the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2021) **108,** 1768–1785 doi: 10.1111/tpj.15542

Genome structure and evolution in the cruciferous tribe Thlaspideae (Brassicaceae)

Soheila Bayat[1,2], Martin A. Lysak[1,2,*] and Terezie Mandáková[1,3,*]
[1]*CEITEC, Masaryk University, Brno 62500, Czech Republic,*
[2]*National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno 62500, Czech Republic, and*
[3]*Department of Experimental Biology, Faculty of Science, Masaryk University, Brno 62500, Czech Republic*

Received 10 May 2021; revised 30 September 2021; accepted 11 October 2021; published online 26 November 2021.
*For correspondence (e-mail terezie.mandakova@ceitec.muni.cz; martin.lysak@ceitec.muni.cz).

SUMMARY

Whole-genome duplications (WGDs) and chromosome rearrangements (CRs) play the key role in driving the diversification and evolution of plant lineages. Although the direct link between WGDs and plant diversification is well documented, relatively few studies focus on the evolutionary significance of CRs. The cruciferous tribe Thlaspideae represents an ideal model system to address the role of large-scale chromosome alterations in genome evolution, as most Thlaspideae species share the same diploid chromosome number ($2n = 2x = 14$). Here we constructed the genome structure in 12 Thlaspideae species, including field pennycress (*Thlaspi arvense*) and garlic mustard (*Alliaria petiolata*). We detected and precisely characterized genus- and species-specific CRs, mostly pericentric inversions, as the main genome-diversifying drivers in the tribe. We reconstructed the structure of seven chromosomes of an ancestral Thlaspideae genome, identified evolutionary stable chromosomes versus chromosomes prone to CRs, estimated the rate of CRs, and uncovered an allohexaploid origin of garlic mustard from diploid taxa closely related to *A. petiolata* and *Parlatoria cakiloidea*. Furthermore, we performed detailed bioinformatic analysis of the Thlaspideae repeatomes, and identified repetitive elements applicable as unique species- and genus-specific barcodes and chromosome landmarks. This study deepens our general understanding of the evolutionary role of CRs, particularly pericentric inversions, in plant genome diversification, and provides a robust base for follow-up whole-genome sequencing efforts.

Keywords: chromosome rearrangements, garlic mustard, genome evolution, pennycress, pericentric inversions, repetitive DNA, Thlaspideae, Brassicaceae.

INTRODUCTION

Flowering plants with about 352 000 species divided in 405 families and 14 559 genera show an outstanding level of biodiversity (APG IV, 2016). This indicates that flowering plants have undergone an extensive diversification and adaptive radiation generating new species with new morphological and physiological traits enabling them to cope with rapid environmental changes (Agrawal et al., 2009; Bateman et al., 1998; Graham et al., 2000). On the genomic level, it has been well documented that whole-genome duplications (WGDs), changes in chromosome number (descending/ascending dysploidy caused by chromosome fusion/fission events) and structure (caused by non-dysploidal chromosome rearrangements, CRs, such as inversions) played the key role in driving the diversification and evolution of plant lineages (De Storme and Mason, 2014; Escudero et al., 2014; Husband, 2004; Madlung, 2013).

Mustard family (Brassicaceae) with 351 genera and 3977 species is one of the most diverse flowering plant families (https://brassibase.cos.uni-heidelberg.de/). The massive species radiation leading to generation of several Brassicaceae lineages is generally considered as a consequence of the family-specific WGD (At-α; 35 million years ago; Walden et al., 2020). The At-α was followed by lineage-specific post-polyploid diploidization and neopolyploidization events. Out of 52 monophyletic Brassicaceae tribes (https://brassibase.cos.uni-heidelberg.de/), at least 10 tribes underwent an independent tribe-specific WGD (mesopolyploid WGDs; Anastaticeae, Biscutelleae, Brassiceae, Cochlearieae, Heliophileae, Iberideae, Microlepidieae, Physarieae, Schizopetaleae and Thelypodieae), mostly followed by species-specific post-polyploid diploidizations and/or neopolyploidizations (Cheng et al., 2014; Guo et al., 2021; Lysak et al., 2005; Mandáková et al., 2010, 2017a; Walden et al.,

 © 2021 Society for Experimental Biology and John Wiley & Sons Ltd

2020). Although the direct link between WGD and diversification of plant lineages is well documented (Barker et al., 2009; Cheng et al., 2014; Mandáková et al., 2012, 2014, 2017a, 2020; Ren et al., 2019; Winterfeld et al., 2018; Yang et al., 2020), there are only sparse cases demonstrating the role of CRs in producing genome variations that might lead to reproductive isolation and consequently speciation in plants (Barb et al., 2014; Fishman et al., 2013; Jang et al., 2013; Lowry and Willis, 2010; Oneal et al., 2014). To improve our understanding of the role of CRs in genome evolution, more comprehensive studies are required.

In this study, we addressed the genome structure and evolution in the cruciferous tribe Thlaspideae. Thlaspideae includes 13 genera and 42 annual and perennial species that are mostly distributed in South West Asia, South and Central Europe. The tribe Thlaspideae comprises a number of ecologically and economically valuable species. Field pennycress (*Thlaspi arvense*) is one of the most important ones. This weedy annual species can produce up to 1300 kg hectare^{-1} seeds (Sedbrook et al., 2014) with high oil content (Moser et al., 2009a). The oil from seeds of *T. arvense* is appropriate for biodiesel production (Chopra et al., 2020; Moser et al., 2009b), and the remaining part of the seed is high in protein and can be used as feedstock. In spite of previous intense and significant efforts (Dorn et al., 2015; Geng et al., 2021; McGinn et al., 2019), the genome structure of field pennycress is still not fully resolved. Further investigation of the field pennycress genome and chromosome structure is needed to transform this wild species to a new crop, especially in areas with severe winters as *T. arvense* is able to tolerate extreme cold weather (Cubins et al., 2019; Warwick et al., 2002).

Another ecologically important representative of the Thlaspideae is garlic mustard (*Alliaria petiolata*). *Alliaria petiolata* is native to West Asia and Europe, but it is also naturalized in North America, Argentina and North Africa (Esmailbegi et al., 2018). In its native area, *A. petiolata* forms relatively small populations, and can be considered as a useful plant to increase the crop yields by breaking the life cycle of soil born diseases and pests (Brown and Morra, 1997; Meekins and McCarthy, 1999). However, in its naturalized area, garlic mustard is considered as one of the most problematic invaders as it decreases the abundance of mycorrhizal fungi in the soil and on plant roots, inhibits growth of native plants, and causes declines in the diversity of native plant communities (Burke, 2008). Phylogenetically, *A. petiolata* is the only polyphyletic species in the tribe, as Asian versus European and North American populations formed two distinct clades (Esmailbegi et al., 2018). While the Asian populations of *A. petiolata* are diploid ($2n = 2x = 14$), the European plus North American populations are hexaploid ($2n = 6x = 42$; Esmailbegi et al., 2018). In spite of being a widespread species with ecologically important impacts, the genome structure of *A. petiolata* is unexplored and the origin of the hexaploid cytotype remains unclear.

In our previous phylogenetic study (Esmailbegi et al., 2018), we reconstructed the relationship of 37 out of 42 Thlaspideae species using one nuclear (ITS) and one chloroplast marker (*trnL-F*), demonstrated the monophyly of the tribe, and uncovered four major well-supported intratribal clades (Figure 1). Thlaspideae species are characterized by having the base chromosome number of $x = 7$ (Al-Shehbaz et al., 2006; Esmailbegi et al., 2018), whereby 72% are diploid ($2n = 2x = 14$). Up to now, chromosome counts were reported for 21 Thlaspideae species (https://brassibase.cos.uni-heidelberg.de/; Esmailbegi et al., 2018). According to the published chromosome counts, 19 out of 21 (90.5%) Thlaspideae species are exclusively diploid, whereas for *Thlaspi alliaceum* ($2n = 14$, 28 and 56), *A. petiolata* ($2n = 14$, 18, 36 and 42) and *Graellsia saxifragifolia* ($2n = 14$ and 28) both diploid and polyploid cytotypes were recorded. Given that approximately 72% of the Thlaspideae species have the same chromosome numbers ($2n = 14$) and 90.5% of species are diploid, we hypothesized that the non-dysploidal CRs and not polyploid and dysploidal chromosome changes probably represent the main genome-diversifying driver in the tribe.

Excluding the tribe Aethionemeae diverged from the rest of Brassicaceae tribes 32 million years ago (Walden et al., 2020), the other 51 tribes are grouped into four (Franzke et al., 2011; Walden et al., 2020) or five (Nikolov et al., 2019) major phylogenetic lineages. In most recent phylogenetic studies of Brassicaceae, the tribe Thlaspideae is placed in lineage II (Nikolov et al., 2019; Walden et al., 2020). The Proto-Calepineae Karyotype (PCK; $n = 7$; Figure 2) has been introduced as an ancestral genome of lineage II (Mandáková and Lysak, 2008). Considering that PCK was identified as an ancestral genome of *Eutrema salsugineum* from the tribe Eutremeae (Mandáková and Lysak, 2008), the sister tribe to Thlaspideae (Guo et al., 2017; Nikolov et al., 2019; Walden et al., 2020), we hypothesized that the ancestral genome of Thlaspideae should resemble PCK.

The major components of the nuclear DNA in most plant species are repetitive sequences, especially LTR retrotransposons (Bennetzen, 2000; Macas et al., 2015; Mehrotra and Goyal, 2014). Repetitive sequences are involved in many processes such as centromeric condensation, chromosome recombination, alteration of gene expression and physiological (Garrido-Ramos, 2015, 2017; Macas et al., 2011; Mehrotra and Goyal, 2014), confirming their important role in speciation, generating phenotypic variation and adaptation (Oliver et al., 2013; Piegu et al., 2006). Repetitive sequences can also be associated with CRs such as inversion, deletion, duplication and translocations, as well as with the formation of dicentric chromosomes and acentric fragments. Transposable elements (TEs) can mediate CRs

