Chromosome-scale genomes of *Dugesia* reveal contrasting modes of genome size evolution and structural stability in flatworms

Daniel Dols-Serrate*[1,2], Ignacio Tenaguillo-Arriola[1,2], Vadim A. Pisarenco[1,2], Marta Olivé-Muñiz[1,2], Julio Rozas*[1,2] & Marta Riutort*[1,2]

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

ORCIDs:

0000-0003-0820-7324 Daniel Dols-Serrate

0000-0001-8011-5286 Ignacio Tenaguillo-Arriola

0000-0002-8085-7674 Marta Olivé-Muñiz

0000-0002-4968-4090 Vadim A. Pisarenco

0000-0002-6839-9148 Julio Rozas

0000-0002-2134-7674 Marta Riutort

* Corresponding authors: M. Riutort mriutort@ub.edu, Daniel Dols-Serrate bcn27101993@gmail.com and Julio Rozas jrozas@ub.edu

Abstract

High-quality, chromosome-scale genomes are crucial for understanding biological processes, yet many metazoan lineages, including most Lophotrochozoa, remain underrepresented in genome databases. Among these, planarians (Platyhelminthes, Tricladida), particularly *Dugesia*, are a globally distributed and phenotypically diverse group that has become an important model in evolutionary biology, notably for investigating the genetic effects of agametic asexuality. However, the lack of chromosome-scale assemblies has limited progress. Here, we present the first chromosome-scale genomes of four Western Mediterranean *Dugesia* species, displaying the first intra- and intergeneric comparisons. Comparison with the regeneration model organism *Schmidtea mediterranea*, rejects a whole-genome duplication as the cause of differences in chromosomal number and genome size between genera. Instead, Dugesia shows extensive lineage-specific and differential expansions of DNA transposable elements, likely contributing to genome size variation during diversification. Despite differences in the dynamics of structural genome rearrangements observed between genera, both groups lack the conservation of ancestral metazoan linkage groups, supporting the idea that genome structural instability is a key feature of flatworm genome evolution. Our newly generated genomic resources and findings offer vital insights into the genetic basis of diversification and establish *Dugesia* as a valuable model for studying metazoan genome dynamics, including the evolution of alternative reproductive systems.

Keywords: *Dugesia* flatworms; PacBio; genome assemblies; chromosomal rearrangements; genome size evolution; *Schmidtea mediterranea*.

1. Introduction

Biological evolution proceeds through cumulative genomic changes that shape genetic diversity and weave the continuous and complex process that ultimately drives speciation. The advent of next-generation sequencing (NGS) has greatly advanced comparative genomics, deepening our understanding of how genomes evolve, regulate biological functions, and give rise to phenotype diversity[1,2,3,4,5]. Chromosome-scale genome assemblies, in particular, are key for elucidating the mechanisms and evolutionary history of structural variation, chromosomal rearrangements (CRs), as well as their consequences for reproductive isolation and speciation[6,7,8,9,10,11]. Although such high-quality genome resources are becoming increasingly available for non-model organisms, their distribution across the Tree of Life remains uneven, with significant taxonomic biases[12,13,14]. This disparity is particularly striking in Lophotrochozoa[15], a major animal superphylum that is severely underrepresented in genomic databases. Despite accounting for approximately 40% of all metazoan phyla[16,17]—including major groups such as Mollusca, the second-largest metazoan phylum—the number of available Lophotrochozoa genomes stands in stark contrast to their immense biodiversity, with less than 0.01% of known species sequenced to date[18,19]. This gap poses a challenge to the field of evolutionary biology and limits our ability to identify innovations unique to this vast portion of animal life, which in turn hinders our comprehensive understanding of metazoan biology. Recent genomic surveys have already underscored the importance of filling this gap, revealing unexpected insights into lophotrochozoan genome architecture and plasticity[20,21]. Planarians (Platyhelminthes, Tricladida), the focus of this study, exemplify this taxonomic imbalance and remain among the most historically overlooked groups in terms of genomic resources.

Planarians are a species-rich taxon encompassing thousands of species distributed worldwide and exhibiting a rich phenotypic diversity across terrestrial, marine, and freshwater environments[22]. Flatworms have been extensively studied in ecology[23], au-

tophagy dynamics[24], and biogeography[25,26], to name a few disciplines. However, planarians are best known for their unmatched regenerative abilities[27,28,29], enabled by their pluripotent stem cells (neoblasts), which have established them as key model systems in regeneration and stem-cell biology. It would be reasonable to expect a wealth of genomic resources for planarians, particularly given that the model species *Schmidtea mediterranea* Benazzi, Baguñà, Ballester, Puccinelli & Del Papa, 1975, was among the early cohort of genomes sequenced using the Sanger-sequence technology[30,31]. Despite this, only five triclad genomes are currently available: those of the four *Schmidtea* species[27], and that of *Dugesia japonica* Ichikawa & Kawakatsu, 1964[32]. Notably, *Schmidtea* genomes have revealed unexpected and peculiar features, including extensive genome scrambling and a complete absence of conserved ancestral metazoan linkage groups (MALGs) across species, prompting the proposal that genome-wide structural instability may be a hallmark of flatworm genome evolution[27].

The genus *Dugesia* Girard, 1850, closely related to *Schmidtea* (Figure 1), comprises approximately 110 freshwater species, roughly half of the family Dugesiidae Ball, 1974, and exhibiting a broad global distribution. In recent years, *Dugesia* has emerged as a model system in evolutionary biology, owing to the interplay between its exceptional species richness in the geologically dynamic Mediterranean region, and the presence of distinct biotypes either different reproductive modes (sexual or asexual by fission) and/or ploidy levels within species (i.e., sexual diploids, fissiparous triploids, and multiple polyploidies or aneuploidies). Extensive research in the Western Mediterranean has leveraged these traits, leading to the discovery of numerous new species, the identification of a unique putative hybridization-driven speciation event, frequent chromosomal translocations in asexual populations, evidence for the mosaic-Meselson effect in fissiparous individuals, and the identification of long-lived asexual lineages[33,34,35,36,37]. These efforts have also established a well-resolved phylogeny linking diversification patterns to abiotic events, progressively advancing from the use of a few molecular markers to reduced-representation genomic datasets, and transcriptomes. Despite its importance

and this extensive body of work, the only genomic reference available for *Dugesia* is a non-chromosome-level draft-assembly of *D. japonica*, comprising 10,017 scaffolds[32]. Generating high-quality genomes for planarians is exceptionally challenging due to three key factors: i) the limited number of already sequenced genomes with the resulting poor understanding of their genomic architecture; ii) the large genome sizes coupled with a high abundance of TEs, which complicate assembly; and iii) the pronounced A/T nucleotide bias, which further hinders sequencing and assembly. Consequently, it is not surprising that, despite their significance in evolutionary biology, planarians remain vastly unrepresented in genomic databases. This critical gap hinders deeper investigation into the molecular mechanisms underlying diversification, as well as the evolutionary significance and long-term fate of sexual and asexual lineages, and more generally limits the potential of planarians as model organisms in Lophotrochozoa evolution.

To address this gap, we generated chromosome-scale genomes of four sexual diploid *Dugesia* species (Figure 1): two endemic insular species from the Corso-Sardinian clade, viz. *Dugesia hepta* Pala, Casu & Vacca, 1981, and *Dugesia xeropotamica* Stocchino, Dols-Serrate & Riutort, 2025, and two continental ones from the Iberia-Apennines-Alps clade, viz. *Dugesia etrusca* Benazzi, 1946, and *Dugesia liguriensis* de Vries, 1988. Additionally, we also report the genome size (GS) estimates for two other Sardinian endemics for which genome assembling was not feasible, viz. *Dugesia benazzii* Lepori, 1951, and *Dugesia leporii* Pala, Stocchino, Corso, & Casu, 2000.

These new genomic resources enable us to: (i) investigate the genomic evolution within the genus, including the role of chromosomal rearrangements (CRs) in speciation (Figure 1); (ii) to explore whether a whole-genome duplication (WGD) could underlie the differences in chromosome number and GS between *Schmidtea* and *Dugesia*, considering both the higher GS and the doubled haploid complement in Western Palearctic *Dugesia* species (Figure 1); (iii) assess whether genomic patterns recently described in *Schmidtea,* including the extensive genome-wide rearrangements and the complete lack of conservation of MALGs[27], are shared by *Dugesia* species. By generating these geno-

mic resources and addressing the three aims outlined here, our study substantially advances population and evolutionary genomics in these two model groups and provides new perspectives on metazoan genome evolution.

Results

2.1. New *Dugesia* genomes: assembly and annotations

We generated chromosome-level genome assemblies for four selected *Dugesia* species. Since wild populations of the selected continental species are known to present coexisting biotypes, we characterised them via karyological analysis and flow cytometry to ensure the selection of only diploid individuals. Subsequently, we generated the draft genomes for our species, referred below as Detr_v1 (*D. etrusca*), Dlig_v1 (*D. liguriensis*), Dhep_v1 (*D. hepta*), and Dxer_v1 (*D. xeropotamica*), using Pacific Biosciences' CLR (Dhep_v1 & Detr_v1) and HiFi (Dlig_v1 & Dxer_v1) reads and Hi-C for scaffolding. The haploid genome assembly sizes yielded between 2.09 and 2.96 Gb (Table 1). Overall, N50 and L50 values, alongside Hi-C contact maps, indicate high contiguity as well (Table 1; Supplementary Fig. S1; Supplementary Fig. S2). Contact maps showed that the number of largest scaffolds per assembly (two or more orders of magnitude larger than any other given scaffold) was equal to the known haploid chromosomal number of each species[38,39,40]. These scaffolds, hereafter referred to as pseudochromosomes, represent between 52.5% and 82.7% of their respective total assembly GSs (Table 1; Supplementary Fig. S2), which in three species yield bigger estimates than those from flow cytometry. Genome completeness was assessed using BUSCO[41], yielding results comparable to those reported for other freshwater planarians (Table 1; Supplementary Fig. S3; cf. Ivanković et al.[27]). This strengthens the idea that BUSCO may be sub-optimal for planarian genomes owing to their substantial loss of genes[42], highlighting the need for a group-specific BUSCO set (e.g., [27,32,36]), as reflected in the high number of missing genes (Supplementary Table S1). However, BUSCO scores improved after using the structurally annotated as-

semblies (Table 2; Supplementary Fig. S3).

Structural annotation using BRAKER2 revealed 35,345-53,829 protein-coding genes in *Dugesia* (Table 2), far exceeding the number reported in *D. japonica* (12,031 genes; cf. Tian et al.[32]). Gene counts for *D. hepta* and *D. xeropotamica* were instead similar to those reported in *S. mediterranea* (~31,000 genes in Grohme et al.[42]; 34,976 genes modelled using Guo et al.[43] assembly). Sequence similarity-based searches indicated that most protein-coding genes (73.5-79.4%; Table 2) had identified matches in the surveyed protein databases.

2.2. Genome size pattern differences between *Schmidtea* and *Dugesia*

2.2.1. Whole-genome duplication

We hypothesized that a WGD event might be the underlying cause for the notable karyotypic differences between both genera and also the GS disparity between *S. mediterranea* and *Dugesia*. We explored this scenario using three approaches: (1) modelling the K_s distribution of each *Dugesia* assembly's paranome using wgd v2[44], (2) assessing the intragenomic collinearity in the assemblies using MCScanX[45], and (3) analysing macrosynteny patterns. All lines of evidence reject a WGD: K_s distributions show an exponential decay consistent with mostly recent duplications (Figure 2a; Supplementary Fig. S4), collinearity is very low (1.40% in Detr_v1, 1.04% in Dhep_v1, 2.19% in Dlig_v1, and 0.40% in Dxer_v1), and macrosynteny (Figure 3b) shows no *Schmidtea* syntenic blocks matching multiple *Dugesia* blocks.

2.2.2. Changes in intron and CDS lengths

To determine the genomic features underlying the GS differences between *Dugesia* and *S. mediterranea*, we analysed intron and CDS length distributions (Figure 2b-c; Table 3) of single-copy orthologues (SOs). These SOs were identified with Broccoli v.1.2[46], yielding 2,266 shared genes between *Dugesia* and *S. mediterranea*, 571 of them being exon-equivalent (exhibiting the same number of exons in all species).

Based on intron length distributions (Figure 2b; Supplementary Fig. S5), we set a threshold value of 2.5 for the logarithmic intron length (316 bp), and classified introns as small-sized (≤ 316 bp) or large-sized (> 316 bp). Small introns showed little difference among *Dugesia* species and *S. mediterranea*, although the latter had lower numbers. Large introns, in contrast, were longer although less abundant in *Dugesia* (Table 3; Figure 2b). In contrast, CDS length distributions were largely similar across species (Figure 2c; Table 3). Overall, since introns represent a small fraction of the genome, our results indicate that GS variation across the genus primarily involves, as expected, changes in intergenic regions.

2.3. Macrosynteny and CRs

We inferred and visualized macrosynteny blocks using GENESPACE v.1.2.3[47]. Overall, macrosynteny is well preserved among the four *Dugesia* species (Figure 3b), although rearrangements were detected on chromosomes 5 to 8 between the Corso-Sardinian and the Iberia-Apennines-Alps clades. Within the latter clade, no major translocations or inversions were observed (Figure 3b) despite that the species shared their last common ancestor ~9.5 Mya[26]. As expected, *D. hepta* exhibits major translocations and fusions, affecting chromosomes 2 and 7, with chromosome 2 sharing major syntenic blocks with multiple chromosomes in the other species (*D. xeropotamica*'s chromosomes 6, 7, 8, and chromosomes 5, 6, 7, and 8 of *D. etrusca* and *D. liguriensis*) (Figure 3b). These findings contrast with observations in *Schmidtea*, where loss of synteny is pervasive among congeners[27]. Finally, ODP analyses confirm that, as in *Schmidtea*, *Dugesia* shows no detectable synteny conservation of MALGs as well (Figure 4).

2.4. Transposable element composition and analysis

We examined the repeat content across all pseudochromosomes (Figure 3a; Supplementary Fig. S6). Transposable elements (TEs) annotations revealed that a large portion of the repetitive content in *Dugesia* corresponds to DNA elements (50-65%), whereas in *S. mediterranea* this category accounts for less than 40% of the repeats (Supplementary Fig.

S6a). In absolute terms, the increase in GS appears largely attributable to an expansion of TEs, particularly DNA/Maverick, DNA/DDE, and, as reported in *Schmidtea*, LTR/Gypsy elements (Figure 3a; Supplementary Fig. S6). When comparing the relative composition of repeats, the proportion of DNA/DDE elements is similar between *Schmidtea* and *Dugesia*. However, *Dugesia* exhibits a clear expansion of DNA/Maverick elements, most notably in *D. etrusca*, which are the largest TEs known in eukaryotes, 15-40 kb[48,49]. Comparisons with the *S. mediterranea* genome also revealed that *Dugesia* and *Schmidtea* share a highly similar LTR composition, with > 90% of LTR TE being Gypsy elements, yet significant differences are found in the LINE and SINE orders (Supplementary Fig. S7d,e). In *Dugesia*, there is a clear expansion of LINE/RTE elements, the smallest non-LTR retrotransposons[50], as well as an expansion of the SINE/5S and SINE/7SL subfamilies, which are seemingly absent in *Schmidtea* (Supplementary Fig. S6e). Notably, ~65% of SINE in *D. hepta* are SINE/5S elements, 3~4 times more than in *D. etrusca* or *D. liguriensis*.

