

Linked by Ancestral Bonds: Multiple Whole-Genome Duplications and Reticulate Evolution in a Brassicaceae Tribe

Xinyi Guo ^{†,1} Terezie Mandáková ^{†,1} Karolína Trachtová,¹ Barış Özüdoğru ² Jianquan Liu ³ and Martin A. Lysak ^{*,1}

¹CEITEC—Central European Institute of Technology, Masaryk University, Brno, Czech Republic

²Department of Biology, Faculty of Science, Hacettepe University, Beytepe, Ankara, Turkey

³Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: martin.lysak@ceitec.muni.cz.

Associate editor: Juliette de Meaux

Abstract

Pervasive hybridization and whole-genome duplications (WGDs) influenced genome evolution in several eukaryotic lineages. Although frequent and recurrent hybridizations may result in reticulate phylogenies, the evolutionary events underlying these reticulations, including detailed structure of the ancestral diploid and polyploid genomes, were only rarely reconstructed. Here, we elucidate the complex genomic history of a monophyletic clade from the mustard family (Brassicaceae), showing contentious relationships to the early-diverging clades of this model plant family. Genome evolution in the crucifer tribe Biscutelleae (~60 species, 5 genera) was dominated by pervasive hybridizations and subsequent genome duplications. Diversification of an ancestral diploid genome into several divergent but crossable genomes was followed by hybridizations between these genomes. Whereas a single genus (*Megadenia*) remained diploid, the four remaining genera originated by allopolyploidy (*Biscutella*, *Lunaria*, *Ricotia*) or autopolyploidy (*Heldreichia*). The contentious relationships among the Biscutelleae genera, and between the tribe and other early diverged crucifer lineages, are best explained by close genomic relatedness among the recurrently hybridizing ancestral genomes. By using complementary cytogenomics and phylogenomics approaches, we demonstrate that the origin of a monophyletic plant clade can be more complex than a parsimonious assumption of a single WGD spurring postpolyploid cladogenesis. Instead, recurrent hybridization among the same and/or closely related parental genomes may phylogenetically interlink diploid and polyploid genomes despite the incidence of multiple independent WGDs. Our results provide new insights into evolution of early-diverging Brassicaceae lineages and elucidate challenges in resolving the contentious relationships within and between land plant lineages with pervasive hybridization and WGDs.

Key words: hybridization, polyploidy, whole-genome duplication, reticulate evolution, diploidization, dysploidy, chromosome rearrangements, phylogenetics.

Introduction

Polyploidy or whole-genome duplication (WGD) followed by diploidization acts as an important evolutionary force promoting diversification of eukaryotes, including land plants (Soltis et al. 2015; Lohaus and Van de Peer 2016; One Thousand Plant Transcriptomes Initiative 2019; Qiao et al. 2019). Whereas autopolyploids originate by crossing between genetically (nearly) identical individuals or by somatic genome doubling (e.g., Spoelhof et al. 2017), allopolyploid genomes are formed by duplication of chromosomes in interspecies hybrids (e.g., Mallet 2007). Thus, hybridization is the first and essential step of an allopolyploid formation, potentially resulting in an allopolyploid speciation event.

Eukaryotic lineages and clades differ in the frequency of natural hybridization, with hybrids in some groups being more frequent than in others. Hybrids should form more

frequently between species in sympatry than between allopatric ones and more easily between congeners than between members of disparate genera or tribes; and within a genus, more frequently among closely related species. However, close genetic relatedness between the parental genomes may lead to homeologous pairing (i.e., formation of multivalents) and reduced fertility or sterility of the newly formed allopolyploids (e.g., Darlington 1937). Conversely, parental genomes genetically divergent to avoid homeologous chromosome pairing may more likely form a new allopolyploid genome (Darlington 1937; Grant 1981; the paradigm often called “Darlington’s rule,” Buggs et al. 2011). To what an extent the genetic or phylogenetic divergence between hybridizing species influences the hybridization frequency and allopolyploid speciation is still debated (Chapman and Burke 2007; Paun et al. 2009; Buggs et al. 2011; Levin 2013; Wagner et al. 2019) and further studies are required to

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

elucidate the role of genetic, chromosomal, and phylogenetic distance in hybrid formation (Soltis and Soltis 2009).

The advent of phylogenetic analyses based on molecular markers has revealed that phylogenetic trees may show a reticulate, rather than a bifurcate structure. This reticulation reflects hybridization and polyploidization events, and it is further amplified by the recurrence of these processes (reviewed by Soltis and Soltis [1999, 2009]; Levin and Soltis [2018]). Polyploids may originate repeatedly at multiple sites by hybridization between the same parental species which, however, may be genetically variable. For instance, multiple individuals of *Arabidopsis halleri* and *A. lyrata* were involved in the origin of the allotetraploid *A. kamchatica* (Brassicaceae; Shimizu-Inatsugi et al. 2009), and at least 46 and 31 independent autotetraploid and triploid origins have occurred in *Galax urceolata* (Diapensiaceae; Servick et al. 2015). Reciprocal hybrids (i.e., with two species being reciprocally maternal and paternal genomes) may result in morphologically and reproductively distinct allopolyploids, as in *Tragopogon* (Asteraceae; Soltis and Soltis 2009). Furthermore, if parental genomes will coexist for hundreds or thousands of years, genetically similar but not identical allopolyploids may originate continuously. As reproductive barriers between polyploid and diploid species do not need to be complete, higher level polyploids, such as triploids, hexaploids, and octoploids, may be formed (Levin and Soltis 2018). This reticulate evolution is further promoted by postpolyploid genome diploidization (e.g., differential subgenome fractionation, different chromosomal rearrangements) acting on similar genomes with different intensities and enforcing speciation/cladogenesis (Dodsworth et al. 2016; Mandáková and Lysak 2018). Altogether recurrent hybridization and polyploidization on timescales in the order of million years formed polyploid complexes comprised diploidizing polyploids of different age, with reticulate relationships in phylogenetic trees (e.g., Estep et al. 2014; Triplett et al. 2014; Glemin et al. 2019; Guo et al. 2019; Huynh et al. 2019; Mandáková et al. 2019; Martin et al. 2020).

Crucifers (Brassicaceae) belong to one of the 15 largest angiosperm families and comprise almost 4,000 species in 351 genera (Koch et al. 2018). The family is well known for high frequency of hybridization (Marhold and Lihová 2006), with more than 43% neopolyploid species (Hohmann et al. 2015). However, this proportion is grossly underestimated by not accounting for older, clade-specific, WGD events (Mandáková et al. 2017a) concealed by postpolyploid diploidization (Mandáková and Lysak 2018). Actually, all crucifers have descended from a paleotetraploid ancestor formed by the At- α WGD dated to \sim 35 Ma (Vision et al. 2000; Edger et al. 2015; Hohmann et al. 2015; Walden et al. 2020a). The paleopolyploid genome has then diverged into four or more major lineages (Franzke et al. 2011; Hohmann et al. 2015; Huang et al. 2016; Nikolov et al. 2019; Walden et al. 2020b) consisting of monophyletic clades, classified as tribes (Al-Shehbaz 2012). Whereas genomes of some lineages and tribes essentially retained their paleotetraploid character (Vision et al. 2000; Mandáková and Lysak 2008; Mandáková et al. 2017b; Mandáková et al. 2020), younger WGDs preceded

the origin of several genera and tribes, such as Brassiceae, Heliophilleae, *Leavenworthia*, Microlepidae (e.g., Mandáková et al. 2010, 2012; Haudry et al. 2013; Hohmann et al. 2015; Lysak et al. 2016; Mandáková et al. 2017a; Mandáková et al. 2017c; Walden et al. 2020a). It was estimated that about a quarter of crucifer tribes have diversified following post-At- α genome duplications (Mandáková et al. 2017a). These younger, mesopolyploid genomes (Mandáková et al. 2010), exhibit different phylogenomic features than paleopolyploid ones. Due to their younger age, mesopolyploid clades are karyologically variable, having multiple base chromosome numbers due to independent diploidizations (Mandáková and Lysak 2018). Their internal phylogenetic structure, as well as the position in the family are contentious (e.g., Mandáková et al. 2017a; Nikolov et al. 2019; Walden et al. 2020a) due to problematic distinction between homeologous and orthologous sequences (Hénocq L, Gallina S, Schmitt E, Castric V, Vekemans X, Poux C, unpublished data).

Complex phylogenomic history of mesopolyploid crucifer clades is perhaps illustrated at its best by long unknown and puzzling genome evolution in the small tribe Biscutelleae. Tribe Biscutelleae (\sim 60 species in five genera, see supplementary file 1, Supplementary Material online) was re-established by German and Al-Shehbaz (2008), including *Biscutella* and *Megadenia* due to their phylogenetic (German et al. 2009) and morphological affinities. Other three Biscutelleae genera have remained unassigned in several molecular phylogenetic studies, until Özüdoğru et al. (2015, 2017) expanded Biscutelleae by the inclusion of *Heldreichia*, *Lunaria*, and *Ricotia*. This treatment was confirmed by a recent family-wide phylogenomic studies (Walden et al. 2020a). However, the contentious phylogenetic placement of Biscutelleae within the family tree remained unresolved even with genome-scale data and improved sampling (supplementary fig. 1, Supplementary Material online). Although Biscutelleae were tentatively retrieved as sister to the crucifer Lineage I in a 1,421-exon tree (Nikolov et al. 2019), it had a basal position in the broadly delimited Lineage II of the whole-family chloroplast tree (Mandáková et al. 2018; Nikolov et al. 2019; Walden et al. 2020a).

A very first insight into genome evolution of Biscutelleae provided Geiser et al. (2016) by revealing that the genome of *Biscutella laevigata* has undergone a WGD followed by biased fractionation, extensive structural diploidization, and reduction of chromosome number (descending dysploidy). The authors assumed that the mesotetraploid ancestral genome of *Biscutella* ($n = 16$ chromosomes) originated by a merger of two structurally very similar ancestral genomes named ancPCK ($n = 8$). Structurally, the ancPCK genome resembled the eight-chromosome Ancestral Crucifer Karyotype (ACK, Schranz et al. 2006), with a single reciprocal translocation differentiating the two genomes. A later study (Mandáková et al. 2018) demonstrated that another Biscutelleae genus, *Ricotia*, had also a polyploid origin, however here the mesotetraploid genome originated through a more distant hybridization between the maternal ancPCK-like genome ($n = 8$) and paternal genome with seven chromosomes ($n = 7$). The latter genome structurally corresponds to the previously

described Proto-Calepineae Karyotype (PCK; Mandáková and Lysak 2008) which originated presumably from the ancPCK through descending dysploidy ($n = 8 \rightarrow n = 7$, Mandáková et al. 2018).

Here, we aimed to reconstruct the origin and genome evolution of all five Biscutelleae genera by combining comparative cytogenetics and transcriptome-based phylogenomics. In addition to the earlier reports on genome evolution in *Biscutella* (Geiser et al. 2016) and *Ricotia* (Mandáková et al. 2018), we expanded our investigation by including remaining three Biscutelleae genera (*Heldreichia*, *Lunaria*, and *Megadenia*). Based on what was known about phylogenetic relationships and genome evolution in Biscutelleae (Özüdoğru et al. 2015, 2017; Geiser et al. 2016; Mandáková et al. 2017a; Mandáková et al. 2018), we aimed to answer the following questions. 1) What causes the repeatedly retrieved incongruences between phylogenetic placements of Biscutelleae?, 2) What are the phylogenetic relationships between Biscutelleae genera, and between the tribe and other Brassicaceae tribes?, 3) Do all Biscutelleae genera have a mesotetraploid origin?, and 4) What genome structure the ancestral parental diploid and polyploid genomes had? By using complementary cytogenomic and phylogenomic approaches, we paint a comprehensive picture of the reticulate evolution in a Brassicaceae tribe. We show that multiple hybridizations and WGDs involving the same or closely related diploid genomes may become obscured by disparate rates of post-polyploid diploidization, and that diploid, autopolyploid, and allopolyploid genomes may constitute a monophyletic clade due to a common ancestral genome. Our results provide new insights into evolution of early-diverging Brassicaceae lineages and elucidate challenges in resolving the contentious relationships within and between land plant lineages with pervasive hybridization and genome duplications.

Results

Comparative Cytogenetic Analysis in Biscutelleae

To analyze genome structure of 11 accessions representing four species of three Biscutelleae genera—*Heldreichia*, *Lunaria*, and *Megadenia* (supplementary table 1, Supplementary Material online)—we employed comparative chromosome painting (CCP) based on localization of contigs of chromosome-specific Bacterial Artificial Clones (BACs) of *A. thaliana* on pachytene chromosomes of target species (see supplementary figs. 2–4, Supplementary Material online, for examples of CCP). The painting probes were designed to reflect the system of 22 ancestral genomic blocks (GBs) of crucifer genomes (Schranz et al. 2006; Lysak et al. 2016). All 22 GBs were unambiguously identified in one or two copies within meiotic chromosome complements of the analyzed Biscutelleae species.

Diploid *Megadenia* Genome Originated through Descending Dysploidy from an ancPCK-Like Genome

The analyzed population of *Megadenia pygmaea* had six chromosome pairs ($2n = 12$). All painting probes hybridizing to only one homeologous region confirmed the diploid status of

the *Megadenia* genome (supplementary fig. 2, Supplementary Material online).

Three chromosomes of *M. pygmaea* (Mp1, Mp3, and Mp4) mirror three ancestral chromosomes (AK1, AK4, and AK7) of the ancestral genome ancPCK ($n = 8$, Mandáková et al. 2018) (fig. 1A). Chromosome Mp5 is homeologous to chromosome AK6/8 (GB association O+P+Wb+R), whereas chromosome Mp2 was formed by a *Megadenia*-specific end-to-end translocation (EET) between chromosomes AK2 and AK3 (supplementary fig. 2, Supplementary Material online), reducing the chromosome number by one.

