The first chromosome-scale *Dugesia* genomes shed light on structural rearrangements and genome size evolution 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 author: M. Riutort mriutort@ub.edu

Abstract

High-quality genomes are indispensable for understanding biological processes, empowering discovery across taxa. Planarian flatworms (Platyhelminthes, Tricladida) are a globally distributed and phenotypically diverse group with a near-complete lack of genomic resources. *Dugesia* flatworms, in particular, have been extensively studied as an emergent key model in biogeography and evolutionary biology, but the advancement of their evolutionary research has been hindered by the absence of chromosome-scale assemblies. Here, we generate the chromosome-scale genomes of *Dugesia etrusca*, *D. hepta*, *D. liguriensis*, and *D. xeropotamica*, the first for this longstanding genus, and provide the first intergeneric comparative perspective with the model species *Schmidtea mediterranea*, a model species in regeneration. Our comparative analyses reveal a striking loss of conserved synteny with ancestral metazoan linkage groups, as well as between *Dugesia* and *Schmidtea* genera, suggesting that pervasive structural instability is a defining hallmark of flatworm genome evolution. However, in contrast to what has been reported in *Schmidtea*, synteny is broadly retained among *Dugesia* species. We also found a widespread expansion of transposable elements, most notably DNA transposons, in *Dugesia*, and reject a whole-genome duplication hypothesis as the underlying cause for its larger genome size and haploid chromosomal set.

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

1. Introduction

Biological evolution proceeds through cumulative genomic changes, subtly weaving the continuous and complex process that leads to speciation. The advent of next-generation sequencing (NGS) has greatly advanced comparative genomics and our understanding of how genomes relate to phenotype, evolve, and are regulated[1,2,3,4,5]. Chromosome-scale genome sequences are key for elucidating the mechanisms and evolutionary history of structural variation, chromosomal rearrangements (CRs), and their consequences for reproductive isolation and speciation[6,7,8,9,10,11]. Fortunately, such resources have become increasingly available for non-model organisms. However, genomic research efforts are not distributed equally across the Tree of Life, leading to significant taxonomic biases[12,13,14]. This disparity is particularly striking in the Lophotrochozoa[15], a superphylum that remains severely underrepresented in genomic databases. Despite accounting for approximately 40% of all animal phyla[16,17], including major groups like Mollusca, the second-largest metazoan phylum, the number of available genomes for Lophotrochozoa stands in stark contrast to their immense biodiversity, with less than 0.01% of known species having been sequenced[18,19]. This significant gap of knowledge poses a challenge to the field of evolutionary biology. Relevant evolutionary innovations unique to this vast portion of animal life risk being overlooked, which in turn limits our comprehensive understanding of metazoan biology. Recent studies already underscore the importance of exploring these phyla, revealing new insights into the architecture and plasticity of lophotrochozoan genomes[20,21]. As subjects of this study, planarians (Platyhelminthes, Tricladida) are among the groups historically overlooked in genomic studies.

Planarians are a species-rich taxon encompassing thousands of species distributed worldwide, while representing a rich phenotypic diversity that can be found in terrestrial, marine, and freshwater environments[22]. Flatworms have been extensively studied in ecology[23] (and references therein), autophagy dynamics[24] (and references

therein), and biogeography[25,26], to name a few disciplines. However, planarians are mostly studied for their unmatched regenerative abilities[27,28,29] and their pluripotent stem cells (neoblasts). Their regenerative prowess has established them as model systems in regeneration and stem cell dynamics. Indeed, 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 Sanger-sequenced genomes[30,31]. Despite this, only five genome assemblies of triclads are currently available for the planarian research community: those of the four *Schmidtea* species known to date[27], and that of *Dugesia japonica* Ichikawa & Kawakatsu, 1964[32]. Nevertheless, *Schmidtea* genomes have already revealed peculiar characteristics, such as widespread genome scrambling and a complete absence of conserved ancestral metazoan linkage groups (MALGs) across their species, prompting the proposal that genome-wide structural instability is a hallmark of flatworm genome evolution[27].

The genus *Dugesia* Girard, 1850, closely related to *Schmidtea* (Figure 1), comprises approximately 110 freshwater planarian species, roughly half of the family Dugesiidae Ball, 1974, with a global distribution. It has become an emergent model system in evolutionary biology over the last decades, particularly due to its high species richness in the geologically active Mediterranean region and the capacity of many species to reproduce both sexually and asexually by fissiparity. A thorough research program in the Western Mediterranean has leveraged these traits to reveal new species, a unique putative hybridization-driven speciation event, frequent chromosomal translocations in asexuals, evidence of the mosaic-Meselson effect in fissiparous lineages, and the discovery of long-term asexual clades[33,34,35,36,37]. These efforts have also established a well-resolved phylogeny linking diversification to abiotic events, progressively advancing from a few molecular markers to reduced-representation genomic data and transcriptomes. Despite this extensive work, the current genomic reference for *Dugesia* remains the non-chromosome-level, haploid consensus assembly of *D. japonica*,

comprised of 10,017 scaffolds[32], highlighting a critical gap between the genus's established phylogenetic and evolutionary framework and the lack of high-quality genomic resources needed to investigate the underlying molecular mechanisms.

In the interest of address this gap and advance *Dugesia* evolutionary research, we sequenced and assembled chromosome-scale genome assemblies of four 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 species 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 of two other Sardinian endemics, for which unfortunately assembling a genome was not possible, viz. *Dugesia benazzii* Lepori, 1951, and *Dugesia leporii* Pala, Stocchino, Corso, & Casu, 2000.

These new genomic resources will enable us to (i) analyze the genomic evolution within the genus, including the interplay of chromosomal rearrangements (CRs) with speciation, given that *D. hepta* is distinguished by its unique haploid complement of seven chromosomes (n = 7), a notable exception to the predominant regional set of n = 8 [38] (Figure 1). Furthermore, (ii) establish a comparative framework between *Schmidtea* and *Dugesia* to explore the notable differences between both genera. Most Western Palearctic *Dugesia* species share a haploid complement number of eight chromosomes (n = 8), which is double the number of chromosomes typically found in the haploid sets of *Schmidtea* (Figure 1). This has led us to hypothesize that a whole-genome duplication (WGD) event may underlie this apparent difference, a possibility that has not been taken into account before. It also offers the opportunity to (iii) deepen our understanding of some of the topics recently brought to light with *Schmidtea,* (iii.a) such as the pervasiveness of genome-wide rearrangements found in the genus and proposed to be potentially a feature of all Tricladida, and (iii.b) the complete lack of conservation of MALGs in planarian flatworms that would suggest an entirely independent genome architecture in triclads[27]. With the focus on these three aims, the development of

these chromosome-level assemblies for *Dugesia* will benefit and contribute to the advancement of population and evolutionary genomics, not only for planarians but to a broader metazoan scope as well.

Results

2.1. New *Dugesia* genomes: assembly and annotations

We generated chromosome-level genome assemblies for four selected *Dugesia* species—a first for this taxon. Wild populations of the selected continental species are known to present coexisting biotypes (i.e., ploidies). Thus, we characterized the biotype of our continental samples via karyological analysis and flow cytometry to ensure that only diploid individuals were used. 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) were equal to the known haploid chromosomal number of each species[39,40,41]. These scaffolds, hereafter referred to as pseudochromosomes, represent between 52.5% and 82.7% of their respective total assembly sizes (Table 1; Supplementary Fig. S2). Genome completeness was assessed using BUSCO[42]. The results obtained were comparable to those previously reported in other freshwater planarians (Table 1; Supplementary Fig. S3; cf. Ivanković et al.[27]), strengthening the idea that BUSCO searches may be sub-optimal when assessing completeness in planarian flatworm genomes owing to their substantial loss of genes[43]. Our results add to the accumulating evidence that planarian flatworms would benefit from having a group-specific BUSCO set (e.g., [27,32,36]) as indicated by the high number of missing

BUSCOs (Supplementary Table S1). However, BUSCO results generally improved after using the structurally annotated assemblies (Table 2; Supplementary Fig. S3).

Structural annotation using BRAKER2 revealed that *Dugesia* assemblies harboured between 35,345 and 53,829 protein-coding genes (Table 2). These values substantially exceed the reported protein-coding genes in *D. japonica* (12,031 genes; cf. Tian et al.[32]). However, results for *D. hepta* and *D. xeropotamica* were similar to those reported in *S. mediterranea* (~31,000 genes in Grohme et al.[43]; 34,976 genes modelled using Guo et al.[44] assembly). Sequence similarity-based searches showed that the majority of protein-coding genes (73.5-79.4%; Table 2) contained hits against the surveyed protein databases for functional annotation.

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

2.2.1. Whole-genome duplication

We hypothesized that a WGD event could be the underlying cause for the notable karyotypic difference between both genera and the GS 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[45], (2) estimating the prevalence of collinear genes in the assemblies using MCScanX[46], and (3) analysing macrosynteny patterns. However, our results indicate that most retained gene duplicates are of recent origin given the exponentially decaying distribution of K_s (Figure 2a; Supplementary Fig. S4). Moreover, our results showed that only 1.40% (Detr_v1), 1.04% (Dhep_v1), 2.19% (Dlig_v1), and 0.40% (Dxer_v1) of genes are collinear. Macrosynteny results (Figure 3b) did not reveal signs of *Schmidtea* syntenic blocks matching multiple syntenic blocks in *Dugesia*.

2.2.2. Changes in intron and CDS lengths

To determine the genomic features that could account for the differences in GS between both genera, we estimated the density distribution of intron and CDS lengths (Figure 2b-c; Table 3) of single-copy orthologues (SOs), based on Broccoli v.1.2[47], and identified

2266 SOs shared between *Dugesia* assemblies and *S. mediterranea*'s, 571 of which were exon-equivalent (same number of exons in all species).

Based on our results (Figure 2b; Supplementary Fig. S5), we set a threshold value of 2.5 for the logarithmic intron length (316 bp) to classify introns in two classes: small-sized (≤ 316 bp) and large-sized (> 316 bp). Differences in small-sized introns were meaningless among *Dugesia* species and *S. mediterranea*, although the latter presented a lesser number of small-sized introns. In contrast, large-sized introns were larger and more abundant in *Dugesia* than *S. mediterranea* (Table 3), a difference that can also be observed by the displacement of the large-sized introns peak in *Dugesia*'s ridgelines (Figure 2b). On the other hand, results for the CDS density distribution were largely similar among all species (Figure 2c; Table 3), indicating that the changes that led to the increase of GS likely occurred in non-coding regions.