We explored the pronounced expansion of SINE/5S elements in *D. hepta*, hypothesizing that it might indicate a role of TEs in CRs detected in the macrosynteny analysis. To test this, we examined the TEs density along *D. hepta*'s multisyntenic chromosome 2 using the R package *karyoploteR*[51], and assessed whether any TEs family was overrepresented within 20kb windows flanking synteny breakpoints using the R package GenomicRanges v.1.46.1[52]. We found no particular TEs family to be significantly enriched or depleted near these breakpoints (Figure 3c,d; Supplementary Table S2).

We investigated the relative TEs activity period by estimating and comparing the distribution of the Kimura 2-parameter (K2P) divergence of TE subclasses and families of Dhep_v1's assembly using EarlGrey[53], against the distribution of four-fold degenerate (4d) sites divergence, which we used as a neutral model of *D. hepta*'s evolution, computed with phyloFit[54]. Our results show that the mode of most TE families and subclasses divergence was larger than that of the 4d divergence (Figure 5; Supplementary Table S3), suggesting that the relative activity of these TEs predates *D. hepta*'s speciation.

2.5. Gene family members' gain, loss, and turnover rates

WGD events involve a rapid loss of gene copies. We explored this effect by estimating the gene gain and gene loss turnover rates using BadiRate v.1.35[55]. We first inferred the orthologous relationships with Broccoli between the four *Dugesia* species and *S. mediterranea*, yielding 18,719 orthogroups (OGs hereafter and referred to as gene families throughout the text). Moreover, we also included all singleton proteins, resulting in a total of 54,011 OGs, needed for the gene family analysis. We compared different branch-rate models under the birth, death, and innovation framework (BDI) of BadiRate. Our results supported a free-rates model (Supplementary Table S4) and showed that gene family gains were concentrated on the terminal branches of the phylogeny (Figure 3e), rather than on internal ones expected under a WGD scenario. This pattern is consistent with gene family turnover rates estimates, which increased in both birth and death rates along the external branches (Figure 3e; Supplementary Table S4), with the largest turnover rates present in the Corso-Sardinian clade.

Discussion

New genomes to contribute to the Tree of Life

Planarian genomes have long posed challenges for assembly owing to their extreme compositional bias, high repeat content, and mosaicism. This has limited the availability of high-quality genomic resources for Tricladida. The high-quality chromosome-scale assemblies presented here directly address this gap, providing a substantial advance. These new *Dugesia* genomes are markedly more contiguous and complete than the reference available for the genus (cf. Tian et al.[32]), representing over a 200-fold improvement in continuity (Table 1). Although the sequence captured in pseudochromosomes is relatively high (Table 1; Supplementary Fig. S1, S2), overall completeness does not attain that achieved for *Schmidtea* (cf. Ivanković et al.[27]). However, these genomes represent a necessary advance toward improving the representation of neglected lophotrochozoan phyla in genomic databases, and constitute a first step toward developing a

group-specific BUSCO set for Platyhelminthes in the future. Our results, nevertheless, should be contextualized in light of the well-known genomic features of planarians: extreme compositional AT-bias (>70% A/T), high repetitive content, and the characteristic mosaicism in *Dugesia* (see mosaic-Meselson effect[33]). These features likely complicate the haplotig collapsing and contribute to the inflated numbers of scaffolds. Indeed, all *Dugesia* assemblies shared a similar high fraction of repeat content (Table 2) that substantially exceeds previously reported values in *D. japonica* (~65% [32]), *S. mediterranea* (~62% [42]), or *Schmidtea polychroa* (Schmidt, 1861) (~58% [56]). All these features reinforce the findings of Ivanković et al.[27], highlighting that planarian genomes remain technically challenging to assemble. In fact, in *D. etrusca* and *D. xeropotamica,* the proportion of the sequence captured in pseudochromosomes was lower than in the other two *Dugesia* species. However, when these proportions are evaluated against their flow cytometry-based GS estimates (Table 1), they increase markedly. This further underscores the technical challenges inherent to assembling planarian genomes using current methodologies.

CRs in *Dugesia* and the origin of *D. hepta*'s karyotype

The availability of these new *Dugesia* assemblies provided, for the first time, the opportunity to explore macrosynteny relationships between genera as well as to assess the origin of the peculiar haploid complement of *D. hepta* (n = 7), a notable exception to the predominant regional set of n = 8[57] (Figure 1). Our results showed several translocations involving chromosomes 5, 6, 7, and 8 between the Corso-Sardinian and the Iberia-Apennines-Alps clades, whereas the four largest chromosomes displayed a well-conserved synteny (Figure 3b). CRs had also shaped *D. hepta*'s karyotype, where we detected relevant translocations and pericentric inversions, particularly in chromosomes 2, 3, and 7.

Chromosome 2 is of special interest, as it is key to explaining the reduction in haploid chromosome number. CRs have long been recognized as a major force in genome evolution, with a relevant role in speciation[58,59]. Specifically, changes in chromosome

number can often lead to reproductive isolation due to heterokaryotypic individuals experiencing reduced fertility (or overall reduced fitness) caused by meiotic segregation problems. In this context, the case of *D. hepta* is unambiguous. Previous genomic analyses showed no evidence of past or present gene flow between *D. hepta* and its congenerics[37], co-occurring in the same localities. While it is difficult to determine whether *D. hepta*'s CRs were concomitant with or followed its speciation, they likely played a subsequent role in reinforcing its reproductive isolation. It is well established that TEs can contribute to the origin of CRs (reviewed in [60]). Evidence for the role of CRs in the speciation could come from linking TEs insertions to CRs breakpoints and the demonstration of a concomitant activity of these TEs with the speciation event. However, our results did not identify any TE family significantly enriched in chromosome 2's synteny breakpoints. Furthermore, our estimates of the relative TEs activity (Figure 5) suggested that most TEs families' mobilization likely predated *D. hepta*'s speciation. Although LINE/R2 and SINE/Alu elements show activity periods concomitant with the speciation event, this evidence is insufficient to implicate them as drivers of the CRs that gave rise to chromosome 2, given the small number of base pairs involved and the absence of enrichment signals (Figure 3d; Figure 5c).

Genome size evolution in planarian flatworms

Little is known about the evolutionary dynamics of GSs in planarian flatworms. A recently published study by Ivanković et al.[27] provided the first comparative view of four species within the genus *Schmidtea*, revealing substantial differences between a clade of small GSs (*S. mediterranea* and *S. polychroa*; 0.82~0.78 Gb, respectively) and a clade of relatively large GSs (*Schmidtea lugubris* Schmidt, 1861 and *Schmidtea nova* Benazzi, 1982; 1.5~1.2 Gb, respectively). The authors proposed that a large portion of this increase could be attributed to TE expansions and an increased length of the protein-coding genes caused by TE insertions within introns.

Incorporating our *Dugesia* results into the comparative framework provides new insights into planarian GS evolution. Many Western Palearctic *Dugesia* species share a

haploid complement of n = 8, and GS estimates for Western Mediterranean species indicate that ~2 Gb is a common size (e.g., *D. benazzii*, 2.05 Gb; *D. leporii*, 2.38 Gb). Owing to the lack of GS estimates for other dugesiid genera, including *Girardia* Ball, 1974, or *Recurva*, it remains difficult to ascertain whether the observed differences reflect a genome expansion in *Dugesia* or a genome reduction in *Schmidtea*. Nevertheless, when comparing the available Dugesiidae GS estimates in the Animal Genome Size Database[61] (http://www.genomesize.com, last accessed in October 2025; cf. *Dugesia subtentaculata* (Draparnaud, 1801), 1.66 Gb; *Girardia tigrina* (Girard, 1850), 1.40~1.88 Gb) with our estimates (Table 1) and those reported previously[27], the data are more consistent with a lineage-specific genome expansion in Western Mediterranean *Dugesia*, potentially coupled with a genome reduction in the *mediterranea-polychroa* clade.

Bearing in mind the aforementioned differences between *Dugesia* and *Schmidtea,* we contemplated the hypothesis of a WGD event as the underlying cause not only for the karyotypic differences but also for the increase of GS in *Dugesia*, in contrast to the proposed TE expansion-driven way of increase within *Schmidtea*. Yet, our results consistently refute the WGD hypothesis. The distribution of K_s values showed exponential decay (Figure 2a), as expected under the assumption of constant gene gain and loss events, without any WGD hallmark[62]. Likewise, collinearity, macrosynteny, and gene family members' gain/loss analyses also support this conclusion. WGDs are expected to leave large blocks of intragenomic collinearity[63], and in *Dugesia*, collinearity is negligible in our genomes. For instance, autopolyploids such as *Arabidopsis thaliana* can retain up to ~27% collinear genes[45]. Furthermore, we observed neither an increase in turnover rates at the base of the genus *Dugesia* (Supplementary Table S4) nor multimatching syntenic blocks in *Schmidtea* (Figure 3b). Together, these observations indicate that the larger GS in *Dugesia* arose from mechanisms other than WGD.

The analysis of TEs revealed that the expansion of multiple TE families in *Dugesia*, as previously reported in the *lugubris-nova* lineage[27], likely underlies the observed increase in GS. However, the specific families involved differ, with DNA/Maverick and DNA/

DDE families showing the most pronounced expansions. In contrast to *Schmidtea*, we did not detect an increase in intronic lengths attributable to TEs insertions. Together, these observations indicate that TE expansions have contributed to genome increases in *Dugesia*, but that the specific TE families involved, and potentially the evolutionary forces driving their proliferation, are not shared across species.

The evolutionary context of the Corso-Sardinian clade is conditioned by the insularity of its populations, and the largest flow cytometry GS estimates in *Dugesia* come from this group. Two of the main mechanisms that may contribute to GS increase are (i) changes in ploidy (ii) and TEs mobilization[60,64]. According to the mutational hazard hypothesis (see Lynch et al.[65]), the fate of DNA gains, including TEs expansions, is governed by the mutation rate and the effective population size (N_e). Indeed, in insular systems, smaller N_e are expected, leading to relaxed selective constraints genome-wide[66,67] (i.e., reduced efficacy of selection and predominance of genetic drift). Under this scenario, TE dynamics may shift toward accumulation, with slightly deleterious TE insertions persisting because purifying selection cannot efficiently purge them. *Dugesia hepta* and *D. xeropotamica* are both endemics from Sardinia and Corsica with very restricted ranges of distribution, and coincidentally, both exhibit the largest genomes according to flow cytometry estimates (Table 1). Moreover, the largest GS estimate known for *Dugesia* belongs to a Sardinian endemic as well, *D. leporii* (2.38 Gb), further reinforcing this pattern. The evolutionary histories of these species are tightly linked to some of the major paleoenvironmental shifts in the Western Mediterranean[37], and it is reasonable to assume that both species experienced fluctuations in N_e, including likely bottlenecks occurring during the Pleistocene glacial phases on both islands[68,69]. Consistent with this view, comparisons of 4d distances of *Dugesia* did not show significant differences among *Dugesia* species or clades (Supplementary Fig. S8), indicating that the neutral evolutionary rates are broadly similar. This, in turn, strengthens the idea that differences in GS might instead reflect historical variations in N_e. Such demographic fluctuations in the insular lineages should have an impact on gene family turnover rates, and indeed,

our analyses reveal a 1.3~2.5-fold and a 2.6~3.8-fold increase in gene birth and death turnover rates, respectively, in *D. hepta* and *D. xeropotamica* relative to the continental counterparts (Supplementary Table S4).

Planarian genome architecture: resilience in front of massive genome scrambling?

A further striking finding is the discrepancy in structural genome stability observed between *Dugesia* and the patterns previously reported for *Schmidtea* (cf. Ivanković et al.[27]). Our results revealed that structural genome rearrangements are widespread between the two genera, resulting in a pattern of near-complete genome scrambling (Figure 3b). In *Dugesia*, macrosynteny analysis showed that two major fusion events underlie the haploid chromosome complement of *D. hepta* (n = 7), as well as several translocations affecting large chromosomal fragments between the two major clades, a phenomenon previously reported in other *Dugesia* species[70,71,72,73], and also frequent in fissiparous individuals[74]. Barring that, the four studied *Dugesia* species retain a largely conserved macrosynteny, despite their deep divergence time (~30 Mya[26]; Figure 1), in stark contrast to the extensive syntenic disruption observed in *Schmidtea*. Based on the patterns reported in the latter genus, it has been proposed that pervasive structural genome instability is a general feature of flatworm genome evolution[27]. However, our findings in *Dugesia* suggest that this feature may not be generalizable. Nevertheless, divergence-time estimates for *Schmidtea* species are currently lacking, leaving open the question of whether their extensive CRs represent a genus-specific feature, a broader characteristic of Dugesiidae, or instead a signature of triclad flatworm lineages that diverged a long time ago. Assuming comparable substitution rates, the lower divergence at four-fold degenerate sites in *Dugesia* (0.06 substitutions per site from their MRCA) relative to *Schmidtea* (~0.3, cf. Ivanković et al.[27]) suggests that the MRCA of the studied *Schmidtea* species predates that of the Dugesia included in this study. It would be most interesting to obtain chromosome-scale genomes of Malagasy *Dugesia*, the earliest diverging clade within the genus (>100 Mya[26]), to assess whether synteny is preserved only at 'shallow' evolutionary timescales in planarian flatworms. Nevertheless, our observation of a com-

plete absence of synteny with the MALGs in *Dugesia* (Figure 4) does not contradict the hypothesis of the pervasive structural instability of the flatworm genome evolution proposed by Ivanković et al.[27].

