Klai *et al. Mobile DNA* (2025) 16:9
https://doi.org/10.1186/s13100-025-00346-z

Mobile DNA

RESEARCH Open Access

Marine vs. terrestrial: links between the environment and the diversity of Copia retrotransposon in metazoans

Khouloud Klai[1†], Sarah Farhat[2†], Laure Lamothe[1], Dominique Higuet[2] and Éric Bonnivard[1*]

Abstract

Background LTR-retrotransposons are widely distributed among the eukaryote tree of life and have extensive impacts on genome evolution. Among the three canonical superfamilies, the Copia superfamily demonstrates the lowest abundances and repartitions among metazoans. To better understand their dynamics, we have conducted the first large-scale study of LTR-retrotransposon diversity in metazoans and we report on the diversity and distribution of the Copia elements.

Results We have identified over than 2,300 Copia elements from 263 metazoan genomes. The sequences were annotated at the clade level based on the classification of their RT/RNaseH domain. Our results confirmed that Copia are scarce in metazoans. However, we observed a great variation in Copia abundance between taxa. Surprisingly, some genomes, had a record number of copies, especially in Squamata. In contrast, terrestrial Deuterostomia display a clear loss of Copia diversity leading to their disappearance in some taxa. Additionally, we identified 18 new clades, tripling the number of previously defined clades. By studying more than 50 widespread taxa, we believe that most metazoan Copia clades have now been identified. The most striking result is that environment appears to be related to Copia distribution. We defined two sets of clades characterizing marine or terrestrial taxa. This two-sided pattern could be partially explained by horizontal transfers within both environments.

Conclusions This research enhances our understanding of transposable element evolution and emphasizes the influence of sharing the same ecological contexts on genomic diversity, and highlights the importance of annotating them at the clade level to characterize their evolutionary dynamics.

Keywords Copia-retrotransposons, Metazoans, Environment, Clade

[†]Khouloud Klai and Sarah Farhat contributed equally to this work.

*Correspondence:
Éric Bonnivard
ebonnivard@sb-roscoff.fr
[1]Sorbonne Université, CNRS, ECOPHY, Station Biologique de Roscoff, Roscoff UMR7144 AD2M, 29680, France
[2]Evolution, Bioaffiliationersité (ISYEB) - Sorbonne Université, Institut de Systématique, Muséum National d'Histoire Naturel, CNRS, EPHE, Université des Antilles, Paris Cedex 05 75252, France

Background

Transposable elements (TEs) are ubiquitous in eukaryotic genomes and play a pivotal role in evolution by generating genetic variation through their mobility [1]. TEs exhibit extensive diversity in terms of structural features, sequences, and replication mechanisms [2, 3], which profoundly influence their dynamics (i.e. their ability to invade, maintain themselves, and evolve within host genomes at the level of individuals, populations or species) and success within genomes (large numbers of copies in many species). LTR-retrotransposons transpose via an RNA intermediate and are similar to retroviruses, possessing direct Long Terminal Repeats (LTR) that flank the coding sequences [4]. The *gag* region, at the 5′ end, encodes proteins forming virus-like particles. The *pol* region encodes protein domains essential for the transposition mechanism, including a reverse transcriptase (RT) and an adjacent RNaseH, often grouped together in the same RT/RNaseH domain. This domain is conventionally used to reconstruct LTR-retrotransposon phylogenies [5]. Since these elements require a multi-compound machinery for mobility, genomic copies are likely to be inactivated by mutations, leading to sequence divergence.

Despite their sequence, structural, and mechanistic similarities, LTR-retrotransposons can be divided into three superfamilies (Copia, BEL/Pao, and Gypsy) [6], mainly based on the phylogeny of their highly conserved domains [6–8]. These superfamilies exhibit uneven relative abundances and distributions across metazoans [6, 9], influenced by both the element type and the host taxon. For example, Copia and Gypsy elements seem absent in birds and mammals, although traces of Gypsy elements persist [10–12]. In mammals, the diversity of LTR-retrotransposons is limited to endogenous and exogenous retroviruses [13–15]. Among metazoans, Gypsy elements are clearly the most abundant, with BEL/Pao elements frequently exceeding Copia retrotransposons in abundance [9, 16]. Various criteria can describe TE abundance, such as the number of copies within a genome, the diversity (i.e., number of different subfamilies or clades in a given species or taxon), and the distribution (i.e., distribution of each family or clade among different host species or phyla). Herein, we define a TE subfamily (an element) as a cluster of very similar copies within a given genome. Additionally, a TE family comprises identical subfamilies shared by different species, while a TE clade denotes a monophyletic group of related families across several host species.