2.3. Macrosynteny and CRs

We inferred and represented macrosynteny blocks using GENESPACE v.1.2.3[48]. Our results showed that overall macrosynteny is well preserved among the four species of *Dugesia* (Figure 3b). However, we detected relevant rearrangements on chromosomes 5 to 8 between the Corso-Sardinian and the Iberia-Apennines-Alps clade. Within the latter clade, there are no relevant signs of major translocations or inversions (Figure 3b) even though both species shared their last common ancestor ~9.5 Mya[26]. As expected, major translocation and fusion events were observed in *D. hepta* affecting chromosomes 2 and 7. Notably, chromosome 2 shared major syntenic blocks with *D. xeropotamica*'s chromosomes 6, 7, 8 and chromosomes 5, 6, 7, and 8 of *D. etrusca* and *D. liguriensis*' (Figure 3b). These findings contrasted with what had been reported in *Schmidtea* where loss of synteny was pervasive and abundant among congeners[27]. Finally, we used the ODP tool (see Methods), which revealed that, just as in *Schmidtea*, *Dugesia* has lost all traces of detectable synteny conservation of the MALGs as well.

2.4. Genome transposable elements composition

We examined the repeat content of all pseudochromosomes (Figure 3a; Supplementary Fig. S6). Transposable element annotations revealed that a large portion of the repetitive content in *Dugesia* was assigned to DNA elements (50-65%), whereas in *S. mediterranea* said type of elements contributed to less than 40% of the repeat content (Supplementary Fig. S6a). In absolute terms, the increase in GS could be attributable to an expansion of transposable elements (TEs), most particularly, DNA/Maverick, DNA/DDE and, as reported in *Schmidtea*, LTR/Gypsy elements (Figure 3a; Supplementary Fig. S6). When comparing the relative composition of the repeat content we find that the proportion of DNA/DDE elements is similar between *Schmidtea* and *Dugesia*. However, in *Dugesia* there is a clear expansion, most notably in *D. etrusca*, of DNA/Maverick TEs, 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 when examining TEs of the LINE and SINE orders (Supplementary Fig. S7d,e). In *Dugesia* there is a clear expansion of LINE/RTE elements, the smallest type of non-LTR retrotransposons[50], as well as an expansion of the SINE/5S and SINE/7SL subfamilies, which we were not able to detect in *Schmidtea* (Supplementary Fig. S6e). Remarkably, we found that ~ 65% of the SINE elements in *D. hepta* are SINE/5S elements, which is 3~4 times more than the prevalence of this type of elements in *D. etrusca* or *D. liguriensis*.

2.5. TEs enrichment analysis

We explored the observed expansion of SINE/5S elements in *D. hepta*. We thought that such a differential expansion could be an indicator of TEs involvement in the CRs detected in the macrosynteny analysis. Consequently, we explored the distribution of TEs density along *D. hepta*'s multisyntenic chromosome 2 using the R package *karyoploteR*[51] and tested whether any TEs family was overrepresented in 20kb flanking windows nearby synteny breakpoints using the R package *GenomicRanges* v.1.46.1[52]. However, we found no particular TEs family was either significantly enriched or depleted along chromosome 2 synteny breakpoints (Figure 3c,d; Supplementary Table

S2).

2.6. Relative TEs activity period

We explored 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 the 4d divergence mode (Figure 5; Supplementary Table S3), suggesting that relative TEs activity period of said families and subclasses might predate *D. hepta*'s speciation.

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

Whole genome duplication events entail not only genome-wide duplication but also a rapid decline of gene copies, which may reflect on the estimation of gene gain and loss turnover rates. We explored the evolution of *Dugesia* genomes under that perspective using BadiRate v.1.35[55]. Orthology inference with Broccoli yielded 18,719 orthogroups (OGs hereafter and referred to as gene families throughout the text). However, Broccoli may leave some proteins unassigned to an OG when handling complex evolutionary scenarios. Thus, we included the unassigned proteins as singleton OGs for this analysis, resulting in a total of 54,011 OGs. We compared different branch rate estimation models under the birth, death and innovation model (BDI). Our results supported a free rates model (Supplementary Table S4) and showed that gene family gains were distributed mostly on the external branches leading to the current species (Figure 3e) and not on the innermost branches as might have been expected under a WGD scenario. This result was in line with the estimates of gene family turnover rates, which showed an increase in both birth and death rates towards the external branches (Figure 3e; Supplementary Table S4). Most notably, the largest turnover estimates were identified in the Corso-Sardinian clade branches.

Discussion

New genomes to contribute to Tree of Life

Here, we presented and analysed four genome assemblies of planarian flatworms of the genus *Dugesia*, which constitute the first of their kind at a chromosome-scale. The new assemblies are more contiguous and complete than the current reference genome available for the taxon prior to this study (cf. Tian et al.[32]). Results for these new *Dugesia* assemblies are promising (Table 1), representing a >200-fold improvement over the prior reference, and the pseudochromosomes covered a large percentage of the total genome assembly sizes (Table 1; Supplementary Fig. S1, S2), our results could not attain the same level of completeness as reported in *Schmidtea* (cf. Ivanković et al.[27]). However, these genomes are a necessary contribution for improving the representation of neglected lophotrochozoan phyla in genomic databases, and constitute a small step towards the attainment of a group-specific BUSCO set for Platyhelminthes in the future. Notably, Detr_v1 and Dxer_v1 were the assemblies presenting the largest number of scaffolds and lowest percentages of total assembly size comprised in the pseudochromosomes (Table 1). However, these results should be contextualized. Planarian flatworms are known to have a strong compositional bias with skewed AT-ratios (>70% A/T). This and the abundance of repetitive elements, as well as the known mosaicism in *Dugesia* (see mosaic-Meselson effect[33]) might explain the high scaffold count, probably due to failure of collapsing haplotigs in the final assemblies. Most notably, all *Dugesia* assemblies shared a similar high fraction of repeat content (Table 2) that substantially exceed previously reported values for *D. japonica* (~65%, Tian et al., 2022), *S. mediterranea* (~62% [43]), or *Schmidtea polychroa* (Schmidt, 1861) (~58% [56]). In all, these results complement the findings of Ivanković et al.[27], highlighting how planarian genomes are, for now, an assembly challenge. In contrast, if we compare the overall size of pseudochromosomes against the estimated GS of *D. etrusca* and *D. xeropotamica* based on flow cytometry estimates (Table 1) then the percentages of total assembly size contained within said chromosomes significantly improve (83.3% and

82.2% respectively, on par with Dhep_v1 results).

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

The availability of these new *Dugesia* assemblies offered us the opportunity to explore for the first time macrosynteny relationships between genera as well as assessing the origin of the peculiar haploid complement of *D. hepta* (n = 7). Our results showed the presence of translocations affecting chromosomes 5, 6, 7 and 8 between the Corso-Sardinian and the Iberia-Apennines-Alps clades, whereas the four larger chromosomes had a well conserved synteny. CRs had also played a relevant role in remodeling the *D. hepta*'s karyotype. We were able to detect translocations and pericentric inversions, most notably in chromosomes 2, 3, and 7. It is of special importance the origin of chromosome 2, as it is key to understand the reduction of the haploid set and, likely, the ultimate responsible for the isolation and speciation of *D. hepta.* CRs have been known since the early stages of genetics and their role in speciation has been lengthily studied and discussed[57,58]. The outcomes of CRs are diverse[57], and they will not always contribute to reinforcing the reproductive isolation between two populations or lineages, especially when small CRs occur[58]. However, the case of *D. hepta* offers little room for doubt. Changes in chromosome number can often lead to reproductive isolation owing to heterokaryotypic individuals experiencing reduced fertility (or overall reduced fitness) caused by meiotic segregation problems, and genomic analyses have demonstrated that there is no gene flow, past or present, between *D. hepta* and its congenerics[37], not even when co-occurring in the same localities. It is acknowledged that TEs may play an important role in the emergence of CRs (reviewed in [59]), however our results could not identify any TEs significantly enriched family in chromosome 2's synteny breakpoints. Furthermore, our relative TEs activity estimates (Figure 5) suggested that most TEs families' mobilization likely predated *D. hepta*'s speciation. Although LINE/R2 and SINE/Alu elements' activity periods could have been concomitant with the speciation, we reckon they may have had a trifling impact on the event itself owing to the small number of base pairs characterized and the absence of

enrichment signaling (Figure 3d; Figure 5c).

Genome size evolution in planarian flatworms

Little is known about the evolution of GSs in planarian flatworms. Recently, a study by Ivanković et al.[27] reported GS estimates of the four known 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). It was proposed that a large portion of said increase could be attributed to a TE expansion and an increased length of the protein-coding genes owing to transposon insertions that would have expanded intron regions. Such hypothesis would reasonably explain why despite the massive occurrence of syntenic block rearrangements within *Schmidtea*[27] the haploid complement (n = 4) is mostly well preserved between both clades[60]. However, a comparative analysis against *Dugesia* may offer us novel views on planarian GS evolution.

Many *Dugesia* species share a haploid complement number of n = 8 and some GS estimates of Western Mediterranean species suggest 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 in general, such as *Girardia* Ball, 1974 or *Recurva*, it is difficult to ascertain whether a gain in GS occurred in *Dugesia* or a reduction in *Schmidtea*. However, if we compare the available GS estimates for Dugesiidae at 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 in other studies[27], it could be suggested that a genome expansion occurred in *Dugesia* and possibly a genome reduction in the *mediterranea-polychroa* lineage. Bearing in mind the aforementioned differences between both genera in the haploid set, we contemplated the hypothesis of a WGD event as the underlying cause not only for the karyotypic differences but for the increase of GS in *Dugesia* as well, in contrast to the proposed

way of increase in *Schmidtea*. Yet, our results consistently dismissed this hypothesis. The distribution of K_s can serve as a proxy for the distribution of the divergence times of all duplication events that have left a trace in the genome of interest. A distribution pattern of exponential decay, as presented in our case, is expected under the assumption of a constant gene duplication rate and a constant rate of loss of duplicated copies[62]. However, a WGD event entails the duplication of all gene copies and a massive loss or divergence of them in a relatively short period of time, which would leave a peak signature of a large number of duplication events with similar divergence times. Likewise, collinearity, macrosynteny and gene family members' gain/loss analyses were in line with K_s results. WGDs are expected to leave large blocks of intragenomic collinearity[63], yet our results showed that collinearity was meaningless in our genomes. To put into perspective, values of collinearity in autopolyploids such as *Arabidopsis thaliana* may reach up to ~27%. Furthermore, our results did not highlight neither an increase of turnover rates at the base of *Dugesia* (Supplementary Table S4) nor multimatching syntenic blocks in *Schmidtea* (Figure 3b). Thus, it appears that *Dugesia* genomes grew larger for a different reason.

Further, when exploring the distribution of lengths of intronic and coding regions of SOs among our genomes and that of *S. mediterranea*, our results indicated that changes might be accumulating within non-coding regions of *Dugesia*, as observed in *Schmidtea*. In fact, our results showed an expansion of multiple TE families in *Dugesia*, as reported in the *lugubris-nova* lineage[27]. However, upon detailed inspection we quickly realized that different TE families had expanded in *Dugesia*. Most notably, DNA/Maverick and DNA/DDE families.