Taken together, our findings, along with those of Ivanković et al.[27], raise questions about the causes and consequences of such extreme genome scrambling in triclad flatworms. While pervasive genome scrambling could potentially disrupt gene regulation and be detrimental, the observation that *S. mediterranea* exhibits tight, proximity-based associations between regulatory elements and their target genes points to a mechanism that may mitigate the negative effects of pervasive CRs. This mechanism may also explain why CRs are so frequent in the wild in fissiparous populations[74]. A similar link between compact regulatory structure and extensive genomic rearrangements has been proposed in tunicates and clitellates[21,75]. But how to explain this pronounced tendency to undergo chromosomal rearrangements? This genomic architecture could actually be a consequence of a constraint linked to one of the most relevant characteristics of planarians: their remarkable regenerative capacity and cell turnover. Many of their striking evolutionary features, such as the mosaic-Meselson effect[33] and the persistence of ancient asexual lineages[35,36], are intrinsically tied to the accumulation of mutational load and the hurdles of Muller's ratchet[76,77,78], and somehow planarians are able to circumvent these limitations[36,74]. High GC content has been shown to prevent double-strand breaks in prokaryotes[79], yet it is also positively correlated with higher mutation rates[80]. Planarians, which exhibit exceptionally low GC content[27,56,81] (Table 1), may have evolved this genomic feature as a compensatory strategy to reduce the burden of non-synonymous mutations accrued through their continuous regeneration and cell turnover. This could result in a genome composition that is resilient to mutational load, yet CRs-prone. Remarkably, *Hydra* (phylum Cnidaria), which exhibits regenerative abilities comparable to those of freshwater planarians, also shows an extreme compositional AT-bias[82]. Under this hypothesis, selective pressures may have favoured the emergence of a gene-centric regulatory architecture capable of mitigating the negative effects of such pervasive

structural instability.

Materials and methods

Samples

All animals used for the analyses were collected from wild populations and reared in laboratory in 1× Montjuïc water (1.6 mM NaCl, 1.0 mM $CaCl_2$, 1.0 mM $MgSO_4$, 0.1 mM $MgCl_2$, 0.1 mM KCl, and 1.2 mM NaHCO3 in DI H_2O, pH 6.9–8.1) at a constant temperature of 14 ºC. *Dugesia benazzii* samples were collected at 40.57498, 9.6746543, near Siniscola, Sardinia (Italy). *Dugesia etrusca* samples were collected at 43.497778, 10.625833, near the town of Rivalto (Italy). *Dugesia hepta* samples were collected at 40.756564, 8.604074, near the city of Sassari, Sardinia (Italy). *Dugesia leporii* samples were collected at 39.436662, 8.5082993, near the town of Fluminimaggiore, Sardinia (Italy). *Dugesia liguriensis* samples were collected at 43.823611, 6.581944, near the village of La Garde (France). *Dugesia xeropotamica* samples were collected at 42.473832, 8.795403, at the Lioli rivulet. Animals were starved for a week prior to DNA/RNA extractions. Animals destined for DNA extractions were flash-frozen in liquid nitrogen and stored at -80 ºC before being sent for sequencing. Final species selected for genome sequencing were chosen based on overall extraction yield and quality.

Flow cytometry

GS estimation was carried out by means of flow cytometry. Samples were prepared by incubating living individuals in a solution of 2% N-acetyl-L-cysteine at pH 7 to remove the mucus and minimize the formation of cell aggregates. Animals were then rinsed for 1 minute using a mixture of distilled water and tap water (1:1) and placed in separated Petri dishes. Thereafter, samples were minced using a sterilized stainless steel razor and incubated in 1 ml Galbraith's isolation buffer[83] supplemented with 10 µl of RNase A (1 mg/µl) for 15 minutes. After incubation, the cells were retrieved and separated by gently pipetting the liquid contents through a nylon mesh (pore size, 75 µm) into an Eppendorf

tube. Further, the cell suspension was stained for 15 minutes with 30 µl of propidium iodide (stock, 1 mg/ml). The quantity of DNA was measured with a Gallios Flow Cytometer at the Unitat de Citometria dels Centres Científics i Tecnològics de la UB (CCiT, UB). Finally, GS was determined by comparing the readings of fluorescence of our samples against those of males of *Blatella germanica*, with a known GS of 2.025 Gb.

DNA/RNA extractions, sequencing, and genome assembly

RNA extractions were performed using Trizol (Thermo Fisher Scientific, USA) following the manufacturer's instructions. Total RNA quantification and integrity were assessed with a Qubit and a Bioanalyzer in the Centres Científics i Tecnològics, Universitat de Barcelona (CCiT, UB). Truseq stranded and ribo-zero libraries were constructed in Macrogen Inc., (Macrogen Europe, Madrid) to obtain Illumina paired-end sequencing.

DNA extraction, sequencing, and genome assembly were performed at Dovetail Genomics (Scotts Valley, CA). High-quality genomic DNA of each species was extracted from pooled samples following the protocol described in Zhang and Alvarado[84]. Genomic sequence data were generated using a single library per species: PacBio CLR (*D. hepta, D. etrusca*) and PacBio HiFi (*D. liguriensis, D. xeropotamica*) libraries. Omni-C libraries (one library per each species) were then built and sequenced using Illumina 150-bp paired-end (HiSeq X platform) to perform the chromosome-level scaffolding using the HiRise software[85]. The assemblies' completeness was assessed by applying the pipeline of Benchmarking Universal Single Copy Orthologues (BUSCO; v.5.4.2[41]), searching the metazoa (odb10; 954 genes), arthropoda (odb10; 1013 genes), and eukaryota (odb10; 255 genes) data sets.

Contaminants sequence removal

Mitochondrial genome sequences were removed from the assembled scaffolds using BLAST+ blastn v.2.12.0[86,87] searches against the mitogenomes of *D. japonica* (GenBank accession number AB618487) and *Dugesia. ryukyuensis* Kawakatsu, 1976 (AB618487).

Any hit of >1kb was removed and replaced by Ns in the corresponding scaffold. We then filtered out potential contaminant sequences and assembly artifacts using several criteria. First, we mapped each assembly against the NCBI nucleotide database using BLAST+ blastn and Blobtools v.3.4.0, as implemented in BlobToolKit[88], and discarded scaffolds with bacterial, virus, or fungal hits. Next, we calculated the mean depth of coverage for the remaining scaffolds and discarded those falling outside the 95% confidence interval that also lacked BUSCO hits.

Identification of repetitive elements (RE)

We analysed repetitive regions of the genome assemblies using a combination of *de novo* TEs identification with RepeatModeler v2.0.3[89] and DeepTE[90], and similarity-based searches using RepearMasker v.4.1.2[91]. We built a RE database for each species. We modelled RE using RepeatModeler and parsed its output to classify RE into identified and unknown elements. Thereafter, we used DeepTE to classify the unknown RE using the metazoan model (-sp M). Lastly, all classified RE were concatenated into a single database to be used to soft-mask each genome assembly with the `-xsmall` option of RepeatMasker. Additionally, we performed *de novo* repeat identification in *D. hepta* using EarlGrey[53] to estimate the repeat landscapes of the major TE families to assess their relative activity in *Dugesia* genomes.

Structural and functional genome annotation

Structural annotation of the soft-masked genome assemblies was performed with BRAKER2[92,93,94,95], using species-specific RNAseq data that had been previously mapped to each assembly using STAR v.2.7.10b[96]. Functional annotation of the resulting gene models was carried out using BLASTP v.2.2 (E-value = 10^{-3}) searches against Swiss-Prot, complemented with searches against the eggNOG database, using eggNOG-mapper v.2.1.9[97,98,99], and protein domain detection with InterProScan v.5.56.89[100]. All functional evidence was integrated. Concurrently, transfer RNA genes (tRNAs) were identified using the tRNAscan-SE v2.0.7[101].

Intron and CDS lengths

We estimated intron and CDS length distribution of *Dugesia* and *S. mediterranea* using only single copy orthologs (SOs; 1:1:1:1:1) from the inferred orthologous groups (OG) determined with Broccoli v.1.2[46]. For the analysis, we used information from the largest scaffolds or pseudochromosomes (*D. etrusca*, *D. liguriensis*, and *D. xeropotamica*, 8 scaffolds; *D. hepta*, 7 scaffolds; *S. mediterranea*, 4 scaffolds, Guo et al. (2022) assembly). Intron and CDS length distributions were visualised using *ggplot2*.

4-fold degenerate sites analysis

We estimated the putative neutral divergence using four-fold degenerate (4d) sites. For that, first, we extracted these sites from the coding sequences (CDS) of single-copy *Dugesia* orthologues identified by Broccoli, in combination with BRAKER's structural annotations. Each orthogroup was aligned with TranslatorX[102], using its implementation in MAFFT[103] to maintain the correct reading frame and ensure positional homology across nucleotides. Finally, we used phyloFit[54] with the default REV substitution model to estimate the neutral model from the 4d sites across all alignments, and visualized the resulting distribution of genetic distances using *ggplot2*.

Detection of whole-genome duplication

We constructed the *K*s distribution to evaluate whether a WGD event could underlie the conspicuous GS differences between *Dugesia* and *S. mediterranea*. First, using wgd v2[44], we inferred the paralogous gene families for each *Dugesia* species from their coding sequences (`wgd dmd`). Thereafter, we estimated (`wgd ksd`) and visualized (`wgd viz`) the *K*s distributions. Furthermore, we identified all potential homologous protein pairs within each species using BLASTP v.2.2 (E-value = 10^{-10} and top 5 matches), and used MCScanX[45] to detect collinear gene pairs.

Macrosynteny analysis

To explore potential CRs in *D. hepta*, we used the R package GENESPACE v.1.2.3[47], which combines OrthoFinder[104] and a reimplementation of the MCScanX algorithm to infer gene-ordered-based syntenic blocks, and visualized them using the built-in riparian plot function. Additionally, we applied the ODP tool[105] to determine whether ancestral metazoan linkage groups (MALGs), highly conserved linkage groups across bilaterians, cnidarians, and sponges[11], are preserved in *Dugesia* genomes. This tool uses hidden Markov models to identify homologs of MALG proteins and applies Fisher's exact test to determine their significant enrichment on specific chromosomes.

TEs enrichment analysis at syntenic breakpoints

We visualized the density distribution of TE along chromosome 2 of *D. hepta* using the R package *karyoploteR*[51] to identify regions potentially involved in CRs detected in the macrosynteny analysis. In addition, we tested whether specific TE families were over-represented nearby synteny breakpoints on the same chromosome using the R package *GenomicRanges* v.1.46.1[52]. Based on GENESPACE syntenic blocks delineation and RepeatMasker's annotation, we established flanking windows of 20kb surrounding each breakpoint and counted all TEs present within these regions. To assess enrichment, we generated a null distribution of repetitive sequences by placing an equal number of randomly distributed windows along the chromosome and then computed 1000 permutation iterations. A two-tailed test was performed to assess whether TE families were significantly enriched or depleted in the flanking windows relative to the null distribution. To account for multiple testing and low-abundance families, we applied two filtering steps: (i) p-values were calculated only for TE families with an average expected count of 10 or more sequences, and (ii) the resulting raw p-values were adjusted using the Benjamini-Hochberg false discovery rate method.

Estimation of gene gain and loss

We estimated gene gains and losses using BadiRate v.1.35[55], under the birth, death, and innovation model (BDI), applying a maximum likelihood (ML) method to estimate

the turnover rates. Furthermore, we also tested multiple branch models to estimate the best-fitting one by comparing the Akaike Information Criterion (AIC) values. The models considered were: (i) global rates (GR), in which all branches evolve under the same turnover rates; (ii) free rates (FR), which allows each branch to evolve under distinct rates; and (iii) branch-specific (BR), where a subset of branches share different turnover rates from the background (here, those leading to *Dugesia*). We used the OG counts obtained from Broccoli and a time-calibrated tree based on the SOs set. Given that Broccoli may leave some proteins unassigned to OGs when handling complex evolutionary scenarios, we also included the singleton OGs for this analysis (eg., 1:0:0:0:0). Since BadiRate is prone to crash in the analysis of highly heterogenous counts, we analyzed these cases (only 0.02% of the OGs) separately using the best-fitting branching model using Wagner parsimony. For the analysis, we inferred a time-calibrated tree with IQ-TREE v2.4.0[106] using an admixture model (LG+C60+F+G+R3) and subsequently calibrated with the *chronos* function from the R package *ape*[107]. We used four calibration points: (i) the *Schmidtea-Dugesia* split (130-160 Mya), (ii) the most recent common ancestor (MRCA) of the featured *Dugesia* (23-36.5 Mya), (iii) the MRCA of the *hepta-xeropotamica* (4.25-8.87 Mya) and (iv) the MRCA of *etrusca-liguriensis* (6.3-13.3 Mya)[26,37].

References

1. Giani, A. M. et al. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.* **18**, 9-19 (2020). https://doi.org/10.1016/j.csbj.2019.11.002

2. Kim, J. et al. Reference-assisted chromosome assembly. *Proc. Natl. Acad. Sci*. **110**(5), 1785-1790 (2013). https://doi.org/10.1073/pnas.1220349110

3. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19,** 789-800 (2018). https://doi.org/ 10.1038/s41576-018-0060-8

4. Satam, H. et al. Next-generation sequencing technology: Current trends and advance-

ments. *Biology*, **12**(7), 997 (2023). https://doi.org/10.3390/biology12070997

5. Slatko, B. E. et al. Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.* **122**, e59 (2018). https://doi.org/10.1002/cpmb.59

6. Kawakami, T. et al. Chromosomal speciation revisited: Modes of diversification in Australian morabine grasshoppers (*Vandiemenella, viatica* species group). *Insects*, **2**, 49-61 (2011). https://doi.org/10.3390/insects2010049

7. King, M. Chromosomal rearrangements, speciation and the theoretical approach. *Heredity* **59**, 1-6 (1987). https://doi.org/10.1038/hdy.1987.90

8. Lucek, K. et al. The impact of chromosomal rearrangements in speciation: From micro- to macroevolution. *Cold Spring Harb. Perspect. Biol.* **15**:a041447 (2023). https://doi.org/10.1101/cshperspect.a041447

9. Mackintosh, A. et al. Chromosome fissions and fusions act as barriers to gene flow between *Brenthis* fritillary butterflies. *Mol. Biol. Evol.* **40**(3):msad043 (2023). https://doi.org/10.1093/molbev/msad043

10. Mérot, C. Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Molecular Ecology* **29**, 2512-2516 (2020). https://doi.org/10.1111/mec.15500

11. Simakov, O. et al. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022). https://doi.org/10.1126/sciadv.abi5884

12. Escuer, P. et al. The chromosome-scale assembly of the Canary Islands endemic spider *Dysdera sylvatica* (Arachnida, Araneae) sheds light on the origin and genome structure of chemoreceptor gene families in chelicerates. *Mol. Ecol. Resour.* **22**, 375-390 (2022). https://doi.org/10.1111/1755-0998.13471

13. Hotaling, S. et al. Toward a genome sequence for every animal: Where are we now? *Proc. Natl. Acad. Sci*. **118**(52), e2109019118 (2021). https://doi.org/10.1073/pnas.2109019118

14. MacLeod, K. L. et al. The first genome of the cold-water octocoral, the pink sea fan, *Eunicella verrucosa*. *Genome Biol. Evol.* **15**(6) (2023). https://doi.org/10.1093/gbe/evad083

15. Halanych, K. M. et al. Evidence from 18S ribosomal DNA that the Lophophorates are protostome animals. *Science* **267**(5204), 1641-1643. https://doi.org/10.1126/science.7886451

16. Kocot, K. M. On 20 years of Lophotrochozoa. *Org. Divers. Evol.* **16**, 329-343 (2016). https://doi.org/10.1007/s13127-015-0261-3

17. Stick, R. & Peter, A. CaaX-less lamins: Lophotrochozoa provide a glance at the playground of evolution. *Protoplasma* **260**, 741-756 (2023). https://doi.org/10.1007/s00709-022-01809-3

18. Lian, S. et al. Marine evo-devo: New frontiers from Lophotrochozoa. *The Innovation* **6**(6), 100869 (2025). https://doi.org/10.1016/j.xinn.2025.100869

19. Nath, R. et al. Lineage-specific Class-A GPCR dynamics reflect diverse chemosensory adaptations in Lophotrochozoa. *Mol. Ecol. Biol.* **42**(3), msaf042 (2025). https://doi.org/10.1093/molbev/msaf042

20. Ivanković, M. et al. A comparative analysis of planarian genomes reveals regulatory conservation in the face of rapid structural divergence. *Nat. Commun.* **15**:8215 (2024). https://doi.org/10.1038/s41467-024-52380-9

21. Vargas-Chávez, C. et al. An episodic burst of massive genomic rearrangements and the origin of non-marine annelids. *Nat. Evol. Evol.* **9**, 1263-1279 (2025). https://doi.org/10.1038/s41559-025-02728-1

22. Sluys, R. & Riutort, M. Planarian diversity and evolution. In *Planarian regeneration: Methods and protocols* 1-56 (Springer New York, New York, NY, 2018).