Copia elements typically exhibit low copy numbers in metazoans, being absent in approximately one-third of the previously analyzed genomes of metazoan [9, 16, 17]. Consequently, their diversity in number of families and clades is limited. Comprehensive phylogenetic studies of Copia elements have identified two major branches [6]. In metazoans, Branch 1 comprises the GalEa clade, which predominates among Copia families in crustaceans but is also widely distributed across metazoans (Ctenophora, Cnidaria, Mollusca, Polychaeta, Echinodermata, Hemichordata, Tunicata, and Teleostei [18]). Branch 2 comprises at least six metazoan clades. The Hydra clade has been observed in Cnidaria, Mollusca, Polychaeta, Amphipoda, amphibians and teleosts [19, 20]. The other five Copia clades are labeled Arthropoda but contain few insect elements. Potentially, other Copia clades may exist, as evidenced by the discovery of a few new clades in novel host phyla. For instance, the new CoMol clade emerged during investigations of mollusks and subsequently polychaetes [17, 21], which highlights the importance of diversifying the hosts analyzed to obtain the most comprehensive view of the diversity of LTR-retrotransposon clades.

These data raise questions about the distribution of known Copia clades and the potential existence of undiscovered clades in metazoans. Investigating this will allow us to determine whether Copia diversity aligns with the dynamic model previously proposed. This model suggests that Copia elements evolve in metazoans not only through an arms race but also *via* the "Domino Day Spreading model " [5, 16], in which only a few clades persist due to amplification bursts in specific taxonomic groups. Despite the increasing number of sequenced genomes from species of evolutionary and/or ecological significance across various taxa and environments (marine, freshwater, and terrestrial ecosystems), transposable elements remain understudied in many of them. Leveraging this growing genomic dataset, we conducted the first large-scale comparative genomic analysis of Copia in metazoans and explored their phylogenetic relationships and distribution.

Methods

Genomic data

The 263 assemblies (Additional file 1) were obtained from the NCBI Assembly database (https://www.ncbi.nlm.nih.gov/datasets/genome/). The selection of these genomes aimed to comprehensively represent all metazoan phyla while ensuring sufficient assembly quality. Genomes assembled at the chromosome level were prioritized. Subsequently, the panel was supplemented with genomes from less-studied taxa, while efforts were made to adhere to specific criteria wherever possible: an L50 less than 300, a number of scaffolds below 50,000, and an N50 of at least 0.5 Mb (approximately a hundred times the maximum size of a canonical Copia element). Ten species were selected even if they did not fully meet all criteria due to their association with taxa that were less widely available and/or considered emblematic (e.g., the sponge *Aplysina aerophoba* or the tardigrade *Hypsibius*

dujardini). Additionally, a few Copia sequences from Repbase [22] were added, including 41 elements derived from 18 species of insects across various representative orders and 18 elements from 12 teleost species selected across diverse species.

De Novo Copia element identification

The initial step involved identifying candidate LTR-retrotransposons in each genome independently using LTRharvest [23]. Parameters utilized included a LTR length ranging from 100 to 1,200 bp, distance between LTRs ranging from 2,500 to 11,000 bp, similarity threshold of 80%. Candidate sequences were subjected to BLASTX [24] against an in-house RT/RNaseH protein database (40 Copia, 60 BEL/Pao, 68 Gypsy, as well as sequences of DIRS elements and polintons as competitors). The aim of these BLAST steps is to discriminate LTR elements from artefactual sequences obtained with LTRharvest (removal of false positives) and to assign the LTR-retrotransposon sequences to one of the three superfamilies, Copia, BEL/Pao or Gypsy. We used a fairly high e-value (lower than 10E-5), which has been defined to be not very stringent to recover as much LTR-retrotransposon sequences as possible. To the same end, the database used contains a limited but sufficient number of well-characterized sequences covering all the diversity of the main Copia clades of metazoans already known. Sequences were assigned to Copia if more than 8 of the first ten matches were allocated to this superfamily, which corrects the sometimes-high number of false positives obtained with LTRharvest. All identified Copia sequences were clustered within each species using the Usearch tool (-cluster_fast -id 0.8 -sort length -strand both") [25]. Orphan sequences were excluded from further analysis in species with 200 or more Copia copies; otherwise, they were grouped together in a single Orphans-cluster. Sequences within each cluster were aligned with the E-INS-i iterative refinement configuration of MAFFT version 7 [26] and inserts cleaned using an in-house program trimming the nucleotides not conserved in at least 80%. This clustering process was performed for four rounds, then clusters and orphans were recovered for each species, and orphans were excluded from classification analysis in species containing ten or more clusters.