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

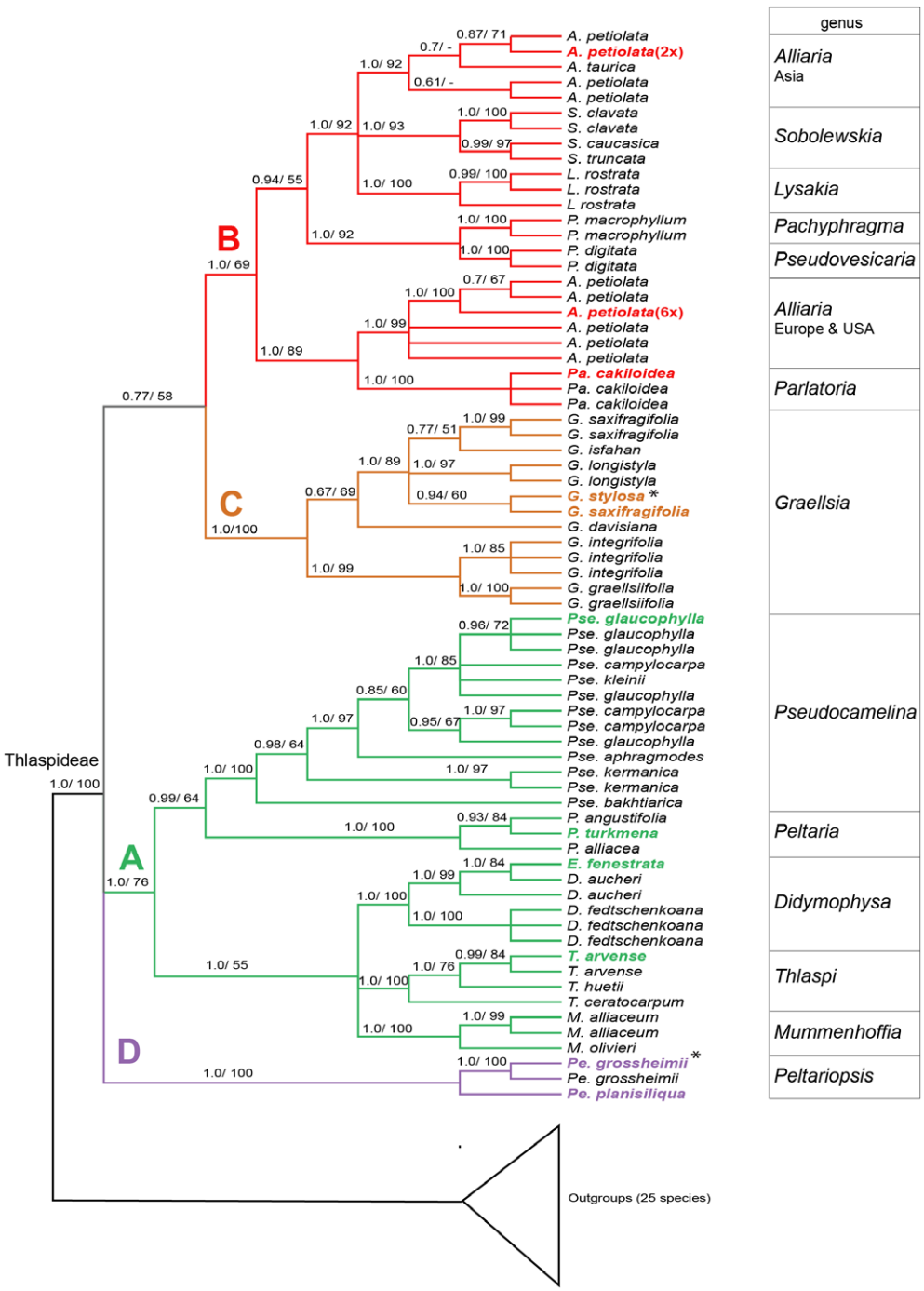

Figure 1. Phylogenetic relationships among the Thlaspideae species.
The tree was reconstructed using two molecular markers (ITS and *trnL-F*). Numbers above the branches are Bayesian posterior probabilities and maximum likelihood supports (Esmailbegi et al., 2018). Highlighted species indicate species analyzed in this study. Stars denote species analyzed only for chromosome structure, while other species were included in both cytogenetic and repeatome analyses. *Pseudocamelina szowitsii* (*Peltariopsis planisiliqua*) was not included in the phylogenetic study of Esmailbegi et al. (2018). Except for the hexaploid *Alliaria petiolata*, all investigated species were diploid ($2n = 2x = 14$).

via two mechanisms: indirectly through recombination between homologous TEs present in a genome; or directly by an alternative transposition (Gray, 2000). The role of the latter mechanism in producing CRs is documented in rice (Ac/Ds elements; Xuan et al., 2011), bacteria (IS 10/TN 10 elements; Chalmers and Kleckner, 1996), *Antirrhinum majus* (Tam 3; Sharma et al., 2020) and *Drosophila* (P elements; Beall and Rio, 1997). A detailed characterization of

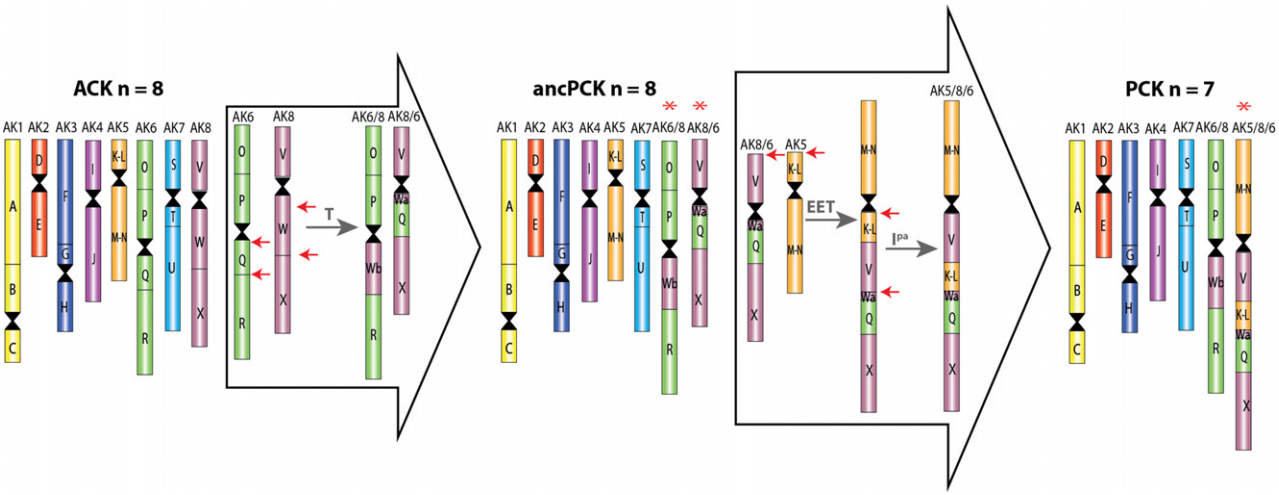

Figure 2. Origin of the ancestral Proto-Calepineae Karyotype (ancPCK, n = 8) and Proto-Calepineae Karyotype (PCK, n = 7) from the Ancestral Crucifer Karyotype (ACK, n = 8).
Chromosome ancPCK originated from the older ACK (AK1–AK8) through a reciprocal translocation between chromosomes AK6 and AK8 (→ AK6/8, AK8/6). The PCK originated from ancPCK via an end-to-end translocation between AK8/6 and AK5, followed by a paracentric inversion (→ PCK-specific chromosome AK5/8/6). T, translocation; EET, end-to-end translocation; I^{pa}, paracentric inversion. Red arrows show breakpoints. Red asterisks denote the chromosomes specific for ancPCK and PCK. adopted from Mandáková et al. (2018).

repetitive sequences in multiple species within a phylogenetic framework could help us to improve our knowledge of processes shaping plant genomes. Up to now, no study has been done to characterize the repeatome composition of Thlaspideae genomes.

In the present study, we aimed to: (i) reconstruct the genome structure in 12 species from the tribe Thlaspideae; (ii) identify the key mechanisms shaping the structure of Thlaspideae genomes; (iii) determine whether diversification within the tribe was accompanied by genus-/species-specific CRs; (iv) reconstruct an ancestral genome of the Thlaspideae; and (v) characterize repeatomes of the investigated Thlaspideae genomes.

RESULTS

Comparative karyotype structures of Thlaspideae species

The karyotype structures of 12 species from eight genera of the Thlaspideae were reconstructed by multicolor Comparative Chromosome Painting (CCP) using *Arabidopsis thaliana* Bacteria Artificial Chromosome (BAC) contigs arranged according to 22 genomic blocks (GBs; A–X) of eight ancestral chromosomes of the Ancestral Crucifer Karyotype (ACK; Lysak et al., 2006, 2016; Schranz et al., 2006; Figure 2). As the ancestral genome of the tribe Eutremeae, the sister tribe to Thlaspideae, originated from the PCK (Mandáková and Lysak, 2008), CCP probes were designed based on the seven chromosomes of the PCK (AK1–AK5/8/6; Figure 2), and the genome structure of the analyzed Thlaspideae species is described here in relation to the PCK.

CCP unambiguously identified all 22 GBs in chromosome complements of the analyzed species and confirmed their ploidy levels reported previously by Esmailbegi et al. (2018). In all diploid species ($2n = 2x = 14$; Table S1), each painting probe hybridized to a single pachytene bivalent (Figure 3a).

Peltaria turkmena. The chromosomes of *P. turkmena* (n = 7; Pt1–Pt7) were mostly colinear to those of the PCK. Among analyzed Thlaspideae species, *P. turkmena* is the only species whose structure of chromosome 1 (Pt1) mirrors the structure of AK1. Moreover, chromosomes Pt2, Pt3, Pt4 and Pt6 have the ancestral structures of AK3, AK4, AK7 and AK6/8, respectively. The only detected difference compared with the PCK was a reciprocal translocation between AK2 and AK5/8/6 forming chromosomes Pt5 and Pt7. Chromosome Pt5 consists of GBs M-N (7.49 Mb) and E (6.39 Mb) from AK5/8/6 and AK2, respectively. The complementary translocation product (Pt7) bears GBs D (3.05 Mb) and V/K–L/Wa/Q/X (12.47 Mb) from AK2 and AK5/8/6, respectively. The AK2–AK5/8/6 translocation was observed in other analyzed species as well (Figure 3a; Table S2).

Parlatoria cakiloidea. In *Pa. cakiloidea* (n = 7, Pc1–Pc7), two out of seven chromosome pairs (Pc2 and Pc3) retained the ancestral structure of AK3 and AK4, respectively. Chromosome Pc7 resulted from a translocation between GB D (3.05 Mb) of AK2 and GBs V/K–L/Wa/Q/X (12.47 Mb) of AK5/8/6. Four remaining chromosomes (Pc1, Pc4, Pc5 and Pc6) originated from ancestral chromosomes by pericentric inversions and centromere repositioning. Chromosomes Pc1 and Pc6 originated through a 7.88-Mb pericentric inversion on ancestral chromosomes AK1 and AK6/8,

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

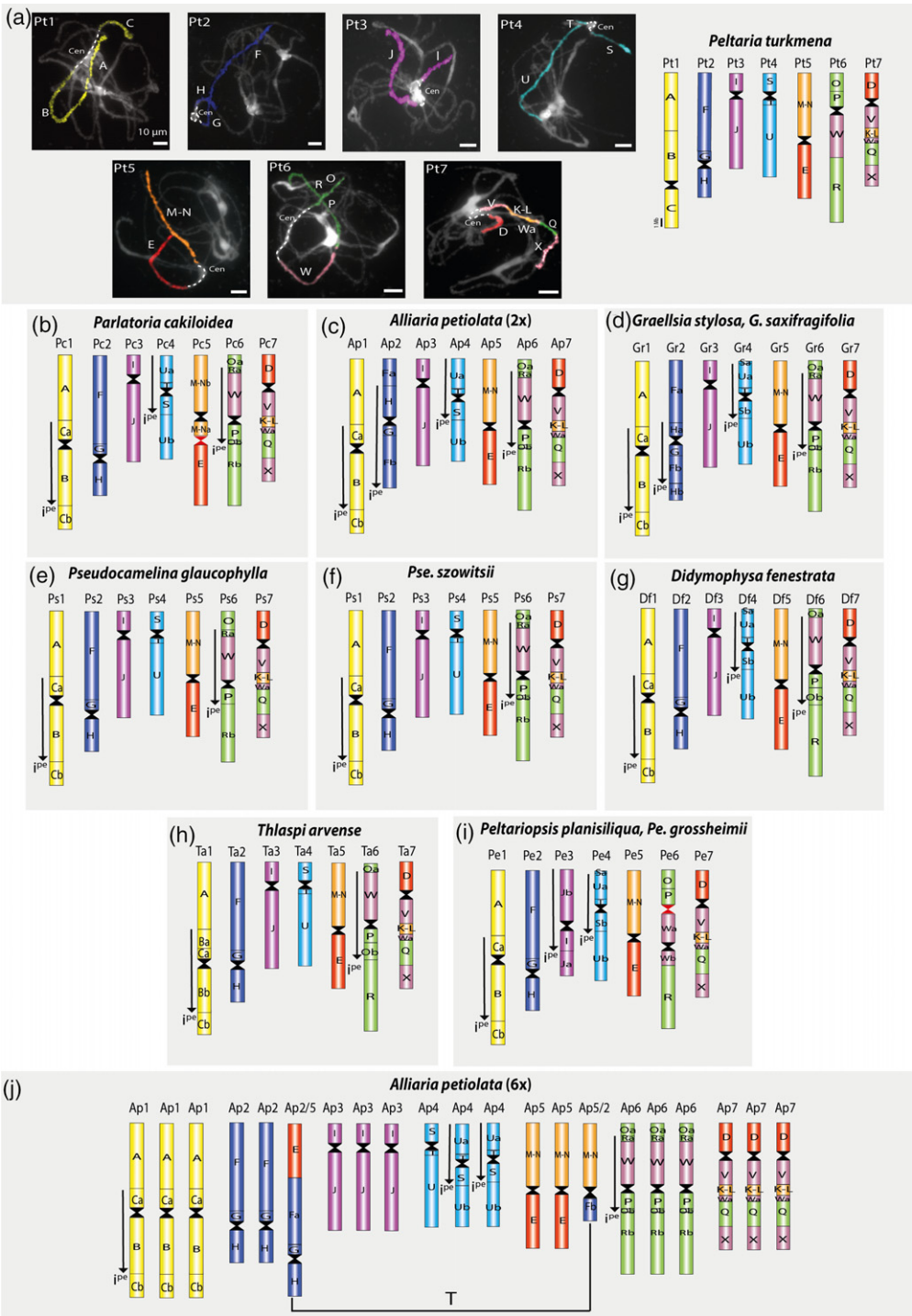

Figure 3. Comparative genome structures of 12 analyzed Thlaspideae species.
(a) Homeologous chromosomes in pachytene complements of *Peltaria turkmena* revealed by comparative chromosome painting (CCP) using *Arabidopsis thaliana* Bacteria Artificial Chromosome (BAC) contigs as painting probes. The fluorescent signals were captured as black and white images, and then pseudocolored to match the chromosome color coding in Ancestral Crucifer Karyotype (ACK). Chromosomes were counterstained with DAPI. Scale bars: 10 µm. (b–j) Cytomolecular comparative maps of 11 analyzed Thlaspideae species based on CCP analyses. The capital letters correspond to 22 genomic blocks (GBs; A–X) of ACK. Blocks split into sub-blocks are labeled as 'a' and 'b'. Downward-pointing arrows denote an inverted orientation of blocks as compared with the Proto-Calepineae Karyotype (PCK). Red hourglass symbols show the original position of centromeres, and black hourglass symbols indicate the new centromere position. All ideograms were scale-drawn, whereby the size of each GB corresponds to the size of the homeologous region in the *A. thaliana* genome (The Arabidopsis Information Resource, TAIR; http://www.arabidopsis.org). i^{pe}, pericentric inversion; T, translocation.