23. Vila-Farré, M. & Rink, J. C. The ecology of freshwater planarians. In *Planarian regeneration: Methods and protocols* 173-205 (Springer New York, New York, NY,

2018).

24. González-Estévez, C. Autophagy in freshwater planarians. In *Methods in Enzymology* 440-465 (2008). https://doi.org/10.1016/S0076-6879(08)03227-8

25. Sluys, R. et al. The genus *Dugesia* in Australia, with its phylogenetic analysis and historical biogeography (Platyhelminthes, Tricladida, Dugesiidae). *Zool. Scr.* **27**(4), 273-289 (1998) https://doi.org/10.1111/j.1463-6409.1998.tb00461.x

26. Solà, E. et al. Three dispersal routes out of Africa: A puzzling biogeographical history in freshwater planarians. *J. Biogeogr.* **49**(7), 1219-1233 (2022). https://doi.org/10.1111/jbi.14371

27. Ivanković, M. et al. Model systems for regeneration: planarians. *Development* **146**, dev167684 (2019). https://doi.org/10.1242/dev.167684

28. Reddien, P. W. The cellular and molecular basis for planarian regeneration. *Cell* **175**(2), 327-345 (2018). https://doi.org/10.1016/j.cell.2018.09.021

29. Wagner, D. E. et al. Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science* **332**, 811-816 (2011). https://doi.org/10.1126/science.1203983

30. Robb, S. M. C. et al. SmedGC 2.0: The *Schmidtea mediterranea* genome database. *Genesis* **53**(8), 535-546 (2015). https://doi.org/10.1002/dvg.22872

31. Robb, S. M. C. et al. SmedGD: The *Schmidtea mediterranea* genome database. *Nucleic Acids Res.* **36**, 599-606 (2007). https://doi.org/10.1093/nar/gkm684

32. Tian, Q. et al. Whole-genome sequence of the planarian *Dugesia japonica* combining Illumina and PacBio data. *Genomics* **114**(2), 110293 (2022). https://doi.org/10.1016/j.ygeno.2022.110293

33. Leria L. et al. Outstanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. *BMC Evol. Biol.* **19**:130 (2019). https://

doi.org/10.1186/s12862-019-1440-1

34. Leria, L. et al. Microplate tectonics and environmental factors as distribution drivers in Western Mediterranean freshwater planarians. *J. Biogeogr.* **49**(6), 1124-1136 (2022). https://doi.org/10.1111/jbi.14373

35. Benítez-Álvarez, L. et al. Phylotranscriptomics interrogation uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean. *Mol. Phylogenet. Evol.* **178**, 107649 (2023). https://doi.org/10.1016/j.ympev.2022.107649

36. Benítez-Álvarez, L. et al. A new asexual scandal: Independent evolution of ancient planarian fissiparous lineages challenges species concept. *Syst. Biol.* Under review (2025).

37. Dols-Serrate, D. et al. Mixed, not stirred: Genomic data confirm the first case of interspecific hybridization in planarian triclads (Platyhelminthes: Tricladida) and raise questions about a possibly novel form of hybrid speciation. *Mol. Phylogenet. Evol.* **207**, 108339 (2025). https://doi.org/10.1016/j.ympev.2025.108339

38. Dols-Serrate, D. et al. New insights into the evolution and biogeography of freshwater planarians on islands in the Tyrrhenian Sea, Western Mediterranean Basin, with the integrative description of a new endemic species from Corsica (Platyhelminthes, Tricladida, *Dugesia*). *Zool. J. Linn. Soc.* **201**, zlae0810 (2024). https://doi.org/10.1093/zoolinnean/zlae080

39. Lázaro, E. M. et al. Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Mol. Phylogenet. Evol.* **52**(3), 835-845 (2009). https://doi.org/10.1016/j.ympev.2009.04.022

40. Pala, M. et al. Stabilized natural interspecific hybrid population of the fresh water planarians *Dugesia gonocephala* s.l. (Turbellaria, Tricladida). *Caryologia* **35**(2), 247-

256 (1982). https://doi.org/10.1080/00087114.1982.10796929

41. Seppey, M. et al. BUSCO: Assessing genome assembly and annotation completeness. In *Gene prediction: Methods and protocols* (2019). https://doi.org/10.1007/978-1-4939-9173-0_14

42. Grohme, M. A. et al. The genome of *Schmidtea mediterránea* and the evolution of core cellular mechanisms. *Nature* **554**(7690), 56-61 (2018). https://doi.org/10.1038/nature25473

43. Guo, L. et al. Island-specific evolution of sex-primed autosome in a sexual planarian. *Nature* **606**(7913), 329-334 (2022). https://doi.org/10.1038/s41586-022-04757-3

44. Chen, H. & Zwaenepoel, A. Inference of ancient polyploidy from genomic data. In *Polyploidy: Methods and protocols* (Springer New York, New York, NY, 3-18, 2023).

45. Wang, Y. et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**(7), e49 (2012). https://doi.org/10.1093/nar/gkr1293

46. Derelle, R. et al. Broccoli: Combining phylogenetic and network analyses for orthology assignment. *Mol. Biol. Evol.* **37**(11), 3389-3396 (2020). https://doi.org/10.1093/molbev/msaa159

47. Lovell, J. T. et al. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* **11**:e78526 (2021). https://doi.org/10.7554/ELIFE.78526

48. Haapa-Paananen, S. et al. Phylogenetic analysis of Maverick/Polinton giant transposons across organisms. *Mol. Phylogenet. Evol.* **78**(1), 271-274 (2014). https://doi.org/10.1016/j.ympev.2014.05.024

49. Bellas, C. & Sommaruga, R. Polinton-like viruses are abundant in aquatic ecosystems. *Microbiome* **9**:13 (2021). https://doi.org/10.1186/s40168-020-00956-0

50. Župunski, V. et al. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol. Ecol. Evol.* **18**(10), 1849-1863 (2001). https://doi.org/10.1093/oxfordjournals.molbev.a003727

51. Gel, B. & Serra, E. karyoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinform.* **33**(19), 3088-3090 (2017). https://doi.org/10.1093/bioinformatics/btx346

52. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**(8):e1003118 (2013). https://doi.org/10.1371/journal.pcbi.1003118

53. Baril, T. et al. Earl Grey: A fully automated user-friendly transposable element annotation and analysis pipeline. *Mol. Ecol. Evol.* **41**(4):msae068 (2024). https://doi.org/10.1093/molbev/msae068

54. Hubisz, M. J. et al. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform.* **12**, 41-51 (2010). https://doi.org/10.1093/bib/bbq072

55. Librado, P. et al. BadiRate: Estimating family turnover rates by likelihood-based methods. *Bioinform.* **28**(2), 279-281 (2012). https://doi.org/10.1093/bioinformatics/btr623

56. Póti, A. et al. Mutational profile of the regenerative process and *de novo* genome assembly of the planarian *Schmidtea polychroa*. *Nucleic Acids Res.* **52**, 1779-1792 (2024). https://doi.org/10.1093/nar/gkad1250

57. Stocchino, G. A. et al. Endemic freshwater planarians of Sardinia: Redescription of *Dugesia hepta* (Platyhelminthes, Tricladida) with a comparison of the Mediterranean species of the genus. *J. Nat. Hist.* **39**(22), 1947-1960 (2005). https://doi.org/10.1080/00222930500060025

58. Faria, R. & Navarro, A. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Cell* **25**(11), 660-669 (2010). https://doi.org/10.1016/j.tree.2010.07.008

59. Lucek, K. et al. The role of structural genomic variants in population differentiation and ecotype formation in *Timema cristinae* walking sticks. *Mol. Ecol.* **28**(6), 1224-1237 (2019). https://doi.org/10.1111/mec.15016

60. Chénais, B. et al. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene* **509**, 7-15 (2012). https://doi.org/10.1016/j.gene.2012.07.042

61. Gregory, T. R. Animal Genome Size Database. https://www.genomesize.com/ (2024).

62. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-1155 (2010). https://doi.org/10.1126/science.290.5494.1151

63. Vanneste, K. et al. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res.* **24**(8), 1334-1347 (2014). https://doi.org/10.1101/gr.168997.113

64. Sproul, J. S. et al. Analyses of 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges. *Genome Res.* **33**(10), 1708-1717 (2023). https://doi.org/10.1101/gr.277387.122

65. Lynch, M. et al. The divergence of mutation rates and spectra across the Tree of Life. *EMBO Rep.* **24**(10):e57561 (2023). https://doi.org/10.15252/embr.202357561

66. Marino, A. et al. Effective population size does not explain long-term variation in genome size and transposable element content in animals. *eLife* **13**:RP100547 (2024). https://doi.org/10.7554/eLife.100574

67. Pisarenco, V. A. et al. How did evolution halve genome size during an oceanic island colonization? *Mol. Biol. Evol.* **42**(9), 1-14 (2025). https://doi.org/10.1093/molbev/msaf206

68. Kuhlemann, J. et al. Würmian maximum glaciation in Corsica. *Austrian J. Earth. Sci.* **97**, 68-81 (2005).

69. Kuhlemann, J. et al. Regional synthesis of Mediterranean atmospheric circulation during the Last Glacial Maximum. *Science* **321**(5894), 1338-1340 (2008). https://doi.org/10.1126/science.1157638

70. de Vries, E. J. On the karyology of *Dugesia gonocephala* s.l. (Turbellaria, Tricladida) from Montpellier, France. *Hydrobiologia* **132**(1), 251-256 (1986). https://doi.org/10.1007/BF00046257

71. Ribas, M. *Cariologia, sistemàtica i biogeografía de les planàries d'aigües dolces als Països Catalans.* Doctoral Thesis, Universitat de Barcelona.

72. Tamura, S. et al. Karyological and taxonomic studies of *Dugesia japonica* from the Southwest Islands of Japan-II. *Hydrobiologia* **227**(1), 157-162 (1991). https:// doi.org/10.1007/BF00027596

73. Tamura, S. et al. Karyology and biogeography of *Dugesia japonica* and *Dugesia ryukyuensis* in Kyushu, Southern Japan. *Hydrobiologia* **383**(1-3), 321-327 (1998). https://doi.org/10.1023/A:1003471207447

74. Leria, L. et al. Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhelminthes, Tricladida): Extreme intraindividual genetic diversity, morphological stasis, and karyological instability. *Mol. Phylogenet. Evol.* **143**:106496 (2020). https://doi.org/10.1016/j.ympev.2019.05.010

75. Plessy, C. et al. Extreme genome scrambling in cryptic *Oikopleura dioica* species. *bioRxiv* (2023). https://doi.org/10.1101/2023.05.09.539028

76. Muller, H. J. Some genetic aspects of sex. *Am. Nat.* **66**(703), 118-138 (1932).

77. Hollister, J. D. et al. Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera. Mol. Biol. Evol.* **29**(1), 401-408 (2015). https://doi.org/10.1093/molbev/msu345

78. Bast, J. et al. Consequences of asexuality in natural populations: Insights from stick insects. *Mol. Biol. Evol.* **35**(7), 168-1677 (2018). https://doi.org/10.1093/molbev/

msy058

79. Weissman, J. K. et al. Linking high GC content to the repair of double-strand breaks in prokaryotic genomes. *PLoS Genetics* **15**(11):e1008493 (2019). https:// doi.org/10.1371/journal.pgen.1008493

80. Vinogradov, A. E. & Anatskaya, O. V. DNA helix: the importance of being AT-rich. *Mamm. Genome* **29**, 455-464 (2017). https://doi.org/10.1007/s00335-017-9713-8

81. Solà, E. et al. Evolutionary analysis on mitogenomes from parasitic and free-living flatworms. *PLoS One* **10**(3) (2015). https://doi.org/10.1371/journal.pone.0120081

82. Chapman, J. A. et al. The dynamic genome of *Hydra*. *Nature* **464**(7288), 592-596 (2010). https://doi.org/10.1038/nature08830

83. Galbraith, D. W. et al. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **220**(4601), 1049-1051 (1983). https://doi.org/10.1126/science.220.4601.1049

84. Zhang, S. & Alvarado, A. S. Planarian high molecular weight DNA isolation by spooling. In *Planarian regeneration: Methods and protocols* 173-205 (Springer New York, New York, NY, 2018).

