communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-07851-0

A comparative genomic analysis at the chromosomal-level reveals evolutionary patterns of aphid chromosomes

Chen Huang[1], Bingru Ji[1,2], Zhaohui Shi[2], Jiangyue Wang[1,2], Jiaqing Yuan[1], Peng Yang[1], Xiao Xu[1], Haohao Jing[1], Lulu Xu[1], Jing Fu[1], Le Zhao[1,3], Yandong Ren[1] ✉, Kun Guo[2] ✉ & Gang Li[1] ✉

Genomic rearrangements are primary drivers of evolution, promoting biodiversity. Aphids, an agricultural pest with high species diversity, exhibit rapid chromosomal evolution and diverse karyotypes. These variations have been attributed to their unique holocentric chromosomes and parthenogenesis, though this hypothesis has faced scrutiny. In this study, we generated a chromosomal-level reference genome assembly of the celery aphid (*Semiaphis heraclei*) and conducted comparative genomic analysis, revealing varying chromosomal evolution rates among aphid lineages, positively correlating with species diversity. Aphid X chromosomes have undergone frequent intra-chromosomal recombination, while autosomes show accelerated inter-chromosomal recombination. Moreover, considering both inter- and intra-chromosomal rearrangements, the increased autosomal rearrangement rates may be common across the Aphidomorpha. We identified that the expansion of DNA transposable elements and short interspersed nuclear elements (SINEs), coupled with gene loss and duplication associated with karyotypic instability (such as *RIF1*, *BRD8*, *DMC1*, and *TERT*), may play crucial roles in aphid chromosomal evolution. Additionally, our analysis revealed that the mutation and expansion of detoxification gene families in *S. heraclei* may be a key factor in adapting to host plant chemical defenses. Our results provide new insights into chromosomal evolutionary patterns and detoxification gene families evolution in aphids, aiding the understanding of species diversity and adaptive evolution.

Genomic variation is a fundamental driving force in biological evolution, encompassing variations in single base pairs, small insertions, or deletions. In certain cases, however, more extensive genomic rearrangements become the primary driver of evolution[1]. These rearrangements involve structural variations such as segmental duplications, deletions, inversions, and others[2], and can lead to the emergence of new genomic structures and functions[3,4]. Transposable elements (TEs) are increasingly recognized as contributors to genomic rearrangements, as they can insert themselves at various loci, promote recombination, and facilitate chromosomal rearrangements, thereby shaping genomic structure and function over evolutionary time[5,6].

Genomic rearrangements are widely recognized as key factors driving biodiversity and differentiation[7]. Insects, being one of the most diverse groups of organisms on Earth, offer a unique opportunity to study the relationship between genomic variation and biodiversity[8]. For instance, the genomic structure of Lepidoptera has remained largely unchanged for at least 140 million years, while *Pieris napi* represents a notable exception with dramatic chromosomal rearrangements[9]. A recent study on Erebia butterflies reveals that repetitive elements within holocentric chromosomes correlate with population differentiation and chromosomal rearrangements, suggesting a role in adaptation and species diversification[10]. Additionally, recent research on aphids has shown significant TE enrichment at gene loci related to xenobiotic resistance, such as cytochrome P450 genes, highlighting the potential of TEs to drive adaptive changes in genomic structure and function[11].

Aphids serve as an emerging model system for investigating the implications of substantial genomic rearrangements on evolutionary

[1]College of Life Sciences, Shaanxi Normal University, Xi'an, 710119, China. [2]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, 100193, China. [3]QinLing-Bashan Mountains Bioresources Comprehensive Development C. I. C., School of Bioscience and Engineering, Shaanxi University of Technology, Hanzhong, 723000, P.R. China. ✉e-mail: renyandong@snnu.edu.cn; kunguoimplad@foxmail.com; gli@snnu.edu.cn

processes[12–14]. The species diversity is remarkable, with over 5000 described species and 23 subfamilies, including the most diverse subfamily Aphidinae, which encompasses more than half of the described aphid species[15]. Aphids exhibit exceptionally rapid chromosome evolution, characterized by a wide range of nuclear chromosome numbers ($2n = 4$ to 72)[16] and a higher rate of chromosomal rearrangements in aphids than in other Hemiptera insects[12,17]. Chromosomal rearrangement events are observed in different subfamilies, even within the same genus[18], which is an intriguing phenomenon that has attracted considerable attention.

Previous research has suggested that holocentric chromosomes and parthenogenesis may be primary factors driving extensive chromosomal rearrangements and persistent genetic diversity[19–21], but these factors do not fully explain the rapid chromosome evolution in aphids. For example, Ruckman et al. conducted a thorough analysis of chromosome numbers in 22 insect orders and found no significant difference in chromosomal evolution rates between clades with monocentric and holocentric chromosomes[22], and species like *Pieris napi*, with holocentric chromosomes but no parthenogenesis, also show significant rearrangements[9]. Mathers et al. have revealed extensive rearrangements in aphid autosomes, while their X chromosomes have shown a degree of conservatism[12]. However, Mathers et al.'s study was confined to only three aphid species, due to the limited availability of chromosomal-level genome assemblies at the time. Incorporating more chromosomal-level aphid genomes into genomic evolutionary analysis is necessary to explore the mechanisms and pattern behind chromosomal rearrangement in aphids, investigate whether different aphid lineages exhibit varying rates of chromosomal rearrangement and the reasons behind these differences, and determine whether a higher rearrangement rate might lead to an increased speciation rate.

The celery aphid, *Semiaphis heraclei*, is a widespread agricultural pest found in East Asia, South Asia, and Hawaii[23]. It primarily infests plants from the *Lonicera* genus (e.g., *Lonicera japonica*) and the *Apiaceae* family (e.g., celery, fennel, and *Angelica* spp.), many of which are important medicinal plants[23]. Understanding how this aphid adapts to its host plants, particularly by detoxifying secondary metabolites, is essential for pest management. Secondary metabolites are key components of plant defense, deterring herbivory and pathogen attacks[24]. In *Lonicera* species, such as *Lonicera japonica*, a variety of metabolites like phenolic acids, flavonoids, saponins, and iridoids have been identified, which help the plant resist environmental stressors[25,26]. Similarly, *Apiaceae* plants produce compounds such as terpenoids, triterpenoid saponins, flavonoids, and coumarins to defend against herbivores[27–30]. Aphids detoxify these compounds through enzyme families, including cytochrome P450 monooxygenases (P450s), carboxyl/cholinesterases (CCEs), UDP-glucosyltransferases (UGTs), and glutathione *S*-transferases (GSTs), allowing them to feed on their host plants[31]. Studying detoxification genes across aphid species can provide insights into the evolutionary adaptations of *S. heraclei* to its hosts, improving our understanding of host–plant interactions and aiding the development of pest control strategies.

Advances in sequencing technology have expanded the availability of aphid genome assemblies, presenting new opportunities for studying aphid genome evolution. In order to gain deeper insights into aphid chromosome evolution and the host adaptation evolution of *Semiaphis heraclei*, we assembled and annotated a chromosomal-level reference genome of this aphid. Furthermore, we performed comparative analyses at the chromosome level for five species within Macrosiphini (*Acyrthosiphon pisum*[12,32], *Myzus persicae*[12,32], *Sitobion miscanthi*[33,34], *Neotoxoptera formosana*[35] and *S. heraclei*), two species from Aphidini (*Aphis gossypii*[36] and *Rhopalosiphum maidis*[37]), and two species from Eriosomatinae (*Eriosoma lanigerum*[38,39] and *Schlechtendalia chinensis*[40]). In this study, we reconstructed the chromosomal evolutionary history of the Macrosiphini tribe and applied comparative genomic analysis, including transposable element assessment and topologically associated domains (TADs) tests, gene adaptive evolution, and gene expression patterns, to explore the genomic characteristics and mechanisms of chromosome rearrangement in aphids. Additionally, we conducted a comparative analysis of detoxification gene families across these species to uncover the genomic basis of host adaptation in *S. heraclei*. This research aims to enrich our understanding of the genetic dynamics underlying karyotype evolution, variations in genomic structures, gene family evolution, and their contribution to speciation and biodiversity in aphids.

Results

Genome assembly and annotation

A chromosome-level reference genome of *S. heraclei* was assembled using a combined sequencing data of 46.4 Gb Illumina paired-end reads, 161.2 Gb PacBio long reads, and 50.3 Gb Hi-C data (Supplementary Table 1). After removing assembly contamination with Blobtools pipeline (Supplementary Fig. 1)[41], the genome assembly was 404.35 Mb in size with scaffold N50 of 109.32 Mb, consistent with the 403.91 Mb genome size estimated by a K-mer analysis (Supplementary Fig. 2). A total of 401.92 Mb (99.4%) assembled sequences were clustered into four chromosomes (Fig. 1a), consistent with the previously reported karyotype $2n = 8$[42]. The completeness of the assembly was assessed by BUSCO v5.3.2[43,44], based on a Hemiptera conserved gene set ($n = 2510$), indicated 98.8% of the genes orthologs were captured (Supplementary Table 2). All these analyses proved the reliability and completeness of the *S. heraclei* genome assembly.

Approximately 33.37% of the genome was consists of repetitive elements, including DNA transposons with a total length of 46,468,311 bp (11.49%). This was followed by SINE retrotransposons, LINE retrotransposons, and LTR retrotransposons with lengths of 16,471,179 bp (4.07%), 9,257,723 bp (2.29%), and 7,769,730 bp (1.92%), respectively (Fig. 1b). Protein-coding genes were predicted by combining ab initio prediction, homology-based prediction, and evidence from mapped RNA-seq data, producing 17,921 protein-coding genes, with a BUSCO completeness of 98.6% (2473 complete genes of the total 2510 Hemiptera conserved orthologus). Of these, 17,530 (97.82%) protein-coding genes have homologs in common public databases, which indicates that the annotated genes were reliable (Supplementary Table 3).

Phylogenomic tree and divergence time

To reconstruct the phylogenomic tree of selected species and estimate their divergence times, we identified orthogroups in the celery aphid, eight other aphid species with chromosome-level assemblies, and six other Hemiptera insects (Supplementary Table 4)[45,46]. After extracting single-copy orthologs from the orthogroups, we constructed a fully resolved phylogenomic tree based on the protein sequences of 2358 highly conserved single-copy genes using the maximum likelihood method (Fig. 1c), which was consistent with the study conducted by ref. 47. Our results indicate that the primary divergences within different aphid subfamilies occurred from the Late Cretaceous to the early Tertiary period. The subfamily Lachninae diverged ~78.5 million years ago (MYA), followed by Eriosomatinae at around 72.6 MYA. The subfamilies Aphidinae and Chaitophorinae also diverged in the Late Cretaceous, around 70.0 MYA, a period when the major host angiosperms of aphids were undergoing rapid radiation and becoming dominant species[48]. These results are consistent with the fossil records that Aphidinae existed in the Late Cretaceous, with most tribes appearing in the early Tertiary[49], which were thought to be driven by the parallel evolution with the rapid radiation of angiosperms[50].