Indeed, TEs expansions seem to have contributed to the enlargement of *Dugesia* genomes, but the factors underlying such evolution may not be shared among our species. In fact, the evolutionary context of the Corso-Sardinian clade is conditioned by the insularity of its populations, and interestingly enough the largest GS estimates of *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[59,64]. According to the mutational hazard hypothesis (see[65]), the fate of DNA gains, among which TEs expansions, is governed by the mutation rate and the effective population size (N_e), although the relevance of this hypothesis is still hotly debated[66,67]. Indeed, in insular systems, smaller N_e are expected, leading to genome-wide relaxation of selective constraints (i.e., reduced efficacy of selection and predominance of genetic drift). This can have an impact in TE dynamics, which in turn could lead to an increase in GS due to the accumulation of slightly deleterious TEs since purging selection cannot act efficiently. *Dugesia hepta* and *D. xeropotamica* are both endemics from Sardinia and Corsica with very restricted ranges of distribution. Coincidentally, both species present the largest genomes according to flow cytometry estimates (Table 1) and, incidentally, the largest estimate of a *Dugesia* genome (that we are aware of) belongs to a Sardinian endemic as well, *D. leporii* (2.38 Gb). The evolutionary history of these species is tightly linked to some of the most drastic changes in paleoenvironmental conditions that affected the Western Mediterranean[37]. It is reasonable to assume that both species experienced fluctuations in N_e, with bottlenecks likely occurring during the Pleistocene glacial phases on both islands[68,69]. In fact, the comparison of 4d distances of *Dugesia* did not show significant differences among species or clades (Supplementary Fig. S8), indicating that the neutral rate of evolution is overall similar in all four species. This, in turn, reinforces the idea that differences in GS might be conditioned by the historical N_e. Such fluctuations in the insular lineages should have an impact on gene family turnover rates, and in fact, our results show a 1.3~2.5-fold and a 2.6~3.8-fold increase of both birth and death turnover rates respectively in *D. hepta* and *D. xeropotamica* when compared to the continental species (Supplementary Table S4).

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

A further interesting result is the discrepancies we found in structural genome stability when comparing *Dugesia* and the findings reported for *Schmidtea* (cf. Ivanković et al.[27]). Our results showed that structural genome rearrangements were surprisingly

common between genera, resulting in a pattern of total genome scrambling. Macrosynteny analysis of *Dugesia* assemblies showed that two major fusion events were responsible for *D. hepta*'s haploid chromosome complement (n = 7), as well as some translocations affecting large fragments of four chromosomes between the two major clades, a phenomenon that has been reported in other *Dugesia* species[70,71,72,73] and is also frequent in fissiparous individuals[74]. Barring that, macrosynteny was generally better conserved across the four *Dugesia* species despite the considerable divergence time between their respective lineages (~30 Mya[26]; Figure 1) than it was among the *Schmidtea* species. Based on the observations in *Schmidtea* genomes, it has been proposed that structural genome-wide instability is a general feature of flatworm genome evolution[27]. Yet, our observations in the four *Dugesia* genomes would seem to indicate that the situation may be different for this genus. However, there are no dedicated dating estimates of divergence times among *Schmidtea* species. Therefore, it is unclear if their pervasive CRs are characteristic of the genus or just a common phenomenon in Dugesiidae or among flatworm lineages that diverged a long (long) time ago. It would be most interesting to sequence chromosome-scale genomes of Malagasy *Dugesia*, the first clade to diverge from the rest of genus more than 100 Mya[26] and displaying haploid chromosome set of n = 5, to assess whether synteny is only preserved at 'shallow' time scales in planarian flatworms. Nevertheless, we have also found the absence of synteny with the MALGs in *Dugesia* (Figure 4), which may support the hypothesis of the structural genome-wide instability of flatworm genome evolution proposed by Ivankovic et al.[27].

Our findings, along with those of Ivanković et al.[27], raise the question of what explains such extreme genome scrambling in flatworms. One potential mechanism comes from the recent observation that *S. mediterranea* exhibits tight associations between regulatory elements and their target genes based on proximity. Such a gene-centric regulatory structure could help prevent possible detrimental consequences of pervasive CRs and explain why they are so frequent in the wild in fissiparous populations[74].

A similar link between tight gene-regulation associations and extensive genomic rearrangements has been proposed in tunicates and clitellates[21,75]. Furthermore, while high GC content has been shown to prevent double-strand breaks in prokaryotes[76], it is also positively correlated with higher mutation rates[77]. Given their fairly low GC content[27,56,78] (Table 1), planarians might be prone to CRs, but they have likely evolved a genomic architecture that mitigates the negative consequences of this scrambling and the non-synonymous mutations arising from their constant regeneration and cell turnover.

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 to DNA extractions were frozen with liquid nitrogen and stored at -80 ºC before being sent to Dovetail Genomics (Scotts Valley, CA). Final species for genome sequencing and assembly were determined based on overall extraction yield and quality.

Flow cytometry

GS estimation was carried out by means of flow cytometry readings. 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 rinsed using a mixture of distilled water and tap water (1:1) for 1 minute 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[79] 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 number were assessed with Qubit and 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 reads.

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 DNA extraction protocol detailed in Zhang and Alvarado[80]. Genomic sequence data was obtained using a single library per species. We used PacBio CLR (*D. hepta, D. etrusca*) and PacBio HiFi (*D. liguriensis, D. xeropotamica*) libraries. Next, Omni-C libraries (one library per each species) were build, followed by Illumina 150-bp paired-end (HiSeq X platform) sequencing to perform the chromosome-level scaffolding using the HiRise software[81]. We determined

the completeness of the new assemblies by applying the pipeline of Benchmarking Universal Single Copy Orthologues (BUSCO; v.5.4.2[42]), 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[82,83] 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 pertinent scaffold. We filtered out potential contaminant sequences and/or errors present in the scaffolds using different criteria. We mapped our genome assemblies against the nucleotide database of the NCBI using BLAST+ blastn and Blobtools v.3.4.0, as implemented in BlobToolKit[84] and discarded any scaffold with bacterial, virus or fungal hits. Thereafter, we calculated the mean depth of coverage for the remaining scaffolds and discarded those that fell outside the 95% confidence interval and did not exhibit BUSCO hits.

Identification of repetitive elements

We analysed repetitive regions of the genome assemblies using a combination of de novo transposable elements (TEs) identification with RepeatModeler v2.0.3[85] and DeepTE[86], and a similarity-based search using RepearMasker v.4.1.2[87]. We built a database of repetitive elements (RE) of 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 metazoans model (-sp M). Lastly, all classified RE were concatenated into a single database and used to soft-mask the repetitive sequences of each genome assembly using the –xsmall option of RepeatMasker. Additionally, we performed de novo repeat identification with *D. hepta* using EarlGrey[53] to estimate the repeat landscapes of the major families to assess the relative TE activity on our genomes.

4-fold degenerate sites analysis

We used phyloFit to compute a neutral model of evolution from four-fold degenerate (4d) sites[54]. We extracted these sites from the coding sequences (CDS) of single-copy *Dugesia* orthologues identified by Broccoli in combination with BRAKER's results. Each orthogroup was aligned with TranslatorX[88], selecting its implementation of MAFFT[89] to maintain a correct reading frame and ensure positional homology of all nucleotides. Finally, we used phyloFit with the default REV substitution model to estimate the neutral model from the 4d sites across all alignments and represented the distribution of genetic distances using *ggplot2*.

Structural and functional genome annotation

Structural annotation of the soft-masked genome assemblies was performed with BRAKER2[90,91,92,93] using RNAseq data of each species, which had been previously mapped to their respective assemblies using STAR v.2.7.10b[94]. To functionally annotate gene models, we used BLASTP v.2.2 (E-value = 10^{-3}) searches against Swiss-Prot. Furthermore, we also conducted searches against the eggNOG database using eggNOG-mapper v.2.1.9[95,96,97] and searched the predicted peptides for specific protein domain signatures within InterProScan v.5.56.89[98]. All functional evidence was integrated. Concurrently, we detected the transfer RNA genes (tRNAs) encoded in the genomic sequences using the tRNAscan-SE v2.0.7 software[99].

Detection of whole-genome duplication

We constructed the *K*s distribution to evaluate the possibility of a whole-genome duplication (WGD) event as the underlying cause for the conspicuous difference of GS between *Dugesia* and *S. mediterranea*. First, using wgd v2[45], we inferred the paralogous gene families of each *Dugesia* species using the coding sequences of each assembly (wgd dmd). Thereafter, the *K*s distribution was inferred (wgd ksd) and represented (wgd viz). Furthermore, we used BLASTP v.2.2 (E-value = 10^{-10} and top 5 matches) to search all potential homologous gene pairs of protein sequences within each species and identified the collinear gene pairs with MCScanX[46].

Intron and CDS lengths

We estimated the distribution of intron and CDS lengths of *Dugesia* and *S. mediterranea* using 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). We inferred orthologous groups (OG) among species using Broccoli v.1.2[47] and selected all single copy orthologues (1:1:1:1:1). We visualized intron and CDS length distributions using *ggplot2* and compared the results for each *Dugesia* species against *S. mediterranea.*

Macrosynteny analysis

To explore the possible chromosomal rearrangement occurred in *D. hepta*, we used the R package GENESPACE v.1.2.3[48], which uses a combination of OrthoFinder[100] 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 used the ODP tool[101] to determine whether ancestral metazoan linkage groups (MALGs), highly conserved linkage groups found in bilaterians, cnidarians, and sponges[11], were preserved in *Dugesia* genomes. This tool employs hidden Markov models to identify homologs of MALG proteins and then applies Fisher's exact test to detect their significant enrichment on specific chromosomes.

TEs enrichment analysis at syntenic breakpoints

We visualized the density distribution of TE within the chromosome 2 of *D. hepta* using the R package *karyoploteR*[51] to explore potentially enriched regions that could have been involved in the rearrangements detected in the macrosynteny analysis. In addition, we tested if specific TE families were overrepresented nearby synteny breakpoints along the same chromosome using the R package *GenomicRanges* v.1.46.1[52]. Based on GENESPACE syntenic blocks delineation on the chromosome of interest and RepeatMasker's annotation, we established flanking windows of 20kb surrounding the breakpoints and counted all elements present in said regions. To assess enrichment,

we created an expected null distribution of repetitive sequences by generating an equal number of randomly distributed windows along the chromosome and performed 1000 permutation iterations. Thereafter, we performed a two-tailed test to assess whether there was statistically significant enrichment or depletion of TE families in the established flanking windows with respect to the expected null distribution. To account for the large number of TE families being tested simultaneously, we applied two layers of filtering and correction: (i) p-values were only calculated for TE families with an average expected count of 10 or more sequences and (ii) the resulting raw p-values were adjusted for multiple testing using the Benjamini-Hochberg method for false discovery rate.