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

respectively. In chromosome Pc1, breakpoints were located within the GBs B [between BAC clone F14P1 (gene At1g19840) and F6F9 (At1g19850)] and C [between F14J22 (At1g49530) and T18C15 (At1g49890)]. In Pc6, breakpoints were detected within GBs O [between F9H3 (At4g03400) and T27D20 (Atg404165)] and R [between F28I16 (Atg519850) and MRN17 (At5g22980)]. Chromosome Pc4 originated from AK7 by a 5.86-Mb pericentric inversion with breakpoint within GBs U [between F20B18 (At4g25900) and T15N24 (At4g26470)]. Chromosome Pc5 originated through a translocation between GB E (6.39 MB) on AK2 and GB M–N (7.49 Mb) on AK5/8/6, followed by a 1.5-Mb centromere repositioning dividing GB M–N into two parts (M-Na and M-Nb; Figure S1). The centromere repositioning on Pc5 was exclusively observed in *Pa. cakiloidea* (Figure 3b and Figure S1; Table S3).

Diploid *Alliaria petiolata*. The genome structure was reconstructed in two populations of *A. petiolata* with different ploidy levels. In the diploid *A. petiolata* ($n = 7$; Ap1–Ap7), chromosome Ap3 mirrors the structure of AK4. Chromosomes Ap5 and Ap7 originated via a translocation between AK2 and AK5/8/6 – Ap5 bears GBs E (6.39 Mb) and M–N (7.49 Mb) from AK2 and AK5/8/6, respectively, while Ap7 consists of GB D (3.05 Mb) of AK2 and GBs V/K–L/Wa/Q/X (12.47 Mb) of AK5/8/6. The other four chromosomes (Ap1, Ap2, Ap4 and Ap6) were shaped by pericentric inversions on AK1, AK3, AK7 and AK6/8, respectively. The structures of Ap1 and Ap6 were identical to Pc1 and Pc6 in *Pa. cakiloidea*. In Ap4 we identified an inversion that resembled that of Pc4 in *Pa. cakiloidea*; however, the position of breakpoints was different in Ap4 and Pc4. Ap4 underwent a 6.28-Mb pericentric inversion with breakpoints in GBs U [between T24A18 (At4g07775) and F27G19 (At4g27310)]. Chromosome Ap2 resulted from an 8.22-Mb pericentric inversion on AK3 with breakpoints in GBs F [between F8A24 (At3g09760) and F13M14 (At3g10390)] and H [within F5H14 (At2g16225)]. This pericentric inversion was exclusively identified in the diploid *A. petiolata* (Figure 3c and Figure S1; Table S4).

Graellsia saxifragifolia and *Graellsia stylosa*. The genome structure of two *Graellsia* species was reconstructed ($n = 7$, Gr1–Gr7). In both analyzed species, the structure of chromosome Gr3 was identical to that of AK4. Chromosomes Gr5 and Gr7 were formed by a translocation between AK2 and AK8/6. In both *Graellsia* species, chromosome Gr5 consists of GB E from AK2 and GB M–N from AK5/8/6, whereas GB D of AK2 and GBs V/K–L/Wa/Q/X of AK5/8/6 form the structure of chromosome Gr7. Chromosomes Gr1 and Gr6 originated from the ancestral chromosomes AK1 and AK6/8, respectively, as a result of the same 7.88-Mb pericentric inversions identified in *Pa. cakiloidea*. Chromosome Gr2 originated from AK3 via a 5.16-Mb pericentric inversion with breakpoints within GBs F [between MRC8 (At3g18035) and K24M9 (At3g18524)] and H [between F16F14 (At2g16225) and F6P23 (At2g16980)]. Another CR shared by both species was a 3.43-Mb pericentric inversion on Gr4 with breakpoints within GBs S [between MYC6 (At5g41310) and K1O13 (At5g41210)] and U [between F24A6 (At4g25040) and F14M19 (At4g25730)]. The last two pericentric inversions were exclusively observed in *Graellsia* (Figure 3d and Figure S1; Table S5).

Pseudocamelina glaucophylla and *Pseudocamelina szowitsii*. In both *Pseudocamelina* species ($n = 7$; Ps1–Ps7), chromosomes Ps2, Ps3 and Ps4 resembled the structure of AK3, AK4 and AK7, respectively. Chromosomes Ps5 and Ps7 resulted from a translocation between AK2 and AK5/8/6, whereby chromosome Ps5 consists of GB M–N from AK5/8/6 and E from AK2, while Ps7 bears GBs D from AK2 and V/K–L/Wa/Q/X from AK5/8/6. In both species, a 7.88-Mb pericentric inversion shaped the structure of chromosome Ps1. CCP revealed species-specific rearrangements on Ps6 differentiating the two species. In *Pse. glaucophylla*, Ps6 originated from chromosome AK6/8 through a 6.93-Mb pericentric inversion with breakpoints within GBs O and P [between T1P17 (At4g12640) and T3H13 (At4g08870)] and R [between K8E10 (Atg522780) and MRN17 (At5g22980)] (Figure S1 and Table S6). Chromosome Ps6 in *Pse. szowitsii* underwent a 7.88-Mb pericentric inversion with breakpoints within GBs O [between F9H3 (At4g03400) and T27D20 (Atg404165)] and R [between F28I16 (Atg519850) and MRN17 (At5g22980); Figure 3e,f and Figure S1; Table S6 and S7].

Didymophysa fenestrata. In *D. fenestrata* ($n = 7$; Df1–Df7), two chromosomes (Df2 and Df3) have the same structure as AK3 and AK4, respectively. Like in other karyotypes described above, chromosomes Df5 and Df7 in *D. fenestrata* originated from AK2 and AK5/8/6 via a translocation. GBs E from AK2 and M–N from AK5/8/6 formed Df5, whereas GBs D from AK2 and V/K–L/Wa/Q/X from AK5/8/6 formed Df7. Similar to other species, a 7.88-Mb pericentric inversion with breakpoints within GBs C and B shaped the structure of Df1. Another reshuffled chromosome is Df4, which originated from AK7 via a 5.48-Mb pericentric inversion with breakpoints within GBs S [between MYC6 (At5g41310) and K1O13 (At5g41210)] and U [between F24A6 (At4g25040) and F14M19 (At4g25730)]. An 8.3-Mb pericentric inversion of AK6/8, with breakpoints within GBs O [between T10M13 (At4g01995) and T2H3 (At4g02190)] and GBs W and R [between MRN17 (At5g22980) and MUP24 (At5g60750)], resulted in the origin of Df6 (Figure 3g and Table S8).

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

Thlaspi arvense. In *T. arvense* ($n = 7$; Ta1–Ta7), three chromosomes (Ta2, Ta3, Ta4) were identical to those of the PCK (AK3, AK4 and AK7). Chromosomes Ta5 and Ta7 originated via a translocation between AK2 and AK5/8/6. Chromosome Ta5 was formed by GBs E (AK2) and M–N (AK5/8/6), and Ta7 consists of GBs D (AK2) and V/K–L/Wa/Q/X (AK5/8/6). Unlike in other analyzed Thlaspideae species, Ta1 originated from AK1 via one *T. arvense*-specific inversion, namely via a 7.53-Mb pericentric inversion with breakpoints within GBs B [between F16L1 (Atg122170) and T22J18 (Atg122610)] and C [between F14J22 (At1g49530) and T18C15 (At1g49890)]. The structure of Ta6 was identical to that of Df6. Chromosome Ta6 is originated from AK6/8 as result of 8.3-Mb pericentric inversion with breakpoints within GBs O [between T10M13 (At4g01995) and T2H3 (At4g02190)] and W/R [between MRN17 (At5g22980) and MUP24 (At5g60750); Figure 3h,g; Tables S8 and S9].

Peltariopsis planisiliqua and *Peltariopsis grossheimii*. In both *Peltariopsis* species ($n = 7$; Pe1–Pe7), the structure of Pe2 was identical to the structure of AK3. Similarly to the Thlaspideae genomes described above, in both *Peltariopsis* species, chromosomes Pe5 and Pe7 were formed by a translocation between AK2 and AK5/8/6. GBs E from AK2 and M–N from AK5/8/6 formed Pe5, whereas GBs D from AK2 and V/K–L/Wa/Q/X from AK5/8/6 constitute Pe7. Pe1 exhibited the same structure as observed in other Thlaspideae species analyzed, with the exception of *T. arvense*. The structure of Pe4 was the same as that of Df4 in *Didymophysa*, shaped by a 5.48-Mb pericentric inversion, with breakpoints within GBs S and U. In chromosome Pe6, GB W of AK6/8 was split via a centromere repositioning into two sub-blocks – Wa (3.27 Mb) and Wb (1.66 Mb), shifting the centromere approximately 3.27 Mb towards the long arm of the chromosome and changing the acrocentric chromosome to a submetacentric. Chromosome Pe3 originated from AK4 via a 7.36-Mb pericentric inversion with two breakpoints within GB J [between T9J22 (At2g26340) and T8I13 (At2g47400); Figure S1]. The last two CRs were only detected in the genus *Peltariopsis* (Figure 3i; Table S10).

Hexaploid *Alliaria petiolata*. The identification of all 22 conserved GBs in six copies in mitotic chromosome complement and three copies in meiotic chromosome complement in *A. petiolata* with $2n = 42$ corroborated its purported hexaploid origin (Figure 3j and Figure S2). In the hexaploid *A. petiolata*, all three copies of Ap3 have the same structure as the ancestral chromosome AK4. In Ap1 and Ap6, a 7.88-Mb pericentric inversion was identified in all three homeologous copies. The position of the inversion breakpoints was the same as in *Pa. cakiloidea* and the diploid *A. petiolata*. From three copies of Ap4, the structure of one copy was identical to that of AK7, whereas the second and third copies were altered by independent pericentric inversions. The second copy of Ap4 was reshuffled by a 5.86-Mb inversion with breakpoints within GBs U [between F20B18 (At4g25900) and T15N24 (At4g26470)] and S [within MPK23 (At5g41500)]. The structure of the second Ap4 homeolog was identical to that of Pc4 in *Pa. cakiloidea*. The structure of the third Ap4 homeolog resulted from a 6.28-Mb pericentric inversion on AK7 with breakpoints in GBs U [between T24A18 (At4g07775) and F27G19 (At4g27310)]. The structure of this copy was the same as the structure of Ap4 in the diploid *A. petiolata*. In the hexaploid, the translocation between AK2 and AK5/8/6 resulted in the origin of Ap5 and Ap7. Ap5 bears GBs E (6.39) from AK2 and M-N (7.49 Mb) from AK5/8/6, and Ap7 consists of GBs D (3.05 Mb) from AK2 and V/K–L/Wa/Q/X (12.47 Mb) from AK5/8/6. Another translocation was identified on one homeolog of Ap5. This translocation occurred between block E (6.39 Mb) of Ap5 and 3.34-Mb part of block F (Fb) from Ap2. The resulting chromosome Ap5/2 consists of GBs M–N from Ap5 and Fb from Ap2. The structure of two out of three Ap2 homeologs has the identical structure as the ancestral chromosome AK3, while the third homeolog (Ap2/5) originated from the translocation between GBs E of Ap5 and Fa/G/H of Ap2 (Figure 3j; Table S11).