Collinearity analysis and reconstruction of the chromosomal evolution history of aphids

To investigate the chromosomal evolution of aphids, we conducted the collinearity tests among the chromosome-level reference genomes of nine aphid species, including five Macrosiphini species, two Aphidini species, and two species from Eriosomatinae (Fig. 2a and Supplementary Table 4). Firstly, the results of this analysis revealed the extensive intra- and inter-chromosomal rearrangements in autosomes, while the X chromosomes of Aphidinae and Eriosomatinae have maintained a conserved intra-chromosomal rearrangements pattern for at least 72.6 million years (Figs. 2a, 1c). As well, after we calculated the chromosomal rearrangement rates of

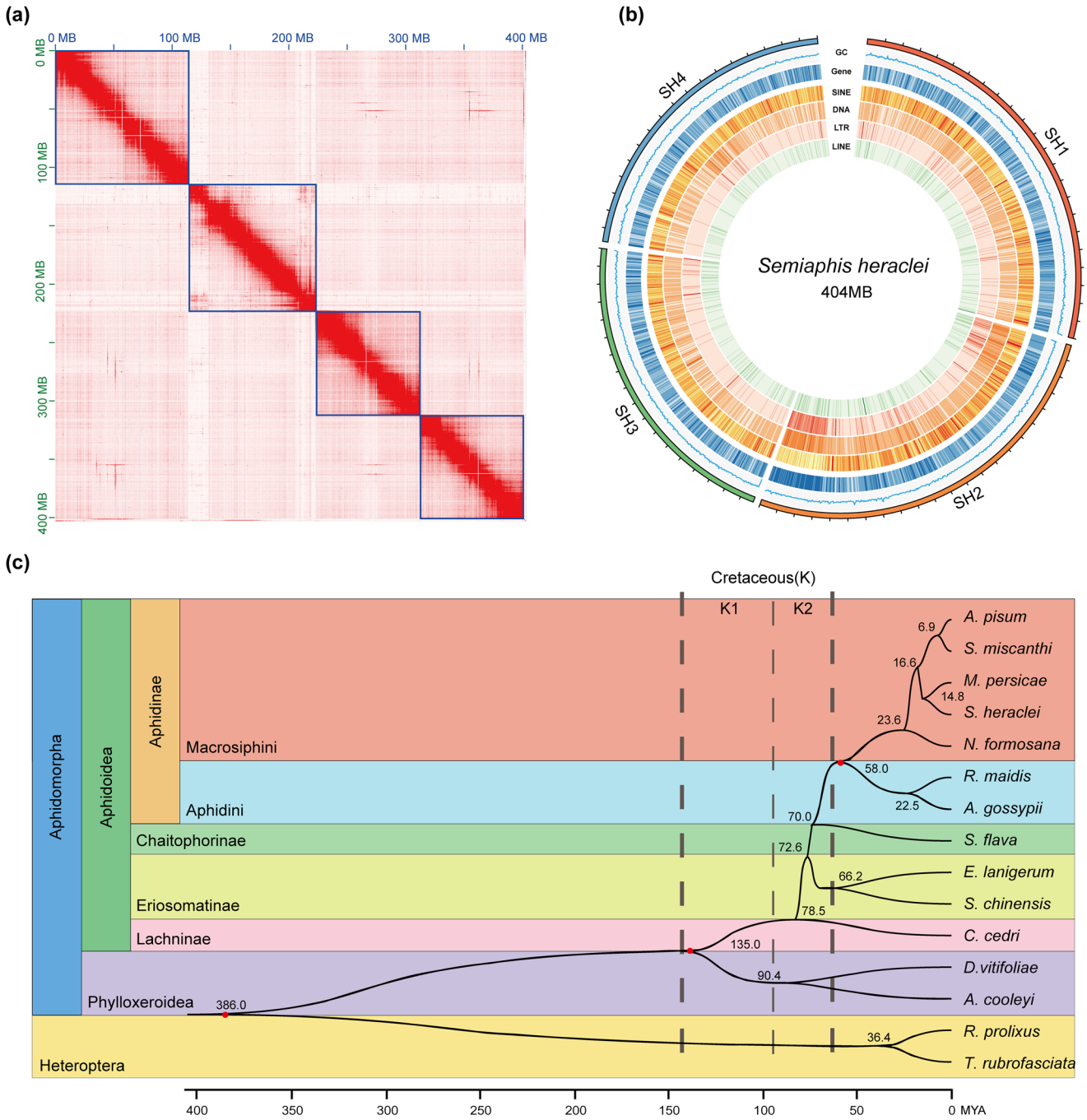

Fig. 1 | Chromosome-scale genome assembly of *S. heraclei* and aphid phylogenetic relationships. a The heatmap displays the Hi-C contact frequency of the *S. heraclei* genome assembly, with blue lines indicating chromosomes. The x-axis and y-axis represent cumulative lengths in millions of base pairs (Mb). **b** The circular diagram illustrates TE density, protein-coding gene density, GC content, and chromosome length of *S. heraclei* genome assembly from the inside out. **c** Time-calibrated phylogenetic tree based on the concatenation of 2358 conserved single-copy genes from all species. All nodes are supported with 100% bootstrap values, and different colors represent different taxa. The thicker dashed lines demarcate the beginning and end of the Cretaceous period, while the thinner dashed lines indicate the boundary between the Early and Late Cretaceous. K1 and K2 represent the Early and Late Cretaceous, respectively. The scale below represents time in millions of years ago (MYA). Numbers beside branches denote divergence times, and calibrated points marked with red dots were used for estimating divergence times.

autosomes and X chromosomes of aphids, we found the rate of autosomal rearrangements in aphids was significantly higher than those of X chromosome (Supplementary Table 5, Mann–Whitney *U*-test, p value = 0.048). It is noteworthy that no obvious inter-autosomal rearrangements were detected between *A. gossypii* and *R. maidis*, with their autosomal rearrangement rates closely resembling those of their X chromosomes (Supplementary Table 5). We speculated that the increased rate of autosomal rearrangements in aphids was mainly due to frequent inter-autosomal rearrangements. To investigate this, we expanded the inclusion of species (*Aphis fabae* and *Rhopalosiphum padi*)[34] to compare rearrangement rates between the X chromosome and autosomes in species without such inter-autosomal rearrangements. Our findings consistently demonstrated that the rearrangement rates of the X chromosome were generally higher than those of autosomes in aphids without inter-autosomal rearrangements (Supplementary Table 6 and Supplementary Fig. 3).

Secondly, to determine whether the high rearrangement rate of autosomes extends to other Hemiptera families, we compared the chromosomal rearrangement rates among each pair of *Daktulosphaira vitifoliae* vs *Adelges*

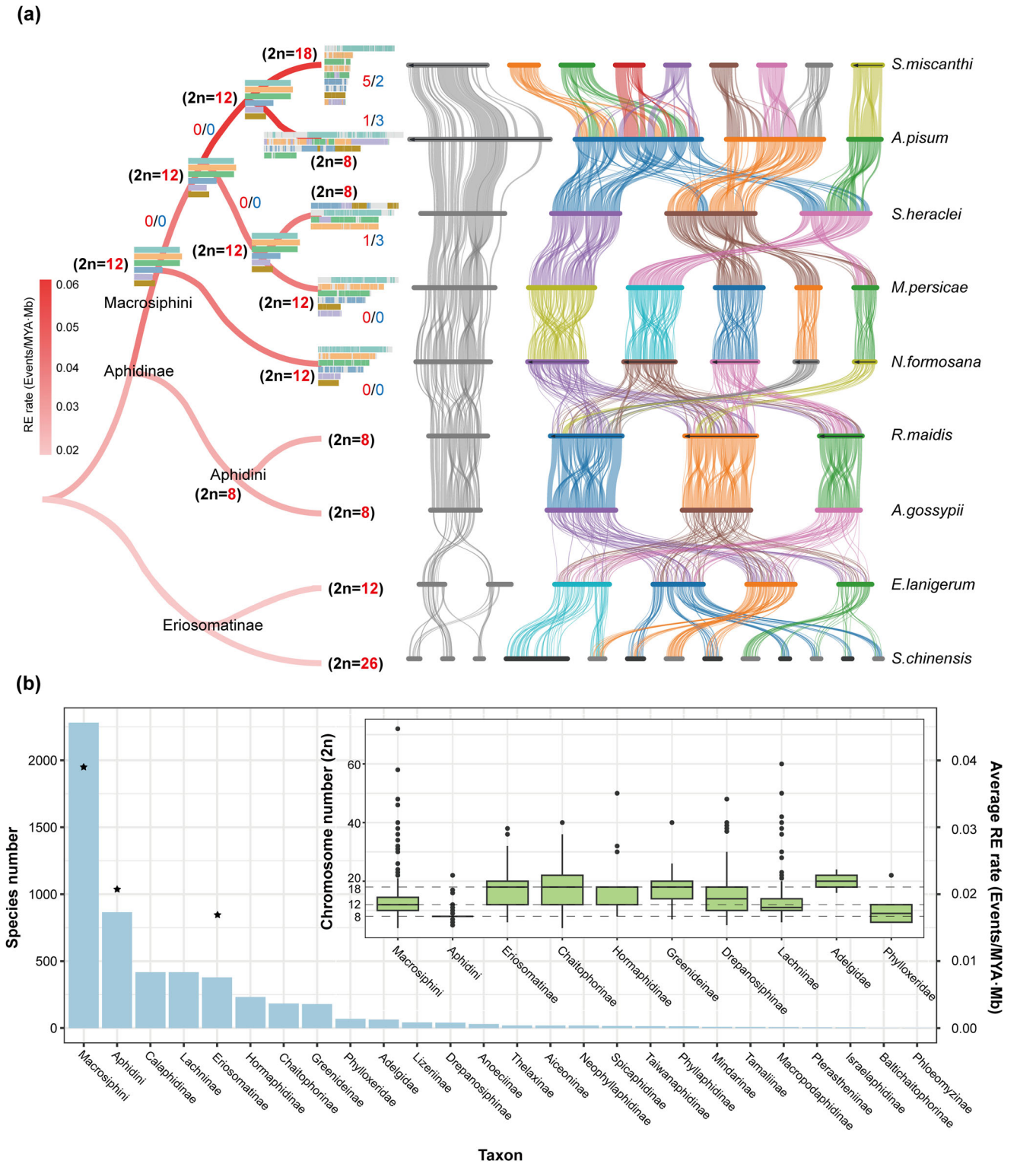

Fig. 2 | Species diversity and chromosomal evolution in aphids. a On the left, the phylogenetic tree illustrates the relationships between the nine aphid species. Branches in the phylogeny are color-coded based on their respective average chromosomal rearrangement rates (supplementary Table 5). Rectangles on nodes represent the ancestral karyotype reconstruction, inferring a common ancestral karyotype ($2n = 12$) for five aphid species (*S. heraclei*, *M. persicae*, *S. miscanthi*, *A. pisum*, and *N. formosana*). Conservative blocks are coded according to the chromosomes from their common ancestor. Different colored numbers indicate the detected occurrences of chromosomal fissions (in red) and fusions (in blue). On the right, the genome syntenic plot is generated based on the gene order. Lines indicate the boundaries of syntenic gene blocks identified by MCScanX, with unconnected regions on the chromosomes possibly lacking syntenic gene blocks. Different chromosomes are color-coded with different links, and the X chromosome for each species is located on the far left and is depicted in gray. **b** The bar graph illustrates the number of species in each group within the Aphidomorpha[15], with pentagrams indicating the average chromosome rearrangement rate for corresponding groups. Additionally, the boxplot depicts the chromosome counts in Macrosiphini, Aphidini, and other major subfamilies of aphids[16] (median line, 25th–75th percentiles as boxes, Dots represent outliers). The dotted line indicates several major karyotypes of aphids[15].

cooleyi (Sternorrhyncha, Phylloxeroidea), *Rhodnius prolixus* vs *Triatoma rubrofasciata* (Heteroptera, Reduvioidea), as well as *Nilaparvata lugens*[51] vs *Laodelphax striatellus*[52] (Auchenorrhyncha, Fulgoroidea). Although *D. vitifoliae* and *A. cooleyi* show better autosomal collinearity than aphids (Supplementary Fig. 4), the higher autosomal rearrangement rate of autosome than that of the X chromosome was also observed, but this trend was not found in the latter two pairs' comparisons (Supplementary Table 5, Mann–Whitney *U*-test, p value = 0.028).

Thirdly, after testing correlations between the rate of chromosomal rearrangements with species diversities of each respective aphid group, the results show the former are positively associated with the reported species numbers in each aphid group (Fig. 2b and Supplementary Table 5, Pearson correlation coefficient, $R = 0.9995$, p value = 0.020). Among aphids, the Macrosiphini tribe exhibits a significantly higher rate of chromosomal rearrangements compared to other aphid groups (Supplementary Table 5, Mann–Whitney *U*-test, p value = 0.048), particularly in species such as *A. pisum* and *S. miscanthi*, where the rearrangement rates are exceptionally high.

Reconstructing ancestral Macrosiphini chromosomes, we can clearly see the evolutionary history of aphid chromosomes, and further reveal the significant changes that have occurred in aphid autosomes during species differentiation. We reconstructed the ancestral karyotype of the Macrosiphini using chromosomal data from *S. heraclei*, *M. persicae*, *S. miscanthi*, *A. pisum*, and *N. formosana*. This analysis revealed the ancestral Macrosiphini (Macro) karyotype, consisting of six pairs of chromosomes ($2n = 12$) at a 50 kb resolution, covering ~73% of the *S. heraclei* genome and 45% of the *A. pisum* genome (*M. persicae*: 71%, *N. formosana*: 55%, *S. miscanthi*: 54%) (Fig. 2a and Supplementary Table 7). Among these species, *N. formosana* diverged earliest, maintaining the ancestral karyotype with only intra-chromosomal rearrangements and sequence expansion. In contrast, *A. pisum* and *S. miscanthi* experienced more chromosomal fusions and fissions, despite their shortest divergence time (Fig. 2a). Due to the limitation of available data, we cannot currently reconstruct the ancestral karyotype of Aphidinae. Nevertheless, based on the counted karyotype in aphid species (Fig. 2b), it is evident that Aphidini possesses a highly stable karyotype[16,53], and its ancestral karyotype is more likely equal to $2n = 8$.