As chromosome Mp6 resembles the PCK-specific chromosome AK5/8/6 (GBs M-N, V, X, Q, Wa, and K-L; Mandáková and Lysak 2008), Yang et al. (2020) inferred the origin of the *Megadenia* genome from the PCK genome based on the fact that chromosome Mp6 in *Megadenia* and AK5/8/6 in PCK differ only by a single paracentric inversion. However, given that the previously published phylogenetic studies (Özüdoğru et al. 2017; Mandáková et al. 2018), as well as the present phylogenetic data (see below), unequivocally confirm the monophyletic origin of Biscutelleae, *Megadenia* genome must have originated from an ancPCK-like ($n = 8$) genome shared by all Biscutelleae genera (fig. 1A). Thus, the formation of chromosome AK5/8/6 must have occurred independently in the PCK and *Megadenia* genomes (supplementary fig. 2, Supplementary Material online). Although the AK5/8/6 chromosome in the PCK and *Megadenia* originated from the same ancestral chromosomes, they were formed by different dysploid rearrangements. In the PCK, chromosome AK5/8/6 originated by an EET followed by a paracentric inversion (Mandáková et al. 2018). In *Megadenia*, a structurally similar chromosome was formed by a nested chromosome insertion (NCI) followed by peri- and paracentric inversions (supplementary fig. 2, Supplementary Material online).

The Five-Chromosome *Heldreichia* Genome Was Duplicated by Autopolyploidy

Heldreichia bupleurifolia has ten chromosome pairs ($2n = 20$). All 22 GBs were found duplicated (fig. 1B and supplementary fig. 3, Supplementary Material online). The presence of two structurally identical haploid chromosome sets (Hb1–Hb5 and Hb1'–Hb5'), and comparable sizes and fluorescence intensity of all painting probes strongly suggest an autotetraploid origin ($2n = 4 \times = 20$) of the genome (supplementary fig. 3, Supplementary Material online). As the five chromosomes of both subgenomes are identical structurally, only five *Heldreichia* chromosomes (Hb1–Hb5) are further described.

Among the five *Heldreichia* chromosomes, chromosome Hb3 has the conserved structure of ancestral chromosome AK3, whereas the remaining chromosomes are composed of GBs originated from three (Hb1, Hb2, and Hb5) or four (Hb4) ancestral chromosomes (fig. 1B and supplementary fig. 3, Supplementary Material online). Origin of chromosomes Hb1 (GBs Rb+T+S+V+Xb+B+A) and Hb5 (U+Ob+P+Wb+Ra+Oa) was most likely initiated by a 10.25-Mb pericentric inversion on ancestral chromosome AK6/8 (O+P+Wb+R \rightarrow Rb+Ob+P+Wb+Ra+Oa; supplementary fig. 3, Supplementary Material online). The

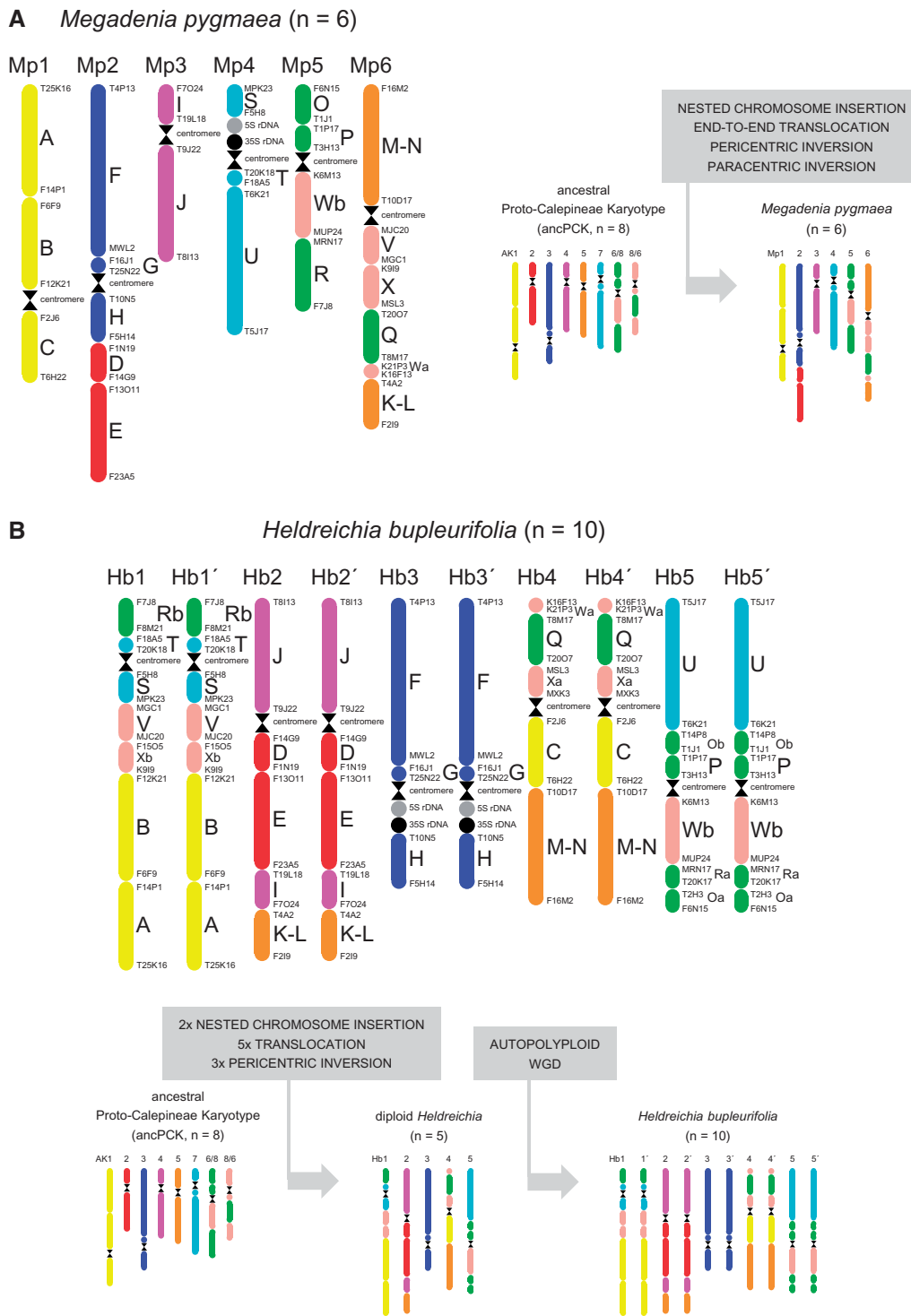


Fig. 1. Genome structure of *Megadenia pygmaea* and *Heldreichia bupleurifolia* based on CCP analysis. (A) Comparative cytogenetic map of *M. pygmaea* ($n = 6$; Mp1–Mp6) and the purported origin of the *Megadenia* genome from the ancestral Proto-Calepineae Karyotype (ancPCK, $n = 8$; see [supplementary fig. 2, Supplementary Material](#) online, for more details). (B) The extant autotetraploid *Heldreichia* genome ($n = 5$; Hb1–Hb5') originated by a WGD of an ancestral $n = 5$ genome derived from ancPCK-like genome ($n = 8$) by descending dysploidy (see [supplementary fig. 3, Supplementary Material](#) online, for more details). The different colors correspond to the eight chromosomes of Ancestral Crucifer Karyotype (ACK), whereas capital letters refer to 22 genomic blocks (A–X). Bacterial artificial chromosome (BAC) clones of *Arabidopsis thaliana* defining each genomic block are listed along the chromosomes. Centromeres are indicated by black hourglass symbols.

“Rb+Ob+P+Wb+Ra+Oa” chromosome lost block Rb through an unequal translocation with the “V+Xb+B+A” telocentric, formed from chromosomes AK1 and AK8/6 (see

the origin of Hb4 below). Finally, the “Ob+P+Wb+Ra+Oa” chromosome received block U through an unequal translocation with chromosome AK7; the resulting *Heldreichia*

chromosome Hb5 contains the AK6/8 centromere. The remaining centric part of AK7 (S+T) participated in an NCI event “inserting” blocks S and T into the “recipient chromosome Rb+V+Xb+B+A.” The resulting structure of chromosome Hb1 consists of blocks Rb+T+S+V+Xb+B+A and the AK7 centromere (fig. 1B and supplementary fig. 3, Supplementary Material online).

The origin of chromosome Hb2 was most likely initiated by an NCI event “inserting” chromosome AK2 into the “recipient” chromosome AK4. The NCI event was followed by a 3-Mb pericentric inversion and an unequal translocation which added the entire top (short) arm of AK5 (K-L) to the chromosome terminus distally from block I (fig. 1B and supplementary fig. 3, Supplementary Material online). The bottom (long) arm of AK5 (M-N) participated in the origin of chromosome Hb4 (see below).

Chromosome Hb4 has originated from ancestral chromosomes AK1, AK5, and AK8/6. The structure of AK8/6 was initially modified by a 7-Mb pericentric inversion changing the original block composition to Xa+Q+Wa+V+Xb. A whole-arm translocation between the inversion-bearing chromosome AK8/6 and chromosome AK1 resulted in the origin of a telocentric (GBs V+Xb+B+A) and submetacentric (Wa+Q+Xa+C) chromosome, respectively. Finally, the submetacentric chromosome received the bottom (long) arm of AK5 (M-N, see above) to form the *Heldreichia* chromosome Hb4 (fig. 1B and supplementary fig. 3, Supplementary Material online).

In *Heldreichia*, no GB associations specific for chromosome AK5/8/6 in PCK were identified. By contrast, both ancPCK-specific chromosomes AK6/8 and AK8/5 were inferred to participate in the origin of the extant *Heldreichia* chromosomes. This strongly suggests that the five *Heldreichia* chromosomes were derived from the eight chromosomes of the ancPCK genome through descending dysploidy ($n = 8 \rightarrow n = 5$). Later, the five-chromosome genome was duplicated by autopolyploidy ($n = 5 \rightarrow n = 10$; fig. 1B), with no large-scale chromosomal rearrangements differentiating both subgenomes.

Lunaria Genome Originated through Hybridization between Ancestral Diploids, Followed by Descending Dysploidy

A uniform chromosome number of $2n = 28$ was identified in six different populations of *Lunaria rediviva*, three populations of *L. annua*, and one population of *L. telekiana*. Several, presumably erroneous, chromosome counts of $2n = 30$ (e.g., (Sharma 1970; Uhrikova 1976; Harriman 1978)) can be most likely attributed to the fragility of the large interstitial nucleolus organizer region (NOR) at the pericentromere of chromosome Lu14. The resulting broken-off chromosome arms of Lu14 were mistaken for an extra chromosome pair (supplementary fig. 4, Supplementary Material online). As all 22 GBs were found in duplicates within the haploid chromosome complement of all *Lunaria* accessions, the genus had a tetraploid origin. Slightly different sizes and fluorescence intensities of the duplicated GBs suggest an

allopolyploid origin of *Lunaria* (supplementary fig. 4, Supplementary Material online).

Among the 14 *Lunaria* chromosome pairs, eight chromosomes (Lu1–Lu3, Lu5, and Lu7–Lu10) are shared with ACK, ancPCK, and PCK ancestral genomes. The *Lunaria* genome contains also homeologs of ancPCK-specific chromosomes AK6/8 (chromosome Lu11) and AK8/6 (Lu13), and PCK-like chromosomes AK6/8 and AK5/8/6 (Lu12 and Lu14). Finally, two *Lunaria* chromosomes (Lu4 and Lu6) were formed by a complex dysploid rearrangement involving ancestral chromosomes AK2, AK3, and AK5 (fig. 2 and supplementary fig. 4, Supplementary Material online). The origin of chromosomes Lu4 and Lu6 was most likely initiated by a whole-arm reciprocal translocation between AK3 and AK5. The larger translocation chromosome (GBs F+G+M-N) has become *Lunaria*'s chromosome Lu6. The smaller translocation chromosome (H+K-L) was subsequently involved in an NCI event as a “recipient chromosome” with “insertion” of AK2. The NCI was followed by a 3-Mb pericentric inversion to form chromosome Lu4 (fig. 2 and supplementary fig. 4, Supplementary Material online).

The presence of both ancPCK- and PCK-specific chromosomal rearrangements in *Lunaria* genomes strongly suggests an allopolyploid origin of the genus. The ancestral tetraploid genome originated through hybridization between ancestral diploid genomes of ancPCK ($n = 8$) and PCK ($n = 7$). The genome merger was followed by descending dysploidy ($n = 15 \rightarrow n = 14$) mediated by a whole-arm translocation and NCI event (fig. 2 and supplementary fig. 4, Supplementary Material online).

