Migrators Within Migrators: Exploring Transposable Element Landscapes, Activity, and Host Gene Associations in the Monarch Butterfly, *Danaus plexippus*

Tobias Baril[1,2]* and Alexander Hayward[1,3]*

[1] Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Cornwall, TR10 9FE, UK

[2] ORCID: 0000-0002-5936-7531

[3] ORCID: 0000-0001-7413-718X

* Corresponding Author information:

Tobias Baril: tb529@exeter.ac.uk

Dr Alexander Hayward: alex.hayward@exeter.ac.uk

Abstract

Background

Lepidoptera (butterflies and moths) are an important model system in ecology and evolution. A high-quality chromosomal genome assembly of the monarch butterfly (*Danaus plexippus*), famous for its North American migration, is available but lacks an in-depth transposable element (TE) annotation. This provides an opportunity to explore host-TE interactions and the impact TEs have in shaping the monarch genome.

Results

We find that 6.47% of the monarch genome is comprised of TEs, a reduction of 6.59% compared to the original TE annotation performed on the draft genome assembly. TE content is low compared to two closely related species, *Danaus chrysippus* (26.70%) and *Danaus melanippus* (11.87%). The biggest contributors to genome size in the monarch are LINEs and *Penelope*-like elements, with 37.7% of TE content is contributed by five newly identified families (two LINE, two *Penelope*-like, and one SINE). There are several recent peaks of TE activity in the monarch, with little evidence for more ancient peaks of activity. LINE fragments demonstrate signatures of genomic deletions as reported by studies on *Heliconius* butterflies, indicating a high rate of TE turnover. Given previous associations reported in other species, we investigated the association of TEs with wing colouration and immune genes. We find a single unclassified element 7kb upstream of the myosin gene locus, associated with wing colouration, and a significant association of TEs with immune genes in comparison to other gene types.

Conclusions

We provide an in-depth TE annotation and analysis of TE diversity and evolution for the monarch genome. We identify highly successful novel DNA TE families, mirroring the activity

profile of the most successful LINEs. We also find evidence of ongoing TE expansion and removal in the monarch, highlighting the dynamic nature of repeat content in genomes over time. We identify immune genes as suitable candidates for interrogation of host adaptation facilitated by TEs. Further in-depth comparative studies across closely related species will be beneficial to our understanding of the evolutionary dynamics of TEs and the processes leading to their contrasting distributions among host genomes.

Keywords

Transposon; TE Annotation; Lepidoptera; *Danaus plexippus*; Butterfly; Genomic Deletion; TE-Host Dynamics; Genome Evolution

Background

Transposable elements (TEs) are autonomous DNA sequences that can move within the genome. TEs are present in nearly all eukaryotes and are important in shaping their genomes [1–8]. TEs also represent a large reservoir of sequences with the potential to contribute to host genomic novelty, and they have been implicated in the evolution of regulatory networks [3,9–13], chromosomal rearrangements [3,14–19], exon shuffling [20–22], and donation of coding sequence [20,22–27].

Due to the great diversity and dynamic nature of TEs, species-specific TE libraries and accompanying TE annotations are required to provide an understanding of their evolution and impact. By examining differences in the diversity and abundance of TEs within a genome, as well as their integration histories, proliferation frequencies, and rate of removal, we can begin to unravel the factors that govern host-TE evolutionary dynamics. Curating TEs in a wide range of species is also an important step more generally, to expand understanding of the interplay between TEs and host genomes, and consequences for host

evolution [28]. Meanwhile, from a practical perspective, accurate and reliable repeat annotation is essential to provide quality genome annotations, ensuring that repeat sequences are not incorrectly annotated as host genes and vice versa, especially given the vast numbers of genomes currently being sequenced across the tree of life.

Lepidoptera (butterflies and moths) are an important model system in ecology and evolution. Lepidopterans have captivated researchers for decades, proving a fascinating and practical system for research into topics such as development, physiology, host-plant interactions, coevolution, phylogenetics, mimicry, and speciation (e.g [29–33]). Over recent decades, Lepidoptera has become an important model system in genomics especially. For example, Lepidoptera genomes have been pivotal in investigations into the genomic basis of migration [34–36], warning colouration [35], adaptive radiations [37], and evolutionary studies into TEs, including TE horizontal transfer [38–40], the influence of TEs on genome size [41], and the role of a TE in the evolution of industrial melanism in the peppered moth [42].

The monarch butterfly, *Danaus plexippus*, is famous for its incredible North American migration, global dispersal, and orange warning colouration [35]. The monarch has been the focus of much research and broad public interest for decades, with a draft genome released in 2011, and work during the last decade revealing the genetic bases of several key traits [35,36]. Most recently, a high-quality chromosomal assembly of the *D. plexippus* genome has been generated [43], providing an opportunity to provide a detailed analysis of TEs in the monarch genome. Yet, despite the monarch being an important research focus, we lack a detailed TE annotation for the species and accompanying understanding of host-TE interactions or the impact that TEs have had on shaping the monarch genome more generally.

Here, we provide a detailed annotation of TEs within the monarch genome, including analyses of TE diversity and abundance, interactions with host genes, genomic distribution including hot and cold spots, and TE evolution. We find evidence of considerable recent TE

activity in the monarch, with two novel Tc1-Mariner element families rapidly expanding over recent timescales (*danPle_TcM-1* and *danPle_TcM-2*). In contrast, we find a pattern of high proliferation rates of LINE and *Penelope-like* elements (PLEs) was historically dominant, and these elements still characterise the monarch TE landscape. We also reveal the role of large genomic deletions in protecting the host genome from costly LINE TEs, which show signatures of removal. Overall, we provide an overview of the broadscale dynamics that shape the TE landscape of the monarch genome and provide a high-quality, curated, and species-specific TE annotation library for use by the scientific community.

Results

Transposable Element Landscape

Repeats are a major determinant of genome size [44–49]. However, in the monarch, we find that repeats comprise just 6.47% of the total genome size (Assembly Scaffold N50 = 9,210kb, 97.5% Complete Busco orthologs) (Figure 1, Table 1). This differs considerably in comparison to the TE annotation performed on the original monarch genome assembly, where repeats were estimated to cover 13.06% of the genome (Table S1; Additional File 1) [36]. When examining the source of this variation, the major difference in coverage can be attributed to unclassified elements, where we find coverage of unclassified elements reduced by 16.47Mb (Table S1; Additional File 1). This is likely due to improvements in TE annotation tools over the last decade, where improved knowledge of TE structure has led to exclusion of erroneous sequences previously annotated as putatively TE, particularly given that the original annotation used similar tools and databases to those applied in this study [36]. Improvements in genome assembly quality will also have improved repeat detection. Consistent with this, a standard ‘no frills’ RepeatMasker run using the *Arthropoda* database and Dfam identified a repeat content of 5.60% (Table S2; Additional File 1). Thus, the sequences annotated as unclassified in the original annotation were most likely not TE

sequences and are no longer annotated as such by current annotation tools. However, we are unable to confirm this for certain, as the original *de novo* repeat library generated for the draft genome was not provided alongside the genome.

We annotated the two additional genomes available in the genus *Danaus* using the same automated TE annotation method that we applied to the monarch. In comparison, the monarch exhibits a much lower TE content, with repetitive elements covering 26.70% of the *D. chrysippus* genome (Scaffold N50 = 1,465kb, 93.5% Complete Busco orthologs), and 11.87% of the *D. melanippus* genome (Scaffold N50 = 890kb, 86.0% Complete Busco orthologs) (Figure 1, Table 1).

Repeat Class	*Danaus plexippus*			*Danaus chrysippus (GCA_004959915.1)*			*Danaus melanippus (GCA_010014825.1)*		
	No. elements	**Total Length (Mb)**	**Percentage sequence (%)**	**No. elements**	**Total Length (Mb)**	**Percentage sequence (%)**	**No. elements**	**Total Length (Mb)**	**Percentage sequence (%)**
Retroelement	**31716**	**9.90**	**3.98%**	**92162**	**37.96**	**11.77%**	**65710**	**23.35**	**6.59%**
SINE	3977	0.90	0.36%	3252	1.45	0.45%	33	0.01	0.00%
LINE	12165	5.20	2.09%	59490	24.68	7.65%	41769	16.21	4.58%
Penelope	14532	3.14	1.26%	15614	5.20	1.61%	6211	1.67	0.47%
LTR element	1042	0.66	0.27%	13806	6.62	2.05%	17697	5.46	1.54%
DNA transposon	4497	2.37	0.95%	31130	14.57	4.52%	19187	7.32	2.07%
Rolling-circle	5454	1.79	0.72%	62062	24.11	7.48%	7484	2.41	0.68%
Unclassified	6658	1.86	0.75%	18949	9.18	2.85%	28069	8.51	2.40%
Other	555	0.17	0.07%	849	0.26	0.08%	2181	0.46	0.13%
Total	48880	16.08	6.47%	20515	86.08	26.70%	12263	42.05	11.87%

repeats				2			1		

Table 1. Transposable element quantification for the three *Danaus* species with available genome assemblies.

We suggest that the reduced TE load observed in the monarch may have arisen through one of two mechanisms: (i) loss of repeat content in the monarch, or (ii) gains in repeat content in both *D. chrysippus* and *D. melanippus* (or in their ancestral lineages). Relatively recent TE activity is observed in all three *Danaus* species, indicated by repeats exhibiting low genetic distance from consensus (Figure 1). The genome sizes of both *D. chrysippus* and *D. melanippus* are also very similar, at 322Mb and 354Mb, whilst the monarch is only 248Mb. Increases in genome coverage are observed for all TE classifications for both *D. chrysippus* and *D. melanippus* in comparison to the monarch, except for a decrease in PLE coverage in *D. melanippus* (Table 1).

The split between the monarch and *D. chrysippus* is estimated at 8.3Mya [50]. Using this, we were able to calculate DNA loss rates for these two species to see if the small genome of the monarch is a result of an increased rate of DNA loss (Table S3; Additional File 1). When comparing the rates of micro- (<30nt) and mid-size (30<nt<10,000) deletions between the monarch and *D. chrysippus*, estimates suggest there is little difference in DNA deletion rate among them (monarch μ=20.68 del_nt/10kb/My; *D. chrysippus* μ=23.59 del_nt/10kb/My), We were unable to calculate DNA loss rates for the split between *D. melanippus* and the monarch as a dated estimate was not available for this species pair, however *D. melanippus* exhibited similar levels of micro- and mid-size deletions as *D. chrysippus* in the deletion analysis (*D. chrysippus* μ=172.38 del_nt/10kb; *D. melanippus* μ=139.76 del_nt/10kb). In this case, the smaller size of the monarch genome, in comparison to *D. chrysippus*, is unlikely to result from a higher rate of DNA loss in the monarch, but rather from DNA gains in *D. chrysippus*. Therefore, we hypothesise that alternative (ii) is

more likely, whereby there have been independent increases in TE content which have led to increased genome sizes for both *D. chrysippus* and *D. melanippus*, rather than high rates of DNA loss in the monarch. The main TE contributors to increases in genome size are LINEs in both *D. chrysippus* (7.7% of genome, 24.7Mb) and *D. melanippus* (4.6% of genome, 16.2Mb) (Table 1).

TEs are unevenly distributed across the monarch genome and between different chromosomes, indicated by a high standard deviation in TE base pairs per 100kb (mean = 6284bp/100kb, SD = 4183bp/100kb) (Figure 2, Figure 3, Table S4; Additional File 1). A larger standard deviation indicates a less even spread of TEs in a genome, where a standard deviation of 0 indicates TEs are evenly distributed across the genome. The largest density of TEs is found on chromosome 28 (10,128bp/100kb), whilst chromosome 9 exhibits the lowest TE load (4604bp/100kb) (Table S4; Additional File 1). We were unable to compare the density across the three *Danaus* genomes as only the monarch assembly was to chromosomal level.