Our data revealed that chromosomes of the hexaploid *A. petiolata* are mostly colinear with those of the diploid *A. petiolata*. In both cytotypes the structure of Ap3 was the same as the ancestral chromosome AK4. Moreover, chromosomes Ap1, Ap6 and one of the three copies of Ap4 originated from the ancestral chromosomes AK1, AK6/8 and AK7, respectively, via identical pericentric inversions. Chromosome Ap7 in the diploid *A. petiolata* and two homeologs of Ap7 in the hexaploid cytotype originated by translocation between AK2 and AK5/8/6. In spite of these similarities, the diploid and hexaploid genomes can be differentiated based on unique CRs: (i) the pericentric inversion on Ap2 in the diploid *A. petiolate*; (ii) the translocation between one homeolog of Ap2 and one homeolog of Ap5 producing chromosomes Ap2/5 and Ap5/2 in the hexaploid *A. petiolata*.

Genus- and species-specific rearrangements in Thlaspideae

Compared with the PCK, a reciprocal translocation between ancestral chromosomes AK2 and AK5/8/6 was observed in all Thlaspideae species analyzed. In addition to this translocation, in diploid Thlaspideae genomes, with the exception of *P. turkmena*, only intra-CRs including pericentric inversions and centromere repositioning were identified, whereas both inter- and intra-CRs were identified in the hexaploid *A. petiolata*. Compared with the PCK, CCP analyses revealed six species-/genus-specific CRs in the analyzed Thlaspideae species, shown in Table 1 and

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

Table 1 Species- and genus-specific CRs identified in analyzed Thlaspideae species.

Species/genus	OC → RC	Type of CR	Size of CR (Mb)
Parlatoria cakiloidea	AK5 → Pc5*	Centromere repositioning	1.5
Psedocamelina glaucophylla	AK 6/8 → Ps6	Pericentric inversion	6.93
Alliaria petiolata (2x)	AK3 → At2	Pericentric inversion	8.22
Graellsia	AK3 → Gr2*	Pericentric inversion	5.16
Graellsia	AK7 → Gr4*	Pericentric inversion	3.43
Peltariopsis	AK3 → Pe2*	Pericentric inversion	7.36
Peltariopsis	AK6/8→ Pe6*	Centromere repositioning	3.27

Asterisks denote genus-specific CRs.
CR, chromosome rearrangement; OC, original chromosome; RC, rearranged chromosome.

Figure S1. Moreover, we could identify a unique 8.22-Mb pericentric inversion in Ap2 in the diploid *A. petiolata* and a unique translocation between blocks E of Ap5 and F of Ap2 in the hexaploid cytotype.

Shared CRs within the tribe Thlaspideae

In addition to species-/genus-specific rearrangements, we identified CRs shared among the investigated Thlaspideae taxa. A 7.88-Mb pericentric inversion on AK1 was shared by 10 analyzed species (Pc1, Ap1, Gr1, Pe1, Df1, Ps1; Figure 4). A 7.88-Mb pericentric inversion on AK6/8 was detected in six studied species, including *G. saxifragifolia* (Gr6), *G. stylosa* (Gr6), *Pse. szowitsii* (Ps6), *Pa. cakiloidea* (Pc6), diploid and hexaploid *A. petiolata* (Ap6). A 5.48-Mb pericentric inversion on AK4 was identified in *D. fenestrata* (Df4), *Pe. planisiliqua* and *Pe. grossheimii* (Pe4). The last identified shared CR was a centromeric repositioning (1.66 Mb) on AK6/8 in *T. arvense* (Ta6), *Pe. planisiliqua* and *Pe. grossheimi* (Pe6; Figure 4). We could identify two CRs on Gr2 and Gr4 in *Graellsia* species that could be considered as synapomorphic CRs for clade C (Figure 1), and two CRs on Pe2 in *Peltariopsis* species that could be considered as synapomorphies for clade D (Figure 1). No synapomorphic CRs were identified in clades A and B.

Comparative analysis of rDNA loci

To complete the structural characterization of chromosomes in Thlaspideae species, we determined the number and position of 5S and 35S rDNA loci by fluorescent *in situ* hybridization (FISH; Figure S3 and Table S12). With the exception of *Pa. cakiloidea* in which one of two identified nucleolus organizer regions (NORs) was terminal, in other analyzed diploid species all NORs were found at heterochromatic pericentromeric regions. In all analyzed Thlaspideae species, 5S rDNA was found at the pericentromeric heterochromatin. With the exception of *Pa. cakiloidea*, in which three 5S rDNA loci were identified, only one 5S rDNA locus was found in other diploid species (Figure S3 and Table S12). In the hexaploid *A. petiolata*, 35S rDNA loci were localized at the pericentromeric heterochromatin of only one homeolog of Ap3 and Ap7. Similar to diploid Thlaspideae species analyzed, in the hexaploid *A. petiolata*, 5S rDNA loci were found at the pericentromeric heterochromatin of two homeologs of Ap2, Ap3 and Ap7 (Figure S3 and Table S12).

Ancestral genome of the tribe Thlaspideae

By comparing the karyotypes of 12 analyzed Thlaspideae species, the ancestral Thlaspideae genome with seven chromosome pairs was inferred (Table S13). Five out of seven chromosomes (Thl1, Thl2, Thl3, Thl4 and Thl6) were colinear with the ancestral chromosomes of the PCK. Two remaining chromosomes (Thl5 and Thl7) originated via a translocation between AK2 and AK5/6/8. Thl5 consists of GB M–N from AK5/6/8 and long arm (block E) of AK2. Short arm of AK2 (block D) along with GBs V/K–L/Wa/Q/X from AK5/6/8 formed chromosome Thl7 (Figure 4).

CR and inversion rates in the Thlaspideae

We calculated the number of evolutionary breakpoints (EBs) and number of inversions for nine Thlaspideae lineages including diploid and hexaploid *Alliaria*, *Didymophysa*, *Graellsia*, *Parlatoria*, *Peltaria*, *Peltariopsis*, *Psedocamelina* and *Thlaspi*. We used the published divergence times (Esmailbegi et al., 2018; Figure S4) to estimate and compare the rate of CRs and inversions during the evolution of Thlaspideae genomes (Table 2). Esmailbegi et al. (2018) estimated the divergence time for Thlaspideae lineages based on both fossil calibration and secondary calibration. As the authors concluded that dating based on secondary calibration was more realistic, we used these divergence time estimates to analyze the frequency of inversion events along Thlaspideae lineages. The tribe Thlaspideae originated 15.58 Mya (Figure S4; Esmailbegi et al., 2018). From the earliest divergence of *Peltaria* to the most recent origin of the hexaploid *Alliaria*, altogether 30 CRs were identified during 5.74 million years. Therefore, the average rate of CRs was estimated as 5.22 per million years. We also calculated the inversion rates for each lineage (Table 2).

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

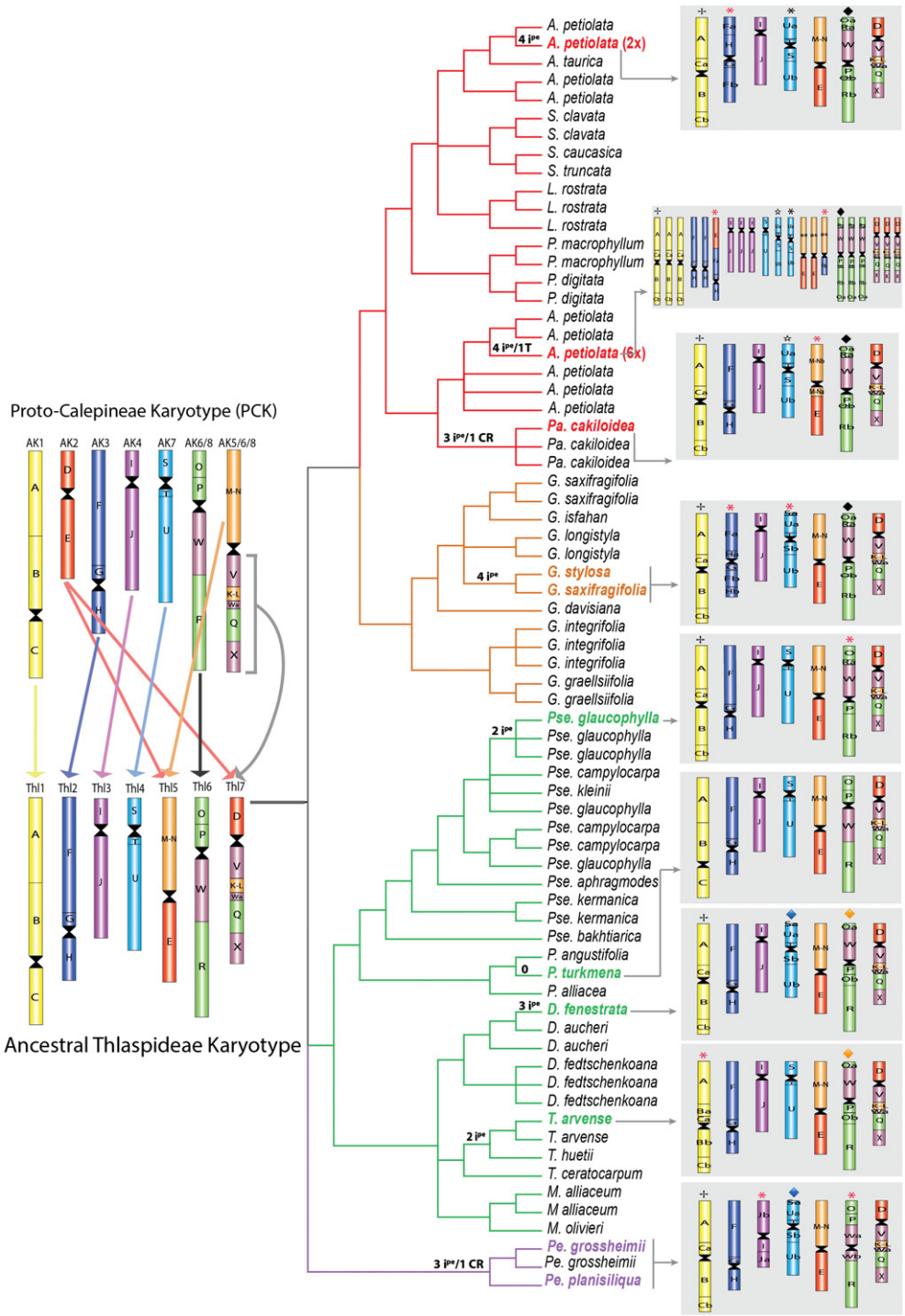

Figure 4. Genome evolution in the tribe Thlaspideae.
The capital letters correspond to 22 genomic blocks (GBs; A–X) of Ancestral Crucifer Karyotype (ACK). Number of pericentric inversions (i^{pe}), centromere repositioning (CR) events and translocations (T) are shown at the nodes. Red asterisks denote genus-/species-specific CRs. The phylogenetic tree is adopted from Esmailbegi et al. (2018). *Pseudocamelina szowitsii* (*Peltariopsis planisiliqua*) was not included in the phylogenetic study of Esmailbegi et al. (2018).

Repeat composition of the Thlaspideae genomes

We carried out characterization of repetitive DNA in nine species to improve our understanding of genome evolution in the Thlaspideae. To reach this goal, whole-genome shotgun sequencing was performed using the Illumina platform, generating 150-nucleotide end-pair reads.

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

Table 2 CR and inversion rates per million years in the Thlaspideae.

Genus (life form)	No. of CR	No. of inversions	No. of EBs	Divergence time (Mya)	CR rate	Inversion rate
Peltaria (perennial)	0	0	0	6.62	0	0
Pseudocamelina (biennial)	2	2	4	5.44	0.74	0.37
Didymophysa (perennial)	3	3	6	4.88	1.23	0.61
Graellsia (perennial)	4	4	8	4.74	1.69	0.84
Thlaspi (annual)	2	2	4	3.72	1.07	0.54
Peltariopsis (biennial/perennial)	4	3	8	2.88	2.78	1.04
Alliaria (2*x*) (biennial)	4	4	8	1.96	4.08	2.04
Parlatoria (annual)	4	3	8	0.91	8.79	3.3
Alliaria (6*x*) (biennial)	5	4	18	0.88	20.45	4.55

Life form data adopted from Esmailbegi et al. (2018).
CR, chromosome rearrangement; EB, evolutionary breakpoint.