Large-Scale Gene Duplications Corroborated Four Genus-Specific WGDs in Biscutelleae

Bursts of gene duplications that are consistent with WGDs were observed in the distribution of synonymous substitutions on synonymous sites (K_s ; fig. 3 and supplementary fig. 5, Supplementary Material online). The mean K_s values for the α -WGD event shared by all Brassicaceae species were estimated to be between 0.616 and 0.770, with an average of 0.670 (supplementary table 2, Supplementary Material online). The presence of additional K_s peaks suggested that younger WGDs have occurred in four Biscutelleae genera, whereas no recent WGD was detected in *Megadenia*. $K_s < 0.05$ inferred for *Heldreichia* points to a young age of the autopolyploidization event identified by cytogenetic analyses. The estimated mean K_s values between 0.248 and 0.435 suggested mesotetraploid events (m-WGDs) in *Biscutella*, *Lunaria*, and *Ricotia* (supplementary table 2, Supplementary Material online). Distribution of the ortholog divergence between species of different genera revealed mean K_s values that are slightly larger than those of the paralog divergence in at least one species (supplementary figs. 5 and 6, Supplementary Material online). Phylogenomic analysis using the Multi-tAxon Paleopolyploidy Search (MAPS) algorithm (Li et al. 2015) found no evidence for bursts of gene duplications being shared by these m-WGDs (supplementary fig. 7, Supplementary Material online). In addition, clustering of annotated GO terms revealed divergent gene retention/loss

Lunaria annua & *L. rediviva* ($n = 14$)

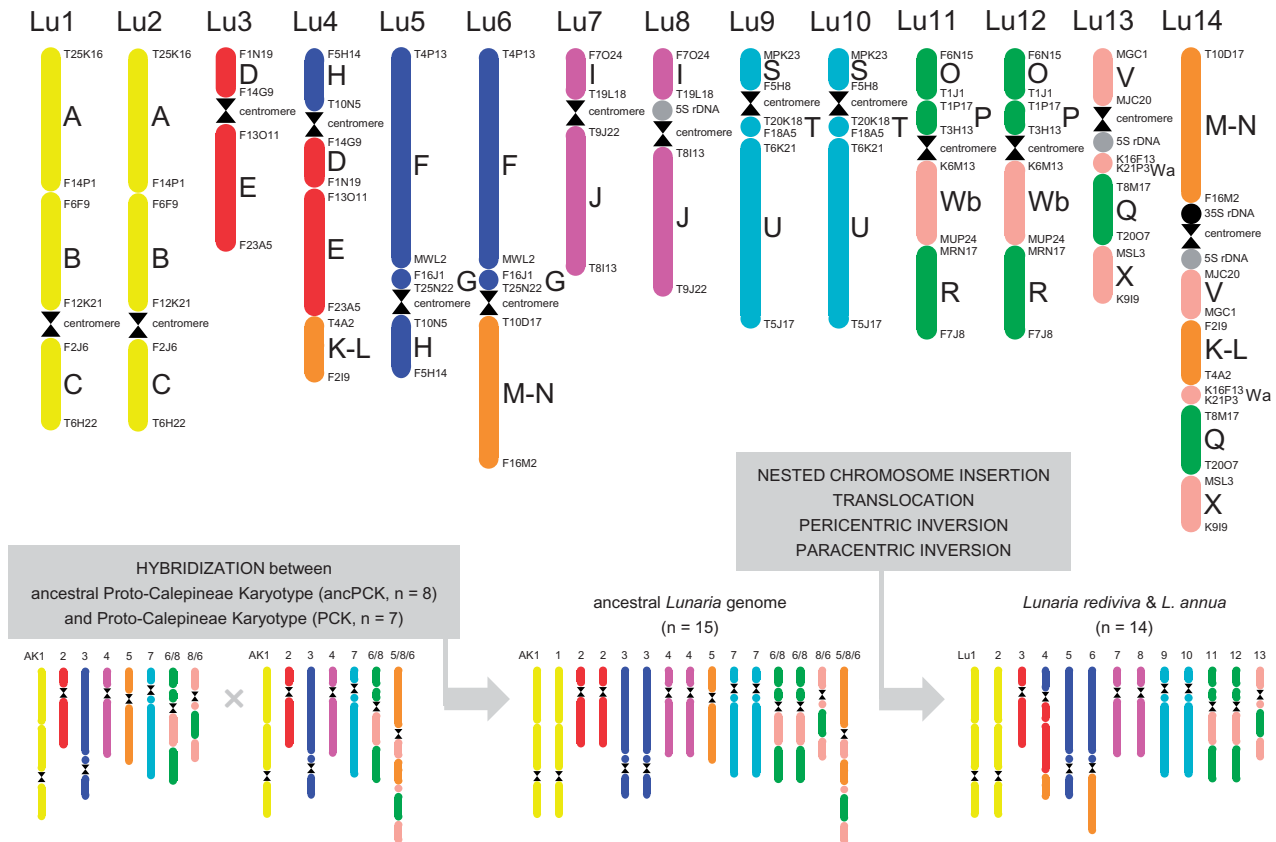


Fig. 2. Comparative cytogenetic map and genome origin of *Lunaria annua* and *L. rediviva* ($n = 14$; Lu1–Lu14) based on CCP analysis. The ancestral allotetraploid *Lunaria* genome ($n = 15$) originated through hybridization between ancPCK ($n = 8$) and PCK ($n = 7$) genomes followed by descending dysploidy (see [supplementary fig. 4, Supplementary Material](#) online, for more details). The different colors correspond to the eight chromosomes of ACK, whereas capital letters refer to 22 genomic blocks (A–X). Bacterial artificial chromosome (BAC) clones of *Arabidopsis thaliana* defining each genomic block are listed along the chromosomes. Centromeres are indicated by black hourglass symbols.

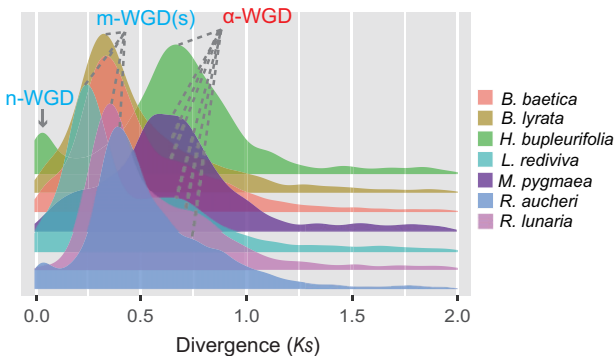


Fig. 3. Comparison of paralog divergence using distribution of synonymous substitutions per synonymous site (K_s). Paralogous gene pairs were identified in transcriptomes of seven Biscutelleae species. An arrow indicates the recent n-WGD in *Heldreichia*. The three mesopolyploid WGD events in *Biscutella*, *Lunaria*, and *Ricotia* are indicated as m-WGDs, whereas α -WGD refers to the ancient duplication shared by all Brassicaceae.

following the m-WGDs, in sharp contrast with the convergent pattern after the α -WGD ([supplementary fig. 8, Supplementary Material](#) online). Collectively, multiple lines of evidence suggest four genus-specific WGDs in Biscutelleae.

Nuclear Multigene Phylogeny Corroborated Biscutelleae as a Monophyletic Clade

To study the genome evolution in tribe Biscutelleae within a phylogenetic framework, we generated two sets of CDS alignments from 12,046 nuclear-encoded homologous gene clusters for 25 Brassicaceae species plus outgroup (*Tarenaya hassleriana*): 1) 1,545 single-copy genes (SCG) that are shared by all species without missing data, and 2) 8,607 “rooted ingroup” (RT) homologs including more than 20 species for each gene cluster. We recovered the same topology with both data sets, hereafter referred to as T1. In T1 ([fig. 4A](#)), the Biscutelleae species formed a monophyletic clade sister to Lineage I species; the four tetraploid Biscutelleae genera were separated into two distinct subclades: 1) *Biscutella* + *Heldreichia* and 2) *Lunaria* + *Ricotia*, whereas the diploid *Megadenia* was sister to the *Biscutella*/*Heldreichia* subclade. Selective exclusion of *Biscutella*, *Lunaria*, and *Ricotia* species has not affected the phylogenetic placement of Biscutelleae ([supplementary fig. 9, Supplementary Material](#) online). In addition, T1 was repeatedly recovered by multiple tree estimation approaches.

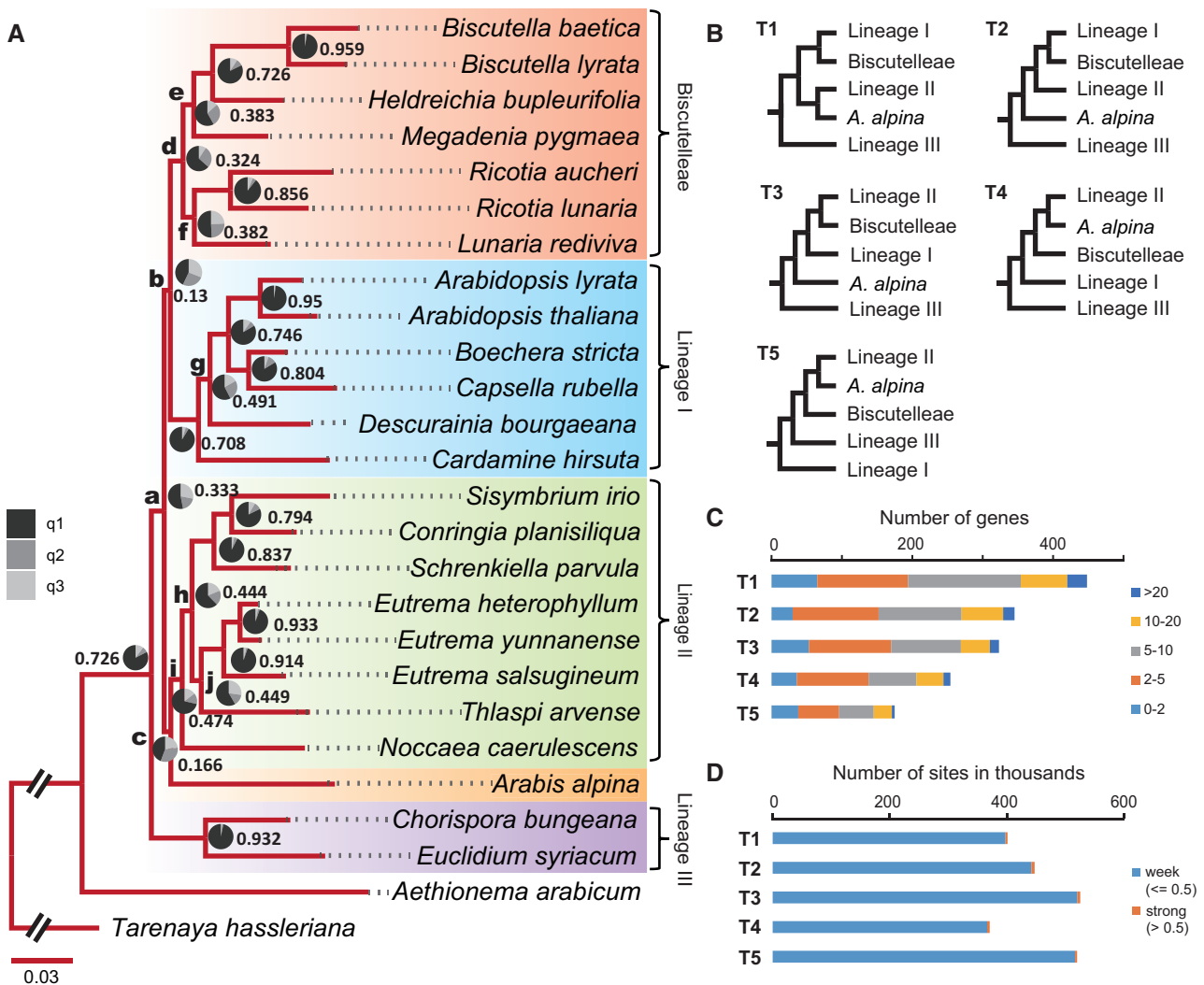


FIG. 4. Species tree inference and phylogenetic signal analyses based on 1,545 single-copy orthologs. (A) Species phylogeny of Brassicaceae, or topology T1. All branches were supported by posterior probabilities of 1.0. Branch lengths indicate numbers of substitution per site estimated from concatenation method. Pie chart at each node indicates ASTRAL quartet scores for the three possible arrangements (q1–q3) for the respective branch leading to the node, with q1 representing the displayed topology. Numbers indicate concordance factors (CF) estimated from Bayesian concordance analysis (BUCKy). Nodes receiving inconclusive support were labeled with lowercase letters. (B) Alternative phylogenetic hypotheses (T2–T5) used to test distribution of phylogenetic signals together with the species tree obtained in this study (referred to as T1). T2 and T5 were recovered by [Nikolov et al. \(2019\)](#) and [Mandáková et al. \(2018\)](#), respectively. T3 and T4 were adapted from [Huang et al. \(2016\)](#) by placing *Arabis alpina* to two possible positions due to the lacking information in the original paper. (C) Number of genes supporting each hypothesis. Colors indicate the range of Δ GSL scores. (D) Number of sites supporting each hypothesis within the single-copy genes. Strong sites were defined as those having Δ SSL scores larger than 0.5.

Contentious Phylogenetic Placement of Biscutelleae

In contrast to the seemingly robust T1 topology, the ASTRAL quartet scores, which represent the proportions of three alternative topologies around an internal branch, provided inconclusive support for multiple nodes corresponding to the early diverging branches within the Brassicaceae (fig. 4A). We observed strong phylogenetic conflicts in two nodes: node b, representing the split between Lineage I and Biscutelleae, and node c, that is, the split of Lineage II and Arabideae (fig. 4A). We further confirmed the discordance by implementing Bayesian concordance analysis (BUCKy), whereby concordance factors (i.e., proportion of gene trees supporting a particular node) for nodes b and c received very low scores of 0.130 and 0.166, respectively (fig. 4A).

To investigate phylogenetic signals underlying the discordance, we focused on the SCG data set and compared the delta gene- and site-wise log-likelihood scores (Δ GSL and Δ SSL) among five species-tree hypotheses (T1–T5) (fig. 4A and B). Analysis of Δ GSL indicated that T1 was favored by 448 out of 1,545 genes (29%), whereas the remaining topologies were supported by 11–22% of genes (fig. 4C). In addition, we failed to recover topologies other than T1 after excluding genes with outlier Δ GSL values. We next asked whether T1 is robust at site-level by comparing Δ SSL values of these genes. Interestingly, the number of sites supporting T1 was only slightly higher than that of T4 but lower than those of the remaining topologies. More than 99% of sites were essentially “weak” sites, with Δ SSL value less than 0.5

(fig. 4D). After 1% of sites with the highest Δ SSL values in each gene was removed, Biscutelleae species coalesced to a deeper node outside Lineages I, II, and III (supplementary fig. 10, Supplementary Material online). Thus, the species tree was not fully resolved, with contentious branches being affected by a few sites widely distributed among the genomes rather than by several outlier genes.