TEs are known to be unevenly distributed between different genomic compartments, with TE densities varying between genic and intergenic space [1]. LINEs make up 2.38% of the genome, which comprises: 2.80% of gene flanking regions, 1.50% of intronic sequences, 0.93% of exonic sequences, and 3.52% of intergenic sequences (Figure 2). Meanwhile, PLEs make up 1.18% of the genome as a whole: 1.43% of gene flanking regions ,1.13% of intronic sequences, 0.17% of exonic sequences, and 1.88% of intergenic sequences (Figure 2). Collectively, the greatest proportion of TEs is found in intergenic regions (total 3.86%: intergenic space 1.16%, and gene flanking regions 2.70%), in comparison to gene bodies (total 2.49%: introns 2.29%, and exons 0.20%) (Figure 2). As expected, exons are largely devoid of TE sequence, with only 1.72% of exonic sequences consisting of TEs [1,51] (Figure 2). Conversely, intronic regions have accumulated a considerably higher proportion of TE sequences, with 5.13% of total intronic sequence derived from TEs (Figure 2).

Overall, we observed a weak negative correlation between gene and TE density in the monarch genome (Spearman's rank, rho = -0.369, $p < 0.01$), suggesting regions that contain a high density of host genes are less likely to contain TEs. We identified 23 hotspots of TE accumulation in the monarch genome, defined as 100kb windows with a TE content above a 99% AUC cut-off (see Methods). This equates to windows where TE density is at least 3.09 times higher than expected (expected = 6.25kb, lowest hotspot density = 19.34kb) (Figure 4). These windows represent regions of the genome able to tolerate large numbers of TE insertions. Conversely, the presence of 38 cold spots, defined as the 100kb windows with TE content below the 1% AUC cut-off, where TE density is at least 12.56 times lower than expected (expected = 6.25kb, highest cold spot density = 498bp), highlights regions of the genome where TE insertions are likely not tolerated [51]. Gene hotspots identified as TE coldspots were found to contain genes involved in diverse host processes (Table S11; Additional File 1).

Transposable Element Activity

The genome of *D. plexippus* is dominated by five relatively young and very successful novel TE families, which collectively comprise 37.7% of total TE content (Table 2, Tables S5 and S6; Additional File 1). Of these families, the consensus sequences of the Penelope-like elements, danPle_penelope-1 and danPle-penelope-2, are intact as both contain both a GIY-YIG endonuclease domain and a reverse transcriptase-like (RT-like) domain, whilst the LINE element danPle_r2hero-1 is putatively transposition-competent with an intact RT-like domain (Table S9, Additional File 1).

TE Family	TE Classification	Weighted Average Age (Mya)	Copy Number
danPle_penelope-1	Penelope-like Element	10.25	7814
danPle_penelope-2	Penelope-like Element	12.84	2033
danPle_penelope-3	Penelope-like Element	10.17	2504

danPle_HaSE2-1	SINE/HaSE2	6.67	1829
danPle_r2hero-1	LINE/R2-Hero	9.91	5405

Table 2. Age and Copy number of the five TE families responsible for 37.7% of total TE content in the monarch.

In total, four families are estimated to have inserted <1Mya (Table S7, Figure S1; Additional Files 1 and 2). The low estimated age of these insertions is a sign of recent host colonisation, potentially via horizontal transposon transfer (HTT) from another host genome [1,52–55]. Peaks of low genetic distance to TE consensus sequence indicate an abundance of closely related elements and thus recent TE proliferation (Figure 1, Table S8; Additional File 1). There has been a recent burst of activity in two Tc1-Mariner elements: *danPle_TcM-1* (499 new copies, from ~0.68Mya), and *danPle_TcM-2* (275 new copies, from ~0.34Mya) (Table S5 and Table S7; Additional File 1). The rapid expansion of these Tc1-mariner elements over such a short time period highlights their great success within the monarch genome when compared to other young TEs, which are found at much lower copy numbers (Table S7, Figure S2; Additional Files 1 and 3). This recent expansion of DNA TEs is interesting given the historic dominance of LINEs in the monarch genome (2.38% of the genome, 38% of total TE content). Whilst there are more DNA TE families than LINE families in the monarch genome (257 vs 207), LINEs have been more successful in proliferating and colonising the genome than their DNA TE counterparts to date, although the reasons for this are unknown. However, it appears that these new DNA TE families challenge this historic trend (Figure 5, Figure S2; Additional File3).

The oldest TE insertions were estimated to be ~86 Mya, predating the emergence of the monarch ~32Mya [56], with the first invaders being Rolling Circle, SINE and LTR elements (Figure 5, Table S8, Figure S1; Additional Files 1 and 2). These TEs likely proliferated around the time that the Nymphalidae family emerged, ~91Mya (CI: 71-112Mya)

[56]. However, the majority of TE families are estimated to have inserted less than ~45Mya (mean TE family age = 25.41Mya), after the speciation of the monarch (~32Mya [56]) (Figure 5). In total, 246 families (35.4%) are estimated to pre-date the emergence of the monarch, whilst 448 families (64.6%) have emerged on the monarch branch (Figure 5). TEs estimated to have inserted <25Mya are predominantly DNA, Rolling Circle, and LINE elements (Figure 5).

Turnover of LINEs and Penelope-like Elements

LINEs and PLEs are reverse transcribed from their 3' end during mobilisation and are often incomplete at the 5' end due to the premature dissociation of reverse transcriptase (RT) or the activity of cellular RNases [57,58]. In contrast, the absence of a complete 3' region is unlikely to arise due to either process and is evidence of a genomic deletion, presumably as a consequence of ectopic recombination between homologous LINE or PLE sequences [57,58].

Consistent with observations in *H. melpomene* [58], we find LINE and PLE fragments across all families exhibiting patterns suggestive of genomic deletions acting to remove these elements from the genome (Table 3, Figure 6). In total, 54.33% of LINE/PLE fragments contain an intact 3' region, while 45.67% of LINE/PLE fragments are truncated at the 3' end. For the *CR1*, *I-Jockey*, *L2*, *RTE-BovB*, and *Penelope* families, most fragments display patterns suggestive of genomic deletions (i.e. truncated at the 3' end), whilst the majority of fragments of the *R1*, *R2-Hero*, and *RTE-RTE* families display patterns indicative of processes associated with degradation of LINEs, such as premature disassociation of RT or breakdown via cellular RNAse activity (Table 3).

Swathes of small fragments across all LINE and PLE families (mean length = 342bp) suggest a high rate of turnover of these non-LTR elements within the monarch genome, in

contrast to patterns observed in mammals, where low rates of DNA loss are observed [58,59].

LINE Classification	Putative Removal Process	Fragment Types	Count	Proportion of Classification Type
CR1	Premature RT Dissociation/Activity of Cellular RNAses	Lone 3', Internal + 3'	16	0.36
	Putative Genomic Deletion	Lone 5', 5' + Internal, Internal Lacking 3'	29	0.64
I-Jockey	Premature RT Dissociation/Activity of Cellular RNAses	Lone 3', Internal + 3'	42	0.20
	Putative Genomic Deletion	Lone 5', 5' + Internal, Internal Lacking 3'	164	0.80
L2	Premature RT Dissociation/Activity of Cellular RNAses	Lone 3', Internal + 3'	61	0.20
	Putative Genomic Deletion	Lone 5', 5' + Internal, Internal Lacking 3'	244	0.80
R1	Premature RT Dissociation/Activity of Cellular RNAses	Lone 3', Internal + 3'	223	0.55
	Putative Genomic Deletion	Lone 5', 5' + Internal, Internal Lacking 3'	181	0.45
R2-Hero	Premature RT Dissociation/Activity of Cellular RNAses	Lone 3', Internal + 3'	2151	0.93
	Putative Genomic Deletion	Lone 5', 5' + Internal, Internal Lacking 3'	166	0.07
RTE-BovB	Premature RT Dissociation/Activity of Cellular RNAses	Lone 3', Internal + 3'	225	0.46
	Putative Genomic Deletion	Lone 5', 5' + Internal, Internal Lacking 3'	268	0.54
RTE-RTE	Premature RT Dissociation/Activity of Cellular RNAses	Lone 3', Internal + 3'	137	0.67
	Putative Genomic Deletion	Lone 5', 5' + Internal, Internal Lacking 3'	68	0.33
Penelope	Premature RT Dissociation/Activity of Cellular RNAses	Lone 3', Internal + 3'	2155	0.41
	Putative Genomic Deletion	Lone 5', 5' + Internal, Internal Lacking 3'	3091	0.59

Table 3. Number of LINEs and PLEs classified by fragment type, and putative removal process.

Despite evidence of loss, LINE and PLE sequences continue to dominate the TE landscape of the monarch. To be autonomous, an intact LINE will contain start and stop codons, and an RT domain [60], whilst an intact PLE will contain an RT domain and a GIY-YIG endonuclease domain [61]. In total, we identified five intact LINEs and two intact PLEs in the monarch (Table S9, Additional File 1), including *danPle_r2hero-1* and *danPle_penelope-1*, which are the two most populous TE families in the monarch genome (Table 2). We also observe a significant accumulation of certain LINE families, with ~20% maintaining a length above 1kb, whilst only ~3% of PLEs have maintained this length (Table S10; Additional File 1). Short elements are more likely to persist in the genome than their longer counterparts, as they are less prone to recombination [19]. This, combined with the high activity level of LINEs and PLEs before and throughout the evolution of the monarch (Figure 2), explains the dominance of LINE fragments across the genomic landscape, despite their high turnover [19,58].

Association of TEs with Monarch Colouration and Immune Genes

Given a previous role for TEs in wing colouration in the peppered moth (*Biston betularia*) [42] and the clouded yellow (*Colias croceus*) [62], we investigated the possibility of TE associations with wing colouration in the monarch. The myosin gene is associated with wing colouration in the monarch [63]. We identified a single unclassified element 7kb upstream of this gene (Figure S3; Additional File 4), contrasting the location of TEs implicated in colouration in the peppered moth and clouded yellow, where the TE insertions were found within an intron and 6kb downstream of the gene body, respectively [42,62]. This is not entirely unexpected, as the colouration differences seen in the peppered moth and

clouded yellow are due to altered gene expression, whilst the white *nivosus* form of the monarch is due to altered pigment transport [35,42,62]. When compared to all monarch genes, the association of a single TE with the myosin gene is within the expected range of TE copy number to find within 20kb of a genic locus (See Methods), and therefore is not significant (Figure S4, Additional File 5).

Immune genes are some of the fastest evolving genes across metazoans [64–67], and so are strong candidates for exploring the role of TEs in host evolution. Across the monarch genome, TEs of all major types were found within immune gene flanks (20kb). Regions surrounding immune genes are significantly enriched for TE-derived base pairs when compared to genes of other functions (Welch's T-Test, $T_{7.0059} = 2.8404$, $p < 0.05$). When comparing association of different TE classifications around immune genes, we find that LINEs are significantly more abundant than rolling circles and unclassified elements (Kruskal-Wallis, $\chi^2_7 = 27.282$, $p < 0.01$) whilst there were no significant pairwise differences between the remaining TE classifications. LINEs are not enriched around immune genes compared to genome-wide levels (genome-wide = 2.09% of genome size; immune gene regions = 2.14% of immune gene compartment size). We also find no significant difference in TE abundance between different types of immune genes (Kruskal-Wallis $\chi^2_{37} = 40.351$, $p = 0.3244$), suggesting no enrichment for TEs around specific classes of immune genes (Figure S5; Additional File 6).