Repeat Explorer pipeline (Novak et al., 2013) was used for evaluating repetitive sequences in the analyzed genomes. Because genome sizes of all analyzed species were not known, a clustering analysis based on the maximum number of reads was performed for individual species to identify the reads derived from repetitive sequences (Table S14). To gain information about repetitive sequences with moderate to high genome proportions, we only characterized clusters representing the repeats that made up at least 0.01% of the genome.

The genome proportions of each type of repetitive sequences for individual species are shown in Table 3 and Figure 5. According to our data, *Pa. cakiloidea* and the diploid *A. petiolata* having the smallest genomes (Table S14) among the analyzed species have the lowest proportion of repetitive DNA of 33.4 and 31.6%, respectively. The highest proportions of repetitive sequences were identified in *G. saxifragifolia* (53.5%), *T. arvense* (52.1%) and *Pe. planisiliqua* (52.3%), which have the largest estimated genome size among the analyzed species. However, because genome sizes for all analyzed species were not estimated due to the lack of fresh leaf material, we were not able to make clear conclusions about repetitive DNA amounts and genome size variation in the Thlaspideae (Table 3).

Although satellites, rDNA, DNA transposons and most of LTR retrotransposons are widely distributed in analyzed species, some LTR retrotransposon elements, such as *Angela* and *SIRE* lineages, were identified in only a few species as species- or genus-specific repeats. For example, *Ogre* and *Ikeros* lineages were detected in *P. turkmena* and *Pse. glaucophylla*, respectively. *Reina* LTR retrotransposon was found in the diploid and hexaploid *A. petiolata*. Another difference between the diploid and hexaploid *Alliaria* genomes and the remaining analyzed Thlaspideae genomes is that *Ivana* LTR retrotransposon was not detected in *Alliaria*, whereas it was found in all other genomes (Table 3). LTR retrotransposons made up the majority of repetitive DNA ranging from 18.5% in the diploid *A. petiolata* to 42.8% in *G. saxifragifolia*. We could identify seven clades of *Ty3/gypsy* family. Out of 16 known lineages of *Ty1/copia* (Neumann et al., 2019), we detected nine lineages in the Thlaspideae (Table 3). In the Thlaspideae genomes, the majority of LTR retrotransposons corresponded to the *Ty3/gypsy* superfamily mainly represented by *Athila* clade, ranging from 15.04% in the diploid *A. petiolata* to 30.06% in *T. arvense*. With the exception of the diploid *A. petiolata*, in which *Tekay* was the most abundant chromovirus clade, in all other species the most dominant chromovirus lineage was the *CRM* clade (Table 3). Compared with *Ty3/gypsy* elements, the abundance of *Ty1/copia* elements was very low, ranging from 0.63% in the diploid *A. petiolata* to 8.8% in *D. fenestrata*. Element *Ale* was the most dominant lineage of *Ty1/copia* in *T. arvense*, *Pe. planisiliqua*, *G. saxifragifolia*, *Pse. glaucophylla* and *D. fenestrata*. In *P. turkmena*, *TAR* represented the dominant superfamily of *Ty1/copia*. In both diploid and hexaploid *A. petiolata* and *Pa. cakiloidea*, *Bianca* made up the majority of the *Ty1/copia* superfamily (Table 3).

Compared with LTR retrotransposons, non-LTR retrotransposons were not found frequently. LINE elements were the only non-LTR retrotransposons identified in four genomes ranging from 0.01% in *T. arvense* to 0.3% in *G. saxifragifolia* (Table 3). From identified lineages of DNA transposons, *MuDR* were the most abundant elements in the diploid *A. petiolata*, *Pe. planisiliqua*, *P. turkmena* and *T. arvense*. In *D. fenestrata*, *G. saxifragifolia* and the hexaploid *A. petiolata*, *CACTA* exhibits the highest genome proportions. *Helitron* and *Mariner* were the most dominant DNA transposons in *Pa. cakiloidea* and *Pse. glaucophylla*, respectively.

The abundance of satellites varied from a very low genome proportion (0.16%) in *Pse. glaucophylla*, followed by 0.28% in *D. fenestrata*, to significantly high genome proportions in the hexaploid *A. petiolata* (6.95%), *Pa.*

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

Table 3 Repeat composition of the analyzed Thlaspideae genomes.

Genome proportion (%)		2*x A. petiolata*	6*x A. petiolata*	*D. fenestrata*	*G. saxifragifolia*	*Pa. cakiloidea*	*Pse. glaucophylla*	*Pe. planisiliqua*	*P. turkmena*	*T. arvense*
	Genome size (Mb)	391.2	440.1	–	440.1	370.662	–	508.56	–	508.56
Repeat family	Lineage									
LTR retrotransposons		18.449	24.867	38.289	42.835	20.244	31.569	38.513	28.677	42.669
	Unclassified LTR	0.0	0.0	0.75	0.11	0.064	0.526	0.024	1.875	2.33
Ty3/gypsy	Non-chromovirus									
	Athila	15.04	18.709	17.963	29.537	16.109	16.091	16.638	17.51	30.065
	Retand	0.043	0.031	0.102	0.219	0.0	0.166	0.055	0.0	0.933
	Ogre	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.99	0.0
	Chromovirus									
	CRM	2.23	3.104	8.205	2.699	2.239	5.808	9.039	1.966	2.824
	Galadriel	0.0	0.084	0.241	0.0	0.084	0.046	0.27	0.048	0.022
	Tekay	0.265	0.937	1.092	2.28	0.168	2.34	1.935	1.843	1.762
	Reina	0.013	0.012	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	unclassified	0.23	0.933	1.113	4.771	0.247	4.069	3.307	1.605	2.107
Ty1/copia		0.628	1.057	8.823	3.219	1.333	2.523	7.245	4.173	2.626
	Ale	0.018	0.0	5.68	1.498	0.011	1.261	3.818	0.0	1.879
	Bianca	0.438	0.72	0.681	0.679	0.868	0.66	0.73	2.024	0.272
	Tork	0.017	0.023	0.0	0.075	0.034	0.0	0.6	0.322	0.066
	Alesia	0.0	0.037	0.29	0.036	0.0	0.0	0.44	0.0	0.1
	Ivana	0.0	0.0	0.513	0.088	0.023	0.042	0.14	0.068	0.038
	TAR	0.0	0.065	0.453	0.35	0.23	0.23	0.233	0.256	0.085
	Ikeros	0.0	0.0	0.0	0.0	0.0	0.15	0.0	0.0	0.0
	Angela	0.0	0.072	0.0	0.045	0.0	0.0	0.0	0.097	0.0
	SIRE	0.0	0.01	0.156	0.0	0.0	0.108	0.0	0.058	0.0
	Unclassified	0.155	0.13	1.05	0.448	0.167	0.072	1.284	0.015	0.186
DNA transposons		1.028	0.911	2.601	4.447	2.183	3.495	3.449	3.885	2.03
	CACTA	0.042	0.255	1.3	2.194	0.394	0.065	0.765	0.99	0.069
	hAT	0.0	0.01	0.019	0.031	0.075	0.06	0.014	0.08	0.0
	Helitron	0.153	0.1	0.018	0.49	0.69	0.022	0.0	0.0	0.032
	Harbinger	0.0	0.0	0.0	0.098	0.041	0.012	0.07	0.011	0.29
	Mariner	0.06	0.093	0.0	0.096	0.411	0.552	0.161	0.411	0.16
	MuDR	0.358	0.13	1.048	0.953	0.373	0.521	1.333	1.146	0.557
	Unclassified	0.415	0.323	0.216	0.585	0.199	2.263	1.106	1.247	0.922
LINE		0.0	0.0	0.0	0.3	0.0	0.0	0.037	0.035	0.011
rDNA		2.4	2.4	4.705	3.252	2.447	3.9	4.51	3.65	3.934
Satellite (high confidence)		8.159	6.946	0.276	1.47	7.012	0.163	2.727	3.014	1.666
Satellite (Low confidence)		0.11	0.0	0.063	0.082	0.0	0.12	0.0	0.09	0.0
Unclassified repeats		1.451	1.659	1.691	1.06	1.529	2.379	3.035	1.755	1.786
Total repeats		31.597	36.783	47.625	53.446	33.415	41.626	52.271	41.106	52.096
Low/single-copy sequences		68.403	63.217	52.375	46.554	66.585	58.374	47.729	58.894	47.904

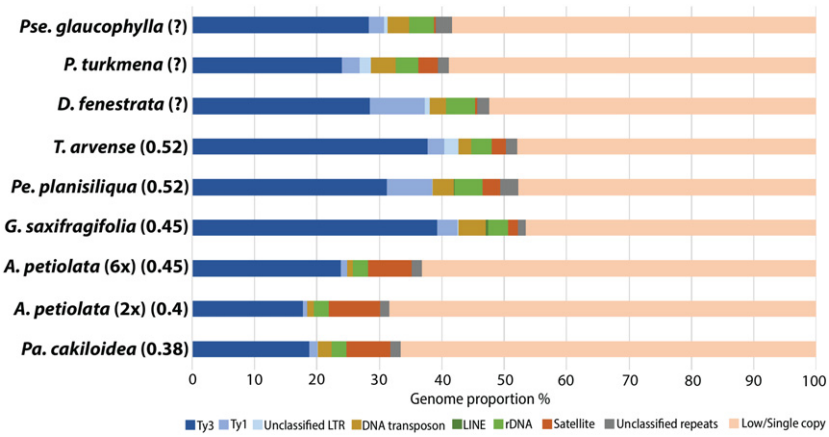

Figure 5. Relative genome proportion of repetitive and low/single copy sequences in analyzed genomes. Numbers in parenthesis denote the estimated genome sizes (pg).

cakiloidea (7.01%) and the diploid *A. petiolata* (8.27%) (Figure 5; Table 3 and Table S15).

Identification of shared repeats within the tribe Thlaspideae

To determine the repeat families shared by multiple species, comparative repeatome analysis was performed. In this approach clusters that represented ortholog repeats from different species were clustered together based on their sequence similarities. As the genome sizes were not estimated for all species, the same number of reads (300 000) for all diploid species and 1 000 000 reads from the hexaploid *A. petiolata* were used to build a combined data set. The results obtained from comparative clustering analysis are shown in Figure S5. Although most clusters identified as LTR retrotransposons and DNA transposons are shared by the majority or even all species, suggesting these repeat elements are widely distributed across the tribe Thlaspideae, we detected 21 clusters shared only by the diploid and hexaploid *A. petiolata* and *Pa. cakiloidea*. From these 21 clusters, 10 clusters were annotated as *Ty3/gypsy* retrotransposons and 11 clusters annotated as DNA transposons (Figure S5). In contrast to LTR retrotransposons and DNA transposons, satellites were the least conserved repeats in the Thlaspideae genomes as their respective clusters contained reads mostly from a single species (e.g. all reads from CL 64 and CL 108 come from *P. turkmena* and *T. arvense*, respectively). However, we could detect three satellite clusters that are shared among *Pa. cakiloidea*, the diploid and hexaploid *A. petiolata* (Figure S6 and Table S15).

Chromosomal localization of the most abundant satellites

The chromosome position of satellite clusters with the highest genome proportion (Table S16) was determined by FISH using labeled oligo probes to mitotic chromosomes. As mentioned above, using comparative repeatome analysis, we could identify three satellite clusters shared by three closely related species (diploid and hexaploid *A. petiolata* and *Pa. cakiloidea*). From these three clusters, cluster 6 (called AlPe6) had the highest genome proportion in all three species and was located in the pericentromeric region of all chromosomes (Figure S7). The most abundant satellite repeats localized within pericentromeric regions of almost all chromosome pairs in the diploid and hexaploid *A. petiolata*, *Pa. cakiloidea*, *P. turkmena*, *G. saxifragifolia* and *T. arvense*. In *Pe. planisiliqua*, the satellite repeat (PePl1) occurred in the pericentromeric regions of one chromosome pair (Pp7). In *D. fenestrata*, the most abundant satellite repeat (DiFe1) was found in the pericentromeric region of two chromosome pairs (Figure S7).