Divergence Time Estimates Revealed Reticulate Evolution of Biscutelleae

Acknowledging that the species phylogeny remained unresolved, we performed molecular dating analyses separately with genes supporting the alternative topologies (supplementary fig. 11, Supplementary Material online), using a single fossil calibration point for the Brassicaceae/Cleomaceae split (Cardinal-McTeague et al. 2016). The estimated divergence times of Lineage I/Biscutelleae and Lineage II/Biscutelleae ranged from 21 and 24.3 Ma, largely overlapping with each other (supplementary fig. 11, Supplementary Material online). To accommodate the conflicts among loci, we used gene trees to estimate divergence times of the most recent common ancestor (MRCA) for species pairs, which do not rely on fixed species trees. The estimated divergence times between Lineage I and the subclade including *Biscutella*, *Heldreichia*, and *Megadenia* were ~ 1 My younger than those between Lineages I and II (fig. 5A). However, this was not the case for the other subclade (*Lunaria* and *Ricotia*), suggesting that the two genera diverged from Lineage II more recently than from Lineage I (fig. 5A).

Network analysis using estimated mean divergence times as the distance showed a large cuboid structure connecting two Biscutelleae subclades to alternative branches with conflicting signals in the center (fig. 5B). HyDe four-taxon tests revealed 1,642 (out of 6,900) triplets that showed significant levels of hybridization (corrected $P = 7.246 \times 10^{-6}$; Bonferroni correction). Among these, Biscutelleae species were frequently detected as hybrids (fig. 5C), although the results varied in different species (fig. 5D). Phylogenetic network analyses revealed complex relationships between Biscutelleae and other early diverged Brassicaceae clades (supplementary fig. 12, Supplementary Material online) and inferred at least three ancient hybridizations (supplementary fig. 13, Supplementary Material online).

Subgenome Divergence Was Impacted by Heterogeneity in Molecular Evolutionary Rates

To estimate the time of subgenome divergence in tetraploid Biscutelleae genomes, we performed molecular dating based on gene tree topologies, using the same calibration point as described above. Clustering analysis of 8,485 nonsingle-copy homologous gene groups revealed 3,132 genes that are duplicated in more than 14 (out of 21) diploid species (supplementary fig. 14, Supplementary Material online), which were likely related to α -WGD and thus removed from further analyses. From the remaining gene groups, we retrieved 799–1,022 pairs of gene duplicates in *Biscutella*, *Lunaria*, and *Ricotia* species, whereas only 44 pairs were retrieved in *Heldreichia*. Due to the low number of retrieved gene pairs,

Heldreichia was excluded from age distribution analysis. The mean time of the subgenome divergence in *Biscutella* was estimated to be between 14 and 15 Ma, whereas the estimated subgenome divergence time in *Lunaria* and *Ricotia* was ~ 19 and 20 Ma, respectively (table 1 and supplementary fig. 15, Supplementary Material online). These estimates were in sharp contrast with K_s analyses, where *Lunaria* had the lowest level of subgenome divergence after m-WGD (fig. 3 and supplementary table 2, Supplementary Material online). By using the same gene pairs for K_s analyses, we confirmed the difference in estimating subgenome divergence between the two approaches (table 1 and supplementary fig. 15, Supplementary Material online).

We next asked if the incongruent estimates of subgenome divergence were caused by different methods. When conducting the same analyses with triplicated genes of *Brassica* genomes, congruent subgenome divergence estimates were obtained (supplementary fig. 16, Supplementary Material online). Thus, the differences observed for the m-WGDs in Biscutelleae likely suggest heterogeneity in molecular evolutionary rates. To test this hypothesis, we calculated ortholog divergence between all Biscutelleae species and *A. thaliana*. The results showed that synonymous substitution rates in *Biscutella* and *Ricotia* genomes (mean $K_s = 0.381$ – 0.435) were 11% to 39% higher than those in *Heldreichia*, *Lunaria*, and *Megadenia* (mean $K_s = 0.314$ – 0.342), indicating different speed of genome evolution among Biscutelleae species (supplementary fig. 17 and table 2, Supplementary Material online). Higher mean K_s values in *Biscutella* and *Ricotia* species were likely caused by higher mutation rates in these genomes. Consequently, the ratio between nonsynonymous substitutions on nonsynonymous site (K_a) and K_s was generally lower in *Biscutella* and *Ricotia* than in other genera (supplementary fig. 18, Supplementary Material online).

Phylogenetic Reconciliation Provides Congruent Estimates of Subgenome Origins in Allotetraploid Species

Phylogenetic reconciliation with singly-labeled Brassicaceae phylogeny revealed accumulation of gene duplication signals mainly around two nodes (supplementary fig. 19, Supplementary Material online): one represents the last common ancestor (LCA) of core Brassicaceae species (i.e., all Brassicaceae species except *Aethionema*), whereas the other one was the LCA of *Biscutella* species. Given the diploid status of most ingroup species, we asked if the unexpected signal at the former node was contributed by m-WGDs in Biscutelleae. To this end, we inferred the mode of WGD in *Biscutella*, *Lunaria*, and *Ricotia* species separately using the recently developed approach based on multilabeled (MUL) species relationships (Thomas et al. 2017), and the results corroborated the allopolyploid origin of these genera (supplementary fig. 20, Supplementary Material online). In *Biscutella* species, both subgenomes were placed within Biscutelleae and represented successive sister lineages to *Heldreichia* (supplementary fig. 20, Supplementary Material online). In *Lunaria* and *Ricotia*, one subgenome was within Biscutelleae, whereas the other was

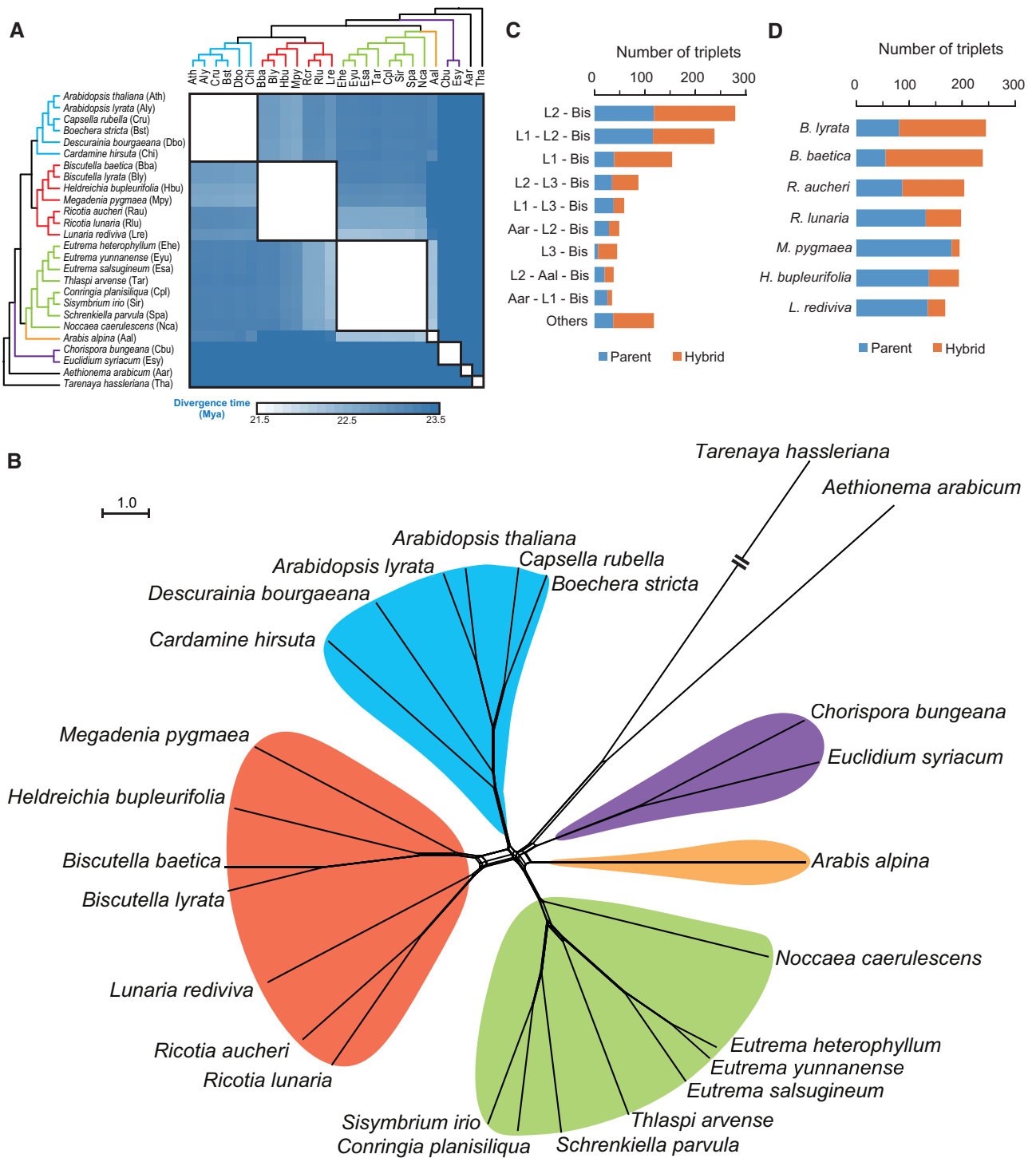


Fig. 5. Reticulate species divergence revealed by molecular dating and HyDe analyses. (A) Matrix of mean age estimates for the most recent common ancestor (MRCA) between a pair of species pairs. Only divergence times between 21.5 and 23.5 Ma are shown. (B) Network analysis using MRCA estimates between species pairs as distances. (C) Summary of significant four-taxon tests for interclade hybridizations involving the Biscutelleae tribe. (D) Summary of significant four-taxon tests in each of the Biscutelleae species. Branches representing different clades were colored following figure 4A. L1, L2, and L3 indicate species of Lineages I, II, and III, respectively. Aal, *Arabis alpina*. Aar, *Aethionema arabicum*. Bis, Biscutelleae.

sister to Lineage II + *Arabis alpina*, thus providing evidence for a distant interclade hybridization (supplementary fig. 20, Supplementary Material online).

Based on these findings, we tentatively assigned duplicated genes to subgenomes according to the local topologies within

gene trees (fig. 6A). Among the gene pairs used in subgenome divergence analyses, 45.1% (404 out of 895) and 52.0% (418 out of 804) fulfilled the required local topology (referred to as “perfect-copy” gene pairs) in *B. baetica* and *B. lyrata*, respectively (fig. 6B). We observed that 46.6% and 38.8% of the gene

Table 1. Estimated Subgenome Divergences with Ks- and Tree-Based Methods in *Biscutella*, *Lunaria*, and *Ricotia*.

Species	Distance	Estimated Age (Ma)
<i>Biscutella baetica</i>	0.359 ± 0.009	14.398 ± 0.362
<i>Biscutella lyrata</i>	0.398 ± 0.009	14.907 ± 0.411
<i>Lunaria rediviva</i>	0.265 ± 0.007	18.727 ± 0.455
<i>Ricotia aucheri</i>	0.389 ± 0.009	20.038 ± 0.466
<i>Ricotia lunaria</i>	0.360 ± 0.007	20.134 ± 0.448

pairs in *B. baetica* and *B. lyrata*, respectively, showed a sister relationship (referred to as “sister-copy” gene pairs) and thus could not be assigned to subgenomes. In *Lunaria* and *Ricotia* species, 55.9% to 61.3% of duplicated genes could be classified as “perfect-copy,” whereas 12.0% to 15.4% belonged to “sister-copy” (fig. 6B). Although the mean divergence time between “sister-copy” gene pairs was much younger than that between gene pairs of other types in all species, the extent to which in *Biscutella* was less pronounced than that in *Lunaria* and *Ricotia* (fig. 6C). ASTRAL analysis of “perfect-copy” gene trees resulted in a topology that largely corroborated the results from gene tree reconciliation analyses (fig. 6D and supplementary fig. 20, Supplementary Material online). Importantly, *Megadenia* was recovered as the genus sister to the remainder of the tribe (fig. 6D).

Discussion

Multiple Hybridization Events Explain Contentious Phylogenetic Placement of Biscutelleae

The position of Biscutelleae within the Brassicaceae family tree remained one of the most unsettled phylogenetic issues (e.g., Al-Shehbaz 2012; Nikolov et al. 2019). A sister relationship between *Biscutella* and *Megadenia* within an ITS phylogeny was recovered by German et al. (2009) and led to the formal re-establishment of Biscutelleae as a bigeneric tribe (German and Al-Shehbaz 2008). Despite expanded taxon sampling in follow-up studies, the tribe was repeatedly retrieved as part of the family's basal polytomy (Couvreur et al. 2010; Warwick et al. 2010; Koch 2012). More recent phylogenetic works expanded the limits of the tribe, to include *Heldreichia*, *Lunaria*, and *Ricotia* (Özüdoğru et al. 2015, 2017), but have not focused on its placement in the family. Although chloroplast phylogenies supported Biscutelleae as the basal clade of Lineage II (Guo et al. 2017; Mandáková et al. 2018; Walden et al. 2020a), family-wide phylogenies based on single-copy nuclear genes have suggested alternative placements of Biscutelleae (Huang et al. 2016; Kiefer et al. 2019; Nikolov et al. 2019).

Here, we consistently inferred a topology that places Biscutelleae as sister to Lineage I, in agreement with the recent phylogenomic studies (Kiefer et al. 2019; Nikolov et al. 2019). As we identified considerable proportion of genes or sites supporting alternative topologies, we reason that the contentious relationships among early diverged Brassicaceae clades, including Biscutelleae, might be caused by multiple evolutionary forces including ancient hybridizations and WGDs (fig. 7).

Biscutelleae genera were repeatedly recovered as a monophyletic clade (Couvreur et al. 2010; Koch 2012; Huang et al.