Discussion

Here, we provide a detailed annotation of TEs in the monarch genome and an analysis of their diversity and evolution. We found a large reduction in TE sequence compared to the original annotation performed on the draft monarch genome. We suggest that this is due to unclassified TE sequences in the original annotation no longer being recognised as TEs with updated TE databases and annotation tools. However, we are

unable to confirm this, due to the original *de novo* TE library and annotation results being unavailable. This raises an important issue in genomics currently. Given TE annotation is a key step in genome assembly and annotation, we suggest that it should be required to include TE libraries alongside the publication of genome projects, ideally in a specific repository, to: (i) reduce duplication of efforts required for reannotation; (ii) improve reproducibility in annotation; and (iii) improve TE consensus libraries to reduce cases of a single TE family being given different names in multiple species. As the number sequenced genomes increases, the wider availability of TE libraries will benefit a range of studies, where gene annotations can be improved by accurate TE annotation, and the availability of large numbers of well-curated TE sequences can facilitate large-scale studies to study TE evolutionary dynamics.

We identified an uneven distribution of TEs across the genome in line with expectations of TE distribution in host genomes [1]. TE insertions in non-coding regions are less likely to have a detrimental impact on host genome function than TE insertions into coding regions [1,21,51,68–70], and so are less stringently removed by selection, enabling TE accumulation in non-coding regions. Indeed, some TEs actively avoid coding regions by targeting regions of the genome unlikely to impact host fitness [51,71], with this process potentially occurring in the monarch, where we find overlapping gene hotspots and TE cold spots.

The most populous regions of the monarch genome for TEs are intergenic regions. The presence of large numbers of TEs in host gene flanks is interesting, given the implication of TEs in the evolution of gene regulatory networks [10–13,72], although this could also be due to the relatively compact nature of the monarch genome leaving TEs with few other loci to integrate within. TEs are implicated in rapidly generating extensive intron gain in eukaryotic genomes, where short non-autonomous DNA transposons independently generated hundreds to thousands of introns across the genomes of eukaryotes [1,51,73]. It is also hypothesised that transposon insertion-associated introns are responsible for the

prevalence of nucleosome-sized exons observed in eukaryotes [73]. The presence of TEs in intronic sequences of the monarch could therefore indicate genomic refugia, where TE insertions are less likely to be detrimental to the host, but they could also indicate ongoing intron generation, where new insertions might lead to the generation of new, or longer, introns over evolutionary time.

The presence of young and highly populous TE families in the monarch genome demonstrates the success of TEs in naïve host genomes [71,74–77]. It should be noted that it is notoriously difficult to accurately estimate mutation rates. Here, we have used divergence from consensus sequence as a proxy for TE age. These estimates are likely to stray from the true value given the differential effects of selection on new TE insertions and other genomic regions, which are ignored in the calculation of a neutral mutation rate. It is also unlikely that any TE insertions can be considered truly neutral and so assuming a neutral rate of mutation is likely to mask differential selection effects acting on different TE insertions at various loci across the host genome. The putative recent arrival of several young DNA TEs, namely two novel Tc1-Mariner families (*danPle_TcM-1* and *danPle_TcM-2*), has led to another wave of TE proliferation in the monarch genomes. These elements are closely related and located in Clade 130 in a recent phylogenetic study of Tc1/mariner diversity [78], which contains elements from diverse insect species, including butterflies, aphids and ants (Figure S6; Additional File 7). Ants and aphids are known to share environments with monarch larvae on milkweed plants [79], increasing the potential for HTT events [54,80]. Elucidating the origin of these young Tc1/mariner elements will require comparison across related butterfly species and other host plants and insects that are known to share environments with the monarch. Over time, it is hypothesised that the activity of these novel families will decrease as the host develops mechanisms to suppress activity, either through the capture of TE sequences in piRNA regions, or through other defensive mechanisms such as methylation [77,81,82]. A reduction in transposition activity of DDE transposase-containing DNA transposons can also be attributed to increases in non-

autonomous copies [78]. In the meantime, however, these TEs are proliferating and dynamically altering the genomic landscape of *D. plexippus*, with potential connotations for ongoing genome structural and functional evolution [1–3,19].

While factors leading to the relatively low TE content observed in the monarch compared to other *Danaus* species are currently unknown, it is clear TEs have been unable to accumulate to the same level as other members of the genus. The maintenance of a small genome in the monarch can likely be partly attributed to the high turnover of TEs, evidenced by the lack of ancient peaks in repeat landscapes and the presence of few intact LINEs and PLEs, many of which show hallmarks of genomic deletions. To remove LINEs and PLEs, ectopic recombination between similar elements is likely to play a role, and it has been postulated that an increased rate of ectopic recombination could be responsible for the differences in accumulation of TEs in insects and mammals [83]. A lower ectopic recombination rate in mammals, despite frequent large TE loads, enables older elements to accumulate, whereas high ectopic recombination rates in insects can lead to the removal of TEs before older copies accumulate in large numbers [83]. Given our results, it is likely that ectopic recombination is occurring in the monarch to remove LINEs and PLEs. Indeed, the only families with an abundance of 3' fragments are, *R1*, *R2-Hero*, and *RTE-RTE* (mean insertion times ~16.02Mya, ~9.9Mya, and ~27.38Mya). It is tempting to suggest that *R1*, *R2-Hero*, and *RTE-RTE* insertions have a lower impact on host fitness, and so are "allowed" to degrade over time, rather than being removed *en masse* through genomic mechanisms. However, it must also be noted that ectopic recombination is a passive process. Whilst it is more likely that LINEs and PLEs are being fragmented and removed through ectopic recombination between copies, this does not account for inter-family differences in the ratio of fragments undergoing degradation through ectopic recombination versus degradation via the activity of RT or cellular RNAses. In the case of the most populous family, *R2-Hero*, it is likely that these elements have so far avoided ectopic recombination, given these elements are young compared to others in the same classification (insertion ~9.9Mya) (Figure 5, Table

S8; Additional File 1)). Our results are consistent with those of previous findings in *Heliconius* species [84], in which a high TE turnover and a lack of ancient TEs were observed, suggesting these characteristics might be widespread among lepidopterans, although further comparative studies will be required to confirm this.

In addition to TE turnover, genome size is likely also maintained through sequestering TE activity via epigenetic mechanisms and piRNAs [1,81,85,86]. Further research to characterise piRNA clusters in the monarch, and interrogation of its transcriptome, will further improve our understanding of how these processes have shaped the TE landscape within the monarch genome.

The lack of closer or more associations of TEs with the myosin gene associated with warning colouration in the monarch is perhaps not unsurprising, as it has been postulated that the white *nivosus* form of the monarch arises from altered pigment transport [35], rather than altered gene expression, as is the case with the peppered moth and clouded yellow [42,62]. The prominence of TEs around immune genes is intriguing, particularly given the significant increase in TE-derived base pairs around immune genes compared to genes of other functions. It is not unexpected to find LINEs are significantly more abundant than rolling circles as they dominate the TE landscape of the monarch, and LINEs are not enriched around immune genes compared to genome-wide levels. Nevertheless, the abundance of TEs around immune genes warrants future research to further understand the impact of TEs in these regions. The use of TE knockout studies in combination with transcriptomic analyses may be able to elucidate the impacts TEs may be having on immune genes in the monarch.

Conclusions

In-depth studies into TE landscapes and dynamics within host species of broad scientific and public interest remain relatively few, even with widespread acknowledgement of the key role that TEs have played in eukaryotic evolution [1,44,87]. We provide a detailed

annotation of TEs in the monarch genome and analysis of TE diversity and evolution, including the identification of highly successful young DNA TE families, which mirror the activity profile of the most successful LINEs found in the monarch, and evidence of ongoing TE expansion and removal in *D. plexippus*, highlighting the dynamic nature of repeat content within genomes over time. We also identify a significant association of TEs with genes of immune function, providing candidates for future studies to elucidate the impact of TEs on immune function in the monarch. Further large-scale comparative studies will be greatly beneficial to further our understanding of the evolutionary dynamics of TEs, and the drivers leading to diverse TE landscapes across closely related species.

Materials and Methods

Transposable Element Annotation

The monarch chromosomal assembly, Dplex_v4, was downloaded from NCBI (Refseq GCF_009731565.1), along with corresponding genome annotations in GFF format [43,88]. The monarch genome assembly is 248.676Mb, with a GC content of 32.1%, and comprises 4,115 scaffolds with a scaffold N50 of 9.21Mb.

The *D. plexippus* genome was annotated using Earl Grey (github.com/TobyBaril/EarlGrey/) [89]. For this, known repeats were first identified and masked using RepeatMasker (version 4.1.0) [90] using the *Arthropoda* library from RepBase (version 23.08) and Dfam (release 3.3) [91,92]. Low complexity repeats and RNA were not masked (*-nolow and -norna*) and a sensitive search was performed (*-s*). Following this, a *de novo* repeat library was constructed using RepeatModeler (version 1.0.11), with RECON (version 1.08) and RepeatScout (version 1.0.5) [93–95]. To generate maximum-length consensus sequences for the *de novo* repeat library, novel repeats identified by RepeatModeler were curated using an automated version of the 'BLAST, Extract, Extend' process [28]. A BLASTn search was performed on the *D. plexippus* genome to obtain up to the top 40 hits for each TE family identified by RepeatModeler [96]. Sequences were extracted with 1,000 base pairs of flanking sequences at the 5' and 3' ends. Each set of family sequences were aligned using MAFFT (version 7.453) [97]. Alignments were trimmed with trimAl (version 1.4.rev22) to retain high-quality positions in the alignment (*-gt* 0.6 *-cons* 60) [98]. Updated consensus sequences were computed with EMBOSS cons (*-plurality* 3) to generate a new TE library featuring consensus sequences with extended flanks [99]. This process was repeated through 5 iterations to obtain maximum-length consensus sequences. The resulting *de novo* repeat library was combined with the RepBase *Arthropoda* library and utilised to annotate repetitive elements using RepeatMasker with a score threshold at the relatively conservative level of 400 (*-cutoff* 400), to exclude poor matches unlikely to be true

TE sequences. Following this, the *D. plexippus* genome was analysed with LTR_Finder (version 1.07), using the LTR_Finder parallel wrapper [100,101] to identify full-length LTR elements. Final TE annotations were defragmented and refined using the loose merge (*-loose*) command in RepeatCraft [102]. Overlapping annotations were removed using MGKit (version 0.4.1) filter-gff, with the longest element of overlapping annotations retained (*-c* length *-a* length) [103]. Finally, all repeats less than 100bp in length were removed before the final TE quantification to decrease spurious hits. Final repeat annotation is provided in dataset S1 (Additional File 8), and *de novo* TE library is provided in dataset S2 (Additional file 9).

Quantifying DNA Loss Rates Between the Monarch and *D. chrysippus*

To estimate DNA loss rates between the monarch and *D. chrysippus*, the DelGet.pl script (https://github.com/4ureliek/DelGet) was used to estimate DNA loss from small (<30nt) and mid-size (30<nt<10,000) deletions, as described previously [46]. Briefly, the script identifies orthologous regions of 10,000bp between three input species (an outgroup and two species with known branch length in Mya) using BLAT (https://genome.ucsc.edu/FAQ/FAQblat.html). These regions are extracted and aligned with MUSCLE [104] before deletions are quantified. Species-specific gaps are identified and quantified. Deletion rates are then calculated between the two species of interest by dividing the normalised deleted nucleotide count (del_nt/10kb) by the corresponding Mya since divergence.