The origin of the hexaploid *Alliaria petiolata*

CCP confirmed the expected hexaploid status of the European population of *A. petiolata*. CCP revealed that from three identified copies of chromosome Ap4, one homeolog mirrors the ancestral structure, one homeolog was identical to the chromosome Ap4 in *Pa. cakiloidea*, and one homeolog has the same structure as that of diploid *A. petiolata* (Figure 2j). Moreover, we could identify 24 repeat clusters shared only by the diploid *A. petiolata*, hexaploid *A. petiolata* and *Pa. cakiloidea*. Furthermore, according to the phylogenetic analysis, these three taxa are placed in the same clade (Figure 1), and the sister relationship between *Pa. cakiloidea* and the hexaploid cytotype of *A. petiolata* is well supported (Esmailbegi et al., 2018). Given the shared CRs and repeat clusters observed in these three species, we inferred the diploid *A. petiolata* and *Pa. cakiloidea* as potential parental genomes of the hexaploid *A. petiolata*. Accordingly, we carried out genomic *in situ* hybridization (GISH) in the hexaploid *A. petiolata* with genomic DNAs of *Pa. cakiloidea* and the diploid *A. petiolata* as probes. However, our GISH experiment resulted in uniform staining of all 42 chromosomes by both genomic probes (data not shown). The GISH method is based on genome-specific repetitive sequences. The low divergence time of analyzed species in the complex can lead to the

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

lack of differentiation of specific repetitive sequences and consequently genetic proximity between analyzed species, which could affect the efficiency of GISH and make it unable to uncover the parental genomes in allopolyploid species (Hodkinson et al., 2002; Silva and Souza, 2013; Souza et al., 2012).

DISCUSSION

In this study we used BAC-based CCP to reconstruct the karyotype of 12 species from nine Thlaspideae genera and analyze the evolution of chromosome alterations in Thlaspideae for which, up to now, no cytogenomic study had been done. By reconstructing the karyotypes of Thlaspideae species, we traced genome evolution in this tribe and detected genus- and species-specific CRs. We also identified the structurally most stable and most altered chromosomes within the Thlaspideae genomes. Furthermore, we analyzed repeatomes of nine Thlaspideae species and identified species-/genus-/clade-specific repeats.

The ancestral genome of Thlaspideae originated from PCK

Our cytogenetic analyses confirmed the hypothesis that like other members of the Brassicaceae lineage II, the ancestral genome of the tribe Thlaspideae with seven chromosomes ($n = 7$) was derived from the PCK ($n = 7$; Mandáková and Lysak, 2008) followed by a reciprocal translocation between chromosomes AK2 and AK5/8/6 (Figure 4). Compared with other analyzed Thlaspideae genomes, no CRs were found in the early diverging species *P. turkmena*. The karyotype of *P. turkmena* resembles the ancestral genome of Thlaspideae, further supporting the ancestral nature of this extant genome.

Inversions as the main driving force in chromosome evolution in Thlaspideae

Species vary widely by the number of chromosomes and their structure. Large-scale CRs, such as translocations and inversions, are associated with species boundaries suggesting a link between CRs and plant diversification (King, 1987; Lai et al., 2005; Stathos and Fishman, 2014). Using CCP we identified a wide range of karyotype variation at the diploid level including genus-/species-specific CRs within the tribe Thlaspideae. We could detect synapomorphic inversions within *Graellsia* and *Peltariopsis* genera, and apomorphic inversions in the diploid *A. petiolata*, *Pa. cakiloidea*, *Pse. glaucophylla* and *T. arvense*. As all these species are diploid with the same chromosome number, our data support the hypothesis that in Thlaspideae intrachromosomal rearrangements, not polyploid and dysploid chromosome changes, have driven genome diversification leading to the divergence of intratribal lineages.

From the earliest diverging genus, *Peltaria* (6.62 Mya), to the youngest lineage, the hexaploid *A. petiolata* (0.88 Mya; Esmailbegi et al., 2018), we identified 27 pericentric inversions, whereas only a single translocation was detected. Our data show that pericentric inversions were the main type of CRs in Thlaspideae, indicating the key role of pericentric inversions in the genome evolution in the tribe. It has been documented that repetitive sequences can endorse DNA breakage and rejoining, and the density of repetitive DNA is higher within proximal (pericentromeric) regions than distal (telomeric) ones (Bennetzen, 2000; Gray, 2000; Kidwell, 2001; Wessler, 2006; Xuan et al., 2011). That might explain why pericentric inversions have occurred more often than paracentric inversion in the Thlaspideae.

The role of inversions in adaptation and plant diversification can be explained based on the recombination suppression model (Noor et al., 2001; Rieseberg, 2001; Trickett and Butlin, 1994). According to this model, inversions reduce recombination between loci that are involved in local adaptation and assortative mating permitting adaptive divergence (Huang and Rieseberg, 2020; Ostevik et al., 2020). Reduced recombination can facilitate the adaptation in the presence of gene flow because regions with low recombination rates increase the possibility of fixation of newly arising mutations with positive fitness effects.

The evolutionary role of inversions has been investigated in Dipteran systems *Drosophila* (Noor et al., 2001), *Anopheles* (Ayala and Coluzzi, 2005) and *Rhagoletis* (Feder et al., 2003), and plants such as *Boechera* (Lee et al., 2017), *Prospero* (Jang et al., 2013), *Helianthus* (Barb et al., 2014), *Mimulus* (Fishman et al., 2013; Lowry and Willis, 2010; Oneal et al., 2014) and *Noccaea* (Mandáková, et al., 2015). However, our study has documented this phenomenon across a whole plant tribe.

Contrasting chromosome stability versus rearrangements

This study represents the most comprehensive cytogenomic study so far performed within a monophyletic plant clade by reconstructing karyotypes of 12 Thlaspideae species covering the phylogenetic diversity of the tribe. Given the divergence time estimated for the tribe (Esmailbegi et al., 2018; Figure S4), it can be concluded that the oldest CR in the genome evolution of the Thlaspideae was a 7.88-Mb pericentric inversion on the ancestral chromosome Thl1, which occurred approximately 5 Mya with the divergence of the second oldest genus, *Pseudocamelina*, and is shared by all analyzed species but *P. turkmena* and *T. arvense* (Figure 3). Our data show that some Thlaspideae chromosomes were involved in more CRs than others. Chromosome Thl7 has not changed in any of the studied species, and the least number of rearrangements was identified for chromosomes Thl2, Thl3 and Thl5 (Figures 3 and 4). The structural integrity preserved in chromosomes Thl2, Thl3, Thl5 and Thl7 might be due to a selection against the fixation of CRs on these chromosomes.

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

In contrast to Thl2, Thl3, Th5 and Th7, we found out that two chromosomes, Thl4 and Thl6, bear the most various pericentric inversions (Figure 3 and Figure S3). Interestingly no CRs were identified on these two chromosomes in analyzed species of the tribes Isatideae, Sisymbrieae and Eutremeae, whose genomes have descended from an ancestral genome identical to that of the Thlaspideae (Mandáková and Lysak, 2008). Thus, CR-prone chromosomes (Thl4 and Thl6) represent the Thlaspideae-specific genome-diversifying drivers.

Correlation between CR rates and divergence times within the Thlaspideae

Extensive studies have been performed to investigate the genome structures of Brassicaceae species but, up to now, no studies had been done to estimate CR rates within Brassicaceae. Using previously published divergence times for Thlaspideae genera (Esmailbegi et al., 2018), we calculated inversion rates for nine Thlaspideae genera and explored the CR rates. Our data show that Thlaspideae species experienced approximately 5.22 CRs per million years. This rate is much higher than that recently reported in Solanaceae (0.03–0.12 CR per million year; Wu and Tanksley, 2010), but it is lower than the CR rates documented in sunflowers (9.7 CRs per million per year; Ostevik et al., 2020) and some grass lineages (up to 35.7 CRs per million year; Dvorak et al., 2018). The number of repetitive sequences and the size of inbred populations are the factors that could affect the rate of CRs (Ostevik et al., 2020). To investigate the effect of these factors on CR rate within the Thlaspideae more analyses are required.

The allohexaploid origin of *Alliaria petiolata*

Hexaploid chromosome number was previously reported for European and North American populations of *A. petiolata* ($2n = 6x = 42$; Esmailbegi et al., 2018). Our CCP data showed that the genome of the hexaploid *A. petiolata* consists of 22 triplicated GBs (Figure 3j and Figure S2), supporting the hexaploid origin of the European populations of *A. petiolata*. Recent molecular phylogenetic study of the Thlaspideae showed that the hexaploid *A. petiolata* is encompassed in a well-supported subclade with *Pa. cakiloidea*. This subclade is part of the Thlaspideae clade where the diploid *A. petiolota* was included as well (Figure 1; Esmailbegi et al., 2018). Diploid and hexaploid *A. petiolata* have similar fruit morphology, but can be distinguished based on the pollen shape and size. On the other hand, the hexaploid *A. petiolata* and *Pa. cakiloidea* can only be differentiated based on fruit morphology, and in the absence of the fruit these two species are so similar that they could be misidentified (Esmailbegi et al., 2018).

Here we reconstructed the genome structure of the hexaploid *A. petiolata*. We uncovered exclusive CRs of both *Pa. cakiloidea* and the diploid *A. petiolata* in the hexaploid. Due to their low tendency toward convergent evolution, CRs can be considered as phylogenetically informative genetic alterations. Considering the CRs and repeats shared among the diploid *A. petiolata*, *Pa. cakiloidea* and the hexaploid *A. petiolata*, we propose an allopolyploid origin of the hexaploid cytotype of *A. petiolata* via hybridization between genomes closely related to the extant diploid *A. petiolata* cytotype and *Pa. cakiloidea*. Although we could not draw clear conclusions about the evolutionary history of the hexaploid *A. petiolata*, our results provide the basis for further studies to understand the origin of this weedy species.

Cytogenetic signatures suggest a taxonomic revision for *Peltariopsis planisiliqua*

In 2012, *Pse. szowitsii* (Boiss.) N. Busch was reduced to a synonym of *Pe. planisiliqua* N. Busch (http://www.theplantlist.org/tpl1.1/record/tro-4104207) and no longer is included in the genus *Pseudocamelina* (Esmailbegi et al., 2018). However, our CCP data showed that the genome of *Pse. szowitsii* is significantly different from that of *Pe. planisiliqua* but almost identical to the genome of *Pse. glaucophylla* (Figures 3 and 4). More importantly, the genus-specific CRs detected in both *Pe. planisiliqua* and *Pe. grossheimii* were not found in *Pse. szowitsii*. Furthermore, the genomes of *Pe. planisiliqua* and *Pse. szowitsii* can be differentiated by the number and chromosomal positioning of 35S rDNA, while the same number and chromosome placement of NORs was observed in *Pse. szowitsii* and *Pse. glaucophylla* (Figure S3). Given our cytogenetic data and the fact that comparative studies rely on strong phylogenetic hypotheses and stable names, we suggest that *Pse. szowitsii* (= *Pe. planisiliqua*) should be placed in the genus *Pseudocamelina* as an accepted species, and therefore a taxonomic revision of *Pseudocamelina* is required.

EXPERIMENTAL PROCEDURES

Experimental plant material

To cover the phylogenetic diversity of the tribe, 12 and nine Thlaspideae species were selected for CCP and repeatome analysis studies, respectively (Figure 1). With the exception of the European population of *A. petiolata*, which was hexaploid ($2n = 6x = 42$), the remaining analyzed species were diploid ($2n = 2x = 14$). With the exception of diploid *A. petiolata*, our study was done on the same accessions that were included in the phylogenetic study of Esmailbegi et al. (2018). The origin of analyzed species is given in Table S1. The analyzed plants were either collected in the wild or grown from seeds in a growth chamber. For CCP, young inflorescences were collected and fixed in freshly prepared fixative (ethanol:acetic acid, 3:1) overnight, transferred into 70% ethanol and then stored at −20°C until used. Genomic DNA was extracted from either fresh, frozen or silica dried leaves using NucleoSpin Plant II kit (Macherey-Nagel) for repeatome analysis.