Classification analyses

We followed procedures outlined in previous studies [5, 17–19, 27] to perform classification of RT/RNaseH amino acid sequences of newly characterized Copia elements. The RT/RNaseH domain was first translated following the alignment generated by BLASTX (e-value < 10E-5) against an in-house database of 40 Copia RT/RNaseH (boundaries determined according to those define in the Gypsy Database [28] (Available: http://gydb.org/index.php/Main_Page) to guide the extraction of the domain before translation. For each cluster, multiple alignments of extracted protein sequences were performed using MAFFT, and consensus sequences were generated. After last Multiple alignments of consensus and orphan protein sequences, the few sequences exhibiting too much gaps, stops or frameshift were manually removed from the dataset. Classification analyses were carried out using Neighbor Joining [29] and the pairwise deletion option of the MEGA11 software [30]. Support for individual groups was assessed using non-parametric bootstrap (100 replicates) [31]. All distance trees include the same reference Copia sequences listed in Additional file 2. It should be noticed that we made an operational classification of the various elements rather than a phylogenetic one. For consistency with the ET clades already described (e.g., those in the Gypsy Database), we have continued to use the term clades, even though they are more accurately clans.

Results

Copia abundance

Using LTRharvest, putative Copia elements were detected in only 167 out of the 263 analyzed genomes (Additional file 1). At this point, we do not exclude that some ancient or remnant sequences of Copia were undetected by our methods. Taxonomy influences the presence of Copia elements, as certain taxa are predominantly devoid of Copia elements, while most species of other taxa possess them (Fig. 1, left). Turtles, flatworms, tardigrades and jawless fishes showed no Copia elements. Copia elements were detected in only a few species of Nematoda, Chondrichthyes, or Echinodermata (Additional file 3). In contrast, Copia elements were found in the majority of Cnidaria (18/25), Acari (16/18), and especially Squamata (23/24). The number of detected copies varied considerably ranging from 1 (*Polyodon spathula*, *Eleutherodactylus coqui*) to over 7,000 in the snake *Hydrophis curtus* (Additional file 3). The number of Copia elements per genome is typically low, with three-quarter of species harboring Copia elements having fewer than 100 copies and only 1/10 having more than 500. The relationship between the number of species versus the number of Copia copies in their genome shows a clear decline, with no particular differences across taxonomic groups (Additional file 4), in other words, for all taxonomic groups, the number of Copia copies is high only in a few species and most species have few Copia copies. This decline is also observed in the 25 species containing 100 to 500 copies, suggesting a copy number limit for certain groups such as Cnidaria or Spiralia. One Gymnophiona, two spiders and 13 Squamata exhibit an impressive number of copies; especially snakes averaging