Identifying Transposable Element Association with Genomic Compartments

The *D. plexippus* gene annotation file was processed to generate a GFF file containing coordinates of exons, introns, and 5' and 3' flanking regions. Gene flanking regions are defined here as 20kb directly upstream or downstream of the gene body. These

flanking regions are determined to identify TEs at a distance that could be described as the proximate promoter region, rather than just accounting for the core promoter region, to include both promoter and more distal enhancer regions for genes [105]. Bedtools intersect (version 2.27.1) was used to identify the number of bases of overlap (*-wao*) between TEs and different genomic compartments [106]. Following this, quantification was performed in R.

TE Hotspots and Coldspots

To identify TE and gene hotspots and coldspots, chromosomal scaffolds were split into 100kb windows, which were defined using the frequency distribution of observations. A polynomial model was fitted to the frequency distribution, to generate a smooth curve through the data points using the lm function in R. Using these models, the total area under the curve (AUC) was calculated using the area_under_curve function of the bayestestR package [107]. From this, the 1% and 99% AUC cutoff values were determined. Starting with the first data point and iteratively adding subsequent datapoints with each round, the AUC for each collection of points was determined. The maximum AUC value not exceeding 99% of the total AUC was taken as the cutoff where windows containing higher TE or gene coverage were considered to be hotspots. The minimum AUC value not exceeding 1% of the total AUC was taken as the cutoff where windows containing lower TE or gene coverage were considered to be coldspots (Figure S7; Additional File 10). Karyoplots of the *D. plexippus* genome, illustrating gene and TE locations, highlighting TE hotspots and coldspots were generated using the KaryoploteR package in R [108].

Estimating TE Age and Activity

TE activity dates were estimated using Kimura2-parameter genetic distances (including CpG sites) calculated between individual insertions and the consensus sequence

of each TE family [109]. A neutral mutation rate of 0.0116 substitutions per site per million years was used for *D. plexippus*, as previously described for monarch [110].

Turnover of LINEs and PLEs

To interrogate LINE and PLE decay patterns, each LINE and PLE sequence was extracted from the genome using bedtools getfasta [106]. Extracted sequences were used to query the repeat library using BLASTn with tabular output specified (*-outfmt* 6). In the event of multiple matches, only the top match was retained (defined by percent identity). Coordinates of each match were scaled to normalise the length of each consensus sequence. Mapped fragments were plotted using ggplot2 of the tidyverse package in R [111,112].

Identification of Intact LINEs

To identify full-length LINEs, all elements longer than 2,700bp were extracted from the monarch genome using bedtools getFasta [106]. Open reading frames (ORFs) were detected using EMBOSS getORF [99]. ORF sequences were filtered to retain all sequences > 600aa. RPSBLAST+ (v2.9.0+) [96] was used to query the ORF sequences against the CDD database [113] to identify conserved protein domains. Sequences were classified as intact LINEs based on the presence of relevant protein domains as previously described [114] (Table S9, Additional File 1).

Identification of Intact PLEs

To identify intact PLEs, all elements longer than 2,400bp were extracted from the monarch genome using bedtools getfasta [106] as full-length PLEs are ~2,500bp long [115].

Open reading frames (ORFs) were detected using EMBOSS getORF [99]. ORF sequences were filtered to retain all sequences >200aa. RPSBLAST+ (v2.9.0+) [96] was used to query the ORF sequences against the CDD database [113] to identify conserved protein domains. Sequences were classified as intact PLEs based on the presence of reverse transcriptase (RT) and a GIY-YIG endonuclease.

Classification of Novel DNA TEs

To further classify the novel DNA elements in the monarch beyond homology to known elements, as used by RepeatModeler. DDE transposases were compared with curated alignments for all DDE TE superfamilies [78]. Specifically, dashes in curated alignments of all known DDE TE sequences for each of the 19 families were removed, and the resultant fasta file was used as the subject in a blastx search against the *de novo* monarch TE library, with an e-value threshold of $1x10^{-20}$ and tabular output specified (-*evalue* 1e-20 -*outfmt* 6) [96]. The best match for each *de novo* query sequence was retained, with a minimum 50% identity over at least 50% of the total length of the curated DDE subject sequence. Following this, novel DNA elements were only detected for the Tc1-Mariner and Sola1 families (Figures S6 and S8; Additional Files 7 and 11). Transposase sequences for the *de novo* monarch sequences were extracted and aligned to the original DDE alignments for their respective families using the profile alignment command in MAFFT [97]. Sequences from related transposon superfamilies were used as outgroups, as described [78]. To estimate phylogenies for the two DDE superfamilies, we performed 1,000 ultra-fast bootstrap replicates in IQ-TREE (v1.6.12) [116] specifying the best-fit amino acid model for each family as indicated by ModelFinder [117]. Phylogenetic trees were visualised in figtree (v1.4.4) (http://tree.bio.ed.ac.uk/software/figtree/).

Association of TEs with Monarch Colouration and Immune Genes

We obtained the transcript sequence in fasta format for the myosin gene responsible for wing colouration from MonarchBase, under accession DPOGS206617 [34]. To identify the matching region in the v4 monarch assembly, a blastn search was performed to query the monarch assembly [96]. The coordinates of the myosin gene were obtained and used in a bedtools window search to identify TEs within 20kb (-*w* 20000)[106]. TE copy number within 20kb of gene bodies was quantified for each monarch gene and a polynomial model was fitted to the frequency distribution to generate a smooth curve through the data points using the lm function in R. Using these models, the total area under the curve (AUC) was calculated using the area_under_curve function of the bayestestR package [107]. From this, the 2.5% and 97.5% AUC cutoff values were determined, equivalent to selecting the central 95% of the data distribution. Starting with the first data point and iteratively adding subsequent datapoints with each round, the AUC for each collection of points was determined. The maximum AUC value not exceeding 97.5% of the total AUC was taken as the cutoff where genes containing higher TE copy number were considered significantly enriched for TEs. The minimum AUC value not exceeding 2.5% of the total AUC was taken as the cutoff where genes containing lower TE copy number were considered significantly depleted for TEs (Figure S4; Additional File 5). A karyoplot of the myosin gene region including 20kb flanks was generated in R using the karyoploteR package [108].

To investigate associations between monarch immunity genes and TEs, nucleotide and protein sequences of monarch immune genes were obtained in fasta format from MonarchBase [34]. To identify corresponding regions in the v4 monarch assembly, blastn and tblastn searches were used to query the monarch genome assembly with the monarch immune gene nucleotide and protein sequences [96]. Matches were filtered by percentage identity and query match coordinates to identify the locus of each immune gene. A bed file was generated, containing the name and location of each immune gene, and a bedtools window search was performed to identify TEs within 20kb of each immune gene(-*w* 20000)[106]. The same method was applied to identify TEs within 20kb of all other genes.

TE bases surrounding each immune gene were quantified in R [118]. As TE count was not normally distributed, a Kruskal-Wallis test was used to compare TE abundance between immune genes and other genes, between different TE classifications around immune genes, and between different immune gene types. Unless otherwise stated, all plots were generated using the ggplot2 package of the tidyverse suite in R [111,112].

Figure Legends

Figure 1. TE content in the monarch, *D. chrysippus* and *D. melanippus*. Major TE types are represented by different colours indicated in the key (A) Pie charts illustrating proportions of each *Danaus* genome comprised of the main TE classifications. (B) Repeat landscapes for each *Danaus* species. The x axis indicated the level of Kimura2-parameter genetic distance observed between TE insertions and their respective consensus sequences in percent. More recent elements are located to the right of the x axis. The y axis indicates the percentage of the genome occupied by TE insertions of each genetic distance class.

Figure 2. TE location in different genomic compartments in the monarch. Major TE types are represented by different colours indicated in the key (A)Number of TE-derived base pairs found in different genomic compartments. (B) Proportion of each genomic compartment contributed by TEs.

Figure 3. Violin plot showing kilobases of TE sequence per 100kb window across the genome for each of the main TE classifications. Red points show the mean kilobases of TE sequence per 100kb.

Figure 4. Karyoplots illustrating monarch chromosomes, with genes represented as black bars. (A) Above each chromosome, TE density is shown for 100kb windows, with major TE types represented by different colours indicated in the key. (B) Above each chromosome, TE hotspots and coldspots are shown, with severity represented by the different colours indicated in the key,

Figure 5. Violin plot illustrating the estimated insertion time (Mya) for each of the major TE types, represented by different colours indicated in the key. The black bar shows the estimated divergence time of the monarch, approximately 32Mya [56].

Figure 6. Plot illustrating LINE and Penelope fragments annotated in the monarch genome. The x axis indicates normalised start and end coordinates for each element relative to its consensus. Elements are organised from longest to shortest, with short elements found towards the top of the plot. Families are represented by different colours indicated by the key.

Additional Files

Additional File 1: Supplementary Tables, including contents page with table legends. Tables S1-S11.

Additional File 2: Figure S1. Scatter plot illustrating estimated TE age for each TE family. Major TE types are represented by different colours indicated in the key. Elements are ordered from oldest to youngest, within each major type.

Additional File 3: Figure S2. Scatter plots illustrating TE age against copy number. Major TE types are represented by different colours indicated in the key. X axis shows estimated age of each TE family, Y axis shows log10 transformed TE copy number identified in the monarch.

Additional File 4: Figure S3. Karyoplot of the monarch gene association with colouration, myosin, obtained from MonarchBase under accession DPOGS206617. Main TE types within 20kb are represented by different colours indicated in the key.

Additional File 5: Figure S4. Plot illustrating the distribution of TE copy number found with 20kb of all genes, showing the 2.5% and 97.5% AUC cutoffs used to determine the TE copy number within 20kb of a given gene for it to be significantly depleted or enriched in comparison to other genes.

Additional File 6: Figure S5. Boxplots illustrating: (A) Normalised TE coverage in 20kb regions surrounding each immune gene type. (B) Normalised TE count in 20kb regions surrounding each immune gene type. Main TE types are represented by different colours indicated in the key.

Additional File 7: Figure S6. Phylogenetic tree of Tc1-Mariner DDE transposases, with novel monarch families added (highlighted in red). Branch support from 1000 ultra-fast bootstrap repetitions. Novel Tc1-Mariner families highlighted in teal are the young TEs found with extremely high copy number.

Additional File 8: Dataset S1. Coordinates and classification of all transposable element sequences identified in the monarch genome using the automated *de novo* library in conjunction with RepBase *Arthropoda* library and Dfam, in bed format.

Additional File 9: Dataset S2. *De novo* transposable element consensus sequences identified in the monarch in FASTA format.

Additional File 10: Figure S7. Plots illustrating: (A) Distribution of genome windows with TE coverage, with lines showing 1% and 99% AUC cutoffs used to define hotspots and coldspots. (B) Distribution of genome windows with host gene coverage, with lines showing 1% and 99% AUC cutoffs used to define hotspots and coldspots.

Additional File 11: Figure S8. Phylogenetic tree of Sola1 DDE transposases, with novel monarch families added (highlighted in red). Branch support from 1000 ultra-fast bootstrap repetitions.

Abbreviations

TE: Transposable element

Non-LTR: Non-long terminal repeat

LTR: Long terminal repeat

LINEs: Long INterspersed Elements

SINEs: Short INterspersed Elements

PLEs: *Penelope*-like Elements

TSS: Transcription Start Site

References

[1] Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol 2018;19:199. https://doi.org/10.1186/s13059-018-1577-z.

[2] Klein SJ, O'Neill RJ. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosom Res 2018;26:5–23. https://doi.org/10.1007/s10577-017-9569-5.