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

Chromosome preparation

Young fixed buds containing immature anthers were used for the preparation of chromosome spreads according to the published protocol (Mandáková and Lysak, 2016a). Slides containing well-spread pachytene or mitotic chromosomes were post-fixed in freshly prepared 4% formaldehyde in distilled water for 7 min and air-dried. Then slides were treated with 100 mg ml^{-1} RNase (Applichem) in 2 x saline sodium citrate (2 x SSC) at 37°C for 60 min, and 0.1 mg ml^{-1} pepsin (Sigma) in 0.01 M HCl at 37°C for 10–60 min, post-fixed in 4% formaldehyde in 2 x SSC for 7 min, washed in 2 x SSC twice for 5 min, and dehydrated in an ethanol series (70, 80 and 96%, 3 min each).

DNA probes

For CCP, in total 674 chromosome-specific BAC clones of *A. thaliana* arranged according to eight ancestral chromosomes (AK1–AK8) and 22 conserved GBs (A–X) were used (Lysak et al., 2016; Schranz et al., 2006). After the initial round of CCP experiments, some BAC contigs were divided into smaller subcontigs and labeled differentially to precisely identify breakpoints. Genomic DNAs of *Pa. cakiloidea* and diploid *A. petiolata* were used as a probe in GISH experiments in hexaploid *A. petiolata*. For *in situ* localization of NORs and 5S rDNA loci, the BAC clones T15P10 (AF167571) and pCT4.2 (M65137) of *A. thaliana* were used, respectively. For *in situ* localization of satellites, oligonucleotide probes (60 bp) were designed from DNA consensus sequences of satellite repeats (Table S11). DNA probes were labeled with biotin-dUTP, digoxigenin-dUTP and Cy3-dUTP by nick translation as described by Mandáková and Lysak (2016b). In CCP experiments, 100 and 500 ng per BAC clone/slide were used in diploid and hexaploid species, respectively. For GISH experiments, 150 ng of each genomic DNA was used for each slide. Labeled DNA probes were pooled following a designed experiment, ethanol precipitated, dried, and dissolved in 20 μl of 50% formamide and 10% dextran sulfate in 2 x SSC per slide (Mandáková and Lysak, 2016b).

In situ hybridization, microscopy and image processing

For *in situ* localization of repeats (rDNAs and satDNA) and CCP, 20 μl of probes was pipetted onto a suitable slide and denatured on a hot plate at 80°C for 2 min immediately. For GISH in hexaploid *A. petiolata*, 20 μl of the labeled gDNA was denatured at 90°C for 10 min, placed on ice for 10 min, pipetted onto a slide, and denatured on a hot plate at 80°C for 2 min. The hybridization process was done in a moist chamber at 37°C for 48 h, followed by post-hybridization washing in 20% and 50% formamide in 2 x SSC at 42°C for CCP and GISH, respectively. The immunodetection of hapten-labeled probes was carried out based on the published protocol (Mandáková and Lysak, 2016b). Chromosomes were counterstained with DAPI (2 μg ml^{-1}) in Vectashield. Zeiss Axioimager Z2 epifluorescence microscope and Coolcube camera (MetaSystems) were used for analyzing and photographing fluorescence signals. Using appropriate excitation and emission filters (AHF Analysentechnik), images were taken for all four fluorochromes, pseudocolored, merged and cropped using the Photoshop CS software (Adobe Systems).

Genome size measurements

Holoploid genome sizes were estimated by flow cytometry for *G. saxifragifolia*, *Pe. planisiliqua* and *T. arvense* as described by Hloušková et al. (2019). For *Pa. cakiloidea* and diploid *A. petiolata* genome sizes were calculated using short sub-sequences (k-mers) of Illumina short read data. For calculation of genome sizes using k-mers, after checking the quality of reads, the Jellyfish k-mer counting program (Marçais and Kingsford, 2011) was used for calculating k-mer distribution. To plot the binned distribution for the selected k-mer lengths, the R statistical package was used. After determining the pick position, the total number of k-mers in the distribution was calculated, and using the peak position the genome size was estimated (Pflug et al., 2020). We could not use this method to estimate genome sizes in *P. turkmena*, *D. fenestrata* and *Pse. glaucophylla*. These three species have high amounts of repetitive sequences, which could have affected the efficiency of K-mers to estimate their genome sizes (Pflug et al., 2020). To measure genome size by flow cytometry the same protocol described in Hloušková et al. (2019) was used.

De novo identification of repetitive sequences

Genome sequencing of nine species including hexaploid *A. petiolata*, diploid *A. petiolata*, *Pa. cakiloidea*, *D. fenestrata*, *P. turkmena*, *Pe. planisiliqua*, *Pse. glaucophylla*, *G. saxifragifolia* and *T. arvense* was performed as described by Hloušková et al. (2019). For each species, the local installation of Repeat Explorer pipeline (Novak et al., 2013) was used for clustering reads into groups of similar reads based on their graph similarities using the default settings. Because the genome sizes of all analyzed species were not known, a clustering analysis based on the maximum number of reads was performed for individual species and repeats with genome proportion >0.01% were annotated in detail.

Repeats with known protein domains were identified by the Repeat Explorer pipeline directly. To identify other repeats, sequence-similarity searches of assembled contigs against GenBank using BlastN and BlastX (Altschul et al., 1990) and the software CENSOR (Kohany et al., 2006) were used. Satellites within contig sequences were analyzed using Tandem Repeat finder (Benson, 1999) and Dotter (Sonnhammer and Durbin, 1995). Tandem repeat analyzer TAREAN (Novák et al., 2017) was used to reconstruct the consensus monomer sequences of satellites (Table S15). Moreover, sequences corresponding to putative mitochondrial and chloroplast contaminations were removed. Then the percentage of reads was used to calculate the genome proportion of each cluster.

To determine repeat families shared by multiple species, a comparative repeatome analysis was performed by building a combined data set containing 300 000 reads from diploid species and 1000000 reads from hexaploid *A. petiolata* labeled with a species code.

ACKNOWLEDGEMENTS

The authors thank Dr Shalva Sikharulidze (Ilia State University, Georgia) for collecting inflorescences of the diploid *Alliaria petiolata*, and Šarlota Štemberová for technical assistance. Core Facility Plants Sciences of CEITEC MU is acknowledged for the cultivation of plants used in this study. Repeat Explorer analyses were supported by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure. This work was supported by research grant from the Czech Science Foundation (grant 19-03442S).

CONFLICT OF INTEREST

No conflicts of interest.

DATA AVAILABILITY STATEMENT

Not applicable.

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,

SUPPORTING INFORMATION

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

Table S1 Origin of the species analyzed in the present study.

Table S2 Genome structure of *P. turkmena* (2*n* = 14).

Table S3 Genome structure of *Pa. cakiloidea* (2*n* = 14).

Table S4 Genome structure of the diploid *A. petiolata* (2*n* = 2*x* = 14).

Table S5 Genome structure of *G. saxifragifolia* (2*n* = 14) and *G. stylosa* (2*n* = 14).

Table S6 Genome structure of *Pse. glaucophylla* (2*n* = 14).

Table S7 Genome structure of *Pse. szowitsii* (2*n* = 14).

Table S8 Genome structure of *D. fenestrata* (2*n* = 14).

Table S9 Genome structure of *T. arvense* (2*n* = 14).

Table S10 Genome structure of *Pe. planisiliqua* (2*n* = 14) and *Pe. grossheimii* (2*n* = 14).

Table S11 Genome structure of hexaploid *A. petiolata* (2*n* = 6*x* = 42).

Table S12 Number and chromosome position of rDNA loci in the analyzed Thlaspideae species. Stars denote chromosomes bearing a terminal NOR.

Table S13 The structure of the ancestral Thlaspideae genome (2*n* = 14).

Table S14 Number of analyzed sequences by Repeat Explorer pipeline and clustering statistics.

Table S15 High confidence satellites detected by Repeat Explorer and TAREAN analyses.

Table S16 Nucleotide sequences of consensus satellites monomers.

Figure S1. Genus- and species-specific CRs in the analyzed Thlaspideae species revealed by CCP using *A. thaliana* BAC contigs as painting probes.

Figure S2. Examples of homeologs pachytene chromosomes in the hexaploid *A. petiolata* revealed by CCP using *A. thaliana* BAC contigs as painting probes.

Figure S3. The number of rDNA loci and their position on chromosomes of the analyzed Thlaspideae species in the phylogenetic context. *Pseudocamelina szowitsii* (*Pe. planisiliqua*) was not included in the phylogenetic study of Esmailbegi et al. (2018).

Figure S4. Time calibrated phylogenetic tree of the tribe Thlaspideae from BEAST analysis based on a concatenated ITS and *trnL-F* loci dataset from Esmailbegi et al. (2018).

Figure S5. The abundance of all clusters representing LTR retrotransposons and DNA transposons from the comparative analysis of repeatomes.

Figure S6. The abundance of satellite clusters from the comparative repetitive analysis.

Figure S7. Chromosome localization of the most abundant tandem repeats in eight Thlaspideae species.

REFERENCES

Agrawal, A.A., Fishbein, M., Halitschke, R., Hastings, A.P., Rabosky, D.L. & Rasmann, S. (2009) Evidence for adaptive radiation from a phylogenetic study of plant defenses. *Proceedings of the National Academy of Sciences*, **106**, 18067–18072.

Al-Shehbaz, I.A., Beilstein, M.A. & Kellogg, E.A. (2006) Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution*, **259**, 89–120.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

APG IV. (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, **181**, 1–20.

Ayala, F.J. & Coluzzi, M. (2005) Chromosome speciation: humans, Drosophila, and mosquitoes. *Proceedings of the National Academy of Sciences*, **102**, 6535–6542.

Barb, J.G., Bowers, J.E., Renaut, S., Rey, J.I., Knapp, S.J., Rieseberg, L.H. *et al.* (2014) Chromosomal evolution and patterns of introgression in *Helianthus*. *Genetics*, **197**, 969–979.

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

Bateman, R.M., Crane, P.R., DiMichele, W.A., Kenrick, P.R., Rowe, N.P., Speck, T. *et al.* (1998) Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. *Annual Review of Ecology and Systematics*, **29**, 263–292.

Beall, E.L. & Rio, D.C. (1997) Drosophila P-element transposase is a novel site-specific endonuclease. *Genes & Development*, **11**, 2137–2151.

Bennetzen, J.L. (2000) Transposable element contributes to plant gene and genome evolution. *Plant Molecular Biology*, **42**, 251–269.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573–580.

Brown, P.D. & Morra, M.J. (1997) Control of soil-borne plant pests using glucosinolate-containing plants. *Advances in Agronomy*, **61**, 167–231.

Burke, D.J. (2008) Effects of *Alliaria petiolata* (garlic mustard; Brassicaceae) on mycorrhizal colonization and community structure in three herbaceous plants in a mixed deciduous forest. *American Journal of Botany*, **95**, 1416–1425.

Chalmers, R.M. & Kleckner, N. (1996) IS10/Tn10 transposition efficiently accommodates diverse transposon end configurations. *EMBO Journal*, **15**, 5112–5122.

Cheng, F., Wu, J. & Wang, X. (2014) Genome triplication drove the diversification of *Brassica* plants. *Horticulture Research*, **1**, 14024.

Chopra, R., Johnson, E.B., Emenecker, R., Cahoon, E.B., Lyons, J. & Kiebenstein, D.J. *et al.* (2020) Identification and stacking of crucial traits required for the domestication of pennycress. *Nature Food*, **1**, 84–91.

Cubins, J.A., Wells, M.S., Frels, K., Ott, M.A., Forcella, F., Johnson, G.A. *et al.* (2019) Management of pennycress as a winter annual cash cover crop. A review. *Agronomy for Sustainable Development*, **39**, 46.

De Storme, N. & Mason, A. (2014) Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. *Current Plant Biology*, **1**, 10–33.

Dorn, K.M., Fankhauser, J.D., Wyse, D.L. & Marks, M.D. (2015) A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Research*, **22**, 121–131.

Dvorak, J., Wang, L.E., Zhu, T., Jorgensen, C.M., Deal, K.R., Dai, X. *et al.* (2018) Structural variation and rates of genome evolution in the grass family seen through comparison of sequences of genomes greatly differing in size. *The Plant Journal*, **95**, 487–5031.

Escudero, M., Martín-Bravo, S., Mayrose, I., Fernández-Mazuecos, M., Fiz-Palacios, O., Hipp, A.L. *et al.* (2014) Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PLoS One*, **9**, e85266.