The differences in transposable elements abundance and genome size in aphids

Transposable elements (TEs) play a significant role in shaping the dynamic landscape of genome evolution[54,55]. The results of the comparative analysis of TEs content showed *R. maidis* and *A. gossypii* from the Aphidini tribe displayed significantly lower TE content (<22%) compared to other aphid species (from 28 to 39%) (Fig. 3a and Supplementary Table 8, Mann–Whitney *U*-test, p value = 0.028). Additionally, the composition of different types of TEs varied significantly among different aphid species. In the genomes of Aphidinae species, the higher proportions of SINE retrotransposons were detected, as well as the Macrosiphini species were found significantly enriched of DNA (Mann–Whitney *U*-test, p value = 0.01587) and SINE retrotransposons (Mann–Whitney *U*-test, p value = 0.01587) than those of other involved aphid species (Fig. 3a and Supplementary Table 8), indicating the association between the events of TEs insertions and phylogenetic signals.

The distribution of TEs in different chromosomes exhibited significant enrichment on the X chromosome of species in Aphidinae, particularly at the ends (Supplementary Fig. 5 and Supplementary Table 9, Mann–Whitney *U*-test, p value = 0.01099). In contrast, the distribution of TEs on autosomes is relatively uniform, with only a few individual chromosomes displaying differences (Supplementary Fig. 5). Notably, no significant TEs enrichment was observed on the X chromosomes of *S. chinensis* and *E. lanigerum* from the Eriosomatinae subfamily (Supplementary Fig. 6), nor in the outgroup species of *Therioaphis trifolii* from subfamily Calaphidinae[56] (Supplementary Table 10). These findings suggest that ancient TEs expansions on X chromosomes in the early ancestor of Aphidinae.

Among different species of Aphidinae species, their genome size shows a significant positive correlation with the proportion of TEs within the genome, specifically, the DNA transposons and SINE retrotransposons (Supplementary Fig. 7 and Supplementary Table 11, Mann–Whitney *U*-test, p value = 0.02334). However, we found that this relationship between TEs and genome size is not consistent across subfamilies. We observed that the two species in Eriosomatinae have shorter single-copy sequence lengths and fewer protein-coding genes than species in Aphidinae (with the exception of *R. maidis*) (Supplementary Table 12, Mann–Whitney *U*-test, p value = 0.036), and despite their much smaller genome sizes, the TEs proportions of their genomes are not significantly different from that of *M. persicae*. Therefore, the difference in genome size between Eriosomatinae and Aphidinae may be influenced by factors beyond TEs, such as sequence expansion in Aphidinae or genome reduction and streamlining in Eriosomatinae.

TEs activity in the evolutionary process of aphids genome

Through analysis of DNA transposon insertion time distributions, we found that five aphid species within Macrosiphini exhibit significantly more recent expansion events compared to four other aphid species (Supplementary Fig. 8 and Supplementary Table 13, Mann–Whitney *U*-test, p value = 0.008). Additionally, despite a short divergence time between *A. pisum* and *S. miscanthi* (Fig. 1c), the former exhibits sustained and significant expansion of DNA transposon, a trend that contributes to its largest genome size (Fig. 3a and Supplementary Table 8).

In the Aphidinae subfamily, all species exhibit a significant peak in SINE retrotransposons age distribution (Fig. 3b and Supplementary Fig. 9). However, the two Eriosomatinae species, *E. lanigerum* and *S. chinensis*, have minimal annotations of SINE retrotransposons, as well as species of Chaitophorinae (*S. flava*) and Lachninae (*C. cedri*) (Supplementary Table 8), suggesting that the specific expansion of SINE retrotransposons in the common ancestor of Aphidinae after its' they diverged with Eriosomatinae. Research on SINE retrotransposons suggests that SINE transposons can provide transcription factor binding sites for neighboring genes, potentially participating in the regulation of temporal gene expression[57], and can cause DNA double-strand breaks through adjacent endonuclease target site duplications[58]. The expansion of SINE retrotransposons may have had a profound impact on aphid species in Aphidinae.

Enrichment of TEs at synteny breakpoints

TEs may trigger genomic rearrangement events through non-allelic homologous recombination[59–61]. To investigate if frequent chromosomal rearrangements in aphids are associated with TEs, we examined TE content at synteny breakpoints in seven Aphidinae species. The enrichment of TEs at these breakpoints varied slightly among species pairs, but DNA, SINE, and unknown elements were significantly enriched at synteny breakpoints in all pairs, indicating their substantial contribution to aphid chromosomal rearrangements (Supplementary Table 14). The most abundant DNA elements were Kolobok-E and hAT-Tip100, which also dominated the genomes. To pinpoint which TE families significantly contribute to genomic rearrangements in aphids, we further analyzed TE enrichment at synteny breakpoints. We found that DNA/Kolobok-E, DNA/MULE-MuDR, DNA/hAT-hAT19, DNA/hAT-Tip100, LINE/I-Jockey, and SINE were all significantly enriched at these breakpoints across all aphids (Fig. 3c and Supplementary Table 15). Compared with five aphid species from Macrosiphini, two species from the Aphidini lineage exhibited fewer rearrangements (Fig. 1c and Supplementary Table 5) and had lower densities of these TE families (only except for SINEs) (Fig. 3d), which may indicate the contributions of these TE families to the genomic rearrangements in aphids.

Chromosomal rearrangement breakpoints and TADs

As functional units of chromosome 3D structures, topologically associating domains (TADs) play a crucial role in gene regulation, expression stability, and other genomic functions[62]. Chromosomal rearrangements can impact the formation and maintenance of TADs, thereby influencing gene

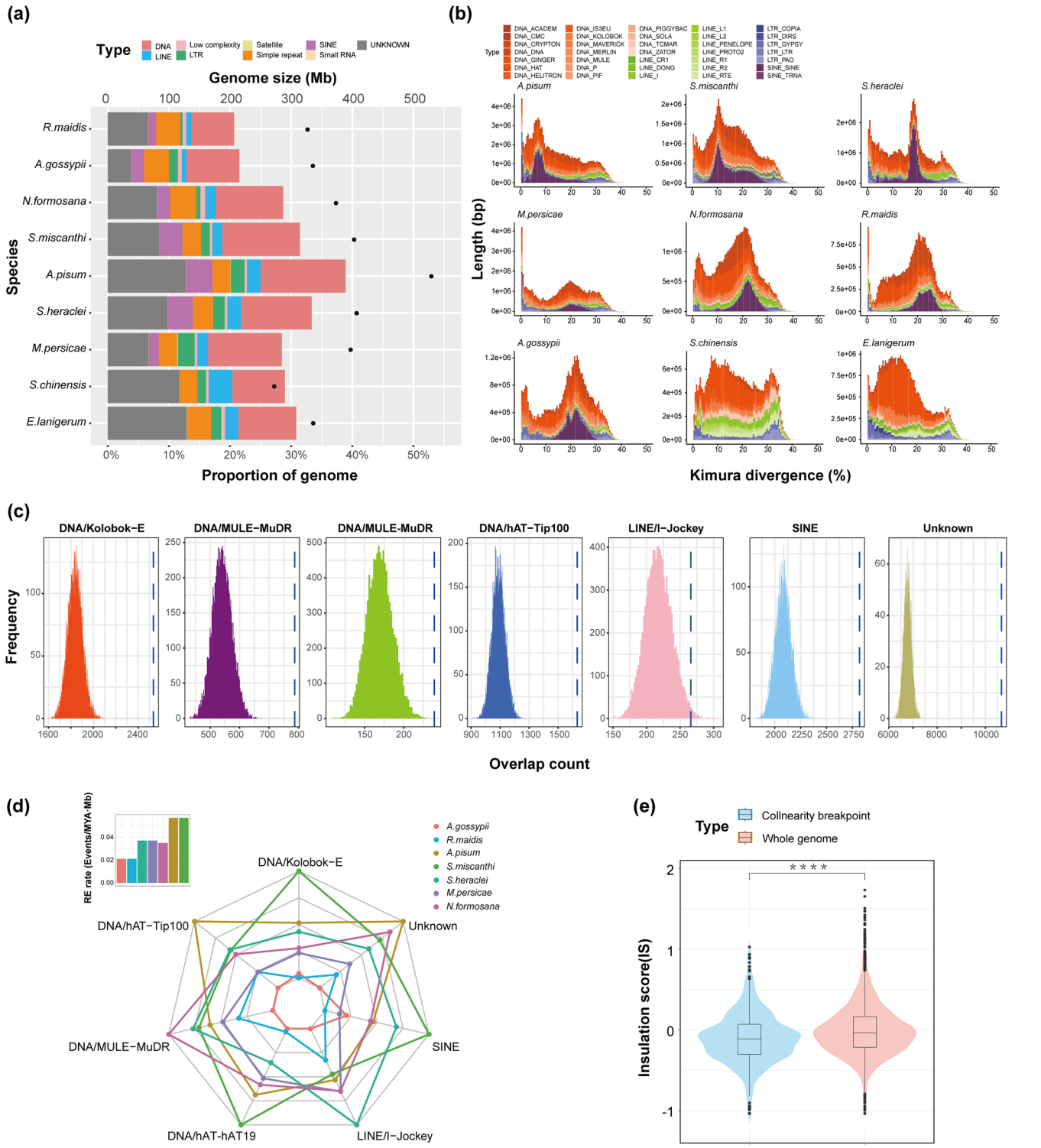

Fig. 3 | Results of the analysis on aphid repetitive sequences and chromosomal rearrangement breakpoints. a Genome size and the proportion of repetitive sequences in the genomes of the nine aphid species. **b** Stacked histogram displaying the age distribution of transposable elements (TEs) in the nine aphid species, with different colors representing different TE families. **c** TE families enriched at all aphid chromosomal rearrangement breakpoints. The histogram displays the distribution of counts for a specific TE family within 10,000 randomly sampled chromosomal regions of equal size to the rearrangement breakpoint regions. The dashed line represents the observed count for the corresponding TE family in rearrangement breakpoint regions. The results shown are for *S. miscanthi* within the collinear species pairs with *A. pisum*. **d** Radar plot illustrating the relative density (/Mb) of TE families from Fig. 3c in the genomes of seven aphid species in Aphidinae. Different colors represent different species, and the intersection of the axes and network lines indicates the relative density of the corresponding TE family in the seven aphid genomes. The histogram in the top left corner shows the chromosomal rearrangement rates for these seven aphid species. **e** Violin plot comparing the IS (Insulation Score) between *S. heraclei* and *M. persicae* for the whole-genome alignment and synteny breakpoint regions (Mann–Whitney *U*-test, p value = 6.549×10^{-12}) (median line, 25th–75th percentiles as boxes, Dots represent outliers).

expression regulation by repositioning genes to new chromatin interaction regions, ultimately affecting species adaptation[63]. In this study, we examined the insulation scores (ISs) of rearrangement sites, and the results revealed enrichment of chromosomal rearrangement breakpoints at TAD boundaries in the seven Aphidinae species (Fig. 3e and Supplementary Fig. 10). This indicates that the locations of chromosomal rearrangement breakpoints are not random but tend to occur near TAD boundaries. Consequently, such breakpoints may have a limited impact on disrupting gene regulation, suggesting that purifying selection rather than positive selection may be driving these chromosomal rearrangements[64]. While chromosomal rearrangement breakpoints tended to occur at TAD boundaries, both chromosome fusions and rearrangements still led to changes in the TAD structure (Supplementary Fig. 11).

X chromosome evolution

A comparative analysis of X chromosome lengths across aphid species revealed significant variability (Supplementary Table 16). The X chromosome of *A. pisum* was 67,744,731 bp longer than that of *S. heraclei*, with 46,995,892 bp (69.34%) of its length comprising repetitive sequences dominated by TEs, which significantly contribute to the chromosome's length (Supplementary Table 17). Furthermore, the *A. pisum* X chromosome contained more genes compared to other aphid species and showed a notable enrichment in genes associated with DNA transposons, heterochromatin, telomeres, and the Kelch-like gene family (Supplementary Fig. 12 and Tables 16, 18). This gene enrichment was linked to the expansion of repetitive sequences. Given the predominance of multicopy gene families on the X chromosome[14], *A. pisum* has undergone extensive gene family amplifications. Overall, the considerable differences in X chromosome lengths among aphid species are primarily due to the cumulative effects of TEs expansion and unique sequence amplification.