[3] Feschotte C, Pritham EJ. DNA Transposons and the Evolution of Eukaryotic Genomes. Annu Rev Genet 2007;41:331–68. https://doi.org/10.1146/annurev.genet.40.110405.090448.

[4] Belyayev A. Bursts of transposable elements as an evolutionary driving force. J Evol Biol 2014;27:2573–84. https://doi.org/10.1111/jeb.12513.

[5] Schrader L, Schmitz J. The impact of transposable elements in adaptive evolution. Mol Ecol 2018. https://doi.org/10.1111/mec.14794.

[6] Casola C, Lawing AM, Betrán E, Feschotte C. PIF-like transposons are common in Drosophila and have been repeatedly domesticated to generate new host genes. Mol Biol Evol 2007;24:1872–88. https://doi.org/10.1093/molbev/msm116.

[7] Jangam D, Feschotte C, Betrán E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends Genet 2017;33:817–31. https://doi.org/10.1016/j.tig.2017.07.011.

[8] Volff JN. Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 2006. https://doi.org/10.1002/bies.20452.

[9] Mita P, Boeke JD. How retrotransposons shape genome regulation. Curr Opin Genet Dev 2016;37:90–100. https://doi.org/10.1016/j.gde.2016.01.001.

[10] Bourque G. Transposable elements in gene regulation and in the evolution of vertebrate genomes. Curr Opin Genet Dev 2009;19:607–12. https://doi.org/10.1016/j.gde.2009.10.013.

[11] Feschotte C. Transposable elements and the evolution of regulatory networks. Nat Rev Genet 2008;9:397–405. https://doi.org/10.1038/nrg2337.

[12] Rebollo R, Romanish MT, Mager DL. Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. Annu Rev Genet 2012;46:21–42. https://doi.org/10.1146/annurev-genet-110711-155621.

[13] Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: From conflicts to benefits. Nat Rev Genet 2017;18:71–86. https://doi.org/10.1038/nrg.2016.139.

[14] Bennetzen JL. Transposable element contributions to plant genome evolution. Plant Mol Biol 2000;42:251–69. https://doi.org/10.1023/A:1006344508454.

[15] Casals F, Cáceres M, Ruiz A. The Foldback-like transposon Galileo is involved in the generation of two different natural chromosomal inversions of Drosophila buzzatii. Mol Biol Evol 2003;20:674–85. https://doi.org/10.1093/molbev/msg070.

[16] Campos-Sánchez R, Cremona MA, Pini A, Chiaromonte F, Makova KD. Integration and Fixation Preferences of Human and Mouse Endogenous Retroviruses Uncovered with Functional Data Analysis. PLoS Comput Biol 2016;12:1–41. https://doi.org/10.1371/journal.pcbi.1004956.

[17] Oliver KR, Greene WK. Transposable elements: Powerful facilitators of evolution. BioEssays 2009;31:703–14. https://doi.org/10.1002/bies.200800219.

[18] Hughes JF, Coffin JM. Human endogenous retroviral elements as indicators of

ectopic recombination events in the primate genome. Genetics 2005;171:1183–94. https://doi.org/10.1534/genetics.105.043976.

[19] Kent T V., Uzunović J, Wright SI. Coevolution between transposable elements and recombination. Philos Trans R Soc B Biol Sci 2017;372. https://doi.org/10.1098/rstb.2016.0458.

[20] Coates BS, Hellmich RL, Grant DM, Abel CA. Mobilizing the genome of lepidoptera through novel sequence gains and end creation by non-autonomous Lep1 Helitrons. DNA Res 2012;19:11–21. https://doi.org/10.1093/dnares/dsr038.

[21] Biémont C, Vieira C. What transposable elements tell us about genome organization and evolution: The case of Drosophila. Cytogenet Genome Res 2005;110:25–34. https://doi.org/10.1159/000084935.

[22] Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 2005;37:997–1002. https://doi.org/10.1038/ng1615.

[23] Philippsen GS, Avaca-Crusca JS, Araujo APU, DeMarco R. Distribution patterns and impact of transposable elements in genes of green algae. Gene 2016;594:151–9. https://doi.org/10.1016/j.gene.2016.09.012.

[24] Wu M, Li L, Sun Z. Transposable element fragments in protein-coding regions and their contributions to human functional proteins. Gene 2007;401:165–71. https://doi.org/10.1016/j.gene.2007.07.012.

[25] Daniel C, Behm M, Öhman M. The role of Alu elements in the cis-regulation of RNA processing. Cell Mol Life Sci 2015;72:4063–76. https://doi.org/10.1007/s00018-015-1990-3.

[26] Vorechovsky I. Transposable elements in disease-associated cryptic exons. Hum Genet 2010;127:135–54. https://doi.org/10.1007/s00439-009-0752-4.

[27] Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, et al. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 2006;441:87–90. https://doi.org/10.1038/nature04696.

[28] Platt RN, Blanco-Berdugo L, Ray DA. Accurate transposable element annotation is vital when analyzing new genome assemblies. Genome Biol Evol 2016;8:403–10. https://doi.org/10.1093/gbe/evw009.

[29] Bates HW. Contributions to an Insect Fauna of the Amazon Valley.—Lepidoptera:—Heliconinæ. J Proc Linn Soc London Zool 1862;6:73–7. https://doi.org/10.1111/J.1096-3642.1862.TB00932.X.

[30] Müller F. Ituna and Thyridia: a remarkable case of mimicry in butterflies. Trans Entomol Soc L 1879;1879:20–9.

[31] Ehrlich PR, Raven PH. Butterflies and Plants: A Study in Coevolution. Evolution (N Y) 1964;18:586. https://doi.org/10.2307/2406212.

[32] Kettlewell, B. Evolution of melanism: the study of a recurring necessity 1973.

[33] Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 2012;487:94–8. https://doi.org/10.1038/nature11041.

[34] Zhan S, Reppert SM. MonarchBase: the monarch butterfly genome database. Nucleic Acids Res 2013;41:D758–63. https://doi.org/10.1093/NAR/GKS1057.

[35] Zhan S, Zhang W, Niitepõld K, Hsu J, Haeger JF, Zalucki MP, et al. The genetics of monarch butterfly migration and warning colouration. Nature 2014;514:317–21. https://doi.org/10.1038/nature13812.

[36] Zhan S, Merlin C, Boore JL, Reppert SM. The monarch butterfly genome yields insights into long-distance migration. Cell 2011;147:1171–85. https://doi.org/10.1016/j.cell.2011.09.052.

[37] Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 2012;487:94–8. https://doi.org/10.1038/nature11041.

[38] Wu M, Sun Z, Luo G, Hu C, Zhang W, Han Z. Cloning and characterization of piggyBac-like elements in lepidopteran insects. Genetica 2011;139:149–54. https://doi.org/10.1007/s10709-010-9542-0.

[39] Sormacheva I, Smyshlyaev G, Mayorov V, Blinov A, Novikov A, Novikova O. Vertical evolution and horizontal transfer of CR1 non-LTR retrotransposons and Tc1/mariner DNA transposons in Lepidoptera species. Mol Biol Evol 2012;29:3685–702. https://doi.org/10.1093/molbev/mss181.

[40] Reiss D, Mialdea G, Miele V, de Vienne D, Peccoud J, Gilbert C, et al. Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot. PLOS Genet 2019;15:e1007965. https://doi.org/10.1371/journal.pgen.1007965.

[41] Talla V, Suh A, Kalsoom F, Dinca V, Vila R, Friberg M, et al. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (leptidea) butterflies. Genome Biol Evol 2017;9:2491–505. https://doi.org/10.1093/gbe/evx163.

[42] Hof AEV t., Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature 2016;534:102–5. https://doi.org/10.1038/nature17951.

[43] Gu L, Reilly PF, Lewis JJ, Reed RD, Andolfatto P, Walters JR. Dichotomy of dosage compensation along the neo Z chromosome of the Monarch butterfly. Curr Biol 2019;29:4071–7.

[44] Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on

eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. Gene 2012;509:7–15. https://doi.org/10.1016/j.gene.2012.07.042.

[45] Osanai-Futahashi M, Suetsugu Y, Mita K, Fujiwara H. Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, Bombyx mori. Insect Biochem Mol Biol 2008;38:1046–57. https://doi.org/10.1016/j.ibmb.2008.05.012.

[46] Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci U S A 2017;114:E1460–9. https://doi.org/10.1073/pnas.1616702114.

[47] Talla V, Suh A, Kalsoom F, Dincă V, Vila R, Friberg M, et al. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (leptidea) butterflies. Genome Biol Evol 2017;9:2491–505. https://doi.org/10.1093/gbe/evx163.

[48] Kidwell MG. Transposable elements and the evolution of genome size in eukaryotes. Genetica 2002;115:49–63. https://doi.org/10.1023/A:1016072014259.

[49] Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. Recent activity in expanding populations and purifying selection have shaped transposable element landscapes across natural accessions of themediterranean grass brachypodium distachyon. Genome Biol Evol 2018;10:304–18. https://doi.org/10.1093/gbe/evx276.

[50] Wiemers M, Chazot N, Wheat CW, Schweiger O, Wahlberg N. A complete time-calibrated multi-gene phylogeny of the European butterflies. Zookeys 2020;938:97. https://doi.org/10.3897/ZOOKEYS.938.50878.

[51] Sultana T, Zamborlini A, Cristofari G, Lesage P. Integration site selection by retroviruses and transposable elements in eukaryotes. Nat Rev Genet 2017;18:292–308. https://doi.org/10.1038/nrg.2017.7.

[52] Oliveira SG, Bao W, Martins C, Jurka J. Horizontal transfers of Mariner transposons between mammals and insects. Mob DNA 2012;3:1–6. https://doi.org/10.1186/1759-8753-3-14.

[53] Ivancevic AM, Kortschak RD, Bertozzi T, Adelson DL. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. Genome Biol 2018;19:1–13. https://doi.org/10.1186/s13059-018-1456-7.

[54] Gilbert C, Schaack S, Pace JK, Brindley PJ, Feschotte C. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 2010;464:1347–50. https://doi.org/10.1038/nature08939.

[55] Wallau GL, Ortiz MF, Loreto ELS. Horizontal transposon transfer in eukarya: Detection, bias, and perspectives. Genome Biol Evol 2012;4:689–99. https://doi.org/10.1093/gbe/evs055.

[56] Espeland M, Breinholt J, Willmott KR, Warren AD, Vila R, Toussaint EFA, et al. A Comprehensive and Dated Phylogenomic Analysis of Butterflies. Curr Biol 2018;28:770-778.e5. https://doi.org/10.1016/J.CUB.2018.01.061.

[57] Ustyugova S V, Amosova AL, Lebedev YB, Sverdlov ED. Cell line fingerprinting using retroelement insertion polymorphism. Biotechniques 2005;38:561–5.

[58] Lavoie CA, II RNP, Novick PA, Counterman BA, Ray DA. Transposable element evolution in Heliconius suggests genome diversity within Lepidoptera. Mob DNA 2013;21:1–10.

[59] Blass E, Bell M, Boissinot S. Accumulation and rapid decay of non-LTR retrotransposons in the genome of the three-spine stickleback. Genome Biol Evol 2012;4:687–702.

[60] Ivancevic AM, Kortschak RD, Bertozzi T, Adelson DL. LINEs between Species: Evolutionary Dynamics of LINE-1 Retrotransposons across the Eukaryotic Tree of

Life. Genome Biol Evol 2016;8:3301–22. https://doi.org/10.1093/GBE/EVW243.