Estimation of gene gain and loss

We reconstructed gene gains and losses per branch per million years using BadiRate v.1.35[55] under the birth, death and innovation model (BDI) using a maximum likelihood (ML) method to estimate the turnover rates. Furthermore, we tested multiple branching models to estimate the best-fitting one by calculating the Akaike Information Creterion (AIC) values under each branching model. The following models were considered: (i) global rates model (GR), in which all branches evolve under the same turnover rates; (ii) free rates model (FR), which assumes all branches evolve under differing rates and (iii) branch-specific model (BR), that assumes that a set of branches share a different turnover rates from the background (in our case, the branches leading to *Dugesia*). We used the OG counts table (obtained from Broccoli) and a time-calibrated tree based on the set of SOs. Since BadiRate is prone to crash with highly heterogenous gene counts, we ran BadiRate separately for the most heterogenous OGs (i.e., only 0.02%) applying the best-fitting estimated branching model but using Wagner parsimony instead to estimate turnover rates, gain and loss of genes. The time-calibrated tree was inferred with IQ-TREE v2.4.0[102] using an admixture model (LG+C60+F+G+R3) and subsequently calibrated using the function *chronos* from the R package *ape*[103]. 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 advancements. *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. Ecol. 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. 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

39. 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

40. 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

41. 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

42. 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

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

44. 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

45. 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).

46. 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

47. 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

48. 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

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

50. 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. 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

58. 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

59. 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

60. 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

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. 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

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

78. 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

79. 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

80. 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).

81. 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

82. 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

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

84. Challis, R. et al. BlobToolKit – Interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics* **10**(4), 1361-1374 (2020). https://doi.org/10.1534/g3.119.400908

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

86. 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

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

88. 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

89. 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

90. 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

91. 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

92. 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

93. 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

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

95. 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

96. 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

97. 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

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

99. 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

100. 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

101. 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

102. 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

103. 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/bioinformatics/bty633.

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

D.D.S., I.T.A., 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 extraction 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 visualisation. 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.

Figures and tables

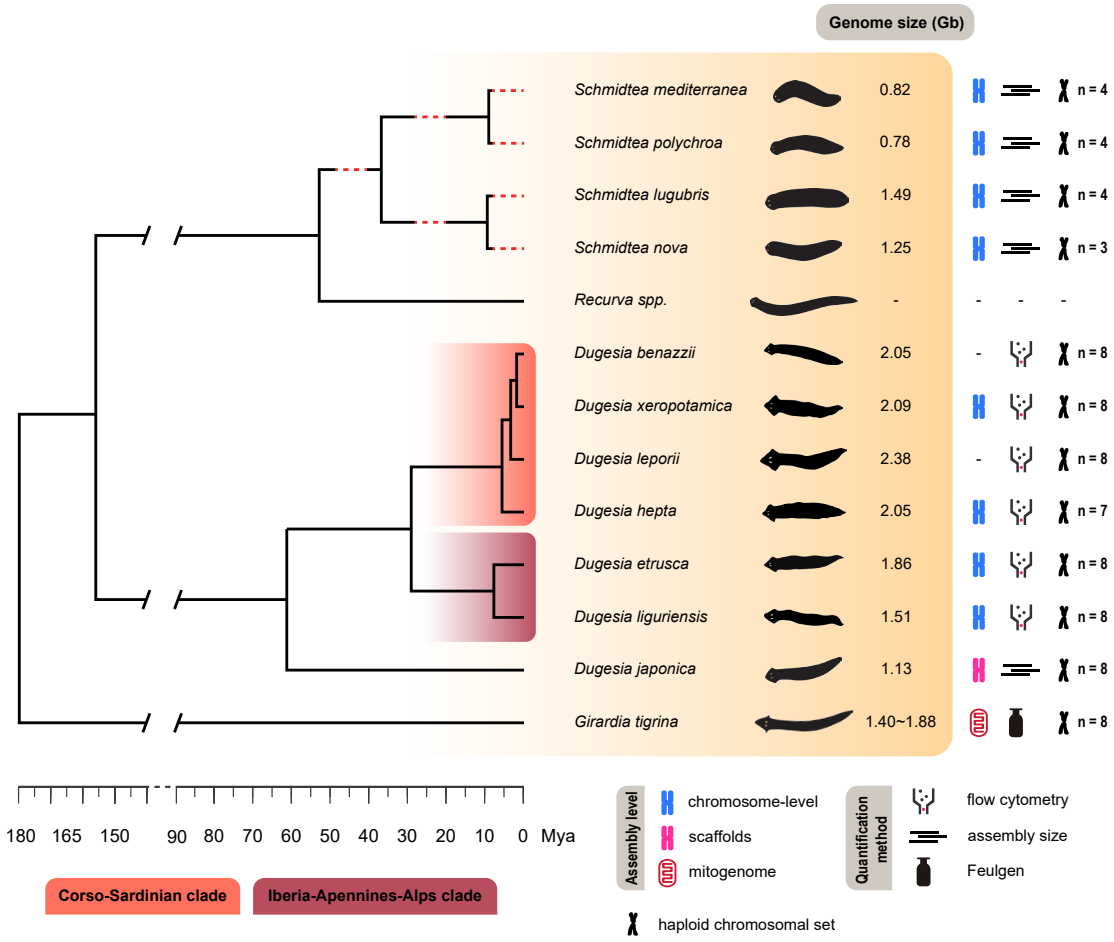


Figure 1. Schematic time-tree showing the phylogenetic relationships between species of *Dugesia*, *Girardia*, *Recurva*, and *Schmidtea* based on Leria et al.[60] and Solà et al.[26]. Red dotted lines in tree branches indicate that time estimate to the most recent common ancestor (MRCA) is unknown.

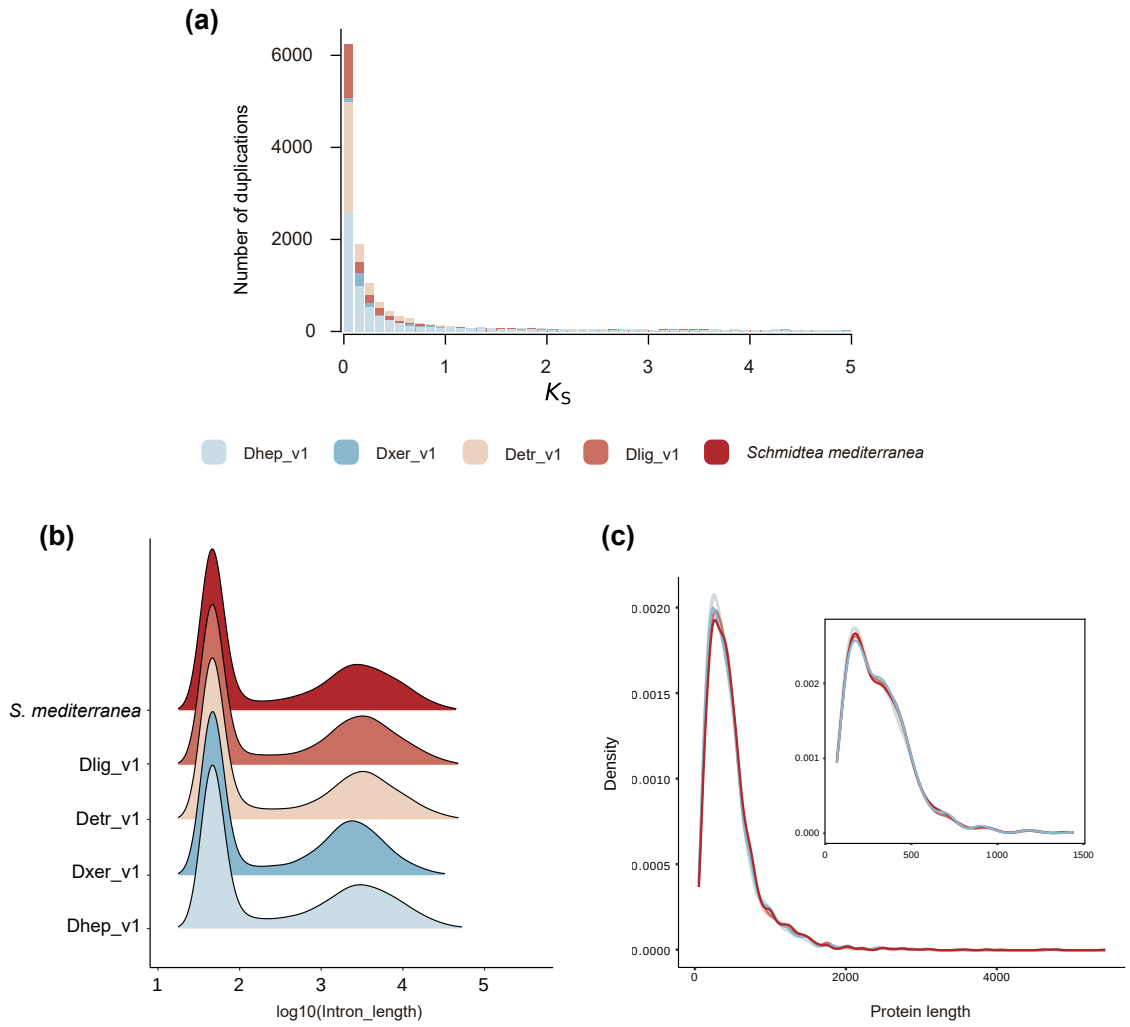


Figure 2. (a) Overlayed paranome K_s distributions of the four *Dugesia* genome assemblies showing a pattern of exponential decay. **(b)** Ridgeline plots showing the density distribution of intron lengths of all introns comprised in the 2266 single-copy orthogroups shared between *D. hepta* (Dhep_v1), *D. etrusca* (Detr_v1), *D. liguriensis* (Dlig_v1), *D. xeropotamica* (Dxer_v1), and *S. mediterranea*. **(c)** Density distribution of CDS lengths of all 2266 single-copy orthologues. Inset shows the same distribution for exon-equivalent orthologues, 571 in total.