2016; Özüdoğru et al. 2017; Mandáková et al. 2018; Kiefer et al. 2019; Nikolov et al. 2019), although the number of sampled genera varied among the studies. Here, we confirmed the monophyly of Biscutelleae based on nuclear transcriptome data representing all its five genera. In the light of their now-known genome structures, the monophyly of Biscutelleae is reflecting the shared ancestry through ancPCK-like ($n = 8$) genome (fig. 7). Indeed, our phylogenetic analyses with genes assigned to subgenomes suggested monophyly of ancPCK-derived (sub)genomes and recovered *Megadenia* as being the most ancestral Biscutelleae genus (fig. 6E), congruently with the plastome phylogeny (Mandáková et al. 2018). The fact that ancPCK genome was most likely an ancestral genome of both Lineage I (ancPCK → ACK) and expanded Lineage II (ancPCK → PCK) (Geiser et al. 2016; Walden et al. 2020b), and that ancPCK × PCK recurrent hybridization preceded the origin of *Lunaria* and *Ricotia* genomes, explains not only the incongruence between phylogenetic trees, inferred with plastid (Mandáková et al. 2018) and nuclear genes (this study) but also occasionally recovered polyphyletic relationships of Biscutelleae (Warwick et al. 2010).

The Origin of the Biscutelleae Diploid–Polyploid Complex

Geiser et al. (2016) and Mandáková et al. (2017a) were first to reveal that diploid-like *Biscutella* genomes have descended from a mesotetraploid ancestor. A similar but independent origin was inferred for the mesotetraploid genus *Ricotia* (Mandáková et al. 2018). Here, we demonstrated that another mesotetraploidy preceded the origin of *Lunaria*, whereas the other two Biscutelleae genera (*Heldreichia* and *Megadenia*) retained a diploid genome structure, despite the recent autopolyploidy in *Heldreichia*.

Based on complementary cytogenomic and phylogenomic analyses, we propose that the Biscutelleae clade represents an assemblage of paleotetraploid, mesotetraploid, and neotetraploid genera, all descending from a single ancestral (At- α) paleotetraploid genome (fig. 7). The paleotetraploid genome diverged into ancPCK ($n = 8$) and more derived PCK ($n = 7$) genome presumably representing the ancient progenitor of Lineage I and expanded Lineage II, respectively. During diversification of the Biscutelleae clade the ancient ancPCK-like genome diversified into several $n = 8$ genomes, some further altered by dysploidial rearrangements to $n = 6$ (*Megadenia*) and $n = 5$ (ancestral diploid genome of *Heldreichia*). The allotetraploid Biscutelleae genomes originated independently, either by hybridization between very similar ancPCK-like ($n = 8$) genomes (8×8 ; *Biscutella*) or between more distantly related ancPCK ($n = 8$) and PCK ($n = 7$) genomes (8×7 ; *Lunaria* and *Ricotia*). As the chloroplast phylogeny does not suggest a sister relationship between *Lunaria* and *Ricotia* (Mandáková et al. 2018), these two similar allotetraploid genomes were most likely formed by recurrent hybridization between genomically very similar parental genomes. Interestingly, the two trajectories of genome evolution differentiated by the absence/presence of the ancestral PCK-like genome are congruent with fruit morphology of Biscutelleae

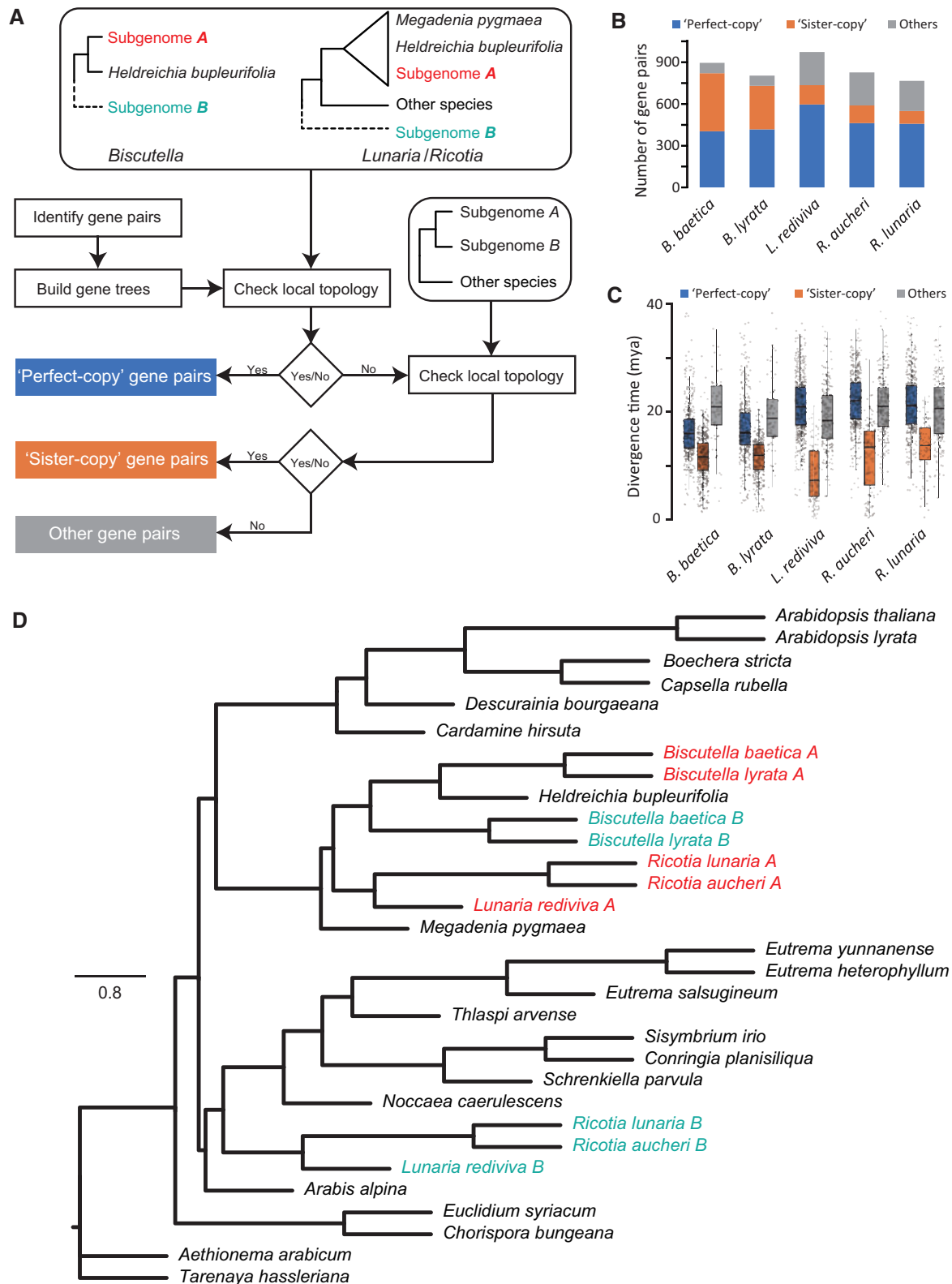


FIG. 6. Subgenome assignment for mesopolyploid Biscutelleae genomes. (A) Pipeline classifying duplicated genes according to local topologies within gene trees. For each pair of “perfect-copy” genes in *Biscutella*, one of the genes is labeled as subgenome A if it is sister to *Heldreichia*, and the other is labeled as subgenome B; in *Lunaria/Ricotia*, one of the genes is labeled as subgenome A if it forms a monophyletic clade with *Heldreichia* and *Megadenia*, the other is labeled as subgenome B with additional requirement that it should not be directly sister to the clade including subgenome A. (B) Contribution of different types of duplicated genes in Biscutelleae species. (C) Distribution of divergence times between duplicated genes. (D) The species tree inferred by ASTRAL with “perfect-copy” genes. Branches representing different Brassicaceae clades were colored following figure 4A.

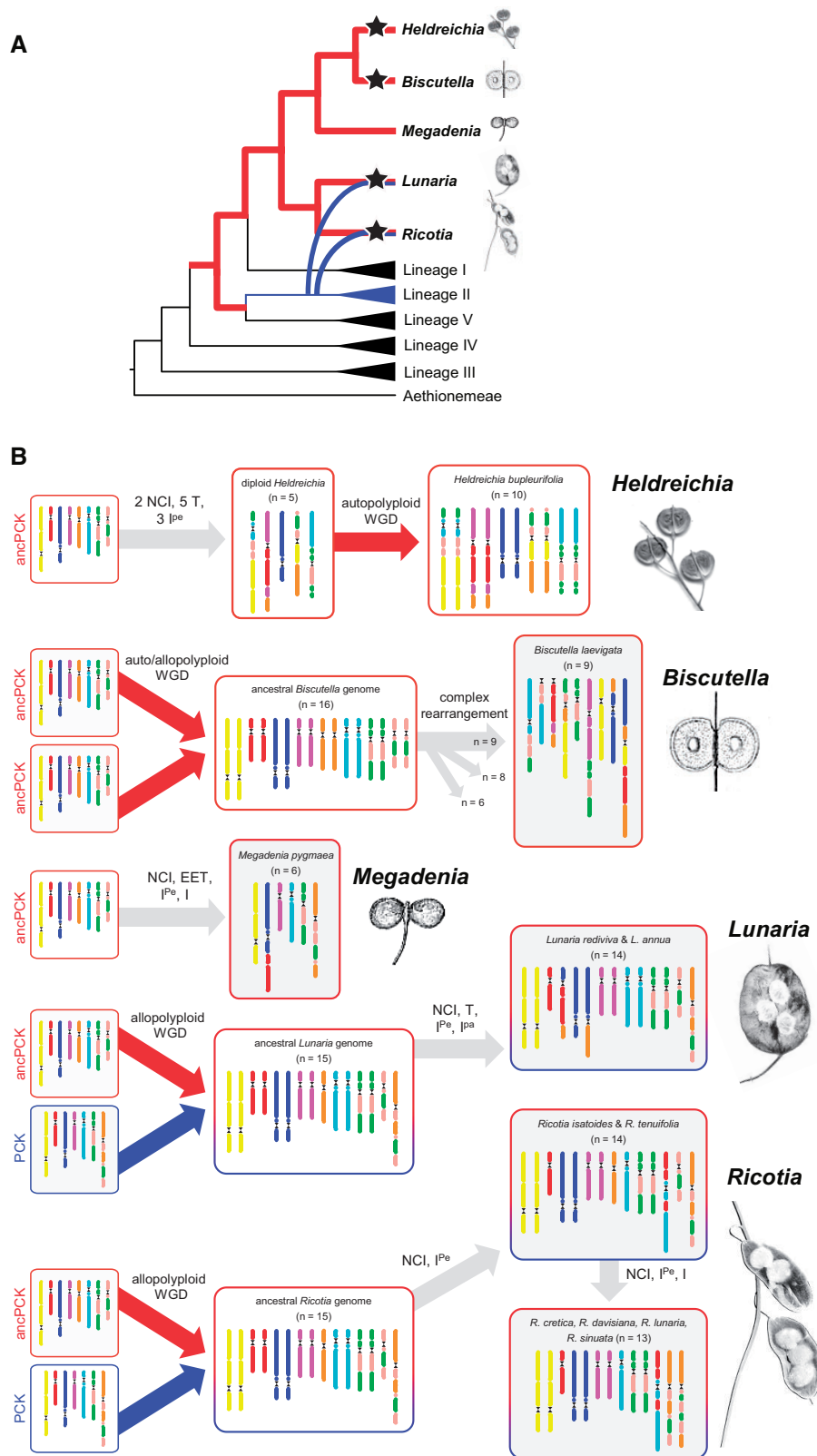


FIG. 7. The origin and evolution of the Biscutelleae diploid–polyploid genome complex. (A) A simplified phylogenetic scheme showing the position of the Biscutelleae in the family Brassicaceae (based on Nikolov et al. 2019). Red- and blue-labeled branches indicate evolutionary trajectory of ancestral ancPCK ($n=8$) and PCK ($n=7$) genomes, respectively. Star symbols indicate the genus-specific WGDs. (B) Reconstructed origin and genome evolution for individual genera of Biscutelleae. All extant Biscutelleae genomes have descended from the ancestral ancPCK-like genome (red contours and arrows). Its divergence led to the origin of ancestral diploid genomes of *Heldreischia* ($n=5$) and *Megadenia* ($n=6$). The extant *Heldreischia* genome originated via autopolyploidization. An ancestral *Biscutella* genome ($n=16$) was formed by hybridization between two ancPCK-like genomes followed by a WGD. Allotetraploid genomes of *Lunaria* and *Ricotia* ($n=15$) originated through recurrent hybridizations between ancPCK ($n=8$) and PCK genome ($n=7$; blue contours and arrows). WGDs in *Biscutella*, *Lunaria*, and *Ricotia* were followed by genus- and species-specific descending dysploidy mediated by nested chromosome insertions (NCI) and end-to-end translocations (EET). Nondysploidal rearrangements included translocations (T), as well as paracentric (I^{Pa}) and pericentric (I^{Pe}) inversions. The different colors correspond to chromosomes and genomic blocks in ancPCK and PCK genomes, centromeres are indicated by black hourglass symbols (see figs. 2 and 3 for details).

genera (Özüdoğru et al. 2017). Whereas three genera (*Biscutella*, *Heldreichia*, and *Megadenia*) containing only ancPCK genome(s) have angustiseptate fruits (flattened perpendicular to septum), *Lunaria* and *Ricotia*, containing both ancPCK and PCK genomes, possess latiseptate fruits (flattened parallel to septum) (fig. 7).

Mode and Tempo of Postpolyploid Diploidization

Although the levels of descending dysploidy in *Megadenia* (8 → 6) and *Heldreichia* (8 → 5) are comparable, both genomes differ dramatically in structure. The *Megadenia* genome originated through recombination involving four ancestral chromosomes, whereas the five *Heldreichia* chromosomes were formed by complex chromosomal rearrangements involving seven ancestral chromosomes. In mesotetraploid genomes with a comparable number of linkage groups ($n = 15$ and 16), the structural diploidization also proceeded with a contrasting intensity. Whereas chromosome numbers were reduced considerably in *Biscutella* (16 → 9, 8 and 6; 1.6- to 2.6-fold reduction), only moderate reduction occurred in *Lunaria* and *Ricotia* (15 → 14, 13; 1.07- and 1.15-fold reduction). Because our taxon sampling (*B. baetica*/*B. lyrata* and *Ricotia aucheri*/*R. lunaria*) covered the deepest split in these genera, the estimated divergence time of species pairs in *Biscutella* (~7–8 Ma) and *Ricotia* (~12–13 Ma) suggested that *Biscutella* might have originated much later than *Ricotia* (supplementary fig. 15, Supplementary Material online). When taking the *A. thaliana* genome as a reference, genomes of *Biscutella* and *Ricotia* species are evolving more than 30% and 20% faster than that of *Lunaria* (supplementary fig. 17, Supplementary Material online), whereas the mutation rates of *Megadenia* and *Heldreichia* are comparable. Thus, the rate of differential structural diploidization is not necessarily correlated with mutation rates and the age of WGDs (Mandáková et al. 2017c). In addition, our results suggest that inferring shared WGDs from K_s values can be misleading, particularly if genome evolution rates differ among the compared species.