[61] Evgen'ev MB, Arkhipova IR. Penelope-like elements – a new class of retroelements: distribution, function and possible evolutionary significance. Cytogenet Genome Res 2005;110:510–21. https://doi.org/10.1159/000084984.

[62] Woronik A, Tunström K, Perry MW, Neethiraj R, Stefanescu C, Celorio-Mancera M de la P, et al. A transposable element insertion is associated with an alternative life history strategy. Nat Commun 2019 101 2019;10:1–11. https://doi.org/10.1038/s41467-019-13596-2.

[63] Zhan S, Zhang W, Niitepõld K, Hsu J, Haeger JF, Zalucki MP, et al. The genetics of monarch butterfly migration and warning colouration. Nat 2014 5147522 2014;514:317–21. https://doi.org/10.1038/nature13812.

[64] Hill T, Koseva BS, Unckless RL. The Genome of Drosophila innubila Reveals Lineage-Specific Patterns of Selection in Immune Genes. Mol Biol Evol 2019;36:1405–17. https://doi.org/10.1093/MOLBEV/MSZ059.

[65] Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. Dynamic evolution of the innate immune system in Drosophila. Nat Genet 2007 3912 2007;39:1461–8. https://doi.org/10.1038/ng.2007.60.

[66] Shultz AJ, Sackton TB. Immune genes are hotspots of shared positive selection across birds and mammals. Elife 2019;8. https://doi.org/10.7554/ELIFE.41815.

[67] Kimbrell DA, Beutler B. The evolution and genetics of innate immunity. Nat Rev Genet 2001 24 2001;2:256–67. https://doi.org/10.1038/35066006.

[68] Jurka J, Kapitonov V V., Kohany O, Jurka M V. Repetitive Sequences in Complex Genomes: Structure and Evolution. Annu Rev Genomics Hum Genet 2007;8:241–59. https://doi.org/10.1146/annurev.genom.8.080706.092416.

[69] Kidwell MG, Lisch DR. Transposable elements and host genome evolution. Trends

Ecol Evol 2000;15:95–9. https://doi.org/10.1016/S0169-5347(99)01817-0.

[70] Biémont C, Vieira C. Junk DNA as an evolutionary force. Nature 2006;443:521–4.

[71] Cosby RL, Chang N-C, Feschotte C. Host–transposon interactions: conflict, cooperation, and cooption. Genes Dev 2019;33:1098–116. https://doi.org/10.1101/GAD.327312.119.

[72] Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol Ecol 2013;22:1503–17. https://doi.org/10.1111/mec.12170.

[73] Huff JT, Zilberman D, Roy SW. Mechanism for DNA transposons to generate introns on genomic scales. Nature 2016;538:533–6.

[74] Gilbert C, Feschotte C. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. Curr Opin Genet Dev 2018;49:15–24. https://doi.org/10.1016/j.gde.2018.02.007.

[75] Galbraith JD, Ludington AJ, Sanders KL, Suh A, Adelson DL. Horizontal transfer and subsequent explosive expansion of a DNA transposon in sea kraits (Laticauda). Biol Lett 2021;17:20210342. https://doi.org/10.1098/RSBL.2021.0342.

[76] Le Rouzic A, Capy P. The First Steps of Transposable Elements InvasionParasitic Strategy vs. Genetic Drift. Genetics 2005;169:1033–43. https://doi.org/10.1534/GENETICS.104.031211.

[77] Kofler R, Senti KA, Nolte V, Tobler R, Schlötterer C. Molecular dissection of a natural transposable element invasion. Genome Res 2018;28:824–35. https://doi.org/10.1101/gr.228627.117.

[78] Dupeyron M, Baril T, Hayward A. Broadscale evolutionary analysis of eukaryotic DDE transposons. BioRxiv 2021:2021.09.26.461848. https://doi.org/10.1101/2021.09.26.461848.

[79] Prysby MD, Oberhauser KS. Temporal and Geographic Variation in Monarch Densities: Citizen. Monarch Butterfly Biol Conserv 2004:9.

[80] Tang Z, Zhang HH, Huang K, Zhang XG, Han MJ, Zhang Z. Repeated horizontal transfers of four DNA transposons in invertebrates and bats. Mob DNA 2015;6:1–10. https://doi.org/10.1186/s13100-014-0033-1.

[81] Kelleher ES, Azevedo RBR, Zheng Y. The evolution of small-RNA-mediated silencing of an invading transposable element. Genome Biol Evol 2018;10:3038–57. https://doi.org/10.1093/gbe/evy218.

[82] Goodier JL. Restricting retrotransposons: A review. Mob DNA 2016;7. https://doi.org/10.1186/s13100-016-0070-z.

[83] Eickbush TH, Furano A V. Fruit flies and humans respond differently to retrotransposons. Curr Opin Genet Dev 2002. https://doi.org/10.1016/S0959-437X(02)00359-3.

[84] Ray DA, Grimshaw JR, Halsey MK, Korstian JM, Osmanski AB, Sullivan KAM, et al. Simultaneous TE Analysis of 19 Heliconiine Butterflies Yields Novel Insights into Rapid TE-Based Genome Diversification and Multiple SINE Births and Deaths. Genome Biol Evol 2019;11:2162–77. https://doi.org/10.1093/gbe/evz125.

[85] Levin HL, Moran J V. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 2011;12:615–27. https://doi.org/10.1038/nrg3030.

[86] Fontanillas P, Hartl DL, Reuter M. Genome organization and gene expression shape the transposable element distribution in the Drosophila melanogaster euchromatin. PLoS Genet 2007;3:2256–67. https://doi.org/10.1371/journal.pgen.0030210.

[87] Parfrey LW, Lahr DJG, Katz LA. The dynamic nature of eukaryotic genomes. Mol Biol Evol 2008;25:787–94. https://doi.org/10.1093/molbev/msn032.

[88] Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database

resources of the National Center for Biotechnology Information. Nucleic Acids Res 2014;42:D7-17. https://doi.org/10.1093/nar/gky1069.

[89] Baril T, Imrie R, Hayward A. TobyBaril/EarlGrey: Earl Grey v1.2 2021. https://doi.org/10.5281/zenodo.5718734

[90] Smit AFA, Hubley RR, Green PR. RepeatMasker Open-4.0. Http://RepeatmaskerOrg 2013.

[91] Jurka J, Kapitonov V V., Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 2005;110:462–7. https://doi.org/10.1159/000084979.

[92] Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res 2016;44:D81–9. https://doi.org/10.1093/nar/gkv1272.

[93] Smit A, Hubley R. RepeatModeler Open-1.0. Http://RepeatmaskerOrg 2015.

[94] Bao Z, Eddy SR. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. Genome Res 2002;12:1269–76. https://doi.org/10.1101/gr.88502.

[95] Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics 2005;21:i351–8.

[96] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. BMC Bioinformatics 2009;10:1–9. https://doi.org/10.1186/1471-2105-10-421.

[97] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 2013;30:772–80. https://doi.org/10.1093/molbev/mst010.

[98] Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2009;25:1972–3. https://doi.org/10.1093/bioinformatics/btp348.

[99] Rice P, Longden L, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 2000;16:276–7. https://doi.org/10.1016/S0168-9525(00)02024-2.

[100] Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 2007;35:W265–8. https://doi.org/10.1093/nar/gkm286.

[101] Ou S, Jiang N. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. BioRxiv 2019:2–6.

[102] Wong WY, Simakov O. RepeatCraft: a meta-pipeline for repetitive element de-fragmentation and annotation. Bioinformatics 2018;35:1051–2. https://doi.org/10.1093/bioinformatics/bty745.

[103] Rubino F, Creevey CJ. MGkit: Metagenomic framework for the study of microbial communities 2014.

[104] Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7. https://doi.org/10.1093/nar/gkh340.

[105] Lenhard B, Sandelin A, Carninci P. Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. Nat Rev Genet 2012;13:233–45. https://doi.org/10.1038/nrg3163.

[106] Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033.

[107] Makowski D, Ben-Shachar M, Lüdecke D. bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. J Open

Source Softw 2019;4:1541. https://doi.org/10.21105/joss.01541.

[108] Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 2017;33:3088–90. https://doi.org/10.1093/bioinformatics/btx346.

[109] Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980;16:111–20.

[110] Pfeiler E, Nazario-Yepiz NO, Pérez-Gálvez F, Chávez-Mora CA, Laclette MRL, Rendón-Salinas E, et al. Population Genetics of Overwintering Monarch Butterflies, Danaus plexippus (Linnaeus), from Central Mexico Inferred from Mitochondrial DNA and Microsatellite Markers. J Hered 2017;108:163–75. https://doi.org/10.1093/jhered/esw071.

[111] Wickham H. ggplot2: elegant graphics for data analysis. Springer; 2016.

[112] Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019;4:1686.

[113] Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 2017;45:D200–3. https://doi.org/10.1093/NAR/GKW1129.

[114] Kapitonov V V., Tempel S, Jurka J. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. Gene 2009;448:207. https://doi.org/10.1016/J.GENE.2009.07.019.

[115] Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, et al. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci 1997;94:196–201. https://doi.org/10.1073/PNAS.94.1.196.

[116] Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective

Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 2015;32:268–74. https://doi.org/10.1093/MOLBEV/MSU300.

[117] Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A von, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 2017 146 2017;14:587–9. https://doi.org/10.1038/nmeth.4285.

[118] Team RC. R: A language and environment for statistical computing 2013.

Declarations

Ethics approval and consent to participate: Not applicable

Consent for Publication: Not applicable

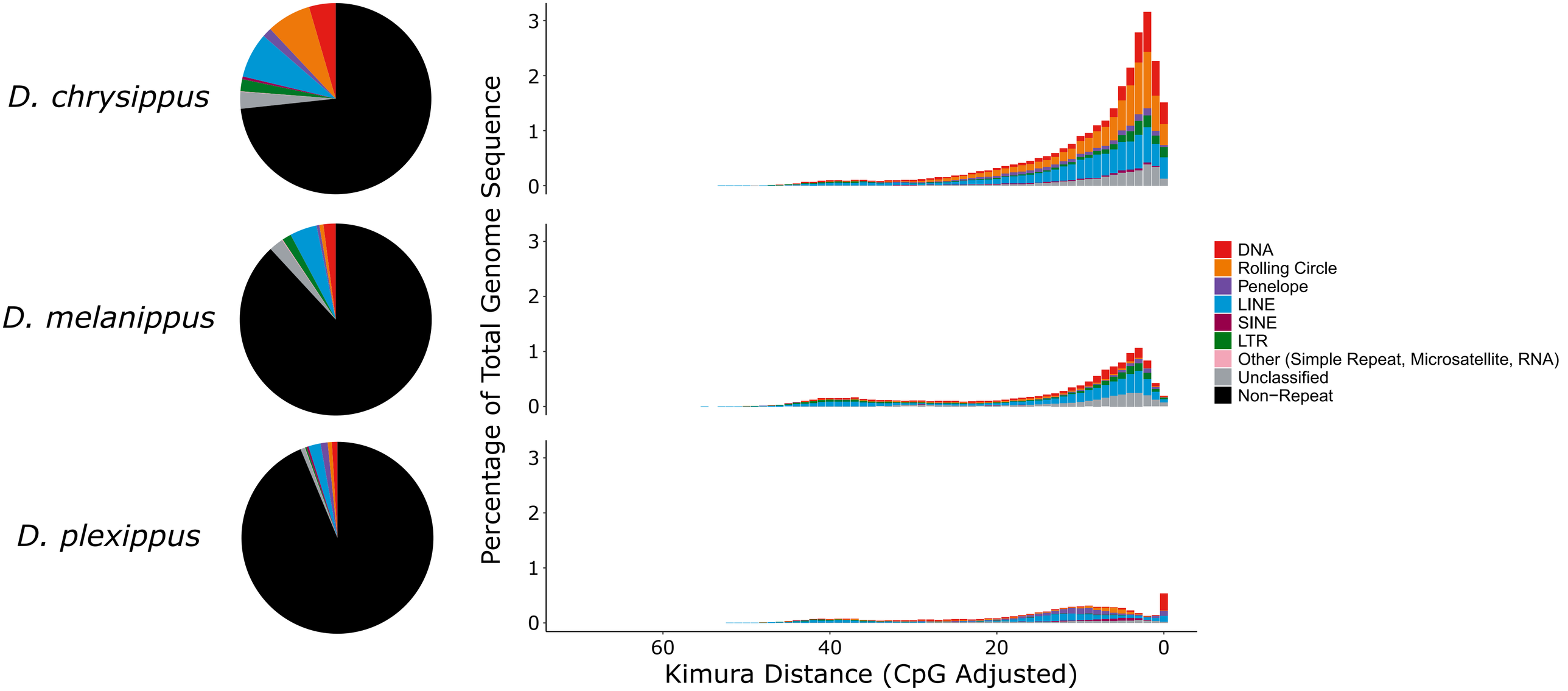
Availability of Data and Materials: All data generated or analysed during this study are included in this article and its supplementary information files.