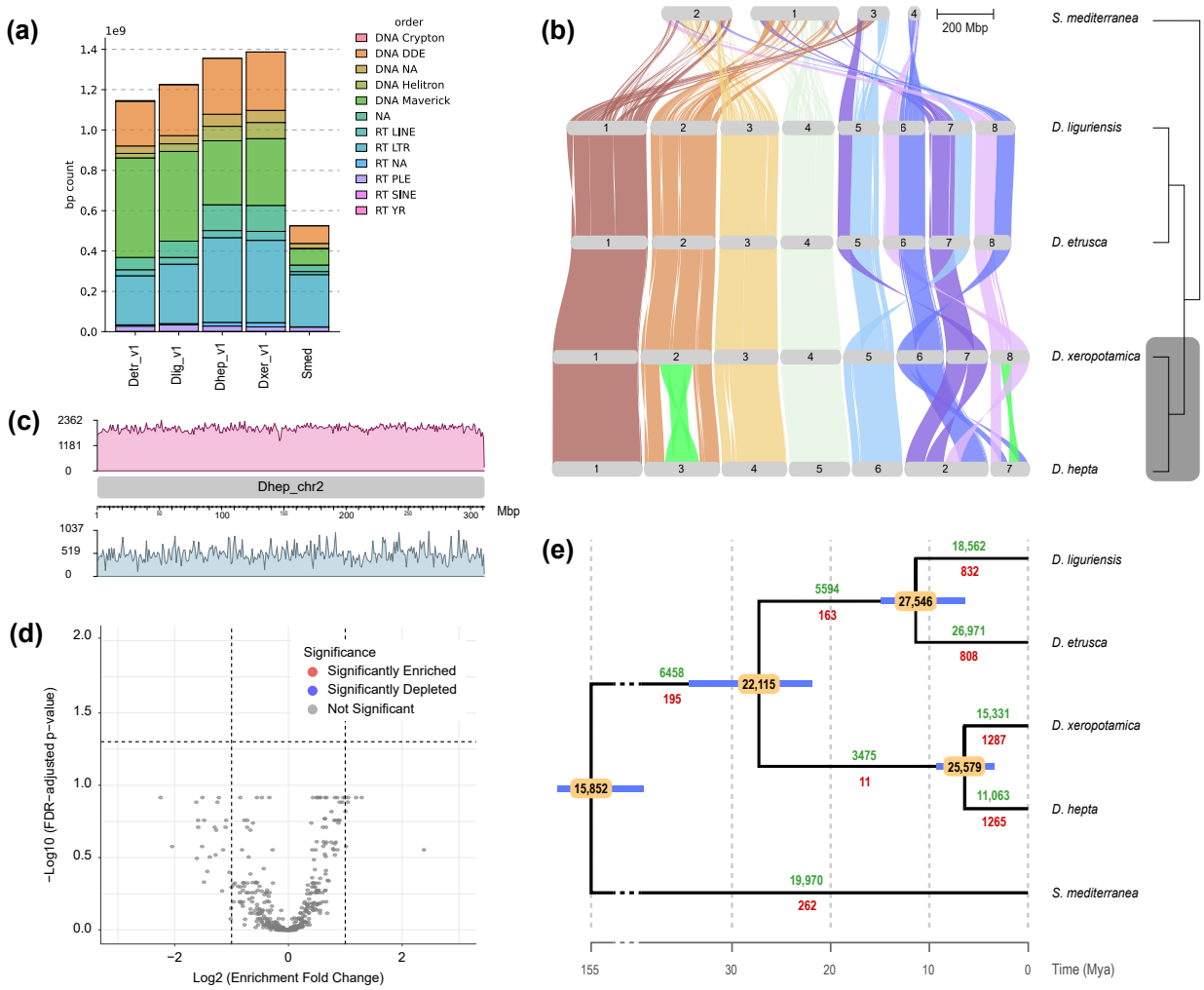


Figure 3. (a) Stacked histograms showing the total amount (bp) of repeat content of each genome assembly classified into TEs orders. **(b)** Synteny analysis between the four *Dugesia* species and *S. mediterranea*. Ribbon coloring on the basis of *D. xeropotamica* (n = 8). Pericentric inversions affecting chromosomes 2 and 7 are highlighted in green. To the right, a cladogram depicting phylogenetic relationships. Highlight in dark grey identifies the Corso-Sardinian clade. **(c)** Density plot of TEs (pink) and genes (cyan) along the chromosomes 2 of *D. hepta*. **(d)** Vulcan plot with enrichment analysis results of 20 kb windows flanking synteny breakpoints across *D. hepta*'s chromosome 2. Horizontal dotted line indicates the threshold of statistical significance. 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 number of genes are indicated within yellow bubbles on top of the tree nodes. Blue bars represent the interval of maximum and minimum time estimates of each node.

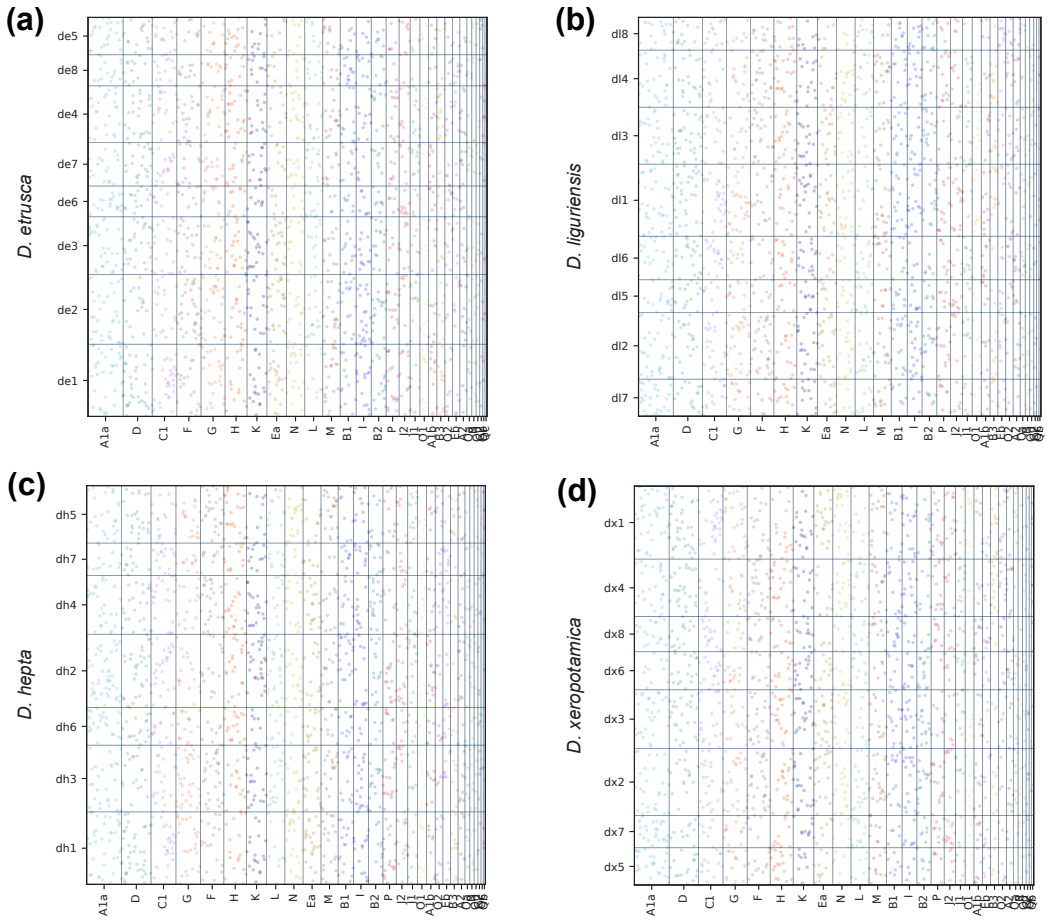


Figure 4. Dotplots between metazoan ancestral linkage groups (MALGs) and **(a)** *D. etrusca*, **(b)** *D. liguriensis*, **(c)** *D. hepta* and **(d)** *D. xeropotamica*. MALGs without enrichment are plotted in light colors. No enrichment for any MALG was detected in *Dugesia.*

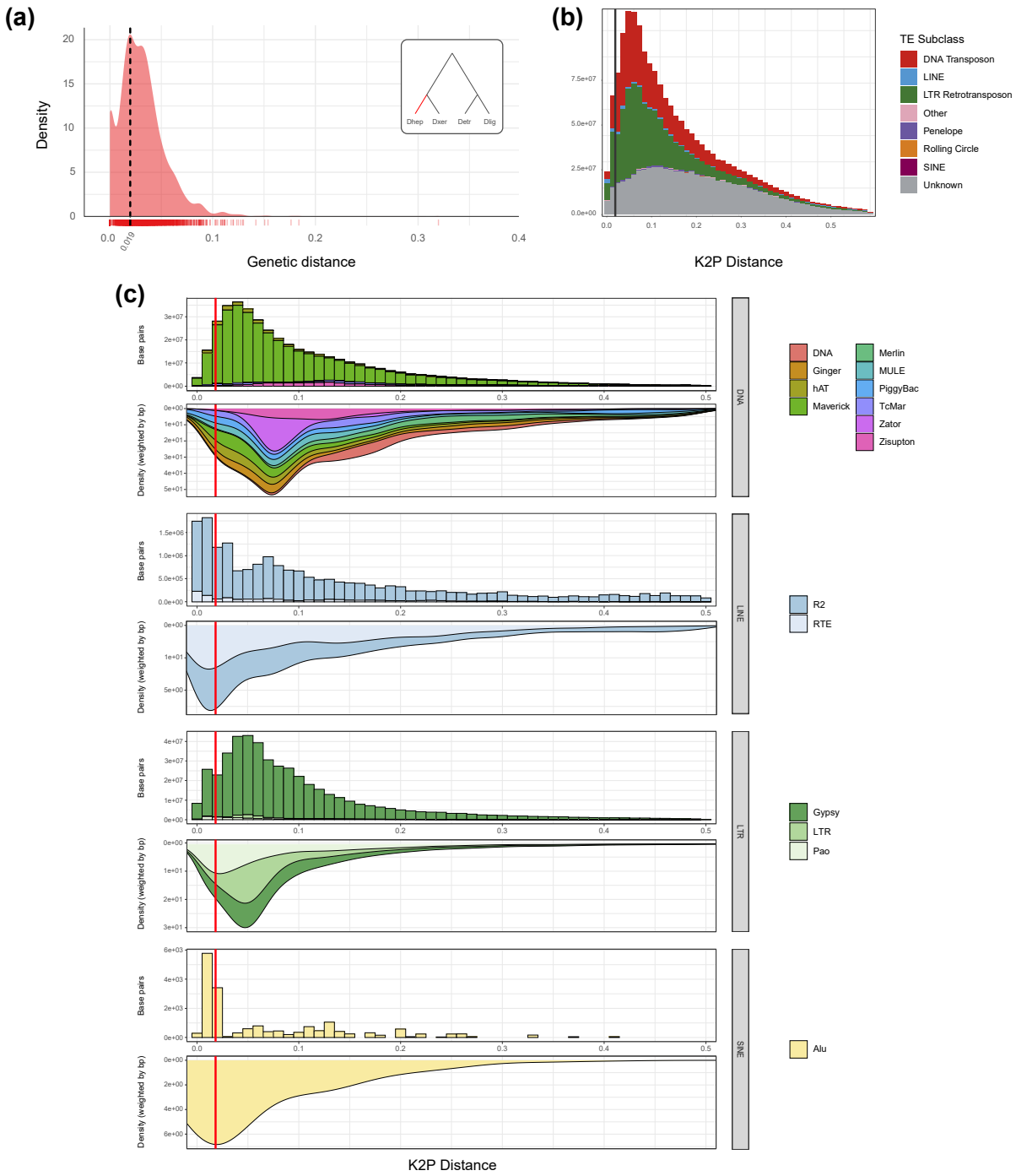


Figure 5. (a) Density distribution of four-fold degenerate (4d) genetic distances of *D. hepta*. Dotted vertical line indicates the mode value of the distribution. **(b)** Repeat landscape plot summarizing relative TEs activity in *D. hepta* using Kimura 2-parameter (K2P) divergence. Recent activity to the left-hand side of the plot. Solid vertical line represents the 4d divergence mode value for comparison. **(c)** Repeat landscapes of DNA, LINE, LTR, and SINE superfamilies in *D. hepta*. Upper plots represent cumulative amount of base pairs grouped by divergence and family (color-coded legends on the right-hand side). Inverted plots show density distribution of those same families. Vertical red lines show the mode value of 4d divergence for comparison.

