea - A. coerulea*: https://genomevolution.org/r/12d2s.
352 *V. vinifera - V. vinifera*: https://genomevolution.org/r/12d14.
353 - **D:A=0:3**
354 *V. vinifera -A. coerulea*: https://genomevolution.org/r/12efq.
355 - **D:A=0:4**
356 *V. vinifera -A. coerulea*: https://genomevolution.org/r/12efw.

357 **Estimating the divergence between synteny block pairs**

358 We used Ks (the number of synonymous substitutions per synonymous site) values
359 provided for each duplicate gene pair by the CoGe portal [52]. We estimated the median
360 Ks of all gene duplicates in a syntenic block after filtering duplicates with Ks>10 due to
361 saturation effect [53].

362 **Quantifying gene order similarity**

363 We first detected three consecutive genes aligning between a pair of columbine and
364 grape chromosomes harboring homologous regions (D:A=0:3). We particularly chose
365 three genes since it is the most stringent value we could use to detect homologous
366 synteny blocks; we detected almost nothing when we required 4 consecutive genes
367 (D:A=0:4). We then looked at the distribution of these genes on a given pair of
368 columbine and grape chromosomes and also on their paralogous counterparts. Once we
369 had the gene order for each chromosome, we assigned a unique word to each synteny
370 block and the genes forming the block to be able to use the text alignment provided by
371 the R package *align_local* [54]. Having each chromosome represented by a sentence, we
372 quantified the gene ("word") similarity as such: for an initial N number of words on a
373 columbine chromosome (N=window size), we did a pairwise alignment between these N
374 words and a grape chromosome (match=4, gap=-1). We repeated the same analysis
375 with the inverted order of N words and picked the maximum alignment score. We
376 repeated these steps by sliding the window by one word and keeping the N constant to
377 get a distribution of scores as in Fig. 5. We used different N values ranging from 4 to 15
378 (3 to 6 for columbine chromosome 7). Note that we excluded columbine chromosomes 3
379 and 4 from this analysis since both have a complex history of lineage-specific

380 chromosomal reshuffling events: fusions and a fission gave rise to columbine
381 chromosome 3 while duplications have shaped the current structure of columbine
382 chromosomes 4.

383 **Statistical testing of gene order similarity**

384 Given the gene order similarity between the two different pairs of columbine and grape
385 chromosomes harboring homologous regions, we performed permutation tests to
386 estimate the probability of observing such a clustering just by chance. To do so, we first
387 combined all the grape genes and sampled the same number of genes ("words") as we
388 observe to reconstruct each of the paralogous grape chromosome. We repeated the
389 quantification step as above to get a permuted distribution of alignment score between a
390 pair of columbine and grape chromosomes. We used Wilcoxon rank sum test (W-
391 statistic) to quantify the shift in the distribution of alignment scores between one of the
392 members of columbine paralogous chromosomes and its best grape hit when combined
393 with the alignment scores between the same columbine chromosome and other grape
394 chromosomes. We repeated the same analysis for the other member of columbine
395 paralogous chromosomes as well. Having these *observed* W-statistics, we counted the
396 number of cases (out of 100) where the permuted distributions generate W-statistics as
397 high as or higher than the observed ones. Note that for columbine chromosome 7, whose
398 structure has been greatly shaped by the fusion of WGD-derived paralog chromosomes
399 ("blue" parts in Fig. S3), we created two paralogous chromosomes using the observed
400 distribution of alignment scores (Fig. S8). Columbine chromosome 7 matches best to
401 grape chromosome 11 for the first 14 "words" and to grape chromosome 4 for most of
402 the remaining "words", which define the putative boundaries of columbine paralogous
403 chromosomes before the fusion event.

404 **GO enrichment analysis**

405 We used gene annotations provided by JGI [1] to test the null hypothesis that the
406 property for a gene to be retained post-WGD and to belong to a given GO category are
407 independent. We created a 2x2 contingency table as shown below and applied Fisher's
408 exact test for each GO category independently. We repeated the same analysis for
409 tandem gene duplicates as identified by SynMap [29,52]; this time testing the null
410 hypothesis that the property for a gene to be tandemly duplicated and to belong to a
411 given GO category are independent. We excluded genes on scaffolds and reported
412 enriched/depleted categories if they remain significant ($p < 0.05$) after multiple test
413 correction (fdr).
414
415 **Table 1:** 2x2 contingency table obtained by classifying genes into 2 categorical
416 variables. The letters denote the number of genes for a given category (e.g. "a" denotes
417 the number of retained genes annotated with the tested GO category).

	GO	not-GO	SUM
retained	a	b	a+b*
not-retained	c	d	c+d
SUM	a+c	b+d	N=total number of genes =29550 (across 7 chromosomes)

*equal to 1302 and 6972 for candidate WGD-derived paralogs and tandem gene duplicates, respectively.