The role of specific gene loss and duplication in the chromosomal evolution of aphids

To investigate the genetic factors underlying extensive autosomal rearrangements in aphids, we conducted a comprehensive gene loss analysis spanning multiple aphid species, along with representative insects from the Prosorrhyncha (*R. prolixus* (obtained from DNA Zoo[46,65]) and *T. rubrofasciata*[66]), Auchenorrhyncha (*Homalodisca vitripennis*[67], *Laodelphax striatellus*[52], and *Nilaparvata lugens*[51]), and Sternorrhyncha (*D. vitifoliae*[68], *A. cooleyi*[69], *Bemisia tabaci*[70], and *Diaphorina citri*[71]) suborders. Our analysis identified the specific loss of 122 genes exclusive to the aphid lineage (Supplementary Table 25). Three genes of particular interest among the lost genes are *RIF1* (Replication timing regulatory factor 1), *BRD8* (Bromodomain containing 8), and *DMC1* (DNA meiotic recombinase 1), were all associated with DNA double-strand break repair (DSB) and genome stability. To confirm their loss, we analyzed Illumina sequencing data for *S. heraclei*, further validating their absence in the aphid lineage. Our investigation dated the loss of *RIF1*, *BRD8*, and *DMC1* genes to the common ancestor of the suborder Aphidomorpha (Fig. 4b). *RIF1* encodes a protein regulating replication timing and interacting with PP1[72]. Additionally, it is a key regulator of *TP53BP1*, which in turn plays a pivotal role in DSB repair by modulating non-homologous end joining (NHEJ) and counteracting homologous recombination (HR) repair mediated by *BRCA1*[73]. Loss of *RIF1* may reduce NHEJ efficiency and increase HR frequency, potentially promoting chromosomal rearrangements and compromising genome stability[74,75]. *BRD8*, part of the NuA4 histone acetyltransferase complex, plays a vital role in DNA DSB repair and genome stability. Studies have demonstrated that *BRD8* depletion impairs genome stability[76,77]. *DMC1*, a conserved meiosis-specific double-strand repair gene, is essential for homologous recombination during meiosis[78]. Its loss has been reported in *Drosophila*, *Anopheles*, *Caenorhabditis elegans*[78], and *Daphnia pulex* (a cyclical parthenogenesis[79]). However, the precise effects of *DMC1* loss remain unclear. Investigating the loss of *DMC1* in eukaryotes will contribute to understanding the evolution of homologous recombination mechanisms in meiosis and potentially in cyclical parthenogenesis, which warrants further investigation.

N. formosana is the earliest diverging species among the five studied Macrosiphini species, retained the ancestral karyotype (Fig. 2a). Gene family expansion analysis 51 gene families exhibited specific expansion in *N. formosana* (Supplementary Table 26). Among these genes, the *TERT* (Telomerase reverse transcriptase) gene displayed tandem duplication ~258 kbp downstream (Fig. 4a and Supplementary Fig. 19). Furthermore, the gene order around *TERT* observed in Aphidinae aphids differed from that in Chaitophorinae, Eriosomatinae, even in Phylloxeridae (Supplementary Fig. 20). By examining the assembly quality around *TERT*, transcriptome data of *N. formosana* and other aphid genomes (Supplementary Fig. 21), we

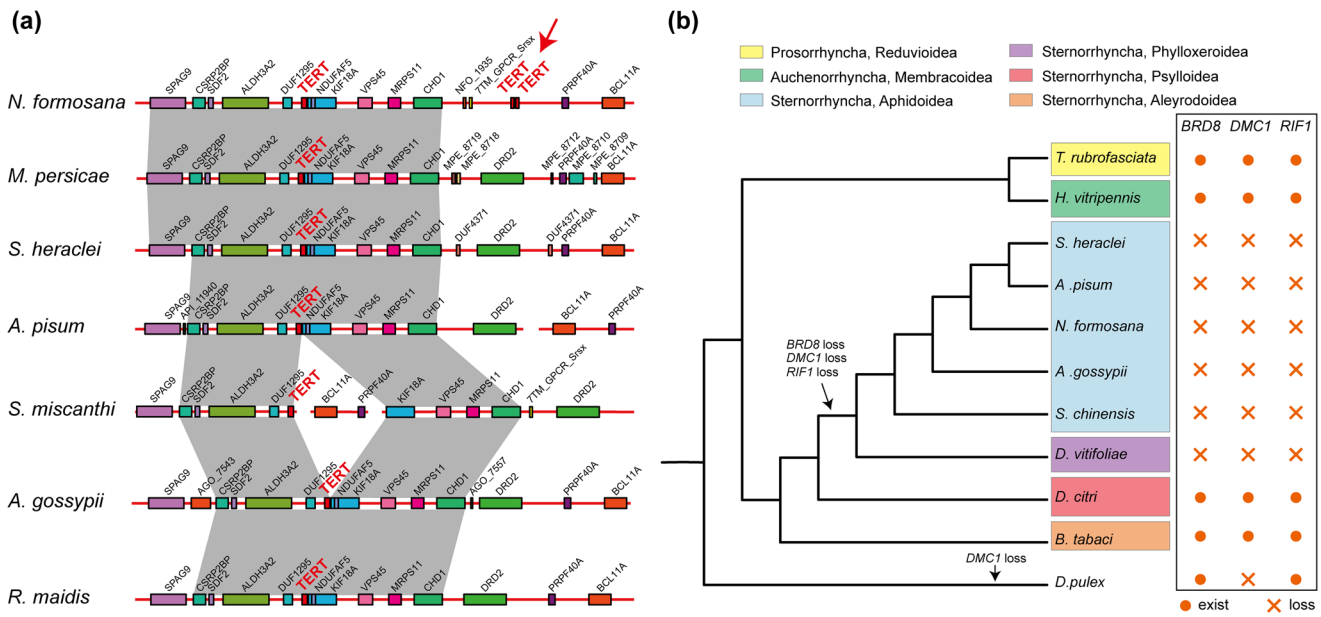

Fig. 4 | Lineage-specific gene loss, duplication, and genome composition in aphids. a Gene order in the genomic region near *TERT*. Gray links represent genomic regions conserved across all species, and *TERT* is highlighted in red. Red arrows point to the concatenated duplication location of the *TERT* gene. A more comprehensive gene order plot for various species is presented in Supplementary Fig. (Supplementary Fig. 19). **b** Timing of gene loss for *BRD8*, *RIF1*, and *DMC1*. Different colors indicate different taxonomic groups.

confirmed that *TERT* duplication in *N. formosana* are not caused by genome assembly errors or annotation errors.

Natural selection on genes linked to chromosomal rearrangements in Macrosiphini

Macrosiphini is the most diverse group in terms of species richness among aphids[15]. To test whether natural selection was associated with the radiation of Macrosiphini, we conducted a positive selection analysis and identified 104 single-copy genes exhibiting evidence of positive selection (Supplementary Tables 27, 28). The DNA mismatch repair gene *spel1*, a homolog of the human gene *MSH2*, is under positive selection in Macrosiphini (Supplementary Table 28). This gene plays a crucial role in post-replication mismatch repair, reducing DNA mutations and maintaining genome stability[80,81]. The positive selection of *spel1* may contribute to preventing the accumulation of genetic mutations, further maintaining genetic stability in organisms.

We further conducted gene family contraction and expansion analysis to investigate their evolutionary significance in Macrosiphini. The results revealed a significant expansion of DNA transposon-related genes in Aphidinae, especially within Macrosiphini (Supplementary Fig. 22 and Supplementary Table 29). This expansion correlated with the higher relative abundance of DNA transposons in their genomes compared to other aphids (Supplementary Table 8).

Detoxification genes and host adaptation evolution in *Semiaphis heraclei*

We used the BITACORA pipeline[82] to reannotate the detoxification gene families (P450s, CCEs, UGTs, and GSTs) in *S. heraclei* and eight other chromosome-level aphid genomes. Functional domains and motifs were identified through visual comparison (Fig. 5a and Supplementary Figs. 23–25), revealing tandem repeat domain characteristics in certain genes, which IGV[83] confirmed as errors in annotating adjacent repeat genes as a single gene. After correction, we enumerated the detoxification gene families across the nine aphid species (Supplementary Table 30).

Notably, *S. heraclei* did not exhibit the highest numbers of P450s, CCEs, UGTs, and GSTs among aphids, which is expected, as species like *M. persicae* and *A. gossypii*—with broader host ranges—likely require larger detoxification gene repertoires to cope with diverse plant secondary metabolites[31]. Species-specific tandem repeats were identified in *S. heraclei*'s CCE Esterase, UGTs Cluster 2, and P450s CYP3/CYP4 families, primarily localized on autosomes (Supplementary Fig. 26). The CCE Esterase plays a key role in insecticide resistance, plant metabolite detoxification, and olfactory degradation[84], while UGTs are involved in detoxifying plant allelochemicals, insecticide resistance, and other physiological functions[85,86]. Similarly, P450 CYP3 and CYP4 are associated with insecticide resistance[87] and detoxification of plant allelochemicals in insects[88,89]. These expansions may enhance the detoxification capacity of *S. heraclei*. Additionally, we compared the expression levels of P450s, CCEs, and UGTs in male and sexual female *S. heraclei* aphids. The analysis showed that most of these genes were upregulated in males, especially those with species-specific expansions (Supplementary Fig. 27), suggesting that males face greater detoxification pressures due to flight and migration.

For GSTs, we classified them into Sigma, Omega, Theta, and Delta classes. Omega and Theta classes showed stable gene counts, while Sigma and Delta class varied significantly (Fig. 5a and Supplementary Table 30). Despite Sigma class expansion, sequences remained highly conserved (Fig. 5a), indicating purifying selection. We identified a GST Sigma gene in *S. heraclei* (SHE.evm.model.SH1.1550) with two insertion events (LP and V), the LP insertion occurring in the conserved N-terminal domain (Supplementary Fig. 28a). Transcriptomic analysis excluded annotation errors (Supplementary Fig. 28b), and SWISS-MODEL predicted that the LP insertion alters the protein structure, potentially affecting gene function (Supplementary Fig. 28c).

Additionally, a species-specific expansion of the GST Delta class was observed in *S. heraclei*, with genes SHE.evm.model.SH1.4558 and SHE.evm.model.SH1.4561 exhibiting unique tandem GST_N/GST_C domain repeats (Fig. 5a). IGV and transcriptomic sequencing confirmed no annotation errors (Fig. 5b), and the transcriptomic assembly validated the gene structure annotation (Supplementary Fig. 29). Both genes were classified into the Glutathione S-transferase family based on InterPro[90] and eggNOG[91] annotations, though no homologs with similar structures were found in NCBI or Uniprot databases. These genes may be novel genes in *S. heraclei* that contain GST_N and GST_C domains, potentially playing key roles in its adaptive evolution. Furthermore, GST genes were predominantly upregulated in males (Fig. 5c), consistent with the expression patterns of other detoxification genes.

Discussion

Aphid genomes can acquire new genetic content through chromosomal rearrangements and transposon insertions, potentially impacting vital adaptive genes and facilitating adaptation to new ecological niches[92]. In parallel, Aphidinae have the widest host–plant range among aphids[93]. Our study supports that the rate of chromosomal evolution varies among different aphid lineages and correlates positively with species diversity, as the most species-rich group of Macrosiphini exhibits significantly higher rates of chromosomal rearrangement compared to other lineages, suggesting the contribution of chromosomal rearrangements in speciation and adaptive evolution of aphid evolution.

In addition to the absence of rearrangements with autosomes, the chromosomal rearrangement rate of aphids' X chromosomes is significantly lower than that of their autosomes. This evolutionary pattern is also evident in aphids' sister groups, *D. vitifoliae* and *A. cooleyi*. Recently, Li et al. compared the inter-chromosomal rearrangements of *D. vitifoliae*, *A. cooleyi*, and *A. pisum*, and they suggested that the increased autosomal rearrangement rates are unique to aphids[17]. However, increasing the lineage sampling, our observations indicate that some aphid lineages also lack inter-autosomal rearrangements, such as *M. persicae* with *N. formosana* and *A. gossypii* with *R. maidis*. Our findings imply that when not distinguishing between intra- and inter-chromosomal rearrangements, the evolutionary pattern of increased autosomal rearrangement rates may not be unique to aphids but may be common across the Aphidomorpha.

We hypothesized that the observed discrepancies in rearrangement rates between the X chromosome and autosomes might be primarily attributed to inter-autosomal rearrangements. To test this, we calculated the rearrangement rates for aphid species that exclusively undergo intra-chromosomal rearrangements and discovered that the X chromosome exhibits a higher rearrangement rate than the autosomes. This finding aligns with our observations of TEs accumulation on the X chromosome, low-level gene expression, and previous studies indicating higher levels of recombination and relaxed purifying selection on the X chromosome[12,14,94].