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%)

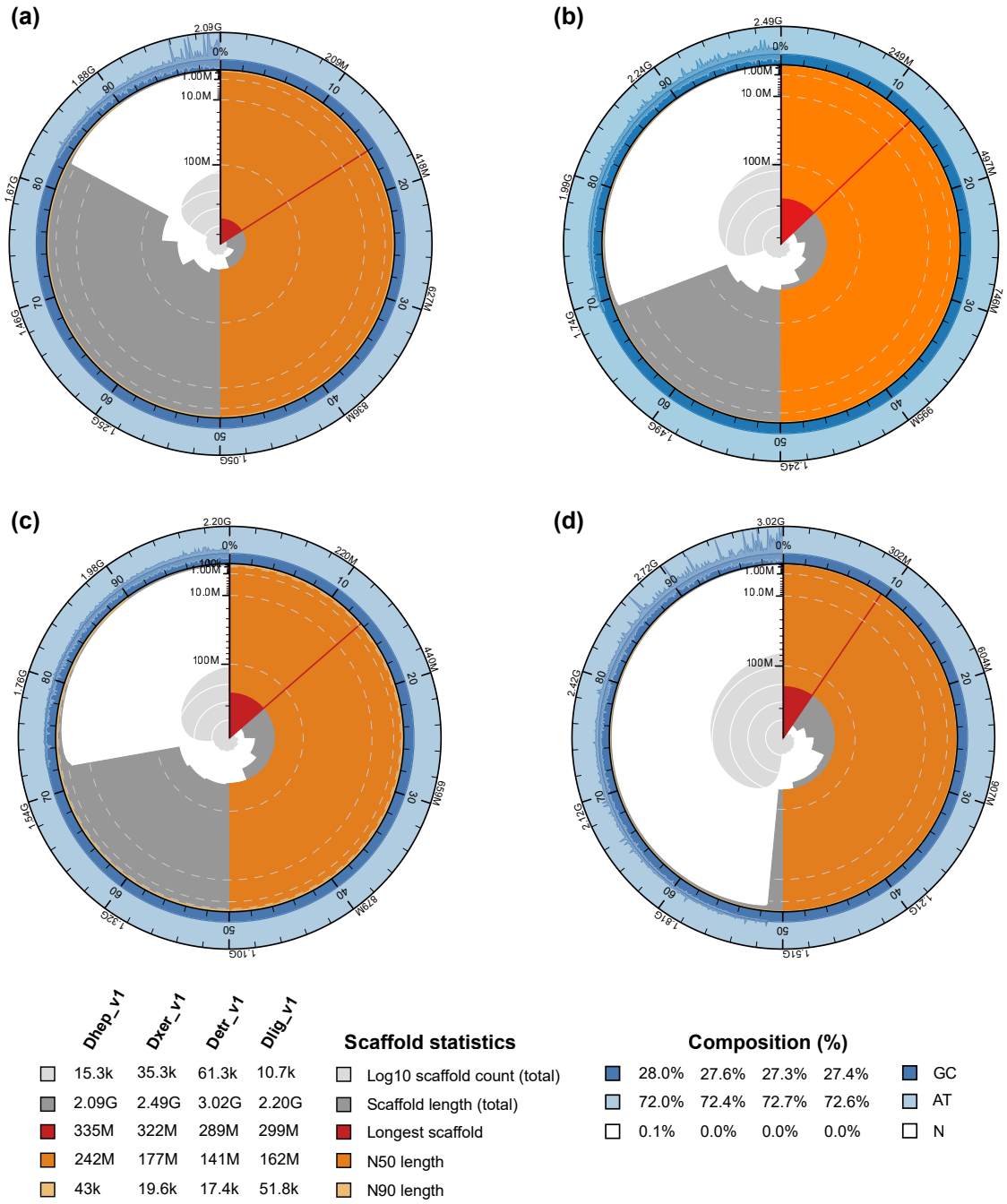
Table 2. Genome annotation statistics.

	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%)

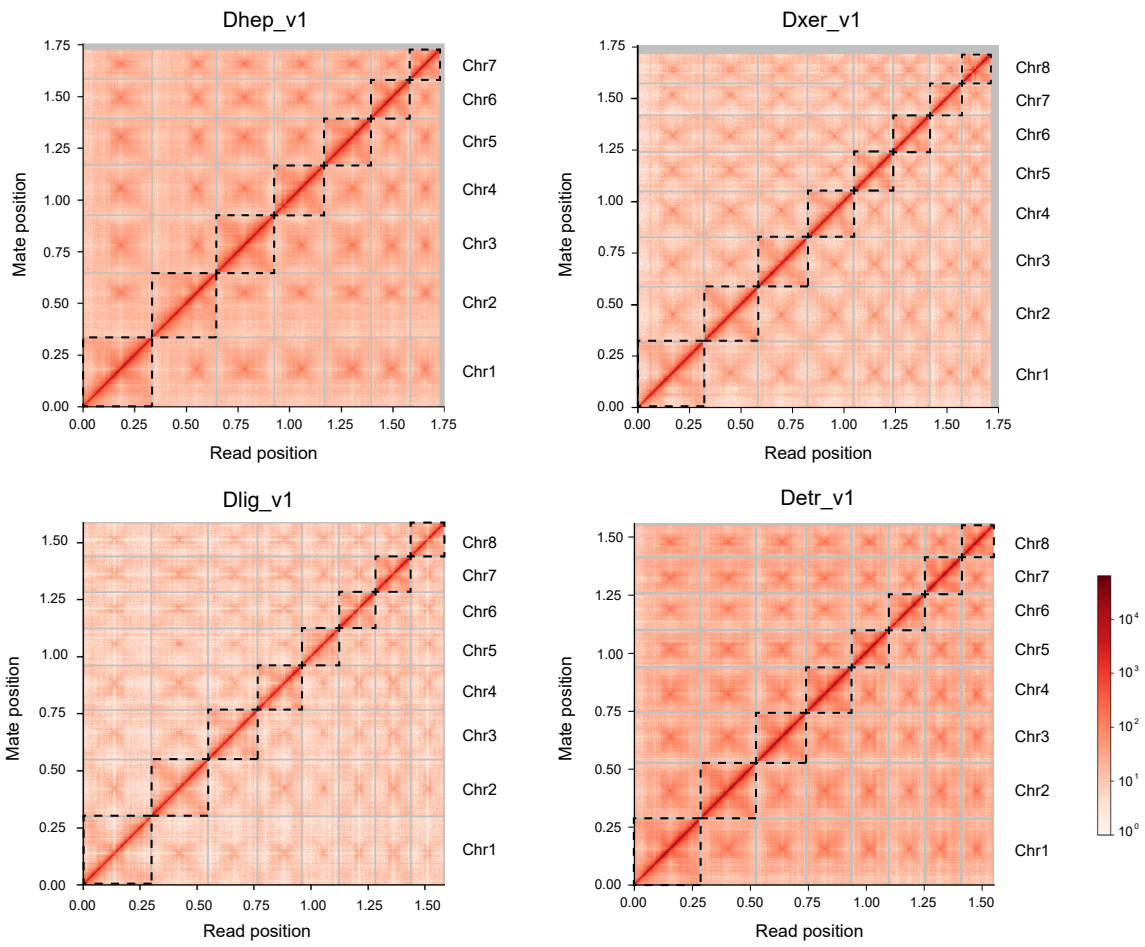
Table 3. Summary length (bp) statistics of introns and protein-coding sequences (CDS) of single-copy orthologues.