List of abbreviations

WGD: whole genome duplication; Ks: the number of synonymous substitutions per synonymous site; GO: Gene Ontology, mya: million years ago.

Availability of data and material

The columbine and grape genomes are available at the CoGE portal for the synteny analyses with the id numbers 28620 and 19990, respectively [52].

Competing interests

The authors declare no competing interests.

Funding

G.A. was supported by the Vienna Graduate School of Population Genetics (Austrian Science Fund, FWF: DK W1225-B20).

Authors' contributions

G.A. performed all the analyses. G.A. and M.N. wrote the manuscript.

Acknowledgements

We thank Robin Burns and Claus Vogl for their comments on the manuscript; Daniel Gómez-Sánchez and Benjamin Jaegle for the fruitful discussions.

References

1. Filiault D, Ballerini E, Mandakova T, Akoz G, Derieg N. The Aquilegia genome: adaptive radiation and an extraordinarily polymorphic chromosome with a unique history. bioRxiv [Internet]. biorxiv.org; 2018; Available from: https://www.biorxiv.org/content/early/2018/02/12/264101.abstract

450 2. Wendel JF. The wondrous cycles of polyploidy in plants. Am J Bot. 2015;102:1753–6.

451 3. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy.
452 Nat Rev Genet. 2017;18:411–24.

453 4. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al.
454 Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473:97–100.

455 5. Leitch IJ, Bennett MD. Genome downsizing in polyploid plants. Biol J Linn Soc Lond.
456 Oxford University Press; 2004;82:651–63.

457 6. Comai L. The advantages and disadvantages of being polyploid. Nat Rev Genet.
458 2005;6:836–46.

459 7. Thomas BC, Pedersen B, Freeling M. Following tetraploidy in an Arabidopsis
460 ancestor, genes were removed preferentially from one homeolog leaving clusters
461 enriched in dose-sensitive genes. Genome Res. 2006;16:934–46.

462 8. Otto SP. The evolutionary consequences of polyploidy. Cell. 2007;131:452–62.

463 9. Renny-Byfield S, Chester M, Kovařík A, Le Comber SC, Grandbastien M-A, Deloger
464 M, et al. Next generation sequencing reveals genome downsizing in allotetraploid
465 Nicotiana tabacum, predominantly through the elimination of paternally derived
466 repetitive DNAs. Mol Biol Evol. 2011;28:2843–54.

467 10. Murat F, Zhang R, Guizard S, Gavranović H, Flores R, Steinbach D, et al. Karyotype
468 and gene order evolution from reconstructed extinct ancestors highlight contrasts in
469 genome plasticity of modern rosid crops. Genome Biol Evol. 2015;7:735–49.

470 11. Murat F, Armero A, Pont C, Klopp C, Salse J. Reconstructing the genome of the most
471 recent common ancestor of flowering plants. Nat Genet. 2017;49:490–6.

472 12. Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, et al. The sunflower
473 genome provides insights into oil metabolism, flowering and Asterid evolution. Nature.
474 2017;546:148–52.

475 13. Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy
476 during Arabidopsis evolution. Plant Cell. 2004;16:1679–91.

477 14. De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y.
478 Convergent gene loss following gene and genome duplications creates single-copy
479 families in flowering plants. Proc Natl Acad Sci U S A. 2013;110:2898–903.

480 15. Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, et al.
481 Following Tetraploidy in Maize, a Short Deletion Mechanism Removed Genes
482 Preferentially from One of the Two Homeologs. PLoS Biol. Public Library of Science;
483 2010;8:e1000409.

484 16. Schnable JC, Springer NM, Freeling M. Differentiation of the maize subgenomes by

485 genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci U S A.
486 2011;108:4069–74.

487 17. Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, et al.
488 Altered patterns of fractionation and exon deletions in Brassica rapa support a two-step
489 model of paleohexaploidy. Genetics. 2012;190:1563–74.

490 18. Hodges SA, Kramer EM. Columbines. Curr Biol. 2007;17:R992–4.

491 19. Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, et al.
492 Angiosperm phylogeny: 17 genes, 640 taxa. Am J Bot. 2011;98:704–30.

493 20. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes.
494 Science. 2000;290:1151–5.

495 21. Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred
496 from age distributions of duplicate genes. Plant Cell. 2004;16:1667–78.

497 22. Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, et al.
498 Widespread genome duplications throughout the history of flowering plants. Genome
499 Res. 2006;16:738–49.

500 23. Vanneste K, Baele G, Maere S, Van de Peer Y. Analysis of 41 plant genomes supports
501 a wave of successful genome duplications in association with the Cretaceous–Paleogene
502 boundary. Genome Res. 2014;24:1334–47.