Mechanisms such as dosage compensation or complete elimination of the X chromosome may explain the lack of inter-chromosomal rearrangements between the X chromosome and autosomes[12,14]. Research indicates that the aphid X chromosome primarily carries sexually antagonistic genes that benefit males[95]. Consequently, during sexual reproduction, rearrangements might disturb the regulation of these genes or change the expression of linked autosomal genes, reducing male aphid fitness. This is supported by the observation that chromosomal translocations present in asexual populations of *M. persicae* are eliminated in the holocyclic population[53]. However, the reasons behind significant inter-chromosomal rearrangements in aphid autosomes are still unclear. Our collinearity analysis reveals that all *A. cooleyi* autosomes exhibit homology with only one *D. vitifoliae* autosome respectively, while in Eriosomatinae, several *S. chinensis* autosomes each match two *E. lanigerum* autosomes (Supplementary Fig. 4). This aligns with the finding of ref. 17, indicating that cyclical parthenogenesis does not fully explain the frequent inter-chromosomal rearrangements observed in aphids. To resolve this issue, future studies could collect more chromosomal-level data from various aphid subfamilies and Phylloxeroidea to reconstruct the chromosomal evolutionary history of their respective taxa, integrating molecular biology and ecological data for comparative analysis.

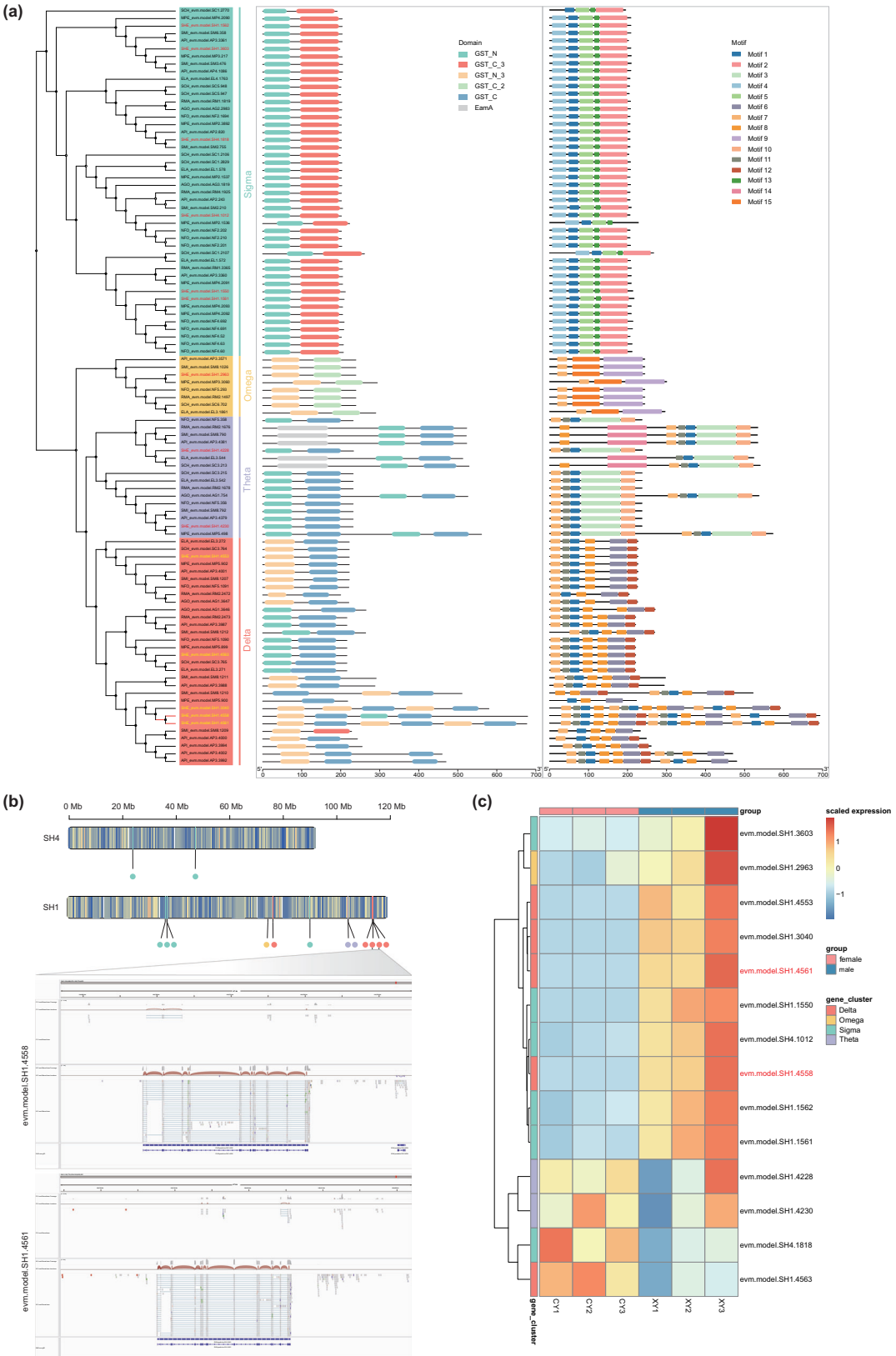

Fig. 5 | Specific expansion and expression patterns of GSTs genes in the *S. heraclei* genome. a Phylogenetic relationships, domain and motif between GSTs from different aphids. Highlight the gene names belonging to *S. heraclei* with distinct colors. The gene branches in *S. heraclei* that underwent species-specific expansion are highlighted in red. **b** GSTs gene distribution in the *S. heraclei* genome and IGV verification of SHE_evm.model.SH1.4558 and SHE_evm.model.SH1.4561 annotations and assembly errors. The upper track represents transcriptome reads from sexual female aphids, while the lower track represents transcriptome reads from male aphids. **c** Expression differences of GSTs genes between sexual females and males.

Our research results demonstrate that TEs have a significant influence on the evolution of aphid chromosomes. We identified a specific expansion of SINE elements within the Aphidinae lineage. This expansion may contribute to variations in genome structural stability and gene expression regulation[57,58] between Aphidinae and other aphid lineages. Furthermore, our observations indicate that Aphidini with stable karyotypes possess a notably reduced proportion of TEs, especially LINE elements, in their genomes compared to other aphids. This finding is supported by the analysis of TE content cross multiple different aphid groups by ref. 11. We speculate that the extensive or recent expansion of TEs may lead to more frequent chromosomal fission or other structural rearrangements that alter chromosome number[96,97], which could be one of the reasons for the differences in karyotype stability among aphids.

This study observed enrichment of certain SINE and DNA elements near the breakpoints of chromosomal rearrangements in aphids, and their density is notably lower in the genomes of Aphidini species with lower chromosomal rearrangement rates, suggesting an association with these rearrangements. Additionally, we noted a significant expansion of gene families related to DNA transposons within the Macrosiphini group, which may lead to an increase in DNA elements in their genomes. Furthermore, we observed that the content of SINE elements in the genomes of Macrosiphini aphids is significantly higher than that in other aphids studied. We hypothesize that the elevated rate of chromosomal rearrangements in Macrosiphini aphids may be related to the expansion of DNA and SINE elements, which may drive chromosomal rearrangements through non-allelic homologous recombination[58,59]. Previous research has also revealed similar trends of chromosomal rearrangement within Erebia subclades and specific genera of Lepidoptera[10].

We observed that the Macrosiphini aphids, compared to other aphid species, have experienced more recent expansion events of DNA elements (Supplementary Table 13), particularly in *A. pisum*, which possesses a greater abundance of young DNA and SINE elements that are currently undergoing accelerated expansion (Supplementary Figs. 8, 9). Given that young repetitive elements are more likely to participate in ectopic recombination[98,99], this elucidates why *A. pisum* and *S. miscanthi* display significantly higher rates of chromosomal rearrangements than other aphids. Additionally, the reconstructed chromosomal evolution history of Macrosiphini aphids indicated that the lineages of *A. pisum* and *S. miscanthi* experienced more chromosomal breakage and fusion compared to other Macrosiphini aphids. It is noteworthy that, compared to the over 2000 species within the Macrosiphini tribe, the number of published chromosome-level genomes remains limited. High-quality chromosome-level genomes with broad taxonomic representation are still lacking, hindering our ability to fully unravel the chromosomal evolution history of the Macrosiphini tribe. Furthermore, there is a significant gap in the availability of genome assemblies for the other subfamilies of aphids except Aphidinae, which obstructs aphidologists from gaining a comprehensive understanding of aphid chromosomal evolutionary history. Addressing this issue requires more available chromosome-level aphid genomes in the future.

Gene loss and duplication may play a crucial role in aphid chromosomal evolution. Specifically, the loss of genes related to DNA double-strand break repair, such as *RIF1*, *BRD8*, and *DMC1*[100], could potentially impact the chromosomal stability of aphids, a similar mechanism observed in parrots[64]. Moreover, the duplication of the *TERT* gene, which encodes the enzyme responsible for elongating DNA telomeres and functionally essential for maintaining chromosome stability after successive rounds of DNA replication, plays a crucial role in chromosome end replication across eukaryotes[101–103], may have a positive effect on the stability of the karyotype in *N. formosana*. Future experimental works are needed to test the influence of gene loss or duplication on the rapid chromosomal evolution in aphids.

Insect detoxification pathways are divided into three stages: in the first stage, P450s and CCEs convert xenobiotics into more hydrophilic products; in the second stage, GSTs and UGTs further conjugate metabolites in preparation for excretion; and in the third stage, transporters expel the transformed products from cells[84,104]. These genes play crucial roles in detoxifying plant secondary metabolites and conferring resistance in insects[105–107], and gene duplication or amplification often leads to changes in detoxification gene expression, one of the most common mechanisms of insect resistance[108,109]. For example, amplification of the *CYP9A* subfamily has been shown to contribute to host adaptation and diverse insecticide resistance in insects such as *Spodoptera exigua* and *Spodoptera frugiperda*[110]. Additionally, amplification of the *CYP6CY* subfamily is closely associated with nicotine adaptation in the tobacco-adapted subspecies *M. p. nicotianae*[92]. Our findings indicate that species-specific expansions of the P450s CYP3 and CYP4, CCEs Esterase, and UGTs gene families in *S. heraclei* may enhance their detoxification capacity against plant secondary metabolites. Further analysis shows that mutations and expansions in the GST gene family, particularly the Sigma and Delta classes, are also closely related to environmental adaptation[111,112]. Notably, the expansion of GST Delta class genes may help *S. heraclei* adapt to specific environmental toxins. Moreover, significant sex-specific differences in the expression of detoxification genes were observed, with detoxification genes generally upregulated in male aphids, suggesting they face greater detoxification pressures during flight and migration. Future studies could further explore the functional roles of these genes in adaptive evolution through techniques such as gene knockout and RNA interference.

In conclusion, the rapid karyotype evolution in aphids, which contributes to their rapid speciation and adaptation, is the result of a combination of various factors. The characteristic expansion of TEs, specific gene duplication and loss, and the alternating sexual and asexual generations in their life cycle may all accelerated the karyotype evolution in aphids. Additionally, the mutation and expansion of detoxification gene families in *S. heraclei* may be a key factor in adapting to host–plant chemical defenses.

Materials and methods

Sample collection and sequencing

The parthenogenetic wingless adult and nymphal aphids of *Semiaphis heraclei* were collected in May 2019 from a planting base of *Lonicera japonica* (variety: 'Siji jinyinhua') located in Zhengcheng Town, Pingyi County, Shandong Province, China (35°16′0″N, 117°38′56″E). In the laboratory, the aphids were reared on cuttings of *Lonicera japonica* for over ten generations under controlled conditions (temperature: 25 ± 1°C, relative humidity: 70 ± 5%, light cycle: 14 h light : 10 h dark)[113]. The specimens are stored at the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences.

We selected 100 aphids of various developmental stages from the population in a laboratory and placed them into cryovials, which were then rapidly frozen in liquid nitrogen. The frozen samples were sent to Biomarker Technologies (Beijing, China) for library preparation and sequencing. DNA was quantified through 0.75% agarose gel electrophoresis, Nanodrop spectrophotometry (Thermo Fisher), and Qubit 3.0 fluorometry (Invitrogen). PacBio long reads were generated from sequencing on the PacBio Sequel II platform, while Hi-C reads and Illumina short reads were produced from sequencing on the Illumina NovaSeq 6000 platform.