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

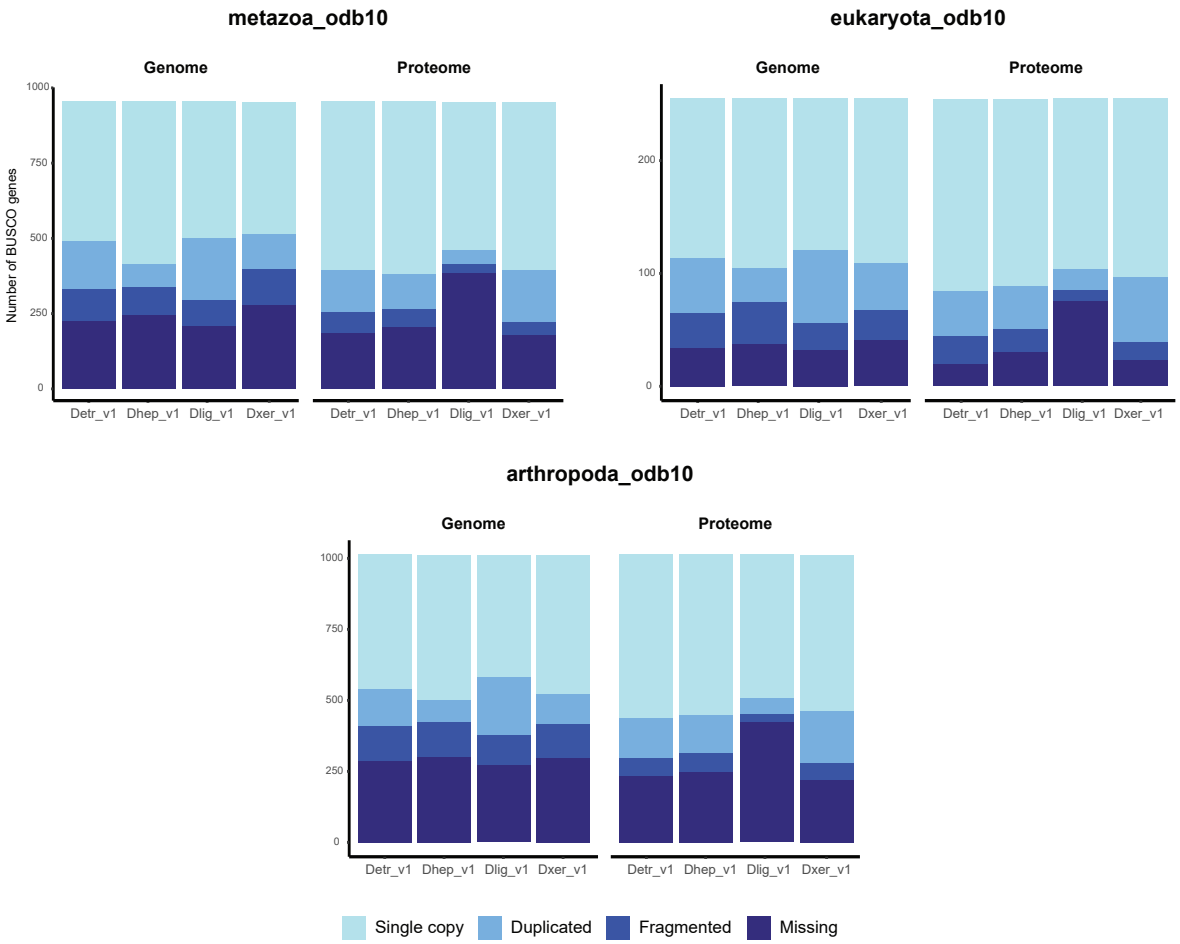
Supplementary Information: Figures



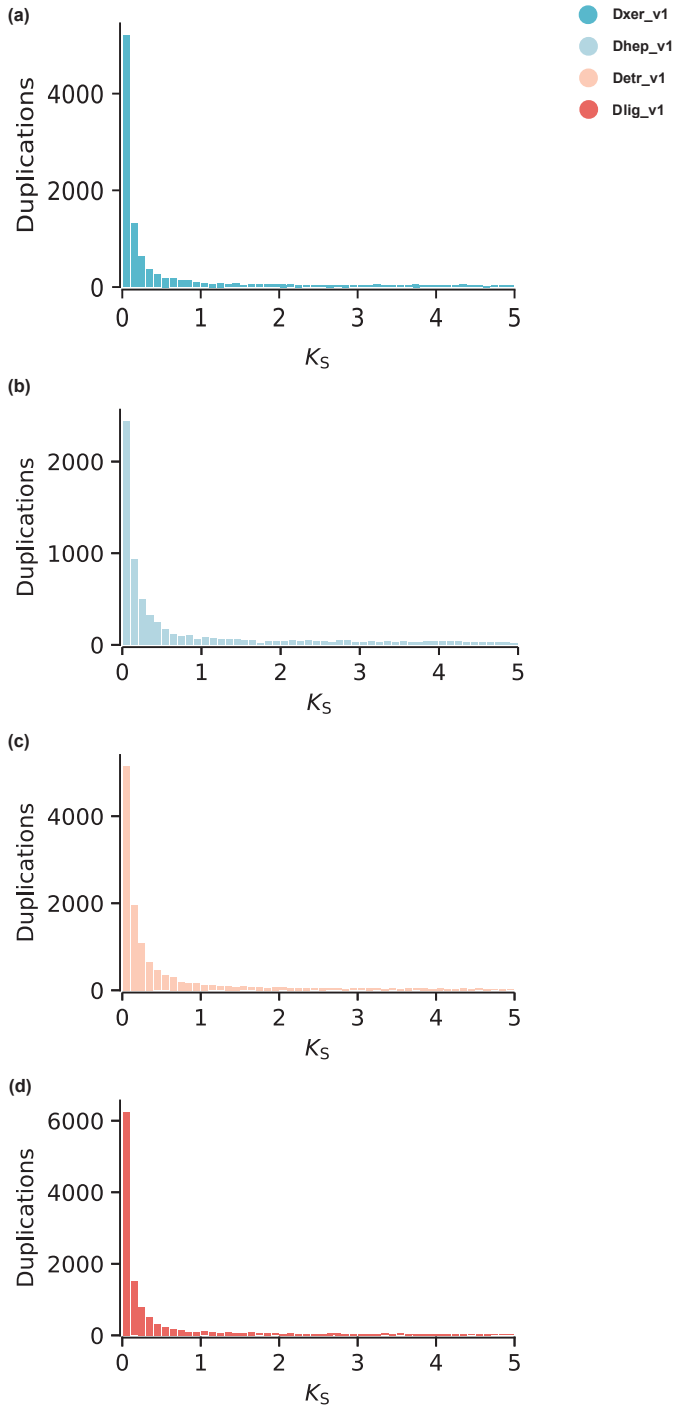
Supplementary Fig. S1. Snail-plots summarizing genome assembly statistics for (a) *D. hepta*, (b) *D. xeropotamica*, (c) *D. liguriensis*, and (d) *D. etrusca*.



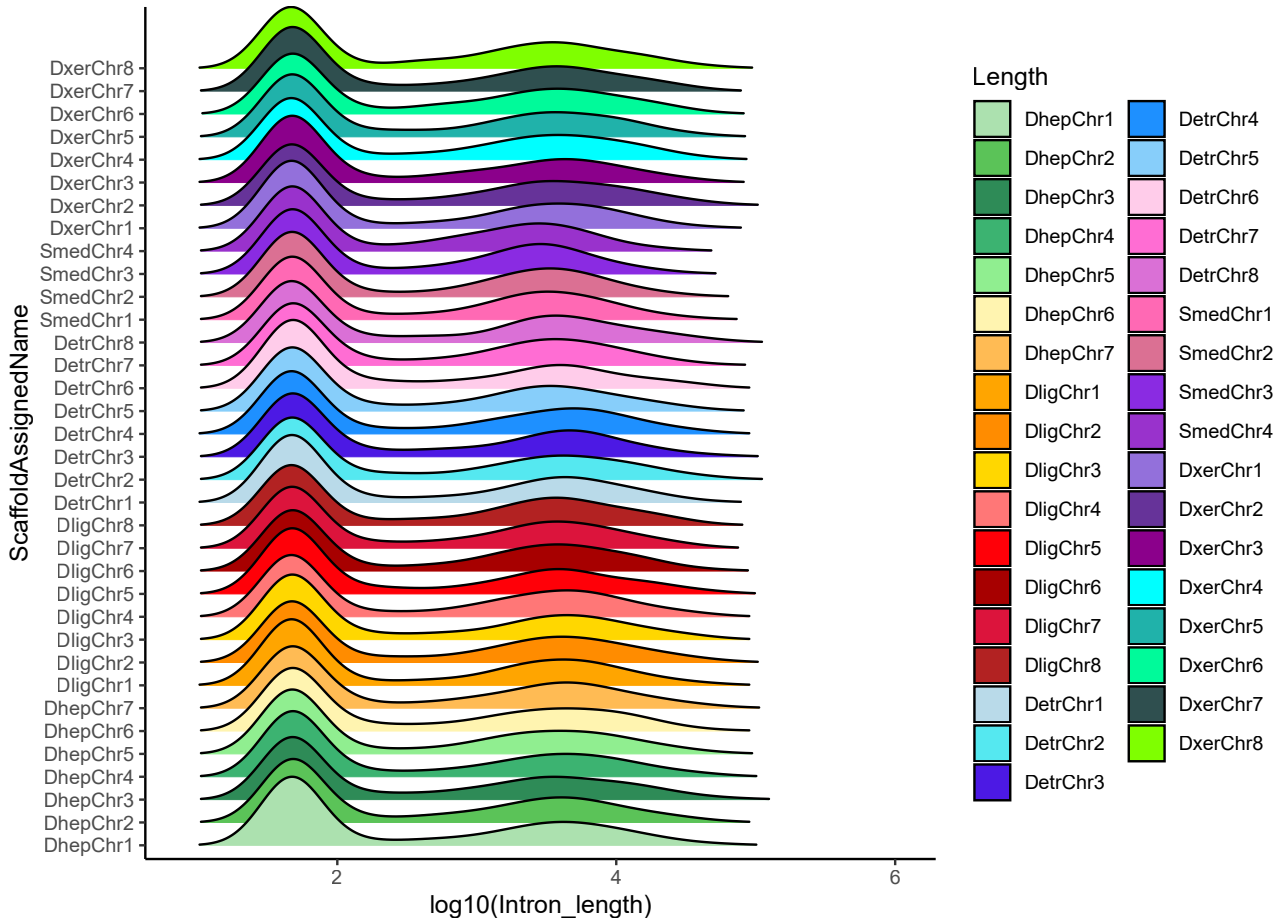
Supplementary Fig. S2. Long-range contact heat-map of paired Omni-C reads depicting the major scaffolds of



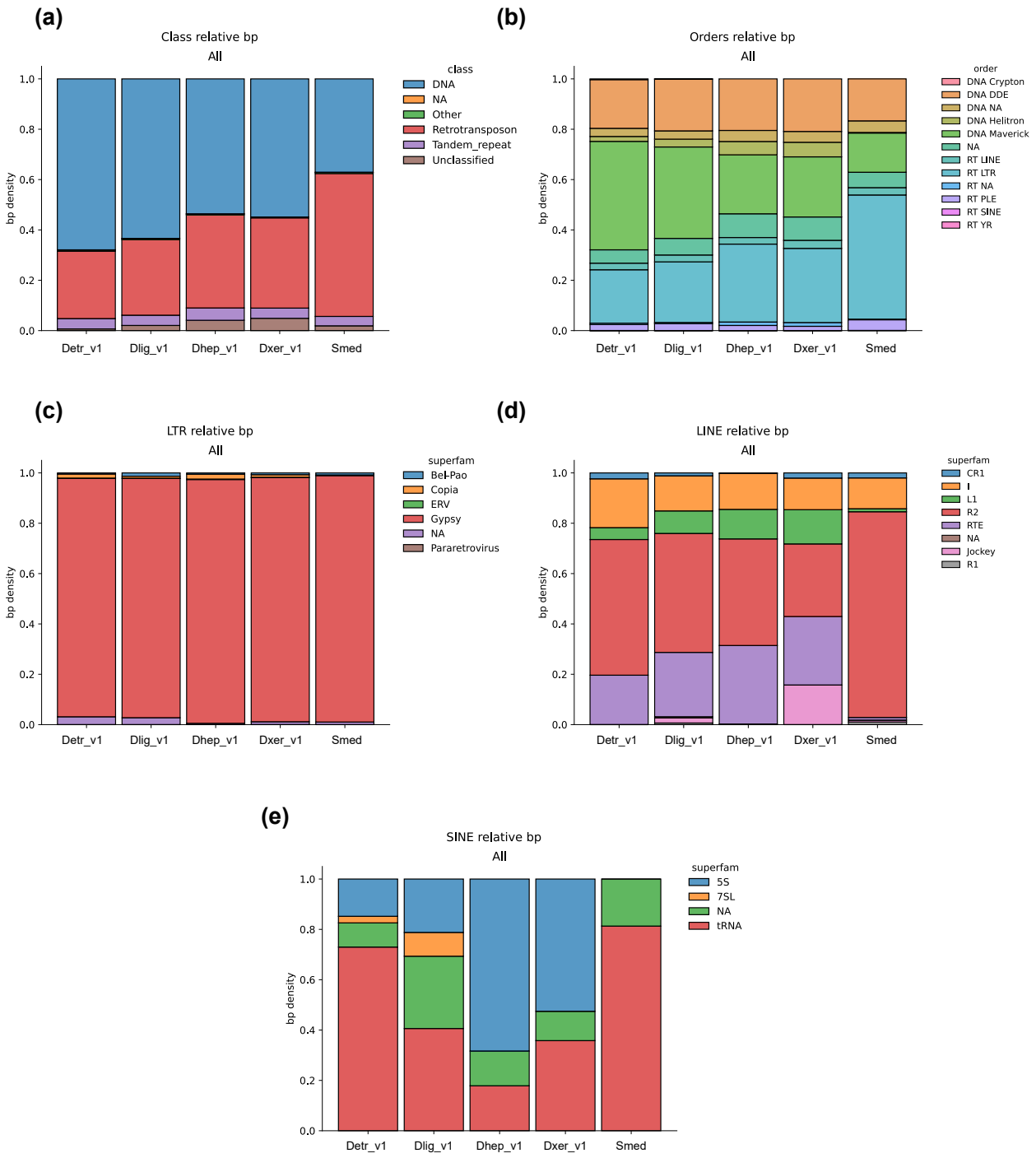
Supplementary Fig. S3. BUSCO assessment of the genome assemblies before (Genome) and after (Proteome) functional annotations.