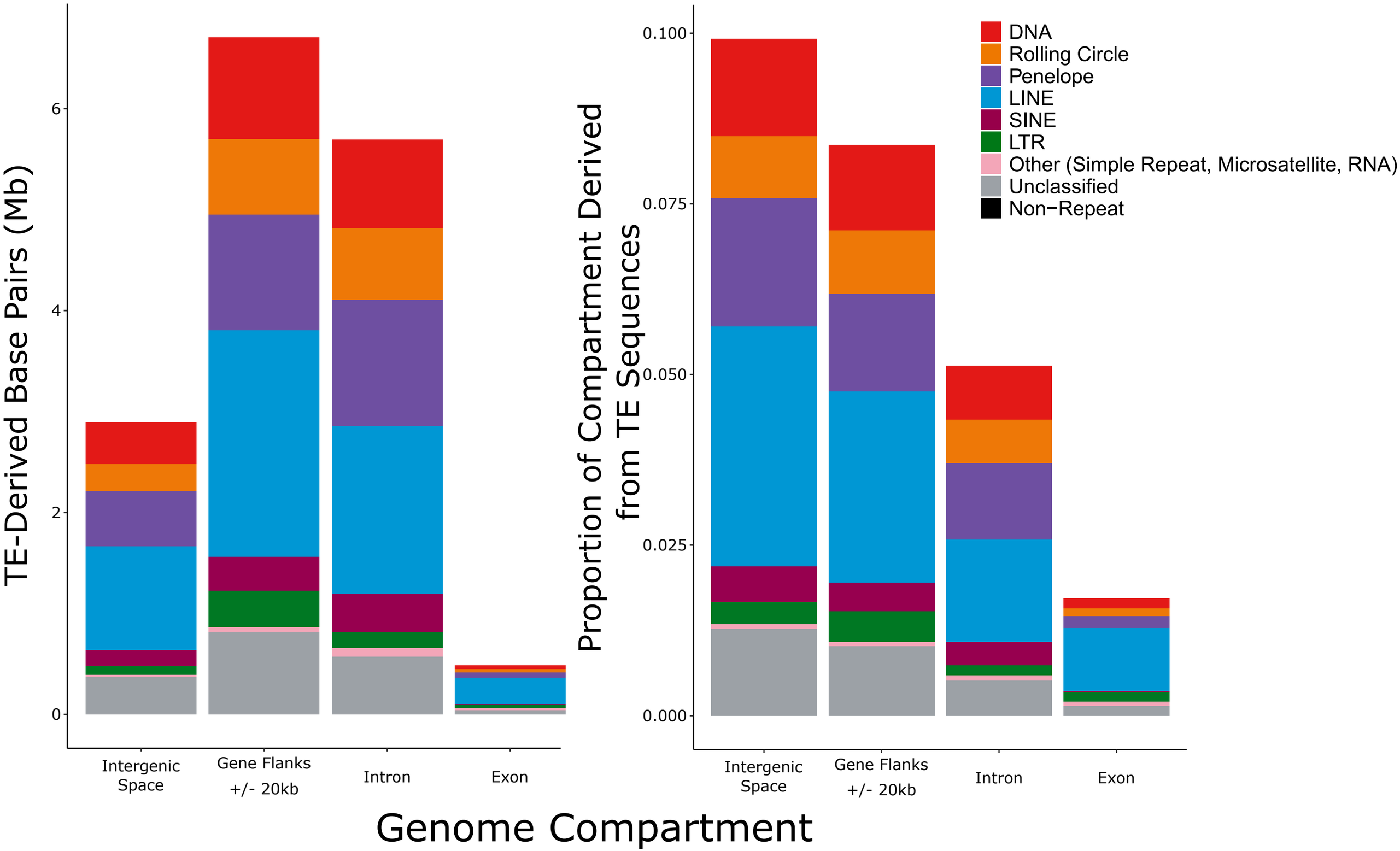
Competing Interests: The authors declare that they have no competing interests.

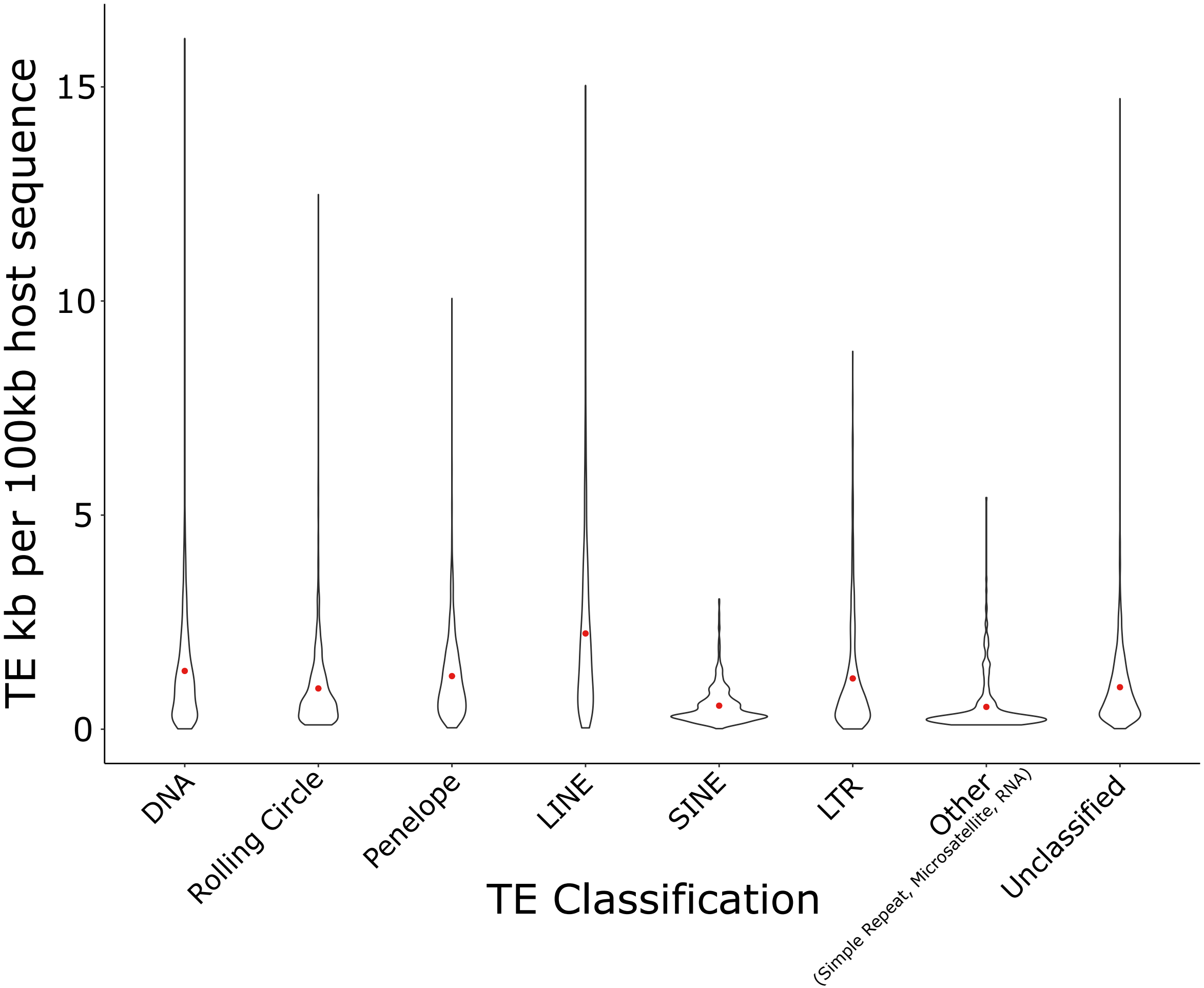
Funding: TB was supported by a studentship from the Biotechnology and Biological Sciences Research Council-funded South West Biosciences Doctoral Training Partnership (BB/M009122/1). AH was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) David Phillips Fellowship (BB/N020146/1).

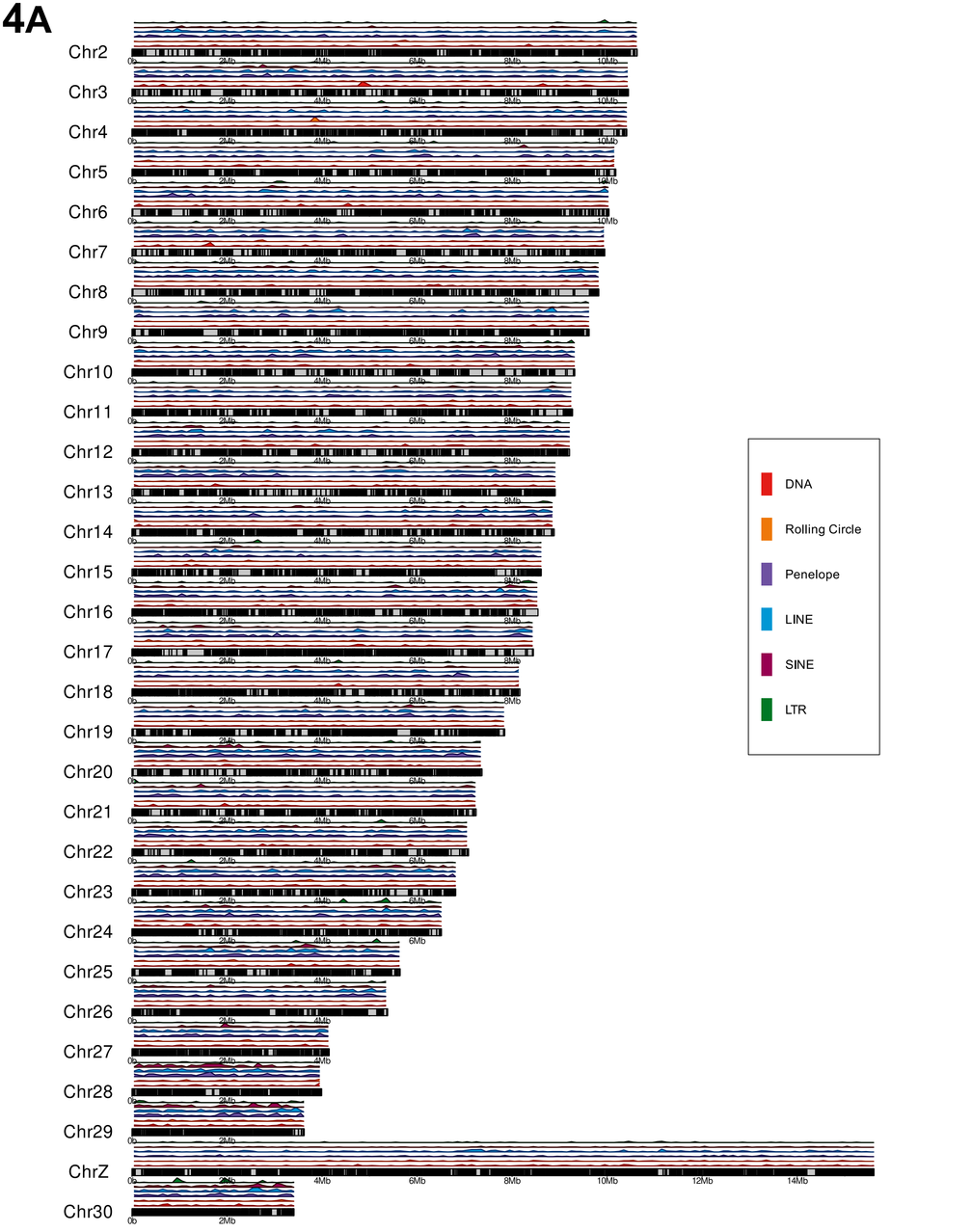
Authors' Contributions: TB developed the pipelines and scripts for automated TE annotation and downstream analyses, produced the figures, and drafted the manuscript. AH conceived and coordinated the study and participated in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements: We would like to thank Dr James Galbraith, Ryan Biscocho, and Ryan Imrie for constructive comments on early drafts of this manuscript, and Dr Aurélie Kapusta for support in running the DNA deletions analysis.