To assist in gene annotation and conduct differential gene expression analysis, we sequenced the transcriptomes of three biological replicates, each of wingless sexual females and winged males, following the method described by ref. 12.

Genome assembly and annotation

The genome size of *S. heraclei* was estimated using k-mer analysis, based on Illumina short paired-end reads (~42 Gb) filtered using the default parameters of fastp v0.23.2[114] and GCE v1.0.0[115] with 17-mer. For the initial assembly, PacBio subreads were assembled using the Canu v2.2[116] with parameters "genomeSize = 400 m" and "-pacbio". We checked the contig assembly for host contamination using the BlobTools pipeline v1.1.1 by generating taxon annotated GC content-coverage plot (Supplementary Fig. 1). Each contig was searched against the NCBI nucleotide database (nt, downloaded on September 10, 2023) using BLASTN v2.13.0. Following the removal of contamination, we used purge_dups v1.2.5[117] to identify and

remove redundant sequences. Subsequently, the contigs were polished by NextPolish v1.4.0[118] with three rounds using pacbio long reads and Illumina short paired-end reads. To obtain a chromosome-level assembly, the assembled contigs were then anchored to super-scaffolds by Hi-C reads applying Juicer v1.6.0[119] and 3D-DNA v180922[65], followed by manual correction using Juicebox Assembly Tools v1.9.9[120]. Finally, the completeness of our genome assembly was evaluated by Benchmarking Universal SingleCopy Orthologs (BUSCO) v5.3.2 with parameters "-m genome -l hemiptera_odb10".

In this study, we employed a comprehensive analysis pipeline to identify repeats. This pipeline involves de novo and homolog-based prediction. For the de novo prediction, we first constructed a de novo repeat custom library for each species with RepeatModeler v2.0.1 (https://github.com/Dfam-consortium/RepeatModeler) and then merged the repeats library with known repeats from the Repbase Insecta library[121] using ReannTE_MergeFasta.pl (https://github.com/4ureliek/ReannTE). For the homology-based method, we used Repeatmasker v4.1.2-p1 (https://www.repeatmasker.org/) to identify repeats across each assembly based on the merged library generated by each species. Through this approach, we identified repeats in all aphid assemblies, including four from Macrosiphini, two from Aphidini, two from Eriosomatinae, one from Chaitophorinae, and one from Lachninae (Supplementary Table 8).

Homology-based, de novo and transcriptome analysis was carried out to identify the protein-coding gene models of *S. heraclei*, following these steps. First, we used the GeMoMa pipeline v.1.8[122] and the MMseq2 v13.45111[123] search engine to identify the protein-coding genes. Our selection of closely related species included *A. pisum*, *M. persicae*, *A. glycines*, *D. noxia*, *A. gossypii*, *S. flava*, *H. cornu*, *N. vitripennis*, and *D. melanogaster*. Subsequently, the annotated gene sets were combined with the model GAF of GeMoMa using the filter parameters "iAA ≥ 0.8 and ce/rce = 1". The protein sequences of these species were aligned to the *S. heraclei* genome using tblastn v2.12.0[124] with an e-value parameter of 1e-5, and the candidate gene regions were further refined with genewise v2-4-1[125] to obtain more accurate gene structures. Second, we mapped RNA-seq reads to the *S. heraclei* assembly using HISAT2 v2.2.1[126] and constructed transcripts using stringTie v2.2.1[127]. Third, we performed de novo prediction of protein-coding genes within *S. heraclei* assembly using Augustus v3.4.0[128], GlimmerHMM v3.0.4[129], and SNAP v2017-03-01[130]. Finally, all of the gene set models generated in the above steps were combined using EvidenceModeler v1.1.1 (https://github.com/EVidenceModeler/EVidenceModeler).

To avoid biases from different annotation strategies and to ensure consistency and accuracy in gene annotation, the protein-coding genes of the other eight chromosome-level aphid genomes were re-annotated using the same de novo and homology-based methods as *S. heraclei* (Supplementary Tables 4, 12). Likewise, EvidenceModeler v1.1.1 was used to combine the gene set models.

Phylogenetic analysis and divergence time estimation

We used protein-coding genes for *S. heraclei* and eight other chromosome-level assembled aphids (*A. pisum*, *M. persicae*, *S. miscanthi*, *N. formosana*, *A. gossypii*, *R. maidis*, *E. lanigerum*, *S. chinensis*), along with the published protein-coding genes from six other insects (*S. flava*, *C. cedri*, *D. vitifoliae*, *A. cooleyi*, *R. prolixus*, *T. rubrofasciata*), to perform orthologue identification (Supplementary Table 4). The longest transcript of each gene was selected to identify orthologue groups among these 15 species using the Orthofinder pipeline v2.5.4[131]. This process resulted in a total of 2358 single-copy orthologues. These single-copy genes were aligned using PRANK v170427[132], and subsequently, TrimAL v1.4.rev15[133] was used to remove gaps and unaligned sites. All amino acid alignments were concatenated into a super-gene alignment, producing a concatenated sequence alignment of 982,857 sites. This protein sequence alignment was used to construct a maximum likelihood species tree using RAxML v8.2.12[134] with the PROTGAMMAWAG model.

R8s v1.81[135] was used to estimate the divergence times of these species, incorporating two fossil calibration points: Aphidomorpha (135 Ma[136,137]) and Aphidinae (70 Ma[137]), along with a previously used secondary calibration point at the root of Hemiptera (386 Ma[138]).

Synteny analysis

We identified syntenic blocks of genes among the nine aphid species and other Hemiptera insects using MCSCANX v1.1[139]. The pairwise comparison of the annotated protein sequences was performed using BLASTP v2.13.0[124,140] with parameters "-outfmt 6 -evalue 1e-10 -max_target_seqs 5" and subsequently ran MCSCANX with parameters "-s 10 -b 2". The results from MCSCANX were visualized using the SynVision website (https://synvisio.github.io/#/).

Due to the extremely rapid genome evolution in aphids[12], we cannot calculate the rearrangement rate by aligning the genomes of all aphids to the genome of a reference species. So, as a compromise, we compared the genomic rearrangement rates between species pairs with close phylogenetic relationships to assess the genomic rearrangement rates in different subfamilies. We used in-house-made Perl scripts to convert syntenic blocks generated by MCSCANX into genomic coordinates. Genomic rearrangements, including inversions, translocations, and inverted translocations, were then detected using the previously described method[141]. The rearrangement rate for each species pair was calculated by dividing the number of rearrangements by the divergence time between the species pair and the average alignment length. To calculate the chromosomal rearrangement rate of *L. striatellus* and *N. lugens*, we added these two species to the aforementioned dataset and re-estimated the divergence time using r8s v1.81. The synteny dot plot was generated using WGDI v0.6.5[142]. Since the genome assembly of *A. cooleyi* is at the scaffold level, we used the 10 longest scaffold sequences to roughly estimate the chromosomal rearrangement rate of *D. vitifoliae* and *A. cooleyi*.

Ancestral karyotype reconstruction

Initially, we selected *A. pisum*, *S. miscanthi*, *S. heraclei*, and *N. formosana* genomes as input and performed alignment against the *M. persicae* genome using LAST v1420. Subsequently, we used UCSC tools (https://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/) to convert the LAST alignment results into chain and net results. The conserved segments of the five Macrosiphini species were obtained using these chain and net results in conjunction with the DESCHRAMBLER algorithm (https://github.com/jkimlab/DESCHRAMBLER). Finally, the branch-and-bound algorithm in ANGES v1.01[143] to infer ancestral nodes on the phylogenetic tree and construct ancestral genome structures.

Repetitive sequence analysis

To assess the enrichment of transposable elements (TEs) in chromosomal breakpoint regions, we used BEDtools v2.30.0[144] and the TE annotation results described above to investigate the content of TEs in chromosomal breakpoint regions of all aphid species, as well as in randomly selected genome regions. We defined chromosomal breakpoint regions as regions located upstream or downstream of the endpoints of syntonic blocks obtained from MCSCANX analysis, with a length of 30,000 bp. To avoid potential interference from highly repetitive telomere sequences near chromosome ends, we excluded these regions from our analysis. BEDtools intersect was used to count the number of TEs within the breakpoint regions. To calculate the significance of TE enrichment (p value), we used the same approach to count the number of TEs in 10,000 sets of random regions within the genomes. Each random region was 30,000 bp in size, and the number of regions equaled the number of breakpoint regions. To prevent potential bias from highly abundant TEs on a particular chromosome, we limited the selection of random regions to each respective chromosome, ensuring that the number of randomly selected regions corresponded to the number of breakpoint regions on that chromosome. The p value for each TE class was calculated by dividing the number of random regions in which the TE count was greater than or equal to the observed count in the breakpoint regions by the number of simulations (10,000). We used parseRM.pl (https://github.

com/4ureliek/Parsing-RepeatMasker-Outputs) to parse the output of RepeatMasker for generating the TE age distribution.

Hi-C contact analysis

Hi-C reads mapping was performed using BWA v0.7.17-r1188[145]. Following this, we filtered, corrected, binned, and normalized the mapping results using the default parameters of HiCExplorer v3.6[146]. Hi-C reads from *S. heraclei* were mapped to their respective genomes, while Hi-C reads from *M. persicae* were mapped to the *S. heraclei* genome, with uniquely mapped reads retained. Subsequently, the data was merged and binned to generate a genome-wide interaction map with a 100 kb resolution. We used hicFindTADs v3.7.2 to calculate insulation scores and identify TAD boundaries, with the parameters "--minDepth 400000 --maxDepth 800000 --numberOfProcessors 4 --step 100000 --thresholdComparisons 0.05 --correctForMultipleTesting fdr".

Gene family analysis

To detect lineage-specific lost genes in aphids, we used the default parameters of the OrthoFinder v2.5.4 software to identify orthologues among nine aphid species, two Prosorrhyncha insects (*R. prolixus* and *T. rubrofasciata*), four Auchenorrhyncha insects (*H. vitripennis*, *M. quadrilineatus*, L. striatellus, and *N. lugens*), and four Sternorrhyncha insects (*D. vitifoliae*, *A. cooleyi*, *B. tabaci*, and *D. citri*). Subsequently, from the orthogroup results, we selected orthologues that were absent in all nine aphid species but present in other insects. These were initially designated as genes lost in aphids. We then used TBLASTN v2.12.0 to search for these potentially lost genes in the genomes of the nine aphid species. Additionally, we validated the presence of these lost genes by examining the transcriptome of *S. heraclei*, as they may be hidden genes resulting from assembly errors. Finally, we manually searched the NCBI[147] and UniProt[148] databases to determine if these genes exist in the aphids and whether other species possess them.

We used the same approach to search for species-specific duplication genes in *N. formosana*. Homologs among the nine chromosome-level aphid species were identified by the default parameters of OrthoFinder v2.5.4. Subsequently, we searched the orthogroup results to identify gene families in which all aphid species except *N. formosana* are single-copy genes. Subsequently, we used the tblastn software to identify the *TERT* genomic regions in the genomes of the other eight aphid species. We also used IGV v2.12.3[83] to examine the genomic assembly continuity of the *TERT* tandem duplication region in the *N. formosana* genome and the transcript support for the *TERT* copies.

Additionally, CAFÉ v4.2[149] was used to detect contractions and expansions of gene families based on the divergence times of the species. Furthermore, we performed GO term and KEGG pathway enrichment analyses for candidate genes using clusterProfiler v4.0.5[150] R package.

Natural selection analysis

The codeml program in PAML v4.5[151] package was used to calculate nonsynonymous and synonymous substitution rates (dN/dS). We applied the branch site model and one- and two-ratio models to detect signatures of natural selection on coding genes in the Macrosiphini lineage. Statistical significance was determined using likelihood ratio tests.

Evolutionary analysis of detoxification gene families

We used the BITACORA pipeline[82] to identify detoxification genes in *S. heraclei* and eight other aphid species with chromosome-level genomes (Supplementary Table 4). This approach integrates popular sequence similarity-based search algorithms and GeMoMa for the annotation of gene family copies within genomes. Subsequently, we manually filtered the gene family annotation results based on sequence length and domain integrity to generate high-confidence annotations of detoxification gene families. The maximum likelihood phylogenetic tree was constructed using MUSCLE v5.1[152] and iqtree v2.3.6[153] with the best-fit model. Additionally, we utilized meme v5.5.7[154] (-nmotifs 15 -maxw 100 -minw 10) and pfam_scan.py v1.0[155] to identify motifs and functional domains within the detoxification protein sequences, and TBtools-II v2.144[156] was used for result visualization. Protein structure prediction and visualization were performed using the SWISS-MODEL web server[157] and PyMOL v3.1.3[158], respectively.