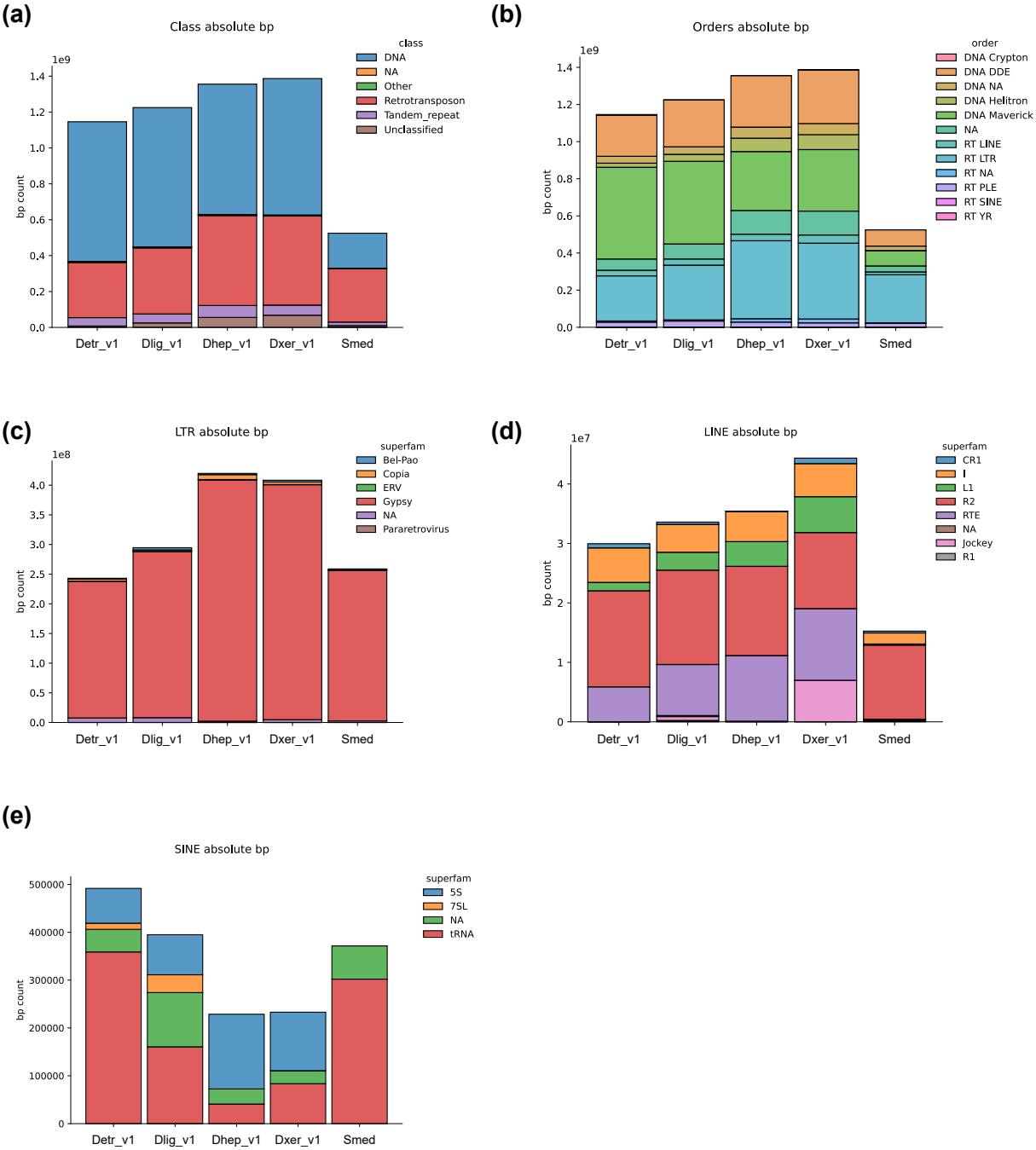
Supplementary Fig. S4. Paranome K_s distributions of the *Dugesia* genome assemblies showing a pattern of exponential decay.



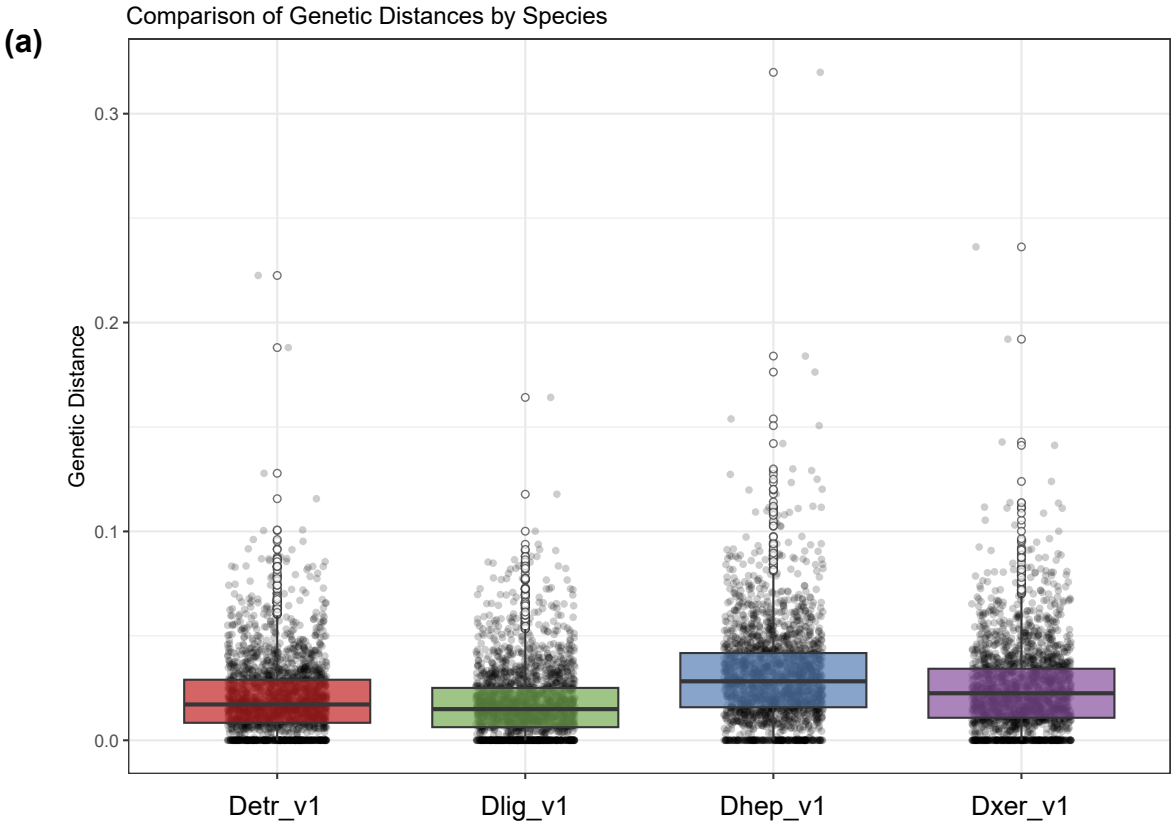
Supplementary Fig. S5. Ridgeline plots showing the density distribution of intron lengths of all introns comprised in the 2266 single-copy orthogroups shared between *S. mediterranea* (Smed), *D. hepta* (Dhep), *D. xeropotamica* (Dxer), *D. etrusca* (Detr) and *D. liguriensis* (Dlig) separated by chromosomes.

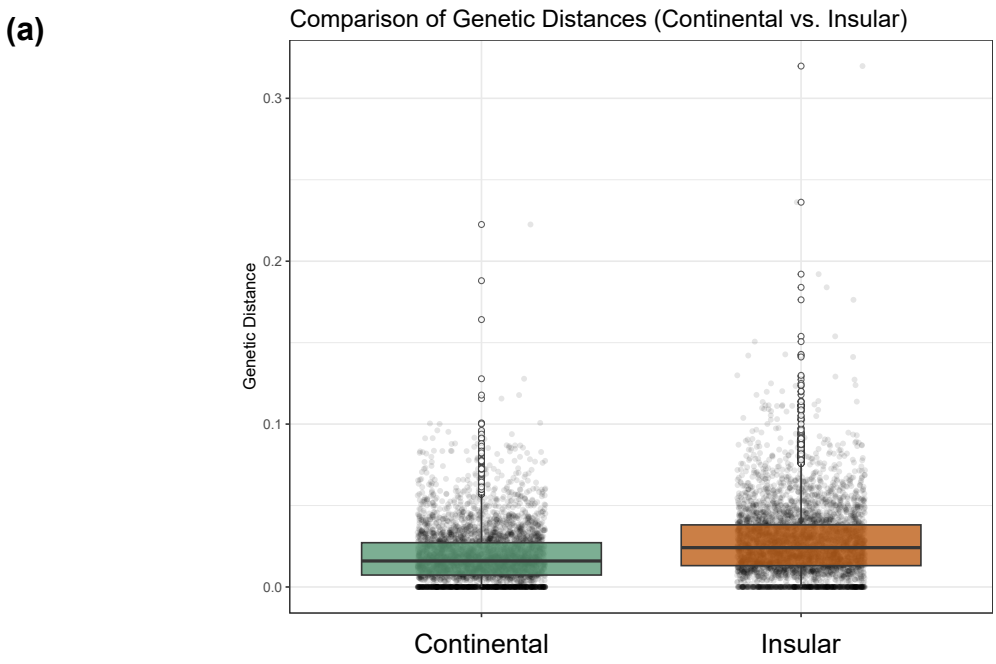


Supplementary Fig. S6. Comparison of the relative contributions (%) of each class (a) and order (b) of repetitive elements to the repeat content of each genome assembly. (c) Relative superfamilies contributions of repetitive elements to the LTR retrotransposon elements. Relative superfamilies contributions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