85. Putnam, N. H. et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342-350 (2016). https://doi.org/10.1101/gr.193474.115.Freely

86. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**(17), 3389-3402 (1997). https://doi.org/10.1093/nar/25.17.3389

87. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**:421 (2009). https://doi.org/10.1186/1471-2105-10-421

88. Challis, R. et al. BlobToolKit – Interactive quality assessment of genome assem-

blies. *G3: Genes, Genomes, Genetics* **10**(4), 1361-1374 (2020). https://doi.org/10.1534/g3.119.400908

89. Smit, A. & Humbley, R. *RepeatModeler Open-1.0* (2008-2015). Retrieved from http://www.repeatmasker.org

90. Yan, H. et al. DeepTE: a computational method for *de novo* classification of transposons with convolutional neural network. *Bioinform.* **36**(15), 4269-4275 (2020). https://doi.org/10.1093/bioinformatics/btaa519

91. Smit, A. & Humbley, R. *RepeatModeler Open-4.0* (2013-2015). Retrieved from http://www.repeatmasker.org

92. Brůna, T. et al. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**(1) (2021). https://doi.org/10.1093/nargab/lqaa108

93. Hoff, K. J. et al. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 62-65 (2019). https://doi.org/10.1007/978-1-4939-9173-0_5

94. Stanke, M. et al. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* **7**, 1-11 (2006). https://doi.org/10.1186/1471-2105-7-62

95. Stanke, M. et al. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioninform.* **24**(5), 637-644 (2008). https://doi.org/10.1093/bioinformatics/btn013

96. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioninform.* **29**(1), 15-21 (2013). https://doi.org/10.1093/bioinformatics/bts635

97. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309-D314 (2019). https://doi.org/10.1093/nar/gky1085

98. Buchfink, B. et al. Sensitive protein alignment at tree-of-life scale using DIAMOND. *Nat. Met.* **18**(4), 366-368 (2021). https://doi.org/10.1038/s41592-021-01101-x

99. Cantalapiedra, C. P. et al. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**(12), 5825-5829 (2021). https://doi.org/10.1093/molbev/msab293

100. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioninform.* **30**(9), 1236-1240 (2014). https://doi.org/10.1093/bioinformatics/btu031

101. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1962**(5), 1-14 (2021). https://doi.org/10.1007/978-1-4939-9173-0

102. Abascal, F. et al. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, 7-13 (2010). https://doi.org/10.1093/nar/gkq291

103. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**(4), 772-780 (2013). https://doi.org/10.1093/molbev/mst010

104. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**:238 (2019). https://doi.org/10.1186/s13059-019-1832-y

105. Schultz, D. T. et al. Ancient gene linkages support ctenophores as sister to other animals. *Nature* **618**, 110-117 (2023). https://doi.org/10.1038/s41586-023-05936-6

106. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**(5), 1530-1534 (2020). https://doi.org/10.1093/molbev/msaa015

107. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinform.* **35**, 526-528 (2019). https://doi.org/10.1093/bi-

oinformatics/bty633.

108. Leria, L. et al. Diversification and biogeographic history of the Western Palearctic freshwater flatworm genus *Schmidtea* (Tricladida: Dugesiidae), with a redescription of *Schmidtea nova*. *J. Zoolog. Syst. Evol. Res.* **56**(3), 335-351 (2018). https://doi.org/10.1111/jzs.12214

Acknowledgements

We are thankful to Longhua Guo and Xinghua Li for providing some of the necessary data for *Schmidtea mediterranea* to carry out some of the analyses of this study. We also express our sincere gratitude to Gema Blasco for her dedication and exceptional care when rearing the planarians in the laboratory. We thank Alejandro Sánchez-Gracia and Sara Guirao-Rico for their invaluable inputs throughout the course of the study's elaboration.

Author contributions

J.R., and M.R. designed the study. D.D.S., I.T.A., and M.R. carried out sampling efforts. D.D.S. and I.T.A. performed the DNA extractions and the main comparative genome analyses. V.A.P. and M.O.M. provided bioinformatics assistance and helped perform some of the analyses. D.D.S. wrote the first version of the manuscript and conceived all data visualization. J.R. and M.R. supervised the research and final draft of the manuscript.

Funding

This research was supported by the Ministerio de Ciencia, Innovación y Universidades (MICIU/AEI 10.13039/501100011033/ project PID2021-125792NB-I00).

Competing interests

The authors declare no competing interests

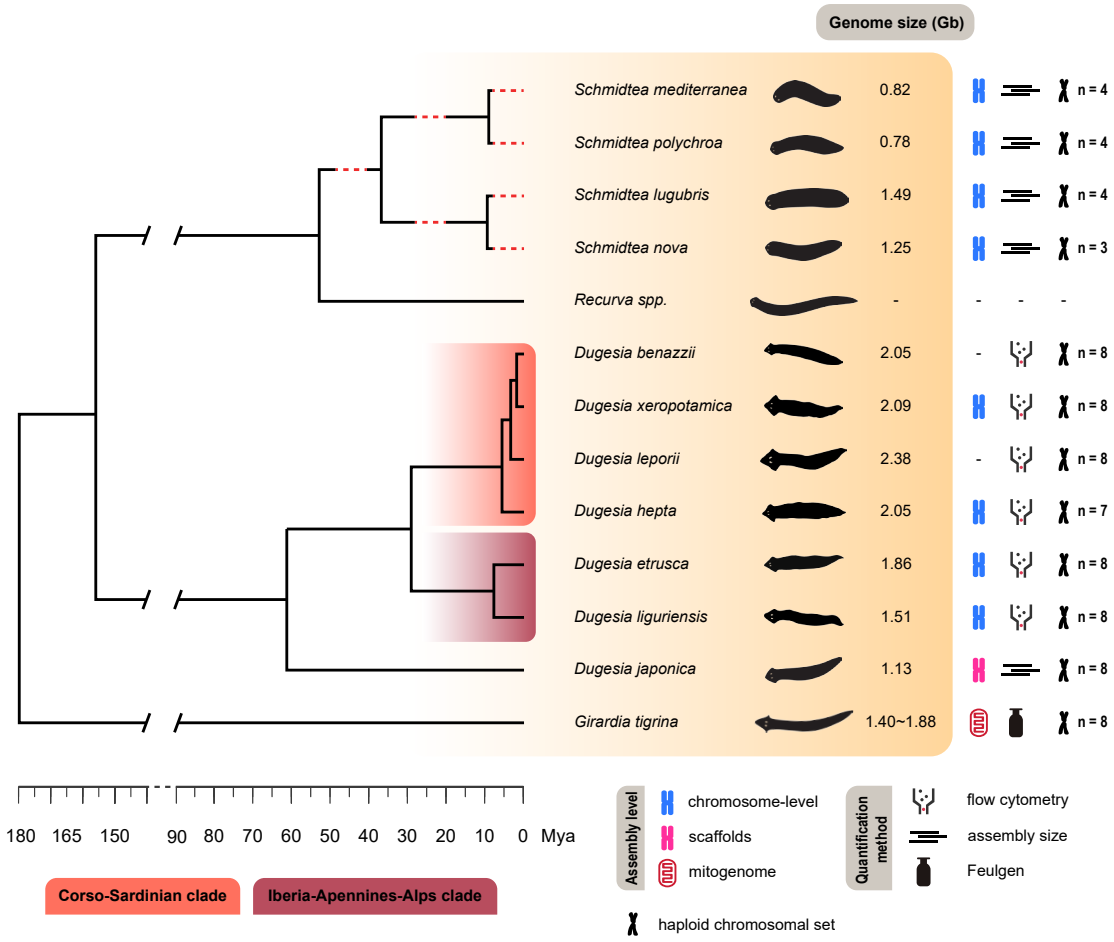


Figure 1. Schematic time-calibrated tree illustrating the phylogenetic relationships among species of *Dugesia*, *Girardia*, *Recurva*, and *Schmidtea* based on Leria et al.[108] and Solà et al.[26]. Red dotted lines indicate branches for which the time to their most recent common ancestor (MRCA) is unknown.

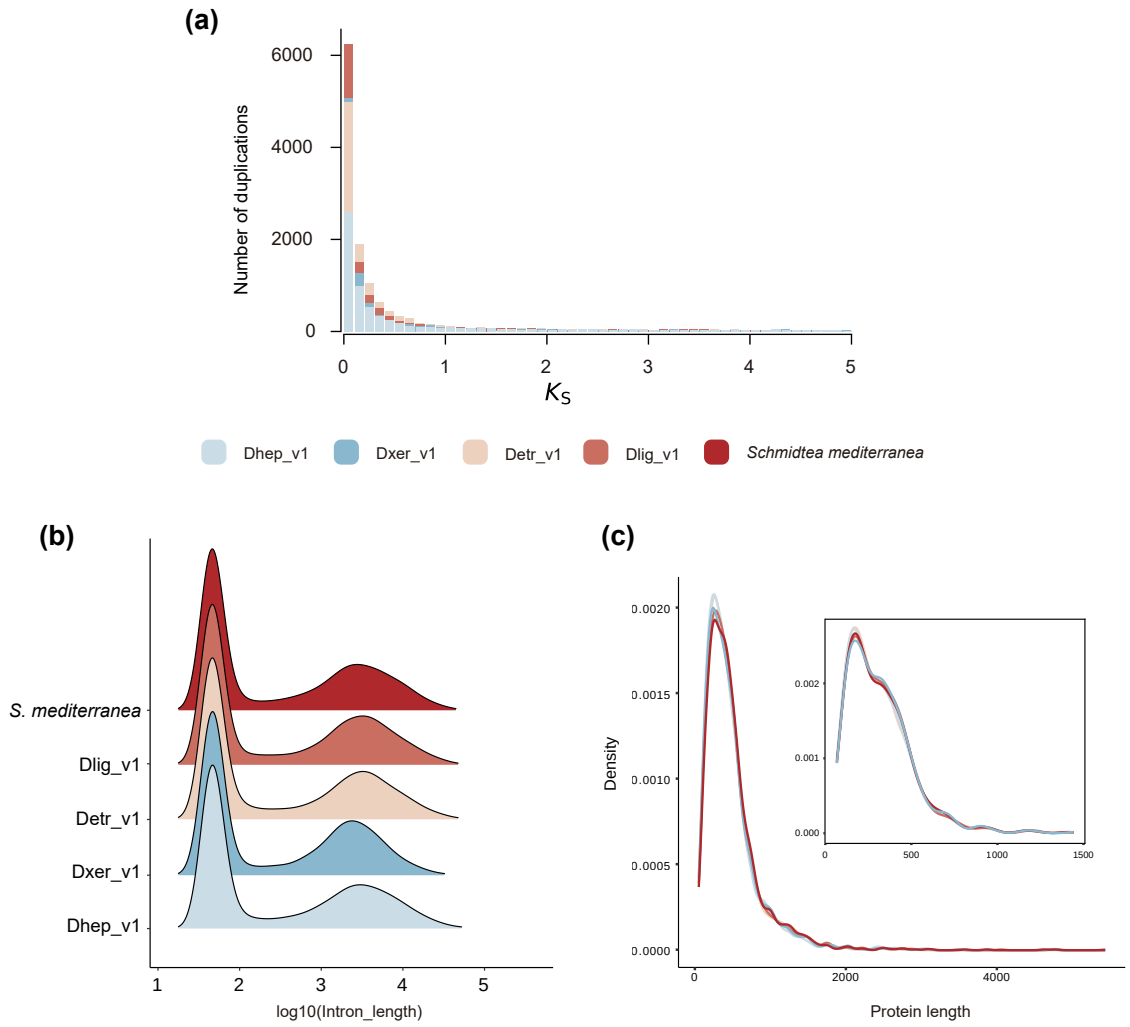


Figure 2. (a) Overlayed paranome K_s distributions of the four *Dugesia* genome assemblies showing an exponential decay pattern. **(b)** Ridgeline plots of intron length distributions for all introns comprised in the 2,266 single-copy orthogroups shared among *D. hepta* (Dhep_v1), *D. etrusca* (Detr_v1), *D. liguriensis* (Dlig_v1), *D. xeropotamica* (Dxer_v1), and *S. mediterranea*. **(c)** Density distribution of CDS lengths of all 2,266 single-copy orthologues. The inset shows the distribution for the 571 exon-equivalent orthologues.

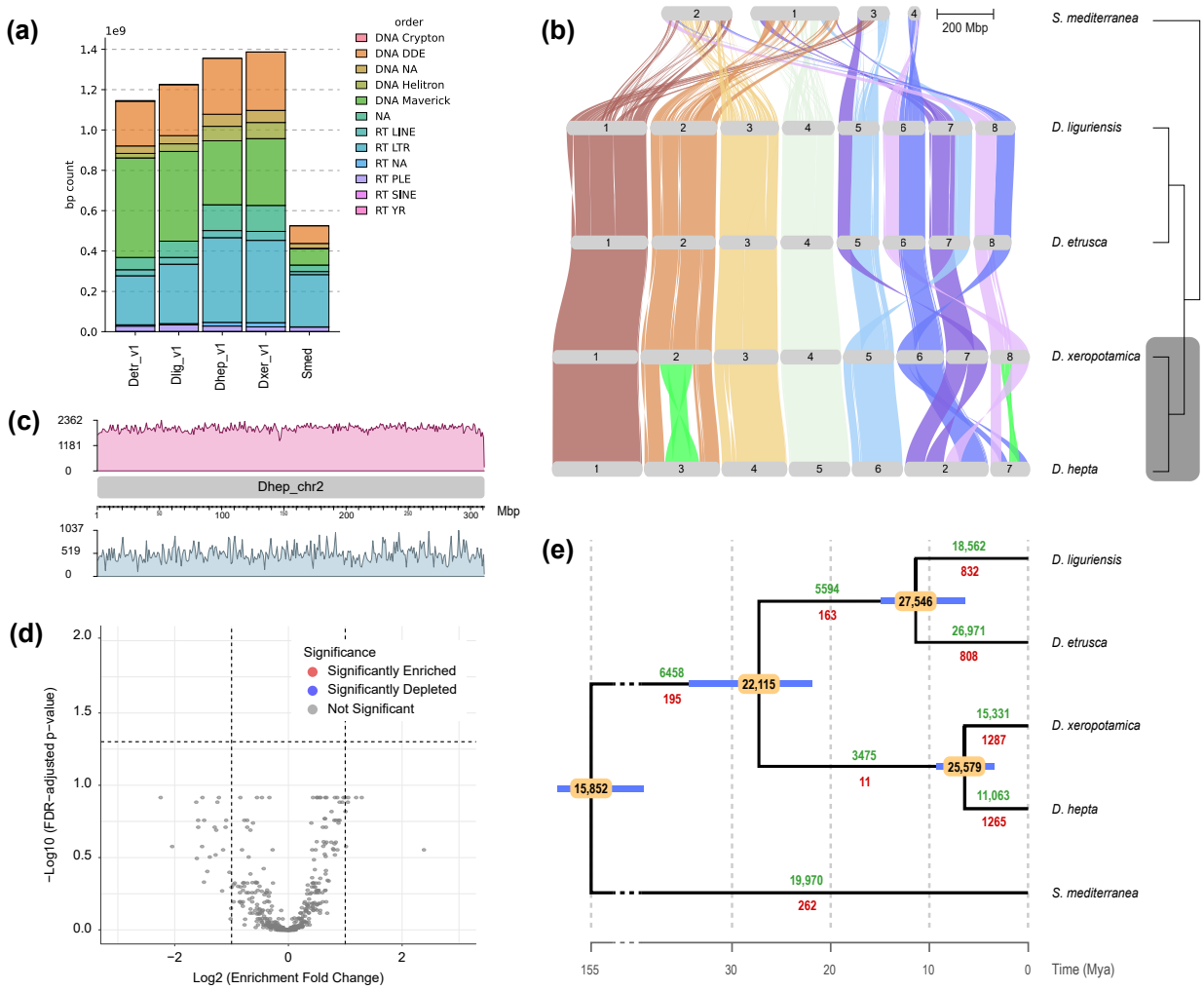


Figure 3. (a) Stacked histograms showing the total repeat content (bp) of each genome assembly, classified by TEs orders. **(b)** Synteny analysis between the four *Dugesia* species and *S. mediterranea*, with ribbons colored based on *D. xeropotamica* (n = 8). Pericentric inversions on chromosomes 2 and 7 are highlighted in green. Right: cladogram depicting phylogenetic relationships. Highlight in dark grey identifies the Corso-Sardinian clade. **(c)** Density plot of TEs (pink) and genes (cyan) along chromosome 2 of *D. hepta*. **(d)** Vulcan plot showing the enrichment analysis for 20 kb windows flanking synteny breakpoints across *D. hepta*'s chromosome 2. Horizontal dotted line indicates the statistical significance threshold. Vertical dotted lines represent a 2-fold depletion (left) and a 2-fold enrichment (right). **(e)** Time-calibrated phylogeny of the studied species showing the minimum number of gene gains (green) and losses (red) estimated with BadiRate. Ancestral gene counts are shown within yellow bubbles on top of the tree nodes, and blue bars represent the maximum and minimum time estimates for each node.