Statistics and reproducibility

For comparisons between two independent groups, we first assessed data normality using the Shapiro–Wilk test. When normality assumptions were violated ($p < 0.05$), non-parametric analysis was performed using the Mann–Whitney *U*-test. This approach was applied to comparisons including chromosome rearrangement rates, transposable element abundance, gene expression levels, and chromosome breakpoint IS versus whole-genome IS analyses. For correlation analyses between two variables, Pearson correlation coefficients were calculated when both datasets satisfied normality criteria. Non-normally distributed data were analyzed using Spearman's rank correlation. These methods were employed to assess relationships between genome size and transposable element abundance, species diversity, and chromosome rearrangement rates, as well as transposable element abundance and gene expression levels. All experiments were conducted with a minimum sample size of $n = 3$ biological replicates, defined as independently prepared samples processed through identical experimental conditions. The critical significance value (α) was set at 0.05, with p values <0.05 considered statistically significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The whole-genome shotgun sequencing data have been deposited in GenBank under accession number JBEIVQ000000000. Raw sequence reads generated in this study are accessible through the NCBI Short Read Archive under BioProject PRJNA1061761. The genome assembly of *S. heraclei* is available at NCBI under accession GCA_043165465.1. For comparative genomic analyses, nine chromosome-level aphid genome assemblies with their genome and re-annotated annotation files have been archived on Figshare[159] with the identifier https://doi.org/10.6084/m9.figshare.27314556. All Supplementary Tables are compiled in Supplementary Data 1, while numerical source data underlying graphs are provided in both Supplementary Data 2 and the aforementioned figshare repository. Additional datasets not specified in this section can be obtained from the corresponding author upon reasonable request.

Code availability

The scripts used for comparative genomics analysis, chromosomal rearrangement rate analysis, and repetitive sequence analysis are available at GitHub (https://github.com/Mrhuangc/aphids_data_analysis).

Received: 14 July 2024; Accepted: 28 February 2025;
Published online: 13 March 2025

References

1. Wellenreuther, M., Merot, C., Berdan, E. & Bernatchez, L. Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. *Mol. Ecol.* **28**, 1203–1209 (2019).
2. Shaw, C. J. & Lupski, J. R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.* **13**, R57–R64 (2004).
3. Ciccarelli, F. D. et al. Complex genomic rearrangements lead to novel primate gene function. *Genome Res.* **15**, 343–351 (2005).
4. Blackman, R. L., Spence, J. M., Field, L. M. & Devonshire, A. L. Chromosomal location of the amplified esterase genes conferring resistance to insecticides in *Myzus persicae* (Homoptera Aphididae). *Heredity* **75**, 297–302 (1995).
5. Fedoroff, N. V. Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758–767 (2012).

6. Chénais, B., Caruso, A., Hiard, S. & Casse, N. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene* **509**, 7–15 (2012).
7. Thompson, F. L., Iida, T. & Swings, J. Biodiversity of vibrios. *Microbiol. Mol. Biol. Rev.* **68**, 403–431 (2004).
8. Mullen, S. P. & Shaw, K. L. Insect speciation rules: unifying concepts in speciation research. *Annu. Rev. Entomol.* **59**, 339–361 (2014).
9. Hill, J. et al. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci. Adv.* **5**, eaau3648 (2019).
10. Cornet, C. et al. Holocentric repeat landscapes: from micro-evolutionary patterns to macro-evolutionary associations with karyotype evolution. *Mol. Ecol*. **33**, e17100 (2023).
11. Baril, T., Pym, A., Bass, C. & Hayward, A. Transposon accumulation at xenobiotic gene family loci in aphids. *Genome Res.* **33**, 1718–1733 (2023).
12. Mathers, T. C. et al. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Mol. Biol. Evol.* **38**, 856–875 (2020).
13. Shigenobu, S. & Yorimoto, S. Aphid hologenomics: current status and future challenges. *Curr. Opin. Insect Sci.* **50**, 100882 (2022).
14. Li, Y., Park, H., Smith, T. E. & Moran, N. A. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Mol. Biol. Evol.* **36**, 2143–2156 (2019).
15. Colin, F. & David, C. E. Aphid species file. Version 5.0/5.0. (2023).
16. Gavrilov-Zimin, I. A., Stekolshchikov, A. V. & Gautam, D. C. General trends of chromosomal evolution in Aphidococca (Insecta, Homoptera, Aphidinea + Coccinea). *Comp. Cytogenet.* **9**, 335–422 (2015).
17. Li, Z. et al. Phylloxera and aphids show distinct features of genome evolution despite similar reproductive modes. *Mol. Biol. Evol*. **40**, msad271 (2023).
18. Monti, V., Lombardo, G., Loxdale, H. D., Manicardi, G. C. & Mandrioli, M. Continuous occurrence of intra-individual chromosome rearrangements in the peach potato aphid, *Myzus persicae* (Sulzer) (Hemiptera: Aphididae). *Genetica* **140**, 93–103 (2012).
19. Drinnenberg, I. A., deYoung, D., Henikoff, S. & Malik, H. S. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife*. **3**, e03676 (2014).
20. Lucek, K., Augustijnen, H. & Escudero, M. A holocentric twist to chromosomal speciation? *Trends Ecol. Evol.* **37**, 655–662 (2022).
21. Mandrioli, M., Zanasi, F. & Manicardi, G. C. Karyotype rearrangements and telomere analysis in Myzuspersicae (Hemiptera, Aphididae) strains collected on Lavandula sp. plants. *Comp. Cytogenet.* **8**, 259–274 (2014).
22. Ruckman, S. N., Jonika, M. M., Casola, C. & Blackmon, H. Chromosome number evolves at equal rates in holocentric and monocentric clades. *PLoS Genet.* **16**, e1009076 (2020).
23. Song, K.-X. et al. Secondary rhinaria contribute to major sexual dimorphism of antennae in the aphid *Semiaphis heraclei* (Takahashi). *Insects* **14**, 468 (2023).
24. War, A. R. et al. Mechanisms of plant defense against insect herbivores. *Plant Signal Behav.* **7**, 1306–1320 (2014).
25. Chen, X. et al. Research progress on the biosynthesis, metabolic engineering, and pharmacology of bioactive compounds from the *Lonicera* genus. *Med. Plant Biol.* **3**, 0–0 (2024).
26. Li, Y., Li, W., Fu, C., Song, Y. & Fu, Q. Lonicerae japonicae flos and Lonicerae flos: a systematic review of ethnopharmacology, phytochemistry and pharmacology. *Phytochem. Rev.* **19**, 1–61 (2019).
27. Thiviya, P., Gunawardena, N., Gamage, A., Madhujith, T. & Merah, O. Apiaceae family as a valuable source of biocidal components and their potential uses in agriculture. *Horticulturae*. **8**, 614 (2022).
28. Pavela, R., Maggi, F. & Benelli, G. Coumarin (2H-1-benzopyran-2-one): a novel and eco-friendly aphicide. *Nat. Prod. Res.* **35**, 1566–1571 (2019).
29. Al-Khayri, J. M. et al. Plant secondary metabolites: the weapons for biotic stress management. *Metabolites*. **13**, 716 (2023).
30. Sayed-Ahmad, B., Talou, T., Saad, Z., Hijazi, A. & Merah, O. The *Apiaceae*: rthnomedicinal family as source for industrial uses. *Ind. Crops Prod.* **109**, 661–671 (2017).
31. Ramsey, J. S. et al. Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. *Insect Mol. Biol.* **19**, 155–164 (2010).
32. Mathers, T. C. et al. Supplementary data for: chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Mol. Biol. Evol*. 38, 856–875 (2020).
33. Mathers, T. C. et al. Hybridisation has shaped a recent radiation of grass-feeding aphids. *BMC Biol.* **21**, 157 (2023).
34. Mathers, T. C. et al. Aphidinae comparative genomics resource. *Zendo* https://doi.org/10.5281/zenodo.5908004 (2022).
35. Ye, S. et al. A chromosome-level genome assembly of *Neotoxoptera formosana* (Takahashi, 1921) (Hemiptera: Aphididae). *G3* **12**, jkac164 (2022).
36. Zhang, S. et al. Chromosome-level genome assemblies of two cotton-melon aphid *Aphis gossypii* biotypes unveil mechanisms of host adaption. *Mol. Ecol. Resour.* **22**, 1120–1134 (2021).
37. Chen, W. et al. Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). *Gigascience* **8**, giz033 (2019).
38. Biello, R. et al. A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Mol. Ecol. Resour.* **21**, 316–326 (2021).
39. Biello, R. et al. A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* (Hausman) (Hemiptera: Aphididae). [Data set]. Zendo https://doi.org/10.5281/zenodo.3797130 (2020).
40. Wei, H. Y. et al. Chromosome-level genome assembly for the horned-gall aphid provides insights into interactions between gall-making insect and its host plant. *Ecol. Evol.* **12**, e8815 (2022).
41. Laetsch, D. R. & Blaxter, M. L. BlobTools: interrogation of genome assemblies. *F1000Research* **6** (2017).
42. Chen, X. & Zhang, G. The karyotypes of fifty-one species of aphids (Homoptera, Aphidoidea) in Beijing area. *Acta Zool. Sin.* **31**, 12–18 (1985).
43. Manni, M., Berkeley, M. R., Seppey, M. & Zdobnov, E. M. BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* **1**, e323 (2021).
44. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
45. Gabaldón, T. et al. Phylogenomics identifies an ancestral burst of gene duplications predating the diversification of *Aphidomorpha*. *Mol. Biol. Evol.* **37**, 730–756 (2020).
46. Mesquita, R. D. et al. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl Acad. Sci. USA* **112**, 14936–14941 (2015).
47. Owen, C. L. & Miller, G. L. Phylogenomics of the *Aphididae*: deep relationships between subfamilies clouded by gene tree discordance, introgression and the gene tree anomaly zone. *Syst. Entomol.* **47**, 470–486 (2022).
48. Wing, S. L. & Boucher, L. D. Ecological aspects of the Cretaceous flowering plant radiation. *Annu. Rev. Earth Planet. Sci.* **26**, 379–421 (1998).
49. Von Dohlen, C. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biol. J. Linn. Soc.* **71**, 689–717 (2000).