As the *Biscutella* mesotetraploid genome originated by hybridization between two very similar ancPCK-like genomes (fig. 7), structurally more similar *Biscutella* subgenomes may have exhibited higher levels of homeologous exchange than between more differentiated subgenomes of the allotetraploid *Lunaria* and *Ricotia* genomes. This hypothesis was supported by the ratio of different gene tree topologies, where in *Biscutella* species putative homeologous genes were more likely found to be sister to each other (fig. 6D). Our phylogenetic reconciliation results also suggested that considerable signals of ancient allopolyploidization could be identified in extant paleo/mesopolyploid genomes. However, characterizing these ancient hybridization events requires sufficient taxon sampling as well as knowledge of genome evolution in closely related lineages or species (Mabry et al. 2020). For instance, the absence of duplication signals near the LCA of *Lunaria*/*Ricotia* could be interpreted as the absence of m-WGDs, whereas the gene duplications near LCA of the core Brassicaceae as a shared WGD postdating the α -WGD

(supplementary fig. 19, Supplementary Material online). This suggests that the inference of ancient WGDs using such methods needs to be complemented by additional evidence, including patterns of gene retention/loss and age distribution of gene duplicates.

Three Independent Origins of a “Fusion” Chromosome

Two independent origins of the AK5/8/6 fusion chromosome were reported for the diploid PCK genome (Mandáková and Lysak 2008) and mesotetraploid *Ricotia* genome (Mandáková et al. 2018). Here, we documented another independent origin of AK5/8/6 chromosome in the diploid *Megadenia*. These chromosomes originated from the same precursors, two ancPCK chromosomes (AK5 and AK8/6), and thus have the same GB composition. However, in each case, the fusion chromosome was formed through different dysploidal rearrangements: an EET followed by paracentric inversion in PCK, and two independent NCIs followed by a pericentric and paracentric inversion in *Ricotia* and *Megadenia*, respectively. We can hypothesize that the recurrent origin of the AK5/8/6 chromosome was due to the increased recombination rates between the ancestral chromosomes AK5 and AK8/6 and/or a reuse of chromosomal breakpoints (fig. 1). Although NCI has been considered to be a rare type of dysploidal chromosomal rearrangements in Brassicaceae (Mandáková and Lysak 2018), our data suggest that NCI represents a dominant pathway of descending dysploidy in Biscutelleae. In addition to multiple NCIs in *Ricotia* (Mandáková et al. 2018), we documented two in *Heldreichia*, and one NCI event in *Lunaria* and *Megadenia*, respectively.

Reticulate Genome Evolution in Brassicaceae and Implication for Other Land Plants

The causes of phylogenetic incongruence during rapid radiations have become a major focus of contemporary evolutionary biology. Here, we have provided clear cyto-phylogenomic evidence for reticulate genome evolution during the early divergence of Brassicaceae lineages. Given the previously inferred hybridogenous origin of tribes Shehbazieae (German and Friesen 2014) and Microlepidieae (Mandáková et al. 2017c), the hitherto documented reticulation in Biscutelleae might be only the tip of the proverbial iceberg, as at least 13 out of the 52 Brassicaceae tribes have been shown to have a mesopolyploid origin (Mandáková et al. 2017a). Phylogenetic placement of many mesopolyploid crucifer clades was contentious, as shown, for example, in a recent large-scale nuclear gene phylogeny (Nikolov et al. 2019). Even for the well-studied whole-genome triplication in *Brassica*/Brassicaceae, the relationship between the subgenomes and closely related diploid genomes remained unresolved (Hénocq L, Gallina S, Schmitt E, Castric V, Vekemans X, Poux C, unpublished data). Based on analysis of six publicly available genomes, Forsythe et al. (2020) proposed that species relationships in Brassicaceae might be obscured by massive nuclear introgression, and population genomic analyses have revealed inter-ploid gene flows in *Arabidopsis* (Monnahan et al. 2019) and *Capsella* (Han et al. 2015) species. Here, we show that the

origin of a monophyletic clade can be more complex than a parsimonious assumption of a single WGD spurring polyploid diversification and cladogenesis. Instead, recurrent hybridization among the same and/or closely related parental genomes may phylogenetically interlink diploid and polyploid genomes despite the incidence of multiple independent genome duplications.

Although the relationship between polyploidy and diversification has been under debate (Mayrose et al. 2011; Levin and Soltis 2018; Ren et al. 2018; Han et al. 2020), many important plant radiations were preceded by ancient allopolyploidization events (e.g., Estep et al. 2014; Aköz and Nordborg 2019; Wang et al. 2019). In addition, some polyploid species represent complex mosaics due to multiple hybridizations (Glemin et al. 2019; Martin et al. 2020) followed by homeologous exchanges between subgenomes (Edger et al. 2018). Despite efforts to study allopolyploidization using low-copy nuclear genes (e.g., Estep et al. 2014; Triplett et al. 2014; Huynh et al. 2019) or partial data from a genome (Kamneva et al. 2017), our results, together with several others (Edger et al. 2019; Guo et al. 2019; Mandáková et al. 2019), suggest that comparing genome structures of related species can be powerful in deconvoluting the reticulate relationships. As increasing number of tools for analyzing hybridization and allopolyploidization become available (e.g., Jones et al. 2013; Jones G, unpublished data; Thomas et al. 2017; Blischak et al. 2018), we suggest that combining multiple approaches may offer more comprehensive understanding of reticulate plant genome evolution.

Materials and Methods

Plant Material

A list of the investigated accessions and their origins is provided as [supplementary table 1, Supplementary Material](#) online. Whole young inflorescences from different individuals were fixed in freshly prepared ethanol:acetic acid (3:1) fixative overnight, transferred to 70% ethanol, and stored at -20°C until further use. Fresh leaves of *H. bupleurifolia*, *M. pygmaea*, and *R. aucheri* were collected and used for RNA isolation.

Chromosome Preparations

Mitotic and meiotic chromosome spreads from fixed young flower buds containing immature anthers were prepared as described previously (Mandáková and Lysak 2016a). Briefly, selected flower buds were rinsed in distilled water (twice for 5 min) and citrate buffer (10 mM sodium citrate, pH 4.8; twice for 5 min), and digested in 0.3% cellulase, cytohelicase, and pectolyase (all Sigma–Aldrich) in citrate buffer at 37°C for 3 h. After digestion, individual anthers were dissected and spread (20 μl of 60% acetic acid, ~ 30 s) on a microscope slide placed on a metal hot plate (50°C). The preparation was then fixed in freshly prepared fixative (ethanol:acetic acid, 3:1) by dropping the fixative around the remaining drop of acetic acid and into it. Chromosome spreads were dried using a hair dryer and checked under a phase contrast for suitable chromosome figures, largely free of cytoplasm. Suitable slides were postfixed in freshly prepared 4% formaldehyde in distilled

water for 10 min and air-dried. Preparations were kept in a dust-free box at room temperature until used.

To remove RNA and the remaining cytoplasm, the preparations were treated with 100 $\mu\text{g}/\text{ml}$ RNase (AppliChem) in $2\times$ sodium saline citrate (SSC; $20\times$ SSC: 3 M sodium chloride, 300 mM trisodium citrate, pH 7.0) for 60 min and 0.1 mg/ml pepsin (Sigma) in 0.01 M HCl at 37°C for 5 min, then postfixed in 4% formaldehyde in $2\times$ SSC for 10 min, washed in $2\times$ SSC twice for 5 min, dehydrated in an ethanol series (70%, 90%, and 100%, 2 min each) and air-dried.

DNA Probes

For comparative chromosome painting (CCP), in total 674 chromosome-specific BAC clones of *Arabidopsis thaliana* grouped into contigs according to eight chromosomes and 22 GBs of the Ancestral Crucifer Karyotype (Lysak et al. 2016) were used. To determine and characterize species-specific chromosome arrangements, after initial CCP experiments, some BAC contigs were split into smaller subcontigs. The *A. thaliana* BAC clone T15P10 (AF167571) containing 35S rRNA genes was used for in situ localization of nucleolar organizer regions (NORs), and the *A. thaliana* clone pCT4.2 (M65137), corresponding to a 500-bp 5S rDNA repeat, was used for localization of 5S rDNA loci.

All DNA probes were labeled by nick translation with biotin-dUTP, digoxigenin-dUTP, or Cy3-dUTP according to Mandáková and Lysak (2016b) as follows: 1 μg DNA diluted in distilled water to 29 μl , 5 μl nucleotide mix (2 mM dATP, dCTP, dGTP, 400 μM dTTP, all Roche), 5 μl $10\times$ NT-buffer (0.5 M Tris–HCl, pH 7.5; 50 mM MgCl_2 , 0.05% bovine serum albumin), 4 μl 1 mM X-dUTP (in which X was either biotin, digoxigenin, or Cy3), 5 μl 0.1 M β -mercaptoethanol, 1 μl DNase I (Roche), and 1 μl DNA polymerase I (Fermentas). The nick translation mixture was incubated at 15°C for 90 min (or longer) to obtain a fragment length of ~ 200 –500 bp. The nick translation reaction was stopped by adding 1 μl 0.5 M EDTA, pH 8.0, and incubation at 65°C for 10 min. Individual labeled probes were stored at -20°C until use.

In Situ Hybridization and Microscopy

Selected labeled probes were pooled to follow the design of a given experiment and precipitated by adding 1/10 volume of 3 M sodium acetate, pH 5.2, and 2.5 volume of ice-cold 96% ethanol and kept at -20°C for 30 min, and centrifuged at 13,000 g at 4°C for 30 min. The pellet was resuspended in 20 μl of hybridization mix (50% formamide and 10% dextran sulfate in $2\times$ SSC) per slide. 20 μl of the probe were pipetted on a chromosome-containing slide. Cover slips were framed by rubber cement. The probe and chromosomes were denatured together on a hot plate at 80°C for 2 min and incubated in a moist chamber at 37°C overnight.

Posthybridization washing was performed in 20% formamide in $2\times$ SSC at 42°C . The immunodetection of hapten-labeled probes was performed according to Mandáková and Lysak (2016b) as follows: biotin-dUTP was detected by avidin–Texas Red (Vector Laboratories) and amplified by goat anti-avidin–biotin (Vector Laboratories) and avidin–Texas Red; digoxigenin-dUTP was detected by mouse

antidigoxigenin (Jackson Immuno Research) and goat anti-mouse–Alexa Fluor 488 (Invitrogen). Cy3-dUTP labeled probes were observed directly. After immunodetection, chromosomes were counterstained with 4',6-diamidino-2-phenylindole (DAPI, 2 µg/ml) in Vectashield (Vector Laboratories).

Inference of Genome Structure

Painted chromosome figures were photographed using an Axioimager Z2 epifluorescence microscope (Zeiss) equipped with CoolCube CCD camera (MetaSystems). Images were acquired separately for the four fluorochromes using appropriate excitation and emission filters (AHF Analysentechnik). The four monochromatic images were pseudocolored, merged, and cropped using Photoshop CS (Adobe Systems) and Image J (National Institutes of Health) software. CCP data were interpreted in the context of 22 ancestral GBs (Schranz et al. 2006; Lysak et al. 2016) and the revealed genome structures compared with some of the known ancestral Brassicaceae genomes, namely with ACK ($n = 8$), ancestral Proto-Calepineae Karyotype (ancPCK, $n = 8$; Geiser et al. 2016), Proto-Calepineae Karyotype (PCK, $n = 7$; Mandáková and Lysak 2008) and translocation Proto-Calepineae Karyotype (tPCK, $n = 7$; Cheng et al. 2013). Similarly, the analyzed genomes were further compared with the known mesotetraploid Biscutelleae genomes, that is, *Biscutella laevigata* subsp. *varia* ($n = 9$; Geiser et al. 2016) and seven *Ricotia* genomes ($n = 13$ or 14 ; Mandáková et al. 2018).

Transcriptome Assembly and Data Processing

Total RNA was isolated from fresh leaves using RNeasy Plant Mini Kit (Qiagen). Sequencing was performed on an Illumina NextSeq500 platform. Raw reads were first corrected with Rcorrector v1.0.4 (Song and Florea 2015) and trimmed by TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, last accessed July 7, 2019) with default parameters. We also retrieved RNA-seq data of additional three species (*Chorisporea bungeana*, *Descurainia bourgaeana*, and *Ricotia lunaria*; SRX155036, SRX1834918, and SRX4225392) from SRA database. All transcriptomes were assembled using Trinity-v2.8.1 (Grabherr et al. 2011) with “-KMER_SIZE 32.” Coding regions were predicted with TransDecoder v5.3.0 (<http://transdecoder.sourceforge.net/>, last accessed July 7, 2019). Sequence redundancy was reduced by cd-hit-est v4.6.1 software (Li and Godzik 2006) with parameters “-c 0.99 -n 10 -r 0.” Gene completeness of each transcriptome was accessed by BUSCO v3 (Simão et al. 2015) with embryophyta_odb9 database. Summary statistics of RNA-seq and transcriptome assemblies are reported in [supplementary tables 3 and 4, Supplementary Material](#) online. The newly generated RNA-seq data have been deposited in NCBI's SRA database under BioProject number PRJNA634714.