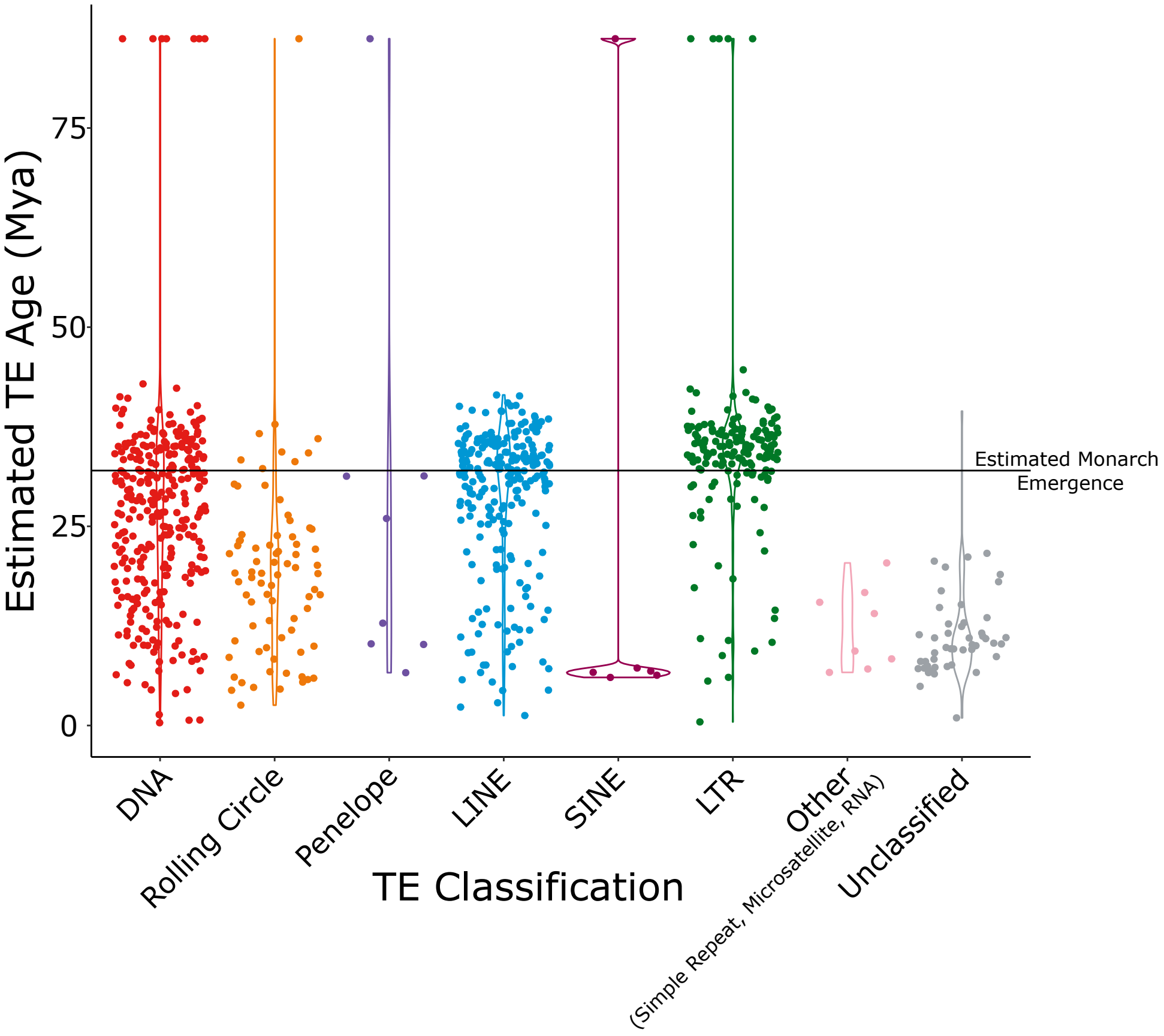

D. chrysippus
D. melanippus
D. plexippus
Percentage of Total Genome Sequence
Kimura Distance (CpG Adjusted)
0
1
2
3
60
40
20
0
DNA
Rolling Circle
Penelope
LINE
SINE
LTR
Other (Simple Repeat, Microsatellite, RNA)
Unclassified
Non−Repeat

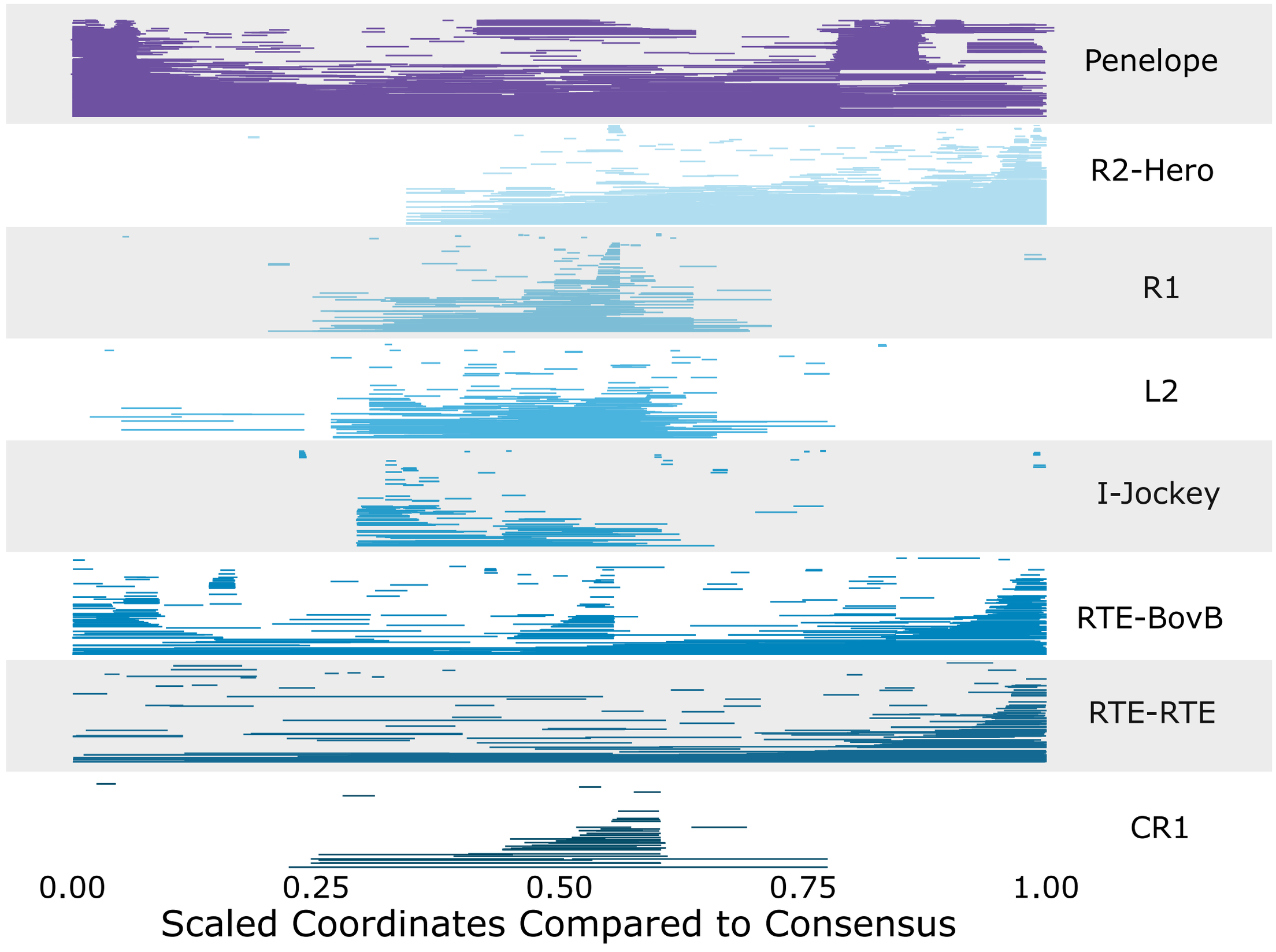

TE-Derived Base Pairs (Mb)
0
2
4
6
Proportion of Compartment Derived from TE Sequences
0.000
0.025
0.050
0.075
0.100
Intergenic Space
Gene Flanks +/- 20kb
Intron
Exon
Genome Compartment
DNA
Rolling Circle
Penelope
LINE
SINE
LTR
Other (Simple Repeat, Microsatellite, RNA)
Unclassified
Non−Repeat


TE kb per 100kb host sequence
15
10
5
0
DNA
Rolling Circle
Penelope
LINE
SINE
LTR
Other
(Simple Repeat, Microsatellite, RNA)
Unclassified
TE Classification


4A
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
Chr23
Chr24
Chr25
Chr26
Chr27
Chr28
Chr29
ChrZ
Chr30
0b
2Mb
4Mb
6Mb
8Mb
10Mb
12Mb
14Mb
DNA
Rolling Circle
Penelope
LINE
SINE
LTR

4B
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
Chr23
Chr24
Chr25
Chr26
Chr27
Chr28
Chr29
ChrZ
Chr30
0b
2Mb
4Mb
6Mb
8Mb
10Mb
12Mb
14Mb
Host Genes
Transposable Elements
TE Hotspot
Host Gene Hotspot
TE Coldspot
Host Gene Coldspot


Estimated TE Age (Mya)
75
50
25
0
Estimated Monarch Emergence
DNA
Rolling Circle
Penelope
LINE
SINE
LTR
Other
(Simple Repeat, Microsatellite, RNA)
Unclassified
TE Classification


Penelope
R2-Hero
R1
L2
I-Jockey
RTE-BovB
RTE-RTE
CR1
0.00
0.25
0.50
0.75
1.00
Scaled Coordinates Compared to Consensus