50. Züst, T. & Agrawal, A. A. Mechanisms and evolution of plant resistance to aphids. *Nat. Plants* **2**, 1–9 (2016).
51. Ma, W. et al. Chromosomal-level genomes of three rice planthoppers provide new insights into sex chromosome evolution. *Mol. Ecol. Resour.* **21**, 226–237 (2021).
52. Zhu, J. et al. Genome sequence of the small brown planthopper, *Laodelphax striatellus*. *Gigascience* **6**, 1–12 (2017).
53. Blackman, R. Chromosome numbers in the *Aphididae* and their taxonomic significance. *Syst. Entomol.* **5**, 7–25 (1980).
54. Gilbert, C., Peccoud, J. & Cordaux, R. Transposable elements and the evolution of insects. *Annu. Rev. Entomol.* **66**, 355–372 (2021).
55. Bohne, A., Brunet, F., Galiana-Arnoux, D., Schultheis, C. & Volff, J. N. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res.* **16**, 203–215 (2008).
56. Huang, T. et al. Chromosome-level genome assembly of the spotted alfalfa aphid *Therioaphis trifolii*. *Sci. Data* **10**, 274 (2023).
57. Lu, J. Y. et al. Genomic repeats categorize genes with distinct functions for orchestrated regulation. *Cell Rep.* **30**, 3296–3311 e5 (2020).
58. Hedges, D. & Deininger, P. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat. Res.* **616**, 46–59 (2007).
59. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368 (2007).
60. Konkel, M. K. & Batzer, M. A. A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. *Semin. Cancer Biol*. **20**, 211–221 (2010).
61. Qi, L. et al. Shuffling the yeast genome using CRISPR/Cas9-generated DSBs that target the transposable Ty1 elements. *PLoS Genet.* **19**, e1010590 (2023).
62. Gonzalez-Sandoval, A. & Gasser, S. M. On TADs and LADs: spatial control over gene expression. *Trends Genetics* **32**, 485–495 (2016).
63. Valton, A. L. & Dekker, J. TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* **36**, 34–40 (2016).
64. Huang, Z. et al. Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nat. Commun.* **13**, 944 (2022).
65. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
66. Liu, Q. et al. A chromosomal-level genome assembly for the insect vector for Chagas disease, *Triatoma rubrofasciata*. *Gigascience* **8**, giz089 (2019).
67. Li, Z. et al. The genomic basis of evolutionary novelties in a leafhopper. *Mol. Biol. Evol*. **39**, msac184 (2022).
68. Rispe, C. et al. The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biol.* **18**, 90 (2020).
69. Dial, D. T. et al. Whole-genome sequence of the Cooley spruce gall adelgid, *Adelges cooleyi* (Hemiptera: Sternorrhyncha: Adelgidae). *G3* **14**, jkad224 (2024).
70. Xie, W. et al. Genome sequencing of the sweetpotato whitefly *Bemisia tabaci* MED/Q. *Gigascience* **6**, 1–7 (2017).
71. Carlson, C. R. et al. High-quality, chromosome-scale genome assemblies: comparisons of three *Diaphorina citri* (Asian citrus psyllid) geographic populations. *DNA Res*. **29**, dsac027 (2022).
72. Mattarocci, S. et al. Rif1 controls DNA replication timing in yeast through the PP1 phosphatase Glc7. *Cell Rep.* **7**, 62–69 (2014).
73. Silverman, J., Takai, H., Buonomo, S. B., Eisenhaber, F. & de Lange, T. Human *Rif1*, ortholog of a yeast telomeric protein, is regulated by ATM and 53BP1 and functions in the S-phase checkpoint. *Genes Dev.* **18**, 2108–2119 (2004).
74. Escribano-Díaz, C. et al. A cell cycle-dependent regulatory circuit composed of *53BP1-RIF1* and *BRCA1-CtIP* controls DNA repair pathway choice. *Mol. Cell* **49**, 872–883 (2013).
75. Ohno, Y., Ogiyama, Y., Kubota, Y., Kubo, T. & Ishii, K. Acentric chromosome ends are prone to fusion with functional chromosome ends through a homology-directed rearrangement. *Nucleic Acids Res.* **44**, 232–244 (2016).
76. Paulsen, R. D. et al. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol. Cell* **35**, 228–239 (2009).
77. Lashgari, A., Fauteux, M., Marechal, A. & Gaudreau, L. Cellular depletion of *BRD8* causes p53-dependent apoptosis and induces a DNA damage response in non-stressed cells. *Sci. Rep.* **8**, 14089 (2018).
78. Lin, Z., Kong, H., Nei, M. & Ma, H. Origins and evolution of the *recA/RAD51* gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc. Natl Acad. Sci. USA* **103**, 10328–10333 (2006).
79. Schurko, A. M., Logsdon, J. M. Jr. & Eads, B. D. Meiosis genes in Daphnia pulex and the role of parthenogenesis in genome evolution. BMC Evol. Biol. 9, 78 (2009).
80. de Wind, N., Dekker, M., Berns, A., Radman, M. & te Riele, H. Inactivation of the mouse *Msh2* gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell* **82**, 321–330 (1995).
81. Traver, S. et al. MCM9 is required for mammalian DNA mismatch repair. *Mol. Cell* **59**, 831–839 (2015).
82. Vizueta, J., Sánchez-Gracia, A. & Rozas, J. BITACORA: a comprehensive tool for the identification and annotation of gene families in genome assemblies. *Mol. Ecol. Resour.* **20**, 1445–1452 (2022).
83. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
84. Cruse, C., Moural, T. W. & Zhu, F. Dynamic roles of insect carboxyl/cholinesterases in chemical adaptation. *Insects*. **14**, 194 (2023).
85. Bull, D. L. & Whitten, C. J. Factors influencing organophosphorus insecticide resistance in tobacco budworms. *J. Agric. Food Chem.* **20**, 561–564 (1972).
86. Ahmad, S. & Hopkins, T. β-glucosylation of plant phenolics by phenol β-glucosyltransferase in larval tissues of the tobacco hornworm, *Manduca sexta* (L.). *Insect Biochem. Mol. Biol.* **23**, 581–589 (1993).
87. Scharf, M., Parimi, S., Meinke, L. J., Chandler, L. & Siegfried, B. D. Expression and induction of three family 4 cytochrome P450 (CYP4) * genes identified from insecticide-resistant and susceptible western corn rootworms, *Diabrotica virgifera virgifera*. *Insect Mol. Biol.* **10**, 139–146 (2001).
88. Liu, B., Fu, D., Ning, H., Tang, M. & Chen, H. Knockdown of *CYP6CR2* and *CYP6DE5* reduces tolerance to host plant allelochemicals in the Chinese white pine beetle *Dendroctonus armandi*. *Pestic. Biochem. Physiol*. **187**, 105180 (2022).
89. Zhang, Y. et al. Three cytochrome P450 CYP4 family genes regulated by the CncC signaling pathway mediate phytochemical susceptibility in the red flour beetle, *Tribolium castaneum*. *Pest Manage Sci.* **78**, 3508–3518 (2022).
90. Paysan-Lafosse, T. et al. InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427 (2023).
91. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
92. Singh, K. S. et al. The genetic architecture of a host shift: an adaptive walk protected an aphid and its endosymbiont from plant chemical defenses. *Sci. Adv.* **6**, eaba1070 (2020).
93. Yan, F., Gexia, Q. & Guangxue, Z. Diversity of host plants and feeding sites of aphids. *Acta Zool. Sin*. **31**, 31–39 (2006).

94. Jaquiery, J. et al. Disentangling the causes for Faster-X evolution in aphids. *Genome Biol. Evol.* **10**, 507–520 (2018).
95. Li, Y., Zhang, B. & Moran, N. A. The aphid X chromosome is a dangerous place for functionally important genes: diverse evolution of hemipteran genomes based on chromosome-level assemblies. *Mol. Biol. Evol.* **37**, 2357–2368 (2020).
96. Carbone, L. et al. Gibbon genome and the fast karyotype evolution of small apes. *Nature* **513**, 195–201 (2014).
97. Sylvester, T. et al. Lineage-specific patterns of chromosome evolution are the rule not the exception in Polyneoptera insects. *Proc. Biol. Sci.* **287**, 20201388 (2020).
98. Renkawitz, J., Lademann, C. A. & Jentsch, S. Mechanisms and principles of homology search during recombination. *Nat. Rev. Mol. Cell Biol.* **15**, 369–383 (2014).
99. Li, L., Jean, M. & Belzile, F. The impact of sequence divergence and DNA mismatch repair on homeologous recombination in *Arabidopsis*. *Plant J.* **45**, 908–916 (2006).
100. Bishop, D. K., Park, D., Xu, L. & Kleckner, N. DMC1: a meiosis-specific yeast homolog of *E. coli recA* required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **69**, 439–456 (1992).
101. Monti, V., Mandrioli, M., Rivi, M. & Manicardi, G. C. The vanishing clone: karyotypic evidence for extensive intraclonal genetic variation in the peach potato aphid, *Myzus persicae* (Hemiptera: Aphididae). *Biol. J. Linn. Soc.* **105**, 350–358 (2012).
102. Sharma, G. G. et al. hTERT associates with human telomeres and enhances genomic stability and DNA repair. *Oncogene* **22**, 131–146 (2003).
103. Yamaguchi, H. et al. Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *N. Engl. J. Med.* **352**, 1413–1424 (2005).
104. BK, S. K., Moural, T. & Zhu, F. Functional and structural diversity of insect glutathione S-transferases in xenobiotic adaptation. *Int. J. Biol. Sci.* **18**, 5713 (2022).
105. Gao, L., Qiao, H., Wei, P., Moussian, B. & Wang, Y. Xenobiotic responses in insects. *Arch. Insect Biochem. Physiol.* **109**, e21869 (2022).
106. Yang, Z. et al. Characterization and functional analysis of UDP-glycosyltransferases reveal their contribution to phytochemical flavone tolerance in *Spodoptera litura*. *Int. J. Biol. Macromol*. **261**, 129745 (2024).
107. Shou-Min, F. Insect glutathione S-transferase: a review of comparative genomic studies and response to xenobiotics. *Bull. Insectol.* **65**, 265–271 (2012).
108. Nauen, R., Bass, C., Feyereisen, R. & Vontas, J. The role of cytochrome P450s in insect toxicology and resistance. *Annu. Rev. Entomol.* **67**, 105–124 (2022).
109. Heckel, D. G. Insect detoxification and sequestration strategies. *Annu. Plant Rev.* **47**, 77–114 (2014).
110. Shi, Y. et al. Divergent amplifications of *CYP9A* cytochrome P450 genes provide two noctuid pests with differential protection against xenobiotics. *Proc. Natl Acad. Sci. USA* **120**, e2308685120 (2023).
111. Song, X. et al. Functional diversification of three delta-class glutathione S-transferases involved in development and detoxification in *Tribolium castaneum*. *Insect Mol. Biol.* **29**, 320–336 (2020).
112. Song, X. W. et al. Characterization of a sigma class GST (GSTS6) required for cellular detoxification and embryogenesis in *Tribolium castaneum*. *Insect Sci.* **29**, 215–229 (2021).
113. Shi, Z.-H. et al. Morphological characteristics for identifying the instars of *Semiaphis heraclei*. *J. Appl. Entomol.* **59**, 457–465 (2022).
114. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
115. Liu, B. et al. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. Preprint at arXiv:1308.2012. (2013).
116. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
117. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
118. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
119. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
120. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
121. Bao, W., Kojima, K. K. & Kohany, O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 1–6 (2015).
122. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89–e89 (2016).
123. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
124. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
125. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).
126. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
127. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
128. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
129. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
130. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 1–9 (2004).
131. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
132. Löytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170 (2014).
133. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
134. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
135. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
136. Havill, N. P., Foottit, R. G. & von Dohlen, C. D. Evolution of host specialization in the Adelgidae (Insecta: *Hemiptera*) inferred from molecular phylogenetics. *Mol. Phylogenet. Evol.* **44**, 357–370 (2007).
137. Heie, O. The evolutionary history of aphids and a hypothesis on the coevolution of aphids and plants. *Boll. Zool. Agrar. Bachic.* **28**, 149–155 (1996).
138. Johnson, K. P. et al. Phylogenomics and the evolution of hemipteroid insects. *Proc. Natl Acad. Sci. USA* **115**, 12775–12780 (2018).

139. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49–e49 (2012).
140. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
141. Liu, J. et al. A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. *Genome Res.* **31**, 497–511 (2021).
142. Sun, P. et al. WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* **15**, 1841–1851 (2022).
143. Jones, B. R., Rajaraman, A., Tannier, E. & Chauve, C. ANGES: reconstructing ANcestral GEnomeS maps. *Bioinformatics* **28**, 2388–2390 (2012).
144. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
145. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
146. Wolff, J. et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).
147. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
148. Consortium, U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
149. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
150. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
151. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
152. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
153. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
154. Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. MEME: discovering an.d analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **34**, W369–W373 (2006).
155. Finn, R. D. et al. The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222 (2010).
156. Chen, C. et al. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* **16**, 1733–1742 (2023).
157. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
158. DeLano, W. L. Pymol: an open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92 (2002).
159. Huang, C. et al. A comparative genomic analysis at the chromosomal-level reveals evolutionary patterns of aphid chromosomes. [Data set]. figshare https://doi.org/10.6084/m9.figshare.27314556 (2025).

Acknowledgements

This work was supported by the National Natural Science Foundation of China 81873095 (K.G.) and 32470445 (G.L.), the Natural Science Basic Research Program of Shaanxi 2020JM-280 (G.L.), Fundamental Research Funds for the Central Universities GK201902008 (G.L.) and GK202304017 (Y.R), the Key Project of Shaanxi Natural Science Basic Research Plan 2024JC-YBMS-152 (Z.L.), the Key Project of Scientific Research Program of Shaanxi Provincial Education Department 23JY020 (Z.L.), the Key projects of Shaanxi University of Technology SLGKYXM2302 (Z.L.) and X20240134 (Z.L.).

Author contributions

G.L. and K.G. designed and supervised the project. K.G., B.J., Z.S., and J.W. prepared the samples. C.H. and L.Z. performed the genome assembly and annotations. C.H., J.Y., X.X., and H.J. carried out the comparative genomic analysis. C.H. and J.F. performed the evolutionary analysis of chromosomes. C.H., P.Y., and L.X. performed the TAD analysis. C.H. performed the transcriptome analysis. J.Y. and Y.R. supported the methodology and the software. K.G., G.L., and Y.R. provided financial support. C.H., G.L., and Y.R. wrote the primary manuscript. G.L., Y.R., K.G., and C.H. edited and revised the manuscript. All of the authors discussed the results and commented on the manuscript. The author(s) read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-07851-0.

Correspondence and requests for materials should be addressed to Yandong Ren, Kun Guo or Gang Li.

Peer review information *Communications Biology* thanks Yazhou Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Aylin Bircan, David Favero. A peer review file is available.