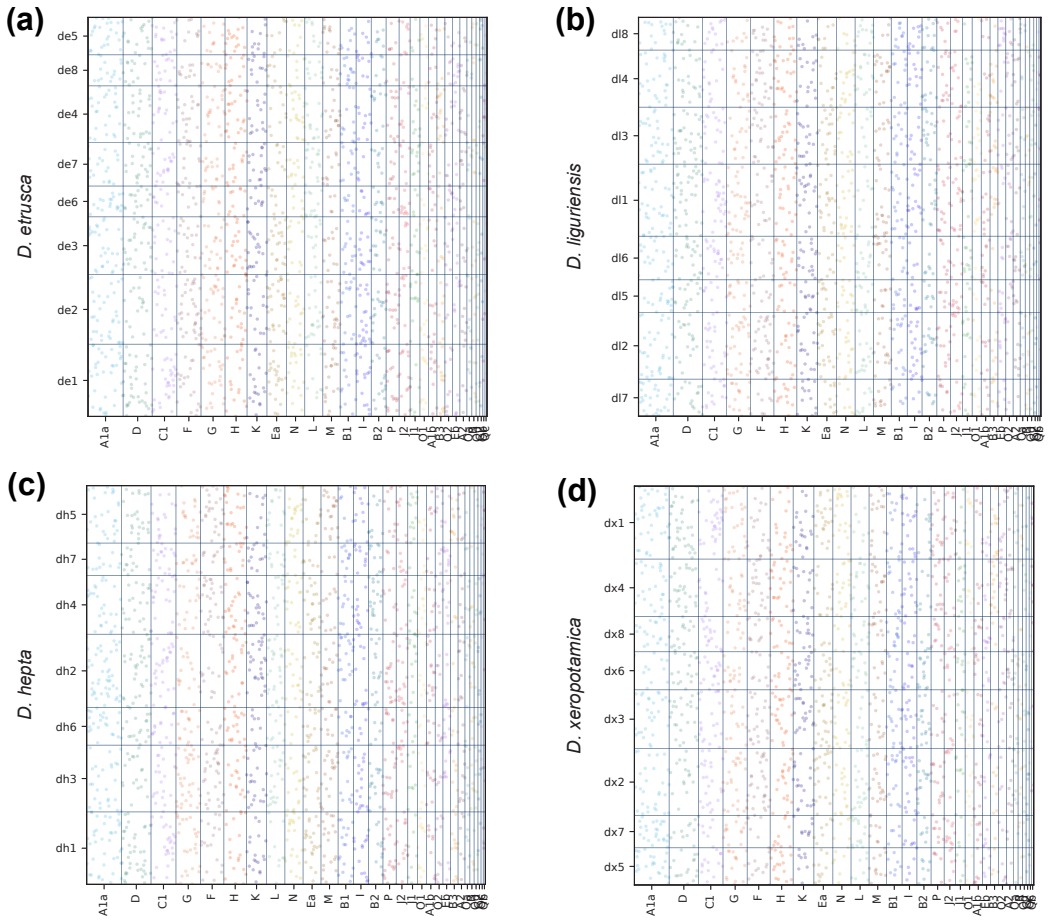


Figure 4. Dotplots showing the relationship between metazoan ancestral linkage groups (MALGs) and **(a)** *D. etrusca*, **(b)** *D. liguriensis*, **(c)** *D. hepta*, and **(d)** *D. xeropotamica*. MALGs without significant enrichment are plotted in light colors. No MALGs enrichment was detected in any *Dugesia* species.

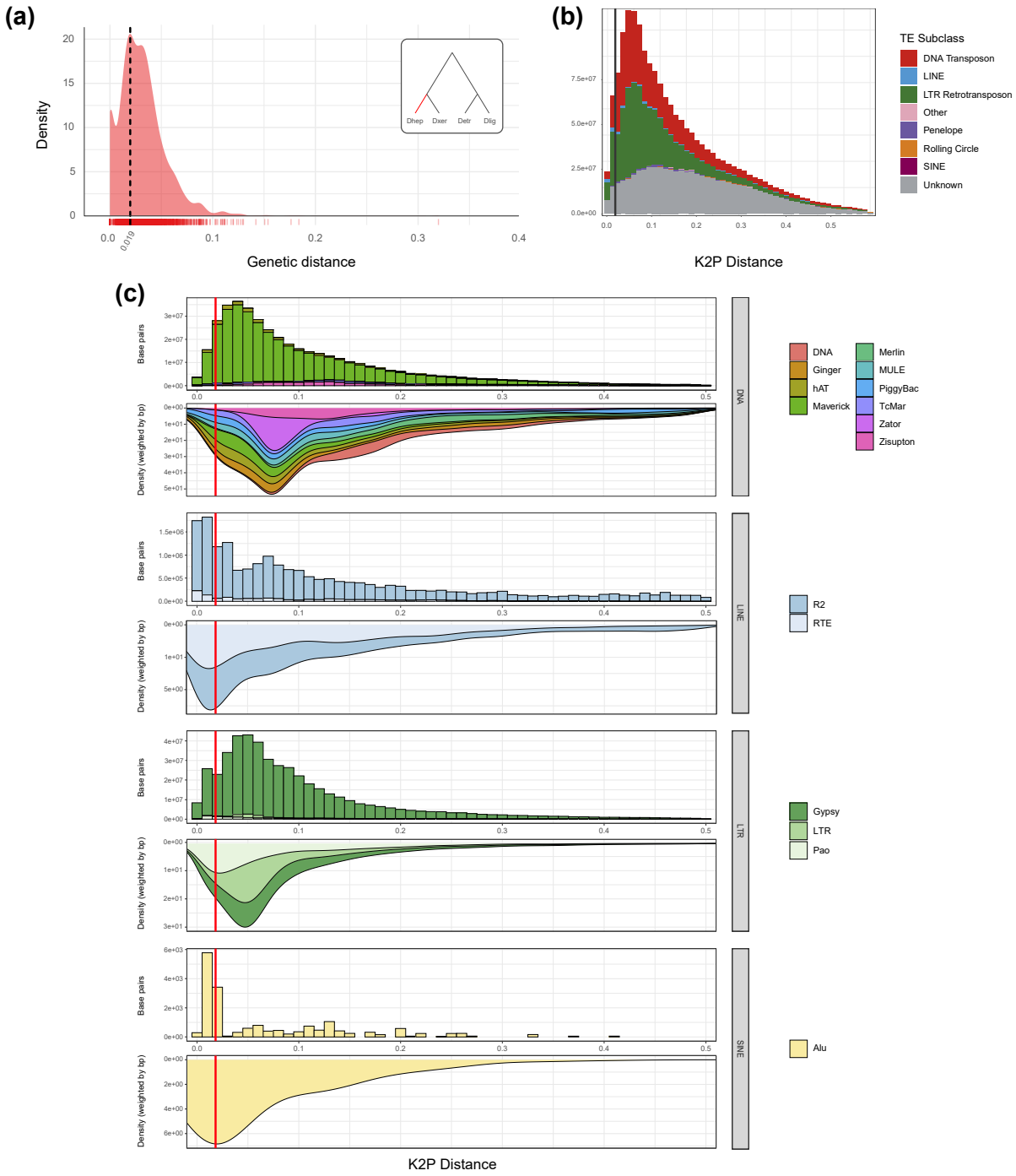


Figure 5. Density and repeat landscape analyses of *D. hepta*. **(a)** Density distribution of four-fold degenerate (4d) site genetic distances. The dotted vertical line indicates the distribution mode. **(b)** Repeat landscape plot summarizing relative TEs activity using Kimura 2-parameter (K2P) divergence; recent activity is shown on the left-hand side of the plot. The solid vertical line represents the 4d divergence mode for comparison. **(c)** Repeat landscapes of DNA, LINE, LTR, and SINE superfamilies. Upper plots represent cumulative base pairs grouped by divergence and family (color-coded), while inverted plots display the corresponding density distributions. Vertical red lines indicate the 4d divergence mode for reference.

Table 1. Genome assembly statistics after contaminants removal.

	Detr_v1	**Dhep_v1**	**Dlig_v1**	**Dxer_v1**
Flow cytometry (Gb)	1.86	2.05	1.51	2.09
Assembly size (bp)	2,961,733,428	2,090,052,221	2,136,556,669	2,484,242,095
GC content (%)	27.22	27.64	27.34	27.94
Number of scaffolds	59,495	15,220	8,935	35,214
Longest scaffold (Mb)	288.8	335.3	299.4	321.6
Sequence in chromosomes (%)	52.50	82.73	74.26	69.18
N50 (Mb)	141.2	242.1	162.2	176.5
L50	8	4	5	6
Complete BUSCOs				
Metazoa data set	553 (58.0%)	616 (64.6%)	659 (69.1%)	556 (58.3%)
Arthropoda data set	599 (59.1%)	591 (58.3%)	636 (62.7%)	597 (58.9%)
Eukaryota data set	189 (74.1%)	180 (70.6%)	199 (78.0%)	188 (73.7%)

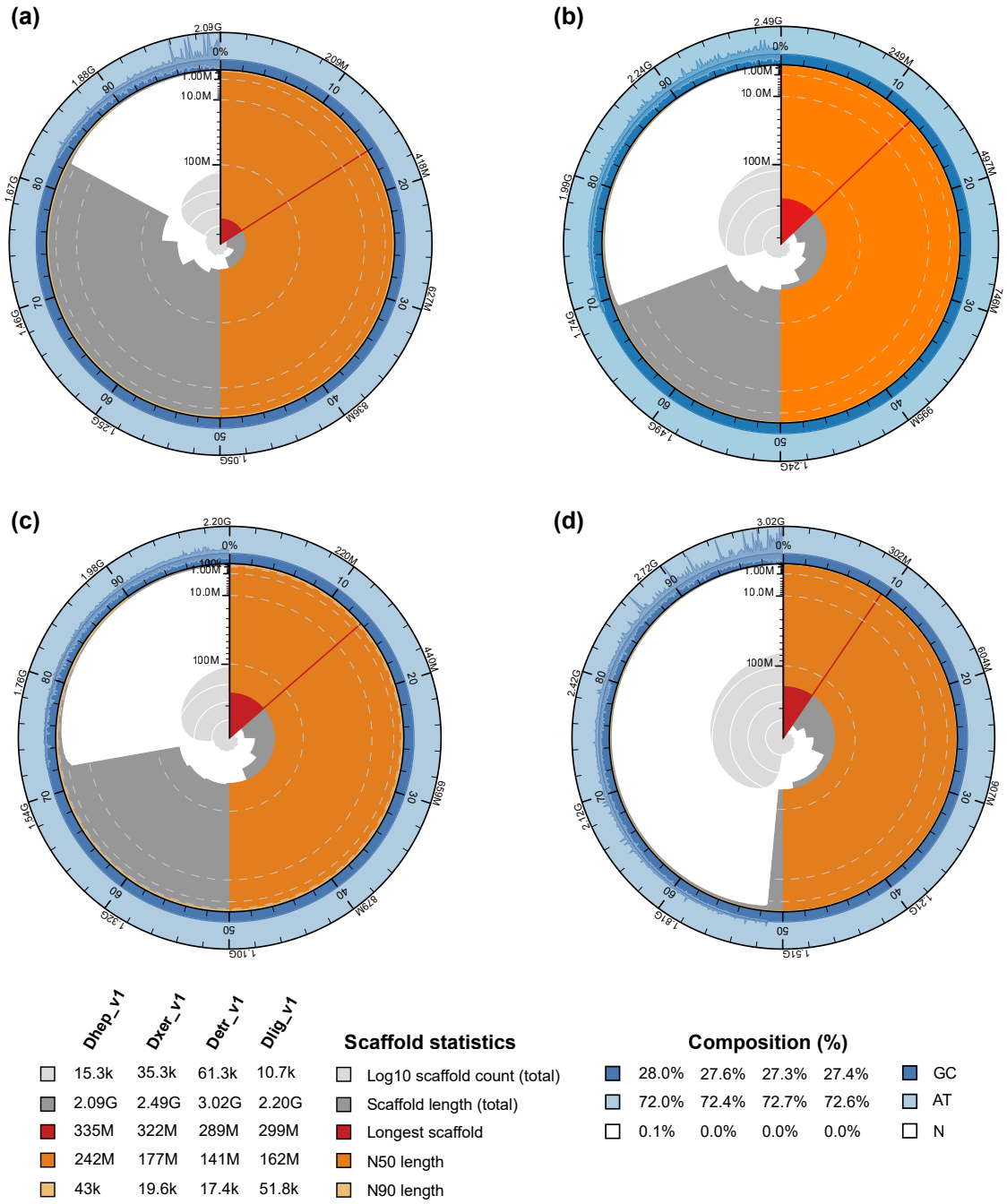
Table 2. Genome annotation statistics.