503 24. Tiley GP, Ané C, Burleigh JG. Evaluating and Characterizing Ancient Whole-
504 Genome Duplications in Plants with Gene Count Data. Genome Biol Evol. 2016;8:1023–
505 37.

506 25. Doyle JJ, Egan AN. Dating the origins of polyploidy events. New Phytol.
507 2010;186:73–85.

508 26. Jiao Y, Paterson AH. Polyploidy-associated genome modifications during land plant
509 evolution. Philos Trans R Soc Lond B Biol Sci [Internet]. 2014;369. Available from:
510 http://dx.doi.org/10.1098/rstb.2013.0355

511 27. Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, et al. Comparative
512 analysis of tandem repeats from hundreds of species reveals unique insights into
513 centromere evolution. Genome Biol. 2013;14:R10.

514 28. Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, et al. The grapevine
515 genome sequence suggests ancestral hexaploidization in major angiosperm phyla.
516 Nature. 2007;449:463–7.

517 29. Lyons E, Pedersen B, Kane J, Freeling M. The Value of Nonmodel Genomes and an
518 Example Using SynMap Within CoGe to Dissect the Hexaploidy that Predates the
519 Rosids. Trop Plant Biol. 2008;1:181–90.

520 30. Potato Genome Sequencing Consortium, Xu X, Pan S, Cheng S, Zhang B, Mu D, et
521 al. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475:189–95.

522 31. Truco MJ, Ashrafi H, Kozik A, van Leeuwen H, Bowers J, Wo SRC, et al. An Ultra-
523 High-Density, Transcript-Based, Genetic Map of Lettuce. G3 . 2013;3:617–31.

524 32. Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, et al. The
525 coffee genome provides insight into the convergent evolution of caffeine biosynthesis.
526 Science. 2014;345:1181–4.

527 33. Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, et al. Insight into
528 the evolution of the Solanaceae from the parental genomes of Petunia hybrida. Nat
529 Plants. 2016;2:16074.

530 34. Salse J. Ancestors of modern plant crops. Curr Opin Plant Biol. 2016;30:134–42.

531 35. Ming R, VanBuren R, Liu Y, Yang M, Han Y, Li L-T, et al. Genome of the long-living
532 sacred lotus (Nelumbo nucifera Gaertn.). Genome Biol. 2013;14:R41.

533 36. Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, et al.
534 A genome triplication associated with early diversification of the core eudicots. Genome
535 Biol. 2012;13:R3.

536 37. Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. Resolution of deep angiosperm
537 phylogeny using conserved nuclear genes and estimates of early divergence times. Nat
538 Commun. 2014;5:4956.

539 38. Zheng C, Chen E, Albert VA, Lyons E, Sankoff D. Ancient eudicot hexaploidy meets
540 ancestral eurosid gene order. BMC Genomics. 2013;14 Suppl 7:S3.

541 39. Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, et al. The genome of
542 Theobroma cacao. Nat Genet. 2011;43:101–8.

543 40. Seoighe C, Wolfe KH. Yeast genome evolution in the post-genome era. Curr Opin
544 Microbiol. 1999;2:548–54.

545 41. Tian C-G, Xiong Y-Q, Liu T-Y, Sun S-H, Chen L-B, Chen M-S. Evidence for an
546 ancient whole-genome duplication event in rice and other cereals. Yi Chuan Xue Bao.
547 2005;32:519–27.

548 42. Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, et al. Global trends of
549 whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature.
550 2006;444:171–8.

551 43. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize
552 genome: complexity, diversity, and dynamics. Science. 2009;326:1112–5.

553 44. Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss
554 SH, et al. Contrasting patterns of evolution following whole genome versus tandem

duplication events in Populus. Genome Res. 2012;22:95–105.

45. Birchler JA, Bhadra U, Bhadra MP, Auger DL. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. Dev Biol. 2001;234:275–88.

46. Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. Nature. 2003;424:194–7.

47. Freeling M, Thomas BC. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. 2006;16:805–14.

48. Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci U S A. 2012;109:14746–53.

49. Malacarne G, Perazzolli M, Cestaro A, Sterck L, Fontana P, Van de Peer Y, et al. Deconstruction of the (paleo)polyploid grapevine genome based on the analysis of transposition events involving NBS resistance genes. PLoS One. 2012;7:e29762.

50. Wolfe KH. Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet. 2001;2:333–41.

51. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, et al. The genome of Eucalyptus grandis. Nature. 2014;510:356–62.

52. Lyons E, Freeling M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 2008;53:661–73.

53. Blanc G, Hokamp K, Wolfe KH. A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res. 2003;13:137–44.

54. Smith DA, Cordell R, Dillon EM. Infectious texts: Modeling text reuse in nineteenth-century newspapers. 2013 IEEE International Conference on Big Data. 2013. p. 86–94.