Esmailbegi, S., Al-Shehbaz, I.A., Pouch, M., Mandáková, T., Mummenhoff, K., Rahiminejad, M.R. *et al.* (2018) Phylogeny and systematics of the tribe Thlaspideae (Brassicaceae) and the recognition of two new genera. *Taxon*, **67**, 324–340.

Feder, J.L., Roethele, J.B., Filchak, K., Niedbalski, J. & Romero-Severson, J. (2003) Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics*, **163**, 939–953.

Fishman, L., Stathos, A., Beardsley, P.M., Williams, C.F. & Hill, J.P. (2013) Chromosomal rearrangements and the genetics of reproductive barriers in Mimulus (monkey flowers): special section of reproductive barriers in Mimulus (monkey flowers). *Evolution*, **67**, 2547–2560.

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

Garrido-Ramos, M.A. (2015) Satellite DNA in plants: more than just rubbish. *Cytogenetic and Genome Research*, **146**, 153–170.

Garrido-Ramos, M. (2017) Satellite DNA: an evolving topic. *Genes*, **8**, 230.

Geng, Y., Guan, Y., Qiong, L.A., Lu, S., An, M., Crabbe, M.J.C. *et al.* (2021) Genomic analysis of field pennycress (*Thlaspi arvense*) provides insights into mechanisms of adaptation to high elevation. *BMC Biology*, **19**, 143.

Graham, L.E., Cook, M.E. & Busse, J.S. (2000) The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proceedings of the National Academy of Sciences*, **97**, 4535–4540.

Gray, Y.H.M. (2000) It takes two transposons to tango:transposable-element-mediated chromosomal rearrangements. *Trends in Genetics*, **16**, 461–468.

Guo, X., Liu, J., Hao, G., Zhang, L., Mao, K., Wang, X. *et al.* (2017) Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics*, **18**, 176.

Guo, X., Mandáková, T., Trachtová, K., Özüdoğru, B., Liu, J. & Lysak, M.A. (2021) Linked by ancestral bonds: multiple whole-genome duplications and reticulate evolution in a Brassicaceae tribe. *Molecular Biology and Evolution*, **38**, 1695–1714.

Hloušková, P., Mandáková, T., Pouch, M., Trávníček, P. & Lysak, M.A. (2019) The large genome size variation in the Hesperis clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. *Annals of Botany*, **124**, 103–120.

Hodkinson, T.R., Chase, M.W., Takahashi, C., Leitch, I.J., Bennett, M.D. & Renvoize, S.A. (2002) The use of dna sequencing (ITS and trnL-F), AFLP, and fluorescent in situ hybridization to study allopolyploid Miscanthus (Poaceae). *American Journal of Botany*, **89**, 279–286.

Huang, K. & Rieseberg, L.H. (2020) Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Frontiers in Plant Science*, **11**, 296.

Husband, B.C. (2004) Chromosomal variation in plant evolution. *American Journal of Botany*, **91**, 621–625.

Jang, T.-S., Emadzade, K., Parker, J., Temsch, E.M., Leitch, A.R., Speta, F. *et al.* (2013) Chromosomal diversification and karyotype evolution of diploids in the cytologically diverse genus Prospero (Hyacinthaceae). *BMC Evolutionary Biology*, **13**, 136.

Kidwell, M.G. (2001) Transposon-induced hotspots for genomic instability. *Genome Research*, **11**, 1321–1322.

King, M. (1987) Chromosme rearrangements, speciation and the theoretical approach. *Heredity*, **59**, 1–6.

Kohany, O., Gentles, A.J., Hankus, L. & Jurka, J. (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*, **7**, 474.

Lai, Z., Nakazato, T., Salmaso, M., Burke, J.M., Tang, S., Knapp, S.J. *et al.* (2005) Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics*, **171**, 291–303.

Lee, C.-R., Wang, B., Mojica, J.P., Mandakova, T., Prasad, K.V.S.K. & Goicoechea, J.L. *et al.* (2017) Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nature Ecology & Evolution*, **1**, 0119.

Lowry, D.B. & Willis, J.H. (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive Isolation. *PLoS Biology*, **8**, e1000500.

Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K. & Schubert, I. (2006) Mechanisms of chromosome number reduction in Arabidopsis thaliana and related Brassicaceae species. *Proceedings of the National Academy of Sciences*, **103**, 5224–5229.

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

Lysak, M.A., Mandáková, T. & Schranz, M.E. (2016) Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Current Opinion in Plant Biology*, **30**, 108–115.

Macas, J., Kejnovský, E., Neumann, P., Novák, P., Koblížková, A. & Vyskot, B. (2011) Next generation sequencing-based analysis of repetitive DNA in the model dioceous plant *Silene latifolia*. *PLoS One*, **6**, e27335.

Macas, J., Novák, P., Pellicer, J., Čížková, J., Koblížková, A., Neumann, P. *et al.* (2015) In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. *PLoS One*, **10**, e0143424.

Madlung, A. (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity*, **110**, 99–104.

Mandáková, T., Guo, X., Özüdoğru, B., Mummenhoff, K. & Lysak, M.A. (2018) Hybridization-facilitated genome merger and repeated chromosome fusion after 8 million years. *The Plant Journal*, **96**, 748–760.

Mandáková, T., Hloušková, P., Windham, M.D., Mitchell-Olds, T., Ashby, K., Price, B. *et al.* (2020) Chromosomal evolution and apomixis in the crucifer tribe Boechereae. *Frontiers in Plant Science*, **11**, 514.

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

Mandáková, T., Li, Z., Barker, M.S. & Lysak, M.A. (2017a) Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *The Plant Journal*, **91**, 3–21.

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

Mandáková, T. & Lysak, M.A. (2016a) Chromosome preparation for cytogenetic analysis in *Arabidopsis*. *Current Protocols in Plant Biology*, **1**, 43–51.

Mandáková, T. & Lysak, M.A. (2016b) Painting of *Arabidopsis* chromosomes with chromosome-specific BAC clones. *Current Protocols in Plant Biology*, **1**, 359–371.

Mandáková, T., Marhold, K. & Lysak, M.A. (2014) The widespread crucifer species *Cardamine flexuosa* is an allotetraploid with a conserved subgenomic structure. *New Phytologist*, **201**, 982–992.

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

Mandáková, T., Singh, V., Krämer, U. & Lysak, M.A. (2015) Genome structure of the heavy metal hyperaccumulator *Noccaea caerulescens* and its stability on metalliferous and nonmetalliferous soils. *Plant Physiology*, **169**, 674–689.

Marçais, G. & Kingsford, C. (2011) JELLYFISH – fast, parallel k-mer counting for DNA. *Bioinformatics*, **27**, 764–770.

McGinn, M., Phippen, W.B., Chopra, R., Bansal, S., Jarvis, B.A. & Phippen, M.E. *et al.* (2019) Molecular tools enabling pennycress (*Thlaspi arvense*) as a model plant and oilseed cash cover crop. *Plant Biotechnology Journal*, **17**, 776–788.

Meekins, J.F. & McCarthy, B.C. (1999) Competitive ability of *Alliaria petiolata* (garlic mustard, Brassicaceae), an invasive, nonindogenous forest herb. *International Journal of Plant Sciences*, **160**, 743–752.

Mehrotra, S. & Goyal, V. (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics Proteomics Bioinformatics*, **12**, 164–171.

Moser, B.R., Shah, S.N., Winkler-Moser, J.K., Vaughn, S.F. & Evangelista, R.L. (2009a) Composition and physical properties of cress (*Lepidium sativum* L.) and field pennycress (*Thlaspi arvense* L.) oils. *Industrial Crops and Products*, **30**, 199–205.

Moser, B.R., Vaughn, S.F. & Knothe, G. (2009b) Production and evaluation of biodisel from field pennycress (*Thalspi arvense* L.) oil. *Energy & Fuels*, **23**, 4149–4155.

Neumann, P., Novák, P., Hoštáková, N. & Macas, J. (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA*, **10**, 1.

Nikolov, L.A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I.A., Filatov, D. *et al.* (2019) Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytologist*, **222**, 1638–1651.

Noor, M.A.F., Grams, K.L., Bertucci, L.A. & Reiland, J. (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, **98**, 12084–12088.

Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P. & Macas, J. (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Research*, **45**, e111.

Novak, P., Neumann, P., Pech, J., Steinhaisl, J. & Macas, J. (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, **29**, 792–793.

Oliver, K.R., McComb, J.A. & Greene, W.K. (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biology and Evolution*, **5**, 1886–1901.

Oneal, E., Lowry, D.B., Wright, K.M., Zhu, Z. & Willis, J.H. (2014) Divergent population structure and climate associations of a chromosomal inversion polymorphism across the *Mimulus guttatus* species complex. *Molecular Ecology*, **23**, 2844–2860.

Ostevik, K.L., Samuk, K. & Rieseberg, L.H. (2020) Ancestral reconstruction of karyotypes reveals an exceptional rate of nonrandom chromosomal evolution in sunflower. *Genetics*, **214**, 1031–1045.

Pflug, J.M., Holmes, V.R., Burrus, C., Johnston, J.S. & Maddison, D.R. (2020) Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). *G3: Genes, Genomes, Genetics*, **10**, 3047–3060.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H. *et al.* (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Research*, **16**, 1262–1269.

Ren, L., Huang, W. & Cannon, S.B. (2019) Reconstruction of ancestral genome reveals chromosome evolution history for selected legume species. *New Phytologist*, **223**, 2090–2103.

Rieseberg, L.H. (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.

Schranz, M., Lysak, M. & Mitchellolds, T. (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends in Plant Science*, **11**, 535–542.

Sedbrook, J.C., Phippen, W.B. & Marks, M.D. (2014) New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (*Thlaspi arvense* L.). *Plant Science*, **227**, 122–132.

Sharma, J., Keeling, K.M. & Rowe, S.M. (2020) Pharmacological approaches for targeting cystic fibrosis nonsense mutations. *European Journal of Medicinal Chemistry*, **200**, 112436.

Silva, G.S. & Souza, M.M. (2013) Review genomic in situ hybridization in plants. *Genetics and Molecular Research*, **12**, 2953–2965.

Sonnhammer, E.L.L. & Durbin, R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, GC1–GC10.

Souza, L.G.R., Crosa, O., Speranza, P. & Guerra, M. (2012) Cytogenetic and molecular evidence suggest multiple origins and geographical parthenogenesis in Nothoscordum gracile (Alliaceae). *Annals of Botany*, **109**, 987–999.

Stathos, A. & Fishman, L. (2014) Chromosomal rearrangements directly cause underdominant F1 pollen sterility in *Mimulus lewisii-Mimulus cardinalis* hybrids. *Evolution*, **68**, 3109–3119.

Tricket, A.J. & Butlin, R.K. (1994) Recombination suppressors and the evolution of new species. *Heredity*, **73**, 339–345.

Walden, N., German, D.A., Wolf, E.M., Kiefer, M., Rigualt, P. & Huang, X.-C. *et al.* (2020) Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. *Nature Communications*, **11**, 3795.

Warwick, S.I., Francis, A. & Susko, D.J. (2002) The biology of Canadian weeds. 9. *Thlaspi arvense* L. (updated). *Canadian Journal of Plant Science*, **82**, 803–823.

Wessler, S.R. (2006) Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences*, **103**, 17600–17601.

Winterfeld, G., Becher, H., Voshell, S., Hilu, K. & Röser, M. (2018) Karyotype evolution in *Phalaris* (Poaceae): the role of reductional dysploidy, polyploidy and chromosome alteration in a wide-spread and diverse genus. *PLoS One*, **13**, e0192869.

Wu, F. & Tanksley, S.D. (2010) Chromosomal evolution in the plant family Solanaceae. *BMC Genomics*, **11**, 182.

Xuan, Y.H., Piao, H.L., Je, B.I., Park, S.J., Park, S.H., Huang, J. *et al.* (2011) Transposon Ac/Ds-induced chromosomal rearrangements at the rice OsRLG5 locus. *Nucleic Acids Research*, **39**, e149.

Yang, Q., Bi, H., Yang, W., Li, T., Jiang, J., Zhang, L. *et al.* (2020) The genome sequence of alpine mmegacarpaea delavayi identifies species-specific whole-genome duplication. *Frontiers in Genetics*, **11**, 812.

© 2021 Society for Experimental Biology and John Wiley & Sons Ltd,