	Detr_v1	**Dhep_v1**	**Dlig_v1**	**Dxer_v1**
Repetitive regions (%)	78.32	78.68	78.91	83.29
Transposable & repetitive elements (bp)	2,318,718,339	1,643,800,512	1,685,580,944	2,067,343,575
Number of elements	5,922,150	4,067,511	4,794,532	5,109,694
Genome annotation				
Protein-coding genes	53,829	35,435	45,367	39,727
Number of transcripts	56,813	37,773	47,772	41,966
Supported genes	27,240	18,588	21,268	29,282
Functionally annotated genes	39,543	28,146	34,991	30,520
Number of tRNA	1,336	6,106	19,549	11,762
Number of tRNA after filtering	842	3,247	16,080	4,258
Complete BUSCOs				
Metazoa data set	699 (73.3%)	689 (72.2%)	539 (56.5%)	732 (76.7%)
Arthropoda data set	718 (70.9%)	701 (69.2%)	561 (55.3%)	736 (72.7%)
Eukaryota data set	211 (82.8%)	205 (80.4%)	170 (66.7%)	216 (84.7%)

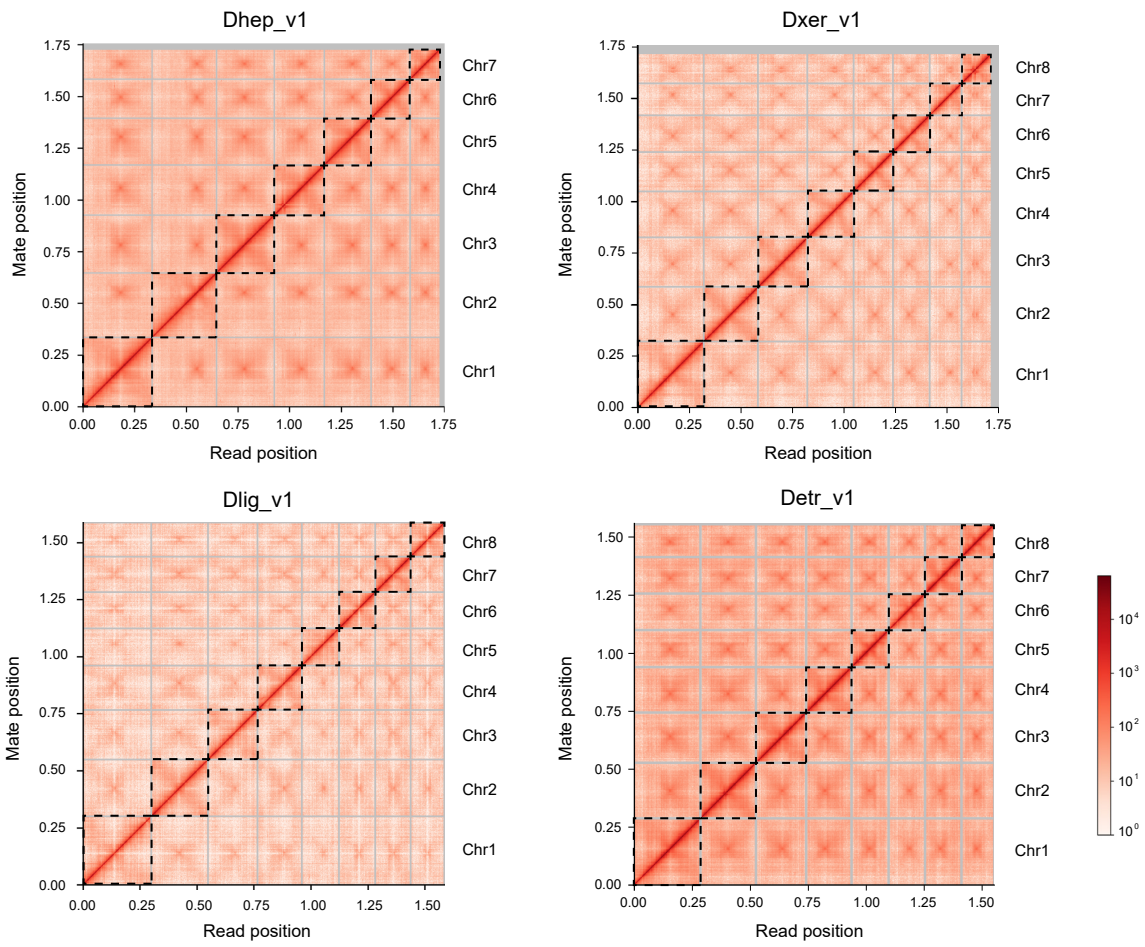
Table 3. Summary length (bp) statistics of introns and protein-coding sequences (CDS) of single-copy orthologues.

All introns	**Detr_v1**	**Dhep_v1**	**Dlig_v1**	**Dxer_v1**	**Smed**
Number of introns	11,032	10,692	10,370	10,312	9,273
Shortest intron	33	38	38	33	34
Median lenght	94	86	123	107	96
Mean length	3326	3398	3375	3253	2435
Longest intron	133,684	114,161	90,611	132,065	248,596
Short introns (≤316 bp)					
Number of introns	6060	5879	5527	5519	4999
Shortest intron	33	38	38	33	34
Median lenght	47	47	47	46	47
Mean length	59.75	58.72	59.02	58.08	58.17
Longest intron	316	316	315	316	316
Large introns (>316 bp)					
Number of introns	4972	4813	4843	4793	5671
Shortest intron	318	317	320	317	317
Median lenght	4165	4041	4140	3866	3118
Mean length	7307	7476	7159	6932	5215
Longest intron	133,684	114,161	90,611	132,065	248,596
All CDS					
Number of CDS	2266	2266	2266	2266	2266
Shortest CDS	52	67	67	62	67
Median length	342	340.5	355	350	364
Mean length	423.6	417.1	437.5	433.3	445.2
Longest CDS	4196	4052	4807	4252	4224
Exon-equivalent CDS					
Number of CDS	571	571	571	571	571
Shortest CDS	71	71	71	71	67
Median length	266	267	273	275	276
Mean length	299.4	300.3	304.5	305	306.4
Longest CDS	1336	1349	1351	1349	1379

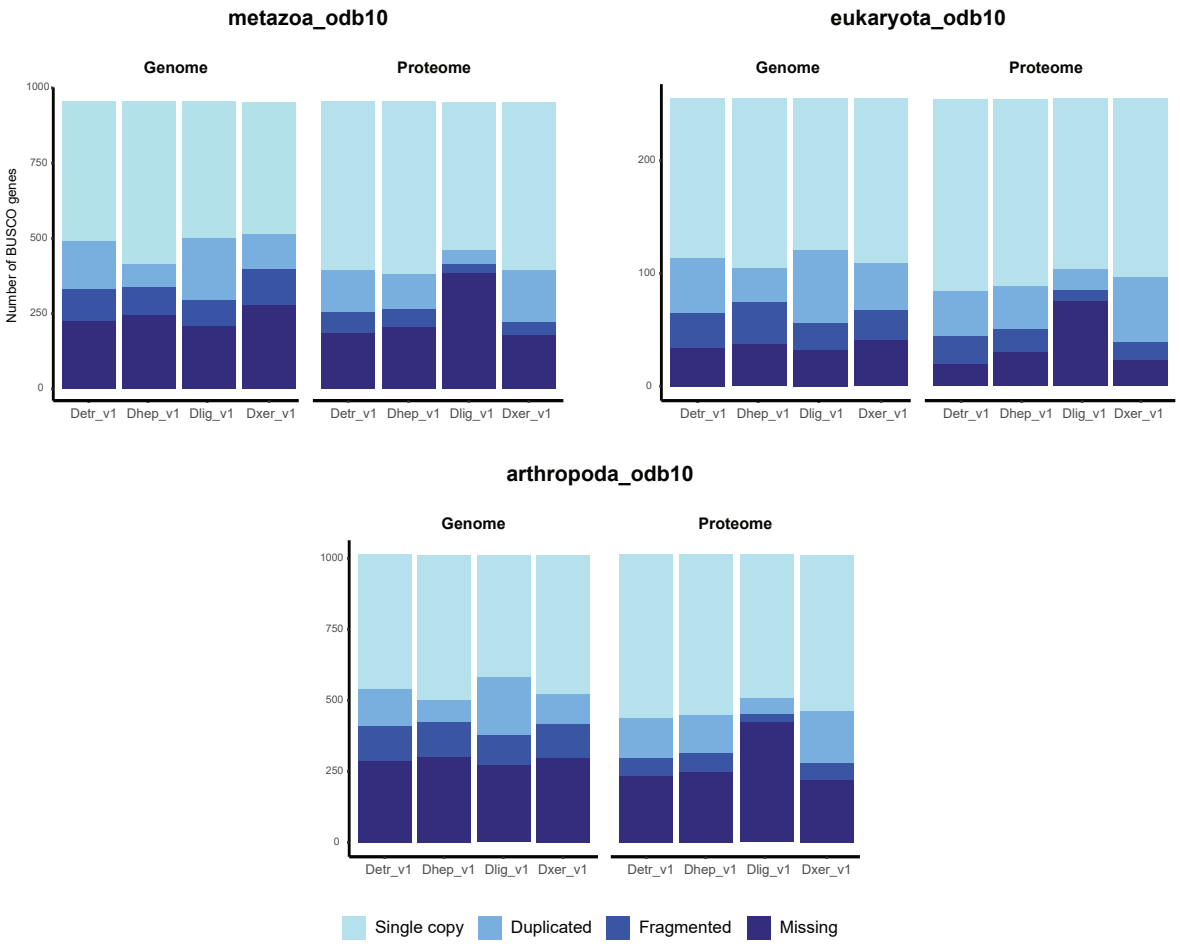
Supplementary Information: Figures



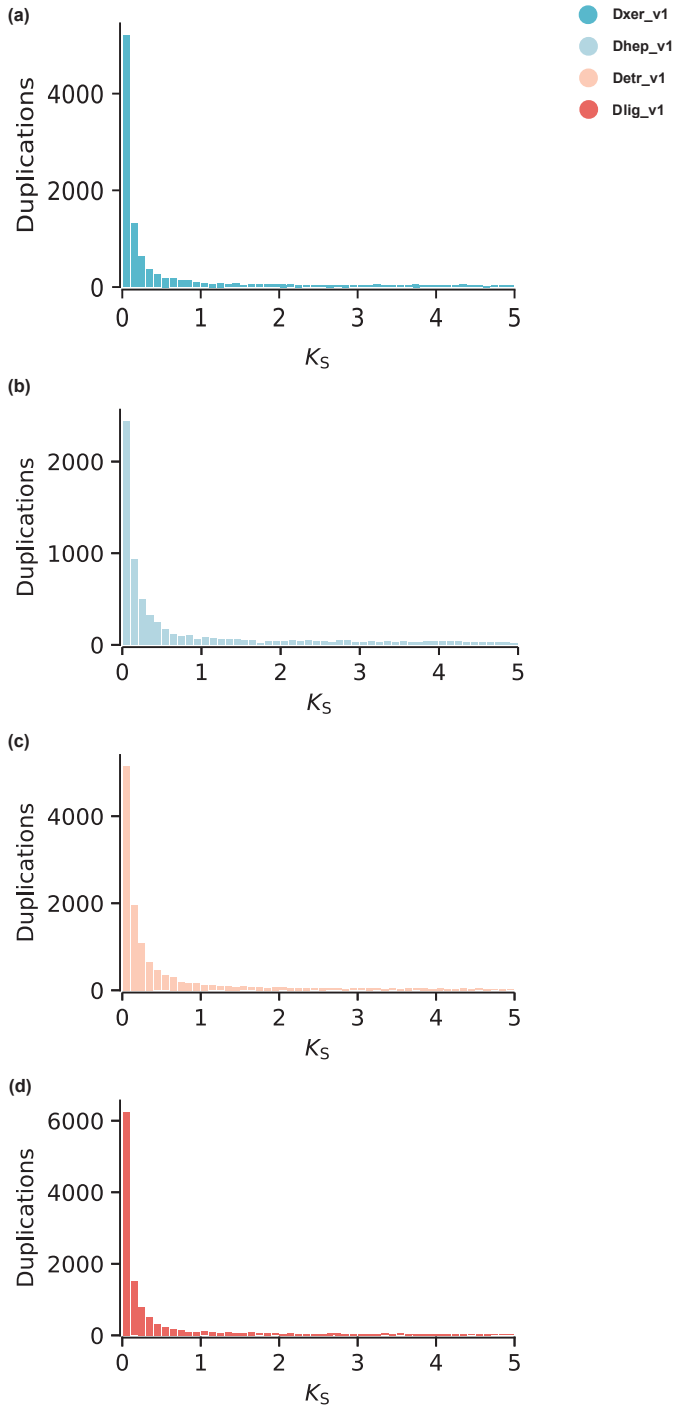
Supplementary Fig. S1. Snail-plots summarizing genome assembly statistics for (a) *D. hepta*, (b) *D. xeropotamica*, (c) *D. liguriensis*, and (d) *D. etrusca*.



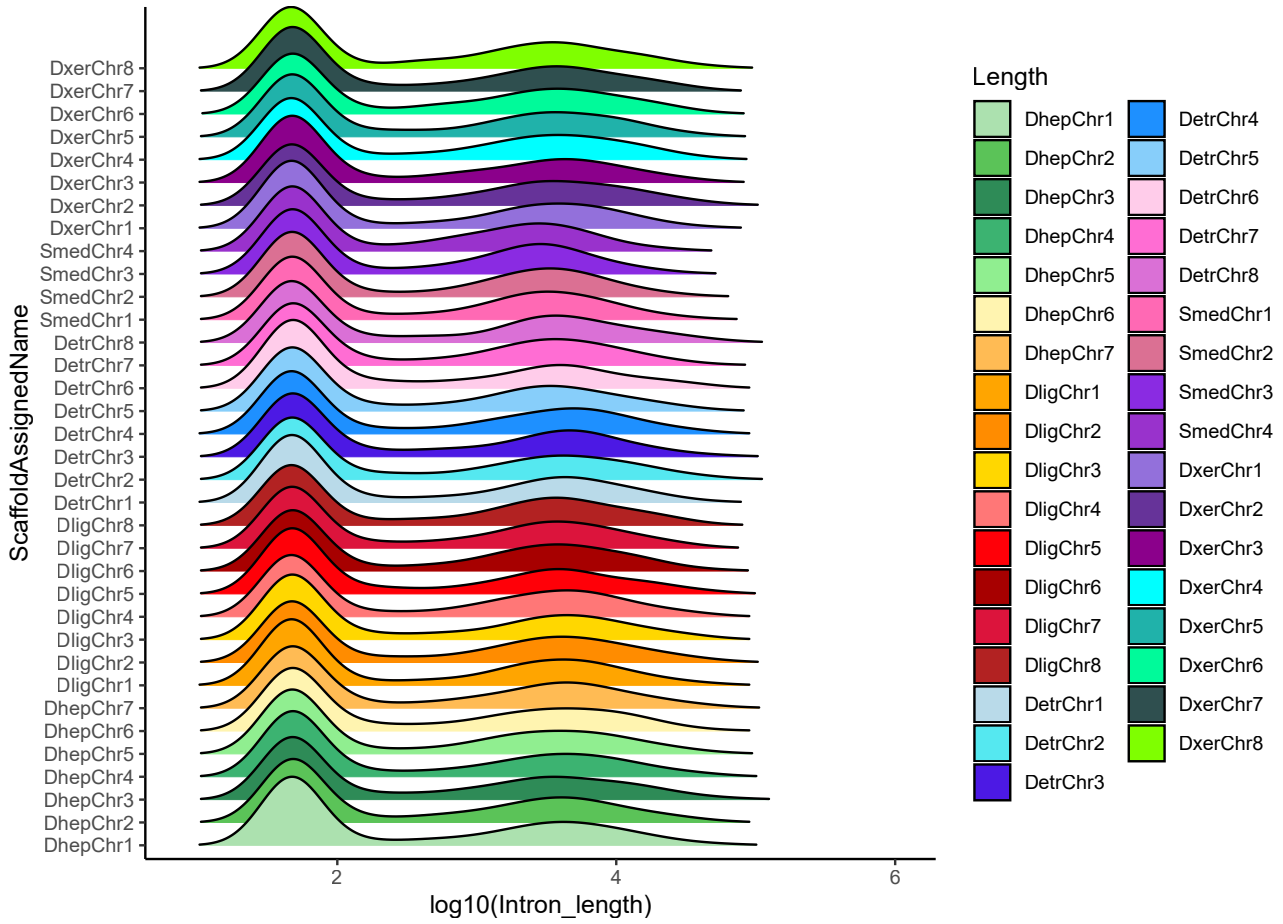
Supplementary Fig. S2. Long-range contact heat-
map of paired Omni-C reads depicting the major scaffolds of