Orthology Inference

Our taxon sampling included 25 species from all major Brassicaceae lineages and one outgroup species (*Tarenaya hassleriana*). Information on newly sequenced species, as well as publically available genomic data, can be found in [supplementary table 1, Supplementary Material](#) online. All-

versus-all homology search was performed for amino acid sequences with “BlastP” function in DIAMOND v0.8.29 software (Buchfink et al. 2015). The resulting sequence pairs were filtered with a hit fraction cutoff of 0.4 and clustered by MCL v14.137 (Enright et al. 2002) with parameters “-tf ‘gq(5)’ -l 1.4.” The resulting 12,046 sequence clusters were aligned by MAFFT v 7.427 (Katoh and Standley 2013) with parameters “-genafpair -maxiterate 1,000.” Alignments were trimmed by trimAl v1.4 (Capella-Gutiérrez et al. 2009) with parameters “-automated1 -splitbystopcodon” and back-translated into aligned nucleotide sequences. Gene trees were inferred by IQ-TREE v1.6.10 (Nguyen et al. 2015) and trimmed with TreeShrink v1.3.1 (Mai and Mirarab 2018). Following Yang and Smith (2014), ambiguous and deep-paralogous tips were trimmed if the internal branch length was longer than 0.4. The process of sequence alignment, tree estimation, and branch-cutting was repeated three times. Finally, we applied both the “monophyletic outgroup” (MO) and the “rooted ingroup” (RT) approaches (Yang and Smith 2014) to recover single-copy orthologous groups that include at least 20 species for each locus.

Inference of WGDs with Large-Scale Gene Duplications

Coding sequences of each Biscutelleae species were searched against themselves in MegaBLAST mode with the BlastN program in BLAST+ suite v2.7.1 (Camacho et al. 2009). We removed sequence pairs that belong to the same gene as identified by Trinity. We further applied a hit fraction cutoff of 0.4 and required the matched region to be larger than 150 bp with identity less than 99.5%. Sequence alignment was performed as described above (see Orthology Inference). In total, 1,687–5,545 duplicated gene pairs were retrieved from each of the transcriptomes, which could be assigned to 8,485 previously identified homologous gene clusters ([supplementary fig. 14, Supplementary Material](#) online). K_s values for pairwise comparisons were calculated with codeml program in the PAML v 4.9e package (Yang 2007). Mean K_s values were determined by Gaussian mixture modeling with “mclust” R package (Scrucca et al. 2016).

To compare the relative timing of speciation and subgenome divergence, we estimated K_s values of ortholog divergence between species pairs of different Biscutelleae genera with OrthoPipes (Barker et al. 2010). We also applied the Multi-tAxon Paleopolyploidy Search (MAPS) algorithm (Li et al. 2015) with default settings to infer the bursts of gene duplication on a simplified topology of Biscutelleae species according to Walden et al. (2020a). Following Mandáková et al. (2017a), we further assigned the identified paralogs to different WGDs based on the K_s ranges from mixture modeling. Gene ontology enrichment analysis was conducted with topGO R package (Alexa et al. 2006).

Species Tree Estimation

We employed both maximum likelihood (ML) and coalescent methods that were based on gene trees to infer species tree.

ML trees were estimated from both concatenated and individual gene alignments by IQ-TREE v1.6.10 (Nguyen et al. 2015), which used ModelFinder (Kalyaanamoorthy et al. 2017) to identify the best-fitted substitution model. Branch support values were provided by 1,000 replicates of ultrafast bootstrap approximation (Hoang et al. 2018). For coalescent-based species tree inference, gene trees were used as input for ASTRAL v 5.6.3 (Mirarab et al. 2014), MP-EST v2.0 (Liu et al. 2010), and STELLS v2.1.0 (Pei and Wu 2017), respectively. Given that bootstrap value may be a poor indicator of branch support in genome-wide data sets (Salichos and Rokas 2013), we annotated branches with quartet scores and posterior probabilities from ASTRAL analysis. In addition, we performed Bayesian concordance analysis with BUCKy v1.4.4 (Ané et al. 2006; Larget et al. 2010) to summarize species tree and calculate concordance factors for each branch.

Phylogenetic Signal Analysis

To compare the support for alternative topologies, we used a previously proposed method (Shen et al. 2017) to calculate gene- and site-wise delta log-likelihood scores (Δ GSL and Δ SSL). Log-likelihood values for genes and sites under different tree hypotheses were calculated using IQ-TREE v1.6.10 (Nguyen et al. 2015). The topology that received the highest score was recorded as the most favored hypothesis. We investigated five topologies recovered in this and previous studies using both nuclear (Huang et al. 2016; Nikolov et al. 2019) and chloroplast genomic markers (Mandáková et al. 2018). Because tribe Arabideae was not included in the study by Huang et al. (2016), we constrained its representative species (here *Arabis alpina*) to be sister either to the clade including Lineage I, Lineage II, and tribe Biscutelleae (T3) or to Lineage II alone (T4). Outlier genes with Δ GSL values that fell outside of the range between upper and lower whiskers of a boxplot were identified in R (Ihaka and Gentleman 1996). Outlier sites were defined as those with the 1% highest Δ SSL scores within a given locus. After excluding outlier genes and sites, we performed phylogenetic analyses with IQ-TREE and ASTRAL analyses as described above.

Molecular Dating

Molecular dating of gene trees was performed with BEAST v2.5.0 (Bouckaert et al. 2014) assuming the monophyly of all crucifer species. The divergence between Brassicaceae and Cleomaceae was set as a normal-distribution prior (mean $47.8 \pm$ SD 2.9 Ma) according to the fossil record for *Palaeocleome lakensis* (Cardinal-McTeague et al. 2016). For all genes, we applied the nucleotide substitution model HKY + Γ , a relaxed log-normal clock model, and a Yule tree prior. The MCMC chain length was set to 2×10^7 with a pre-burnin value of 500,000 and sample frequency of 2,000. The maximum clade credibility tree was summarized with TreeAnnotator v2.5.0. Pairwise species divergence times from each time tree were extracted using the “cophenetic” function in the “ape” package (Paradis and Schliep 2019) in R. Mean divergence times of species pairs were estimated by Gaussian mixture modeling with R package “mclust” (Scrucca et al. 2016). The network based on pairwise mean divergence

times was visualized with Splitstree v4.15.1 (Huson et al. 1998).

For molecular dating with a fix species tree, we performed analyses separately for each of the five alternative topologies mentioned above, using concatenated alignment of genes that supported the hypothesis. Divergence time was estimated using three codon partitions with MCMCTREE in the PAML v 4.9e package (Yang 2007), assuming clock model of independent rates. The root age was fixed at a range between 40 and 60 Ma according to Guo et al. (2017). The overall substitution rate (rgene gamma) was set at G (1, 1.42) based on the estimation by BASEML (in PAML). The rate-drift parameter (sigma2 gamma) was set at G (1, 10). After a burn-in period of two million generations, the MCMC run was sampled every 800 generations until a total of 10,000 samples were collected. The analysis was considered to reach a stationary phase with Tracer v1.7.1 (Rambaut et al. 2018) if its effective sampling size (ESS) of parameters was over 200.

Species Network and Hybridizations

We applied the NeighborNet algorithm as implemented in Splitstree v4.15.1 (Huson et al. 1998) to build a phylogenetic network with uncorrected distances (Bryant and Moulton 2003), with 1,000 bootstrap replicates. Four-taxon tests of hybridization were conducted with the HyDe software (Blischak et al. 2018). Putative hybridizations were also investigated based on gene trees in which more than 75% of nodes received a high bootstrap value of at least 75. We inferred phylogenetic networks under zero to five reticulation scenarios with the pseudolikelihood method “InferNetwork MPL” (Yu and Nakhleh 2015) implemented in PhyloNet v5.7.1 (Wen et al. 2018). Analysis of each scenario was repeated three times to test the stability of the recovered reticulation nodes.

Phylogenomic Analysis of Mesopolyploid WGDs

Here, we combined *Ks*- and phylogeny-based methods to further investigate the meso-tetraploidization in *Biscutella*, *Lunaria*, and *Ricotia*. To avoid introducing paralogs derived from α -WGD, a subset of orthologous groups (4,106 out of 12,046) was selected by requiring that genes should present as single-copy in diploid species. Thus, the resulting multicopy genes in mesopolyploid species were most likely derived from subsequent WGD event(s). *Heldreichia* was treated as diploid species here because it represented a recently formed auto-tetraploid and should not affect the result. To further mitigate the effect of sequence redundancy, we recalculated distribution of *Ks* values for the identified gene groups for each species using a node-weighted approach (Zwaenepoel and Van de Peer 2019).

Phylogenetic reconciliation analyses were performed with PUG (McKain et al. 2016) and GRAMPA (Thomas et al. 2017). Because GRAMPA works with multilabeled (MUL) gene trees but focuses on one WGD event, we separately prepared data sets for each species by keeping only one pair of duplicates in the orthologous groups. For each alignment, gene tree inference and molecular dating were performed simultaneously with BEAST v2.5.0 (Bouckaert et al. 2014) as described above

(see Molecular Dating). The topology of T1 was used as the species tree. Because of the ambiguous placement of *Arabis alpina*, we also performed analyses with alternative species tree hypotheses. Divergence times between duplicated genes were extracted for Gaussian mixture modeling. To allow direct comparison of estimated subgenome divergence between different methods, we also calculated *Ks* values for the same gene pairs. In addition, we applied our pipeline to compare the time of subgenome divergence in three *Brassica* species, using the well-annotated gene triplicates representing the three subgenomes (Parkin et al. 2014; Zhang et al. 2018; Perumal et al. 2020). To evaluate mutation rates heterogeneity in Biscutelleae species, we performed *Ks* analyses with orthologous gene pairs identified by reciprocal-best BLAST hit between Biscutelleae species and *A. thaliana*.

Due to the lack of reference genomes, we used a custom Perl script to assign the duplicated genes to their corresponding subgenomes based on the reconciliation results. Here, we aimed to identify “perfect-copy” genes (Guo et al. 2019) which represented homoeologous genes from different subgenomes. According to the results from GRAMPA analyses, we labeled gene duplicates according to local topologies: in *Biscutella*, the copy that is sister to *Heldreichia* was defined as representing subgenome A and its counterpart represented subgenome B; in *Lunaria* and *Ricotia*, the copy that forms a monophyletic Biscutelleae clade with *Heldreichia* and *Megadenia* genes was labeled as subgenome A, and the other copy was labeled as subgenome B with an additional requirement that the gene is not a direct sister to the Biscutelleae clade containing subgenome A (see fig. 6A). We noticed that not all duplicated genes can be properly assigned to subgenomes. For a pair of genes that are sister to each other, we referred to as “sister-copy” (fig. 6A) and discarded them in phylogenetic analysis. The remaining patterns were referred to as “others” and also discarded. Then, we inferred gene trees for orthologous groups with subgenome-labeled genes as described above. Finally, we summarized a nuclear multigene phylogeny from gene trees with ASTRAL v5.6.3 (Mirarab et al. 2014).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was financially supported by the CEITEC 2020 Project (Grant No. LQ1601) and a research grant from the Czech Science Foundation (20-03419Y). We thank Dr K. Skokanová for providing the Slovak populations of *Lunaria rediviva*, and Prof. D. Lakušić for providing seeds of *L. telekiana*. Core Facility Plants Sciences of CEITEC MU is acknowledged for the cultivation of plants used in this paper. Computational resources were supplied by the project “e-Infrastruktura CZ” (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

Data Availability

Raw sequencing data generated in this study have been deposited in NCBI’s SRA database under BioProject number PRJNA634714.

References

- Aköz G, Nordborg M. 2019. The *Aquilegia* genome reveals a hybrid origin of core eudicots. *Genome Biol.* 20(1):256.
- Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22(13):1600–1607.
- Al-Shehbaz IA. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* 61(5):931–954.
- Ané C, Larget B, Baum DA, Smith SD, Rokas A. 2006. Bayesian estimation of concordance among gene trees. *Mol Biol Evol.* 24(2):412–426.
- Barker MS, Dlugosch KM, Dinh L, Challa RS, Kane NC, King MG, Rieseberg LH. 2010. EvoPipes.net: bioinformatic tools for ecological and evolutionary genomics. *Evol Bioinform Online.* 6:143–149.
- Blischak PD, Chifman J, Wolfe AD, Kubatko LS. 2018. HyDe: a Python package for genome-scale hybridization detection. *Syst Biol.* 67(5):821–829.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 10(4):e1003537.
- Bryant D, Moulton V. 2003. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol.* 21(2):255–265.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 12(1):59–60.
- Buggs RJA, Soltis PS, Soltis DE. 2011. Biosystematic relationships and the formation of polyploids. *Taxon* 60(2):324–332.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
- Cardinal-McTeague WM, Sytsma KJ, Hall JC. 2016. Biogeography and diversification of Brassicales: a 103 million year tale. *Mol Phylogenet Evol.* 99:204–224.
- Chapman MA, Burke JM. 2007. Genetic divergence and hybrid speciation. *Evolution* 61(7):1773–1780.
- Cheng F, Mandáková T, Wu J, Xie Q, Lysak MA, Wang X. 2013. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *Plant Cell* 25(5):1541–1554.
- Couvreur TL, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA, Mummenhoff K. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol Biol Evol.* 27(1):55–71.
- Darlington CD. 1937. Recent advances in cytology. 2nd ed. Philadelphia (PA): P. Blakiston’s.
- Dodsworth S, Chase MW, Leitch AR. 2016. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot J Linn Soc.* 180(1):1–5.
- Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci U S A.* 112(27):8362–8366.
- Edger PP, McKain MR, Bird KA, VanBuren R. 2018. Subgenome assignment in allopolyploids: challenges and future directions. *Curr Opin Plant Biol.* 42:76–80.
- Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson ADL, Wai CM, et al. 2019. Origin and evolution of the octoploid strawberry genome. *Nat Genet.* 51(3):541–547.

- Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30(7):1575–1584.
- Estep MC, McKain MR, Vela Diaz D, Zhong J, Hodge JG, Hodkinson TR, Layton DJ, Malcomber ST, Pasquet R, Kellogg EA. 2014. Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc Natl Acad Sci U S A.* 111(42):15149–15154.
- Forsythe ES, Nelson AD, Beilstein MA. 2020. Biased gene retention in the face of introgression obscures species relationships. *Genome Biol Evol.* 12(9):1646–1663.
- Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K. 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* 16(2):108–116.
- Geiser C, Mandáková T, Arrigo N, Lysak MA, Parisod C. 2016. Repeated whole-genome duplication, karyotype reshuffling, and biased retention of stress-responding genes in buckler mustard. *Plant Cell* 28(1):17–27.
- German DA, Al-Shehbaz IA. 2008. Five additional tribes (Aphragmeae, Biscutelleae, Calepineae, Conringieae, and Erysimeae) in the Brassicaceae (Cruciferae). *Harvard Pap Bot.* 13(1):165–170.
- German DA, Friesen N, Neuffer B, Al-Shehbaz IA, Hurka H. 2009. Contribution to ITS phylogeny of the Brassicaceae, with special reference to some Asian taxa. *Plant Syst Evol.* 283(1–2):33–56.
- German DA, Friesen NW. 2014. *Shehbazia* (Shehbazieae, Cruciferae), a new monotypic genus and tribe of hybrid origin from Tibet. *Turczaninowia* 17(4):17–23.
- Glemin S, Scornavacca C, Dainat J, Burgarella C, Viader V, Ardisson M, Sarah G, Santoni S, David J, Ranwez V. 2019. Pervasive hybridizations in the history of wheat relatives. *Sci Adv.* 5(5):eaav9188.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652.
- Grant V. 1981. Plant speciation. New York: Columbia University Press.
- Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, Zhang D, Ma T, Hu Q, Al-Shehbaz IA, et al. 2017. Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* 18(1):176.
- Guo Z-H, Ma P-F, Yang G-Q, Hu J-Y, Liu Y-L, Xia E-H, Zhong M-C, Zhao L, Sun G-L, Xu Y-X, et al. 2019. Genome sequences provide insights into the reticulate origin and unique traits of woody bamboos. *Mol Plant.* 12(10):1353–1365.
- Han TS, Wu Q, Hou XH, Li ZW, Zou YP, Ge S, Guo Y. 2015. Frequent introgressions from diploid species contribute to the adaptation of the tetraploid Shepherd's purse (*Capsella bursa-pastoris*). *Mol Plant.* 8(3):427–438.
- Han TS, Zheng QJ, Onstein RE, Rojas-Andrés BM, Hauenschild F, Muellner-Riehl AN, Xing YW. 2020. Polyploidy promotes species diversification of *Allium* through ecological shifts. *New Phytol.* 225(1):571–583.
- Harriman NA. 1978. IOPB chromosome number reports LX. *Taxon* 27:228.
- Haudry A, Platts AE, Vello E, Hoen D, Leclercq M, Williamson R, Forczek E, Joly-Lopez Z, Steffen J, Hazzouri KM, et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat Genet.* 45(8):89–8.
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35(2):518–522.
- Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* 27(10):2770–2784.
- Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP, et al. 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol Biol Evol.* 33(2):394–412.
- Huson DH, Klopper T, Bryant D. 1998. SplitsTree 4.0 – computation of phylogenetic trees and networks. *Bioinformatics* 14(1):68–73.
- Huynh S, Marcussen T, Felber F, Parisod C. 2019. Hybridization preceded radiation in diploid wheats. *Mol Phylogenet Evol.* 139:106554.
- Ihaka R, Gentleman R. 1996. R: a language for data analysis and graphics. *J Comp Graph Stat.* 5:299–314.
- Jones G, Sagitov S, Oxelman B. 2013. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. *Syst Biol.* 62(3):467–478.
- Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14(6):587–589.
- Kamneva OK, Syring J, Liston A, Rosenberg NA. 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evol Biol.* 17(1):180.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kiefer C, Willing E-M, Jiao W-B, Sun H, Piednoël M, Hümann U, Hartwig B, Koch MA, Schneeberger K. 2019. Interspecies association mapping links reduced CG to TG substitution rates to the loss of gene-body methylation. *Nat Plants.* 5(8):846–855.
- Koch MA. 2012. Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). *Taxon* 61(1):76–92.
- Koch MA, German DA, Kiefer M, Franzke A. 2018. Database taxonomics as key to modern plant biology. *Trends Plant Sci.* 23(1):4–6.
- Large BR, Kotha SK, Dewey CN, Ané C. 2010. BUCKY: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26(22):2910–2911.
- Levin DA. 2013. The timetable for allopolyploidy in flowering plants. *Ann Bot.* 112(7):1201–1208.
- Levin DA, Soltis DE. 2018. Factors promoting polyploid persistence and diversification and limiting diploid speciation during the K–Pg interlude. *Curr Opin Plant Biol.* 42:1–7.
- Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS. 2015. Early genome duplications in conifers and other seed plants. *Sci Adv.* 1(10):e1501084.
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
- Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol.* 10(1):302.
- Lohaus R, Van de Peer Y. 2016. Of dups and dinos: evolution at the K/Pg boundary. *Curr Opin Plant Biol.* 30:62–69.
- Lysak MA, Mandáková T, Schranz ME. 2016. Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Curr Opin Plant Biol.* 30:108–115.
- Mabry ME, Brose JM, Blischak PD, Sutherland B, Dismukes WT, Bottoms CA, Edger PP, Washburn JD, An H, Hall JC, et al. 2020. Phylogeny and multiple independent whole-genome duplication events in the Brassicales. *Am J Bot.* 107(8):1148–1164.
- Mai U, Mirarab S. 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* 19(S5):272.
- Mallet J. 2007. Hybrid speciation. *Nature* 446(7133):279–283.
- Mandáková T, Guo X, Özüdoğru B, Mummenhoff K, Lysak MA. 2018. Hybridization-facilitated genome merger and repeated chromosome fusion after eight million years. *Plant J.* 96(4):748–760.
- Mandáková T, Hloušková P, German DA, Lysak MA. 2017b. Monophyletic origin and evolution of the largest crucifer genomes. *Plant Physiol.* 174(4):2062–2071.
- Mandáková T, Hloušková P, Koch MA, Lysak MA. 2020. Genome evolution in Arabideae was marked by frequent centromere repositioning. *Plant Cell* 32(3):650–665.
- Mandáková T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA. 2010. Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell* 22(7):2277–2290.
- Mandáková T, Li Z, Barker MS, Lysak MA. 2017a. Diverse genome organization following 13 independent mesopolyploid events in

- Brassicaceae contrasts with convergent patterns of gene retention. *Plant J.* 91(1):3–21.
- Mandáková T, Lysak MA. 2008. Chromosomal phylogeny and karyotype evolution in $x = 7$ crucifer species (Brassicaceae). *Plant Cell* 20(10):2559–2570.
- Mandáková T, Lysak MA. 2016a. Chromosome preparation for cytogenetic analyses in Arabidopsis. *Curr Protocols Plant Biol.* 1(1):43–51.
- Mandáková T, Lysak MA. 2016b. Painting of Arabidopsis chromosomes with chromosome-specific BAC clones. *Curr Protoc Plant Biol.* 1(2):359–371.
- Mandáková T, Lysak MA. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Curr Opin Plant Biol.* 42:55–65.
- Mandáková T, Mummenhoff K, Al-Shehbaz IA, Mucina L, Mühlhausen A, Lysak MA. 2012. Whole-genome triplication and species radiation in the southern African tribe Heliophileae (Brassicaceae). *Taxon* 61(5):989–1000.
- Mandáková T, Pouch M, Brock JR, Al-Shehbaz IA, Lysak MA. 2019. Origin and evolution of diploid and allopolyploid *Camelina* genomes were accompanied by chromosome shattering. *Plant Cell* 31(11):2596–2612.
- Mandáková T, Pouch M, Harmanová K, Zhan SH, Mayrose I, Lysak MA. 2017c. Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Mol Ecol.* 26(22):6445–6462.
- Marhold K, Lihová J. 2006. Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. *Plant Syst Evol.* 259(2–4):143–174.
- Martin G, Cardí C, Sarah G, Ricci S, Jenny C, Fondi E, Perrier X, Glaszmann J-C, D'Hont A, Yahiaoui N, et al. 2020. Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *Plant J.* 102(5):1008–1025.
- Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333(6047):1257–1257.
- McKain MR, Tang H, McNeal JR, Ayyampalayam S, Davis JJ, dePamphilis CW, Givnish TJ, Pires JC, Stevenson DW, Leebens-Mack JH, et al. 2016. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biol Evol.* 8(4):1150–1164.
- Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30(17):i541–i548.
- Monnahan P, Kolář F, Baduel P, Sailer C, Koch J, Horvath R, Laenen B, Schmickl R, Paajanen P, Šrámková G, et al. 2019. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat Ecol Evol.* 3(3):457–468.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, Bailey CD, Tsiantis M. 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytol.* 222(3):1638–1651.
- One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679.
- Özüdoğru B, Akaydin G, Erik S, Al-Shehbaz IA, Mummenhoff K. 2015. Phylogeny, diversification and biogeographic implications of the eastern Mediterranean endemic genus *Ricotia* (Brassicaceae). *Taxon* 64(4):727–740.
- Özüdoğru B, Al-Shehbaz IA, Mummenhoff K. 2017. Tribal assignment of *Heldreichia* Boiss. (Brassicaceae): evidence from nuclear ITS and plastidic *ndhF* markers. *Plant Syst Evol.* 303(3):329–335.
- Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3):526–528.
- Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al. 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol.* 15(6):1–18.
- Paun O, Forest F, Fay MF, Chase MW. 2009. Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytol.* 182(2):507–518.
- Pei J, Wu Y. 2017. STELLS2: fast and accurate coalescent-based maximum likelihood inference of species trees from gene tree topologies. *Bioinformatics* 33(12):1789–1797.
- Perumal S, Koh CS, Jin L, Buchwaldt M, Higgins EE, Zheng C, Sankoff D, Robinson SJ, Kagale S, Navabi Z-K, et al. 2020. A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral *Brassica* genome. *Nat Plants.* 6(8):929–941.
- Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 20(1):38.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 67(5):901–904.
- Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol Plant.* 11(3):414–428.
- Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497(7449):327–331.
- Schranz ME, Lysak MA, Mitchell-Olds T. 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* 11(11):535–542.
- Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J.* 8(1):289.
- Servick S, Visger CJ, Gitzendanner MA, Soltis PS, Soltis DE. 2015. Population genetic variation, geographic structure, and multiple origins of autopolyploidy in *Galax urceolata*. *Am J Bot.* 102(6):973–982.
- Shen XX, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat Ecol Evol.* 1(5):0126.
- Sharma AK. 1970. Annual report, 1967–1968. Res Bull Univ Calcutta Cytogenetics Lab. 2:1–50.
- Shimizu-Inatsugi RIE, Lihová J, Iwanaga H, Kudoh H, Marhold K, Savolainen OUTI, Watanabe K, Yakubov VV, Shimizu KK. 2009. The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Mol Ecol.* 18(19):4024–4048.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Soltis DE, Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends Ecol Evol.* 14(9):348–352.
- Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. *Curr Opin Genet Dev.* 35:119–125.
- Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol.* 60(1):561–588.
- Song L, Florea L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 4(1):48.
- Spoelhof JP, Soltis PS, Soltis DE. 2017. Pure polyploidy: closing the gaps in autopolyploid research. *J Syst Evol.* 55(4):340–352.
- Thomas GW, Ather SH, Hahn MW. 2017. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Syst Biol.* 66(6):1007–1018.
- Triplett JK, Clark LG, Fisher AE, Wen J. 2014. Independent allopolyploidization events preceded speciation in the temperate and tropical woody bamboos. *New Phytol.* 204(1):66–73.
- Uhriková A. 1976. In: Májovský J, et al., Index of chromosome numbers of Slovakian flora. Part 5. *Acta Fac Rerum Nat Univ Comeniana, Bot.* 25:1–18.
- Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in Arabidopsis. *Science* 290(5499):2114–2117.
- Wagner F, Ott T, Zimmer C, Reichhart V, Vogt R, Oberprieler C. 2019. 'At the crossroads towards polyploidy': genomic divergence and extent

- of homoploid hybridization are drivers for the formation of the ox-eye daisy polyploid complex (*Leucanthemum*, Compositae-Anthemideae). *New Phytol.* 223(4):2039–2053.
- Walden N, German DA, Wolf EM, Kiefer M, Rigault P, Huang X-C, Kiefer C, Schmickl R, Franzke A, Neuffer B, et al. 2020a. Nested whole-genome duplications coincide with diversification and high morphological disparity over geological times in the Brassicaceae. *Nat Commun.* 11:1–12.
- Walden N, Nguyen TP, Mandáková T, Lysak MA, Schranz ME. 2020b. Genomic blocks in *Aethionema arabicum* support Arabideae as next diverging clade in Brassicaceae. *Front Plant Sci.* 11:719.
- Wang J, Qin J, Sun P, Ma X, Yu J, Li Y, Sun S, Lei T, Meng F, Wei C, et al. 2019. Polyploidy index and its implications for the evolution of polyploids. *Front Genet.* 10:807.
- Warwick SI, Mummenhoff K, Sauder CA, Koch MA, Al-Shehbaz IA. 2010. Closing the gaps: phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. *Plant Syst Evol.* 285(3–4):209–232.
- Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. *Syst Biol.* 67(4):735–740.
- Yang W, Zhang L, Mandáková T, Huang L, Li T, Jiang J, Yang Y, Lysak MA, Liu J, Hu Q. 2020. The chromosome-level genome sequence and karyotypic evolution of *Megadenia pygmaea* (Brassicaceae). *Mol Ecol Resour.* Available from: <https://doi.org/10.1111/1755-0998.13291>.
- Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol Biol Evol.* 31(11):3081–3092.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yu Y, Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics* 16(S10):S10.
- Zhang L, Cai X, Wu J, Liu M, Grob S, Cheng F, et al. 2018. Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Horticulture Res.* 5:1–11.
- Zwaenepoel A, Van de Peer Y. 2019. wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35(12):2153–2155.