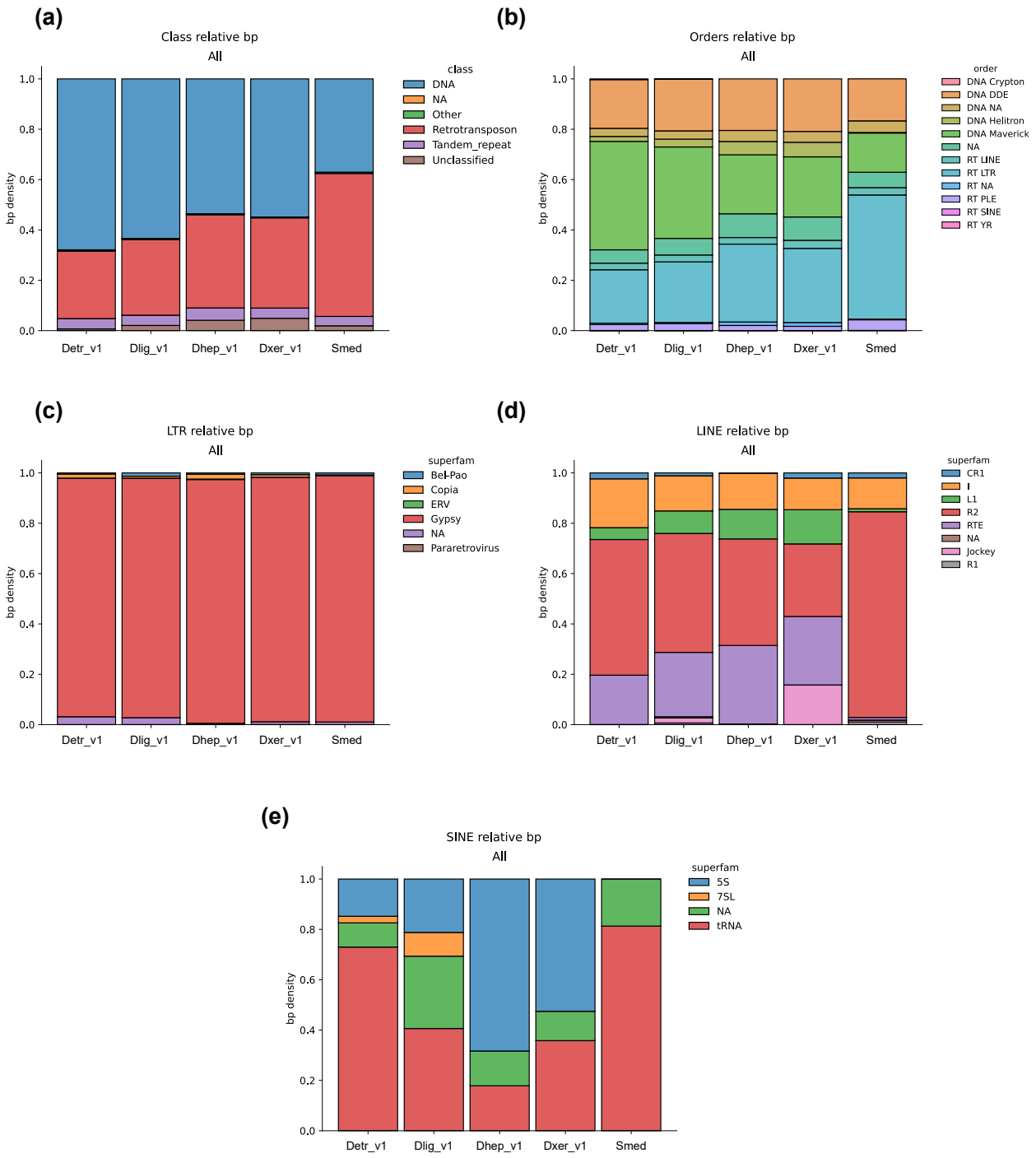
Supplementary Fig. S3. BUSCO assessment of the genome assemblies before (Genome) and after (Proteome) functional annotations.



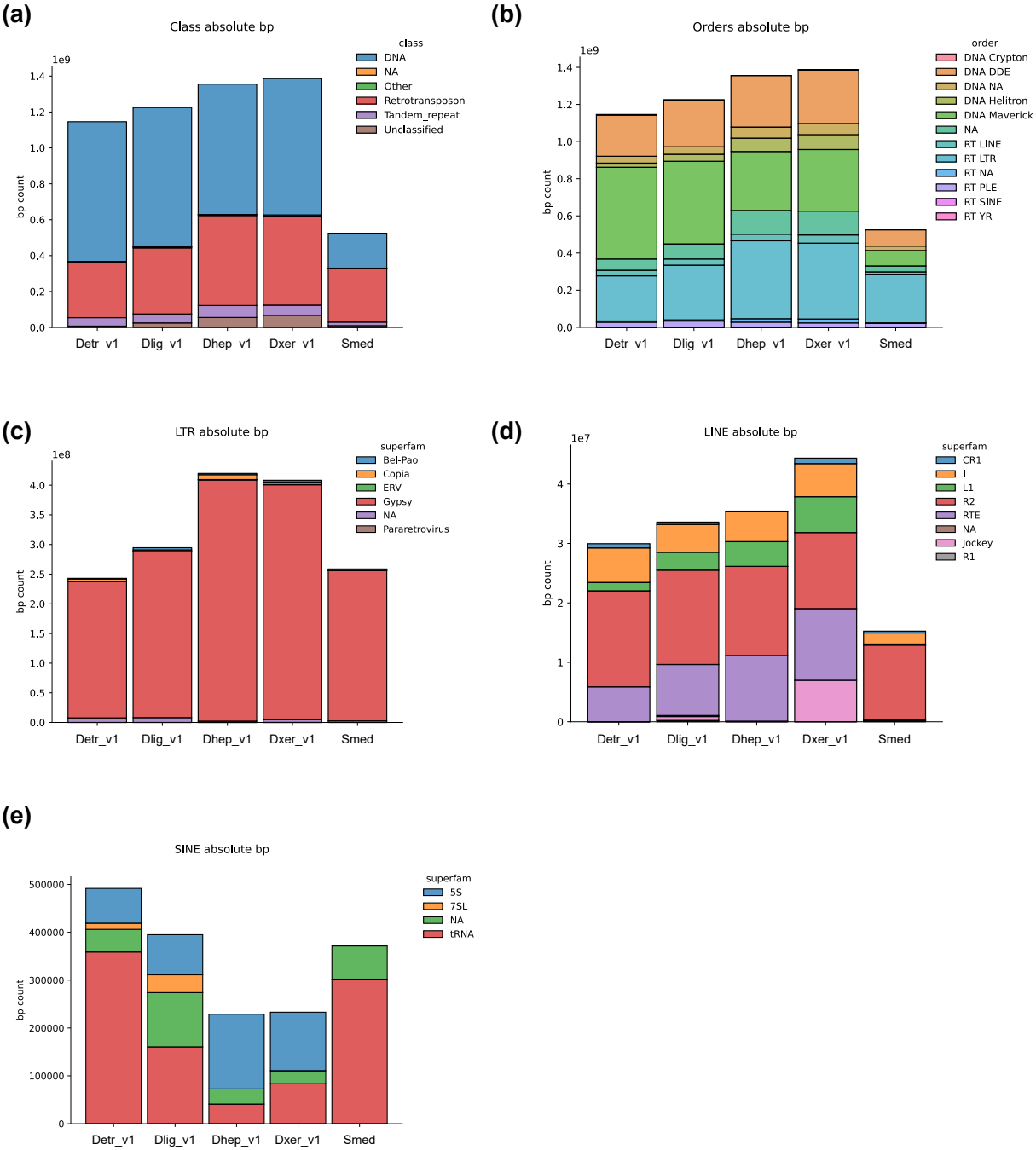
Supplementary Fig. S4. Paranome K_s distributions of the *Dugesia* genome assemblies showing a pattern of exponential decay.



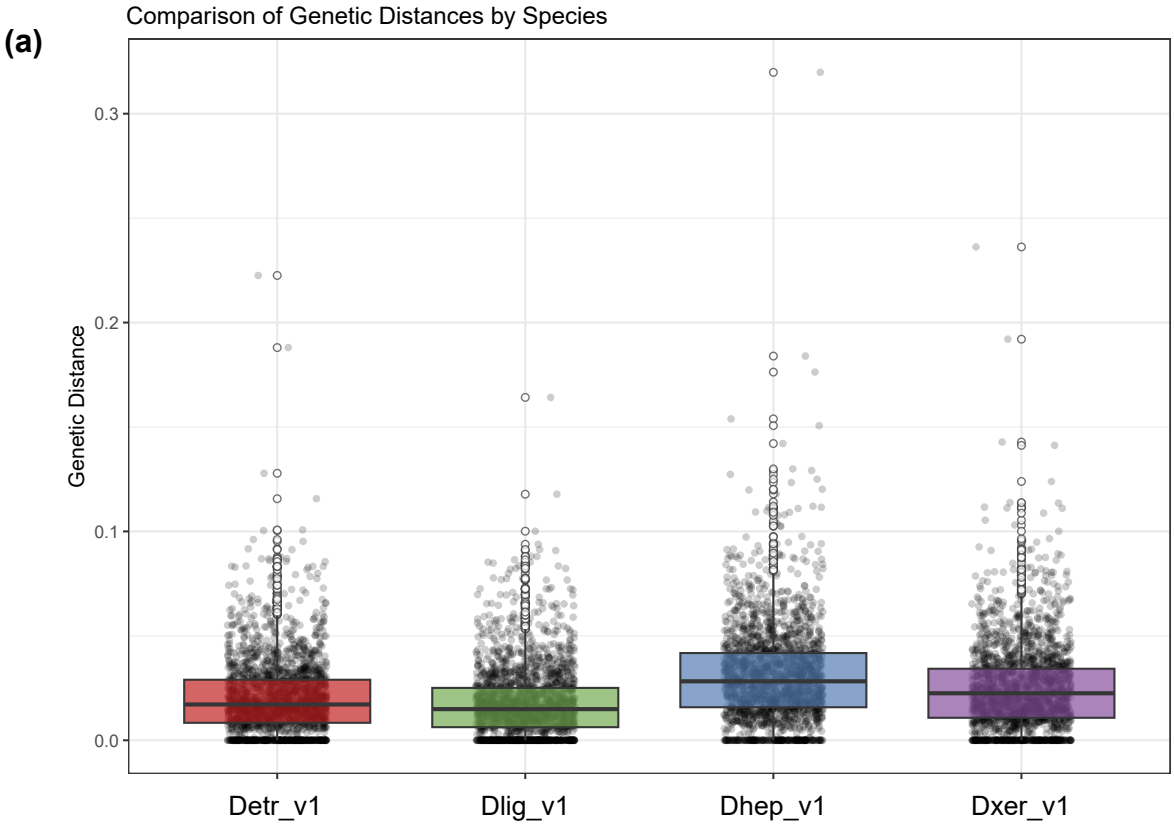
Supplementary Fig. S5. Ridgeline plots showing the density distribution of intron lengths of all introns comprised in the 2266 single-copy orthogroups shared between *S. mediterranea* (Smed), *D. hepta* (Dhep), *D. xeropotamica* (Dxer), *D. etrusca* (Detr) and *D. liguriensis* (Dlig) separated by chromosomes.

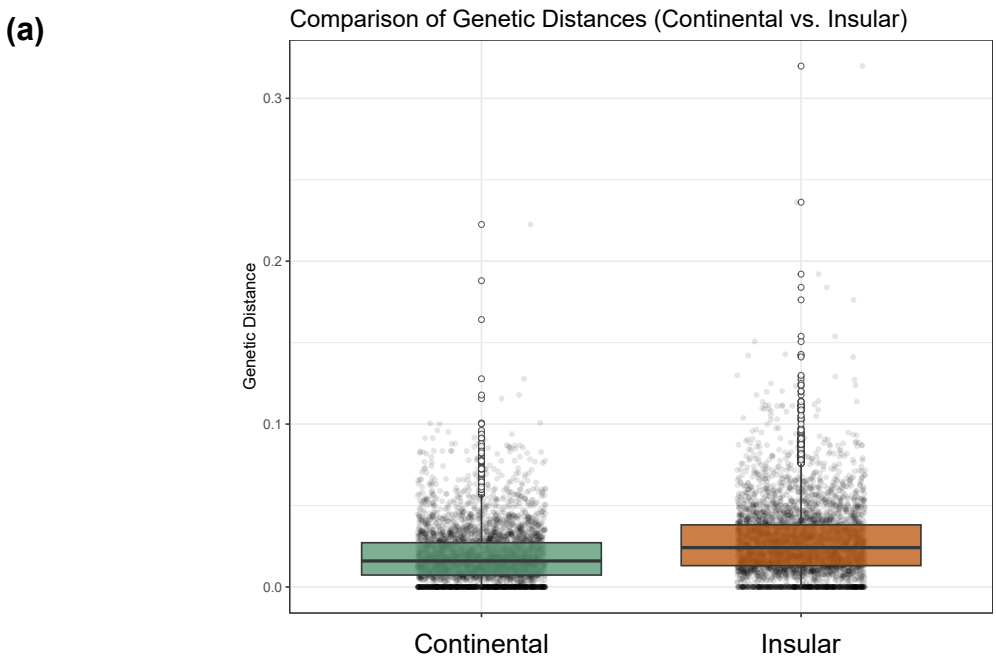


Supplementary Fig. S6. Comparison of the relative contributions (%) of each class (a) and order (b) of repetitive elements to the repeat content of each genome assembly. (c) Relative superfamilies contributions of repetitive elements to the LTR retrotransposon elements. Relative superfamilies contributions of repetitive elements to the (d) LINE retrotransposon and (e) SINE retrotransposon elements.



Supplementary Fig. S7. Stacked histograms showing the total amount (bp) of repeat content of each genome assembly classified into (a) classes of repetitive elements and (b) orders within said classes. (c) Close up look at the proportion of LTR retrotransposon elements. (d) Close up looks at the proportion of LINE retrotransposon elements and (e) SINE retrotransposon elements.





Supplementary Fig. S8. Comparison of the distributions of four-fold degenerate (4d) distances per species (a) and grouping continental and insular lineages (b).