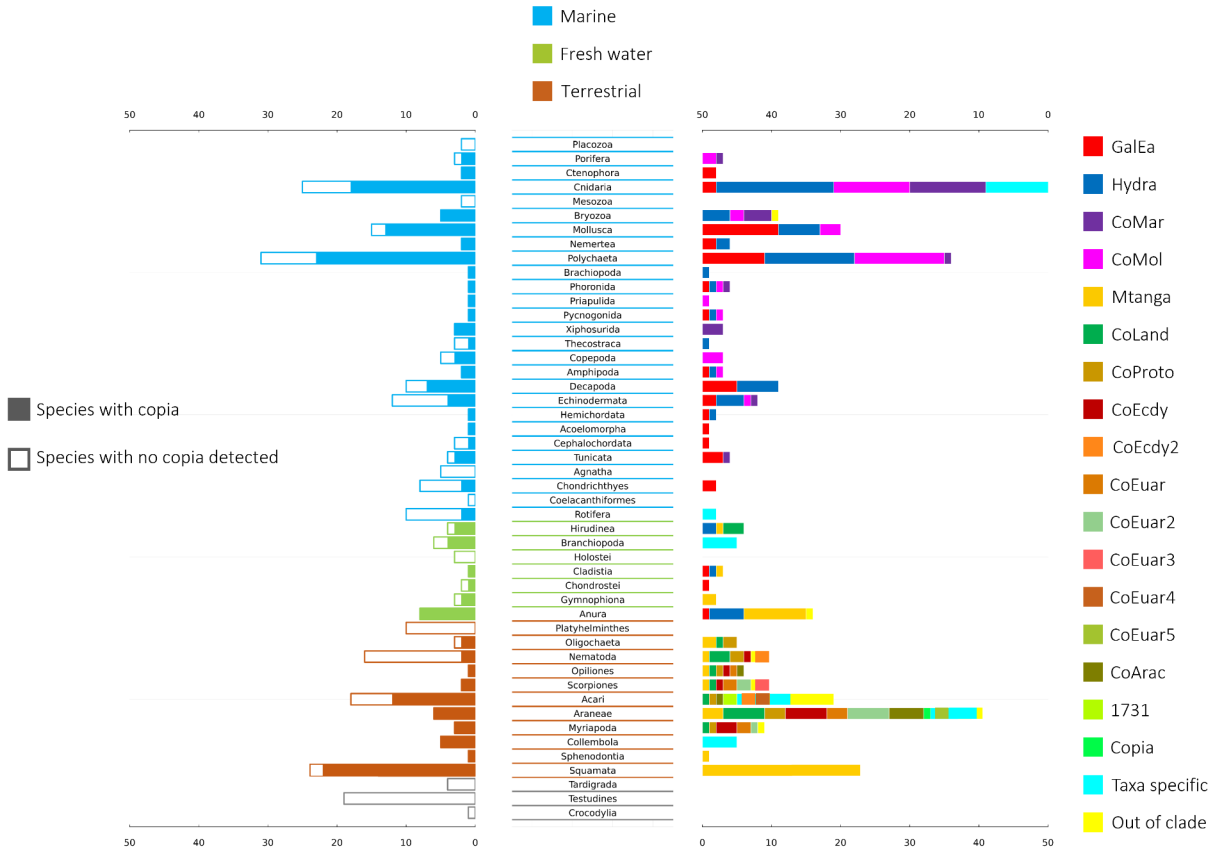

Fig. 1 Distribution of Copia clades among metazoan taxa. Taxa are separated by their overall environment. Left: Number of species with (color) or without (white) Copia elements per taxon, with species considered as harboring Copia elements even if their elements could not be annotated in distance trees. Right: Number of species presenting different clades within each taxon. Since a species may contain multiple clades, the cumulative number may logically be larger than the number of species analyzed

over 2,500 copies per genome (compared to only 740 for spiders, the second-best endowed taxon), constituting more than half of the 49,920 Copia copies detected in all the analyzed species (66% considering all Squamata). The two *Hydrophis* snakes, fully marine species since their divergence from terrestrial relatives (~20Mya) [32], contain the highest number of Copia elements. There is indeed a 7-fold increase in Copia number compared to their closest sequenced terrestrial relative, *Notechis scutatus*. Additionally, a high increase in copy number could be also underlined between the sea snake *Laticauda colubrina* (2,121 copies) and its closest terrestrial relatives *Naja naja* and *Notechis scutatus* (621 and 989 copies, respectively).

All the identified copies were distributed either in a cluster, representing a subfamily, or as an orphan sequence (copy not included in a cluster). As expected, the number of clusters generally increases with the number of copies detected by LTRharvest in each species. However, because Copia elements are rare, the number of copies is often low, resulting in a limited number of clusters. Notably, 80% of species have fewer than 10 clusters. Only 16 species have more than 40 clusters. The two snakes of the *Hydrophis* genus have about the same number of clusters (111 and 91 clusters, respectively, for approximately 7,000 copies) and exhibit several very large clusters with over 500 copies each (the largest clusters in *H. curtus* contain 2,227, 1,086, 807, 686, and 539 copies). For other species, the size of the largest clusters remains more limited, with clusters containing 313, 232, and 186 copies in the snake *Arizona elegans*, and 165 and 75 copies in the spider *Dolomedes plantarius*. The amphibian *Microcaecilia unicolor* (4,102 copies grouped into 14 clusters) and the shark *Scyliorhinus canicula* (719 copies grouped into 8 clusters) have a very low number of clusters relative to their number of copies as they possess a very large cluster containing 95% and 44% of the copies, respectively. Overall, in the remaining species, Copia copies were grouped into small clusters, with less than 10 sequences. The number of orphans is quite high, accounting for approximately 19% of the Copia sequences and representing, on average, 35% of the copies in each genome (Additional file 3). Some species have very few orphan sequences (e.g., less than 6% in

the spider *Hylyphantes graminicola*, the cnidarian Hydra vulgaris, the barnacle *Sacculina carcini* or the snake *H. curtus*) while others exhibit a high proportion of orphan sequences (e.g., 87% in the squamate *Phrynosoma platyrhinos* or 71% in the great white shark *Carcharodon carcharias*).

Copia diversity

In order to identify new Copia clades and to define which Copia clades are present in the various species and taxa, all Copia elements were annotated using distance trees based on translated RT/RNaseH domain. In addition to the analyzed sequences, each tree includes the same 134 reference Copia elements, representing all the major Copia clades currently known for the metazoans (Additional file 2). For 21 species, which contain only orphan sequences or a single cluster, RT/RNaseH sequences were too corrupt to be included in the tree (Additional file 3). The Copia content in these species therefore remains to be determined, for example, by studying the Integrase [21]. Consequently, more than 2,300 Copia elements from 146 genomes were annotated. For illustrative purposes, several trees were first constructed by separating species based on their taxonomy (Fig. 2 and Additional file 5). For seven species having a large number of clusters, specific trees were constructed beforehand (Additional file 6) to select a few representative clusters, which were then integrated into the general trees. Finally, a tree covering all the metazoans was created (Fig. 3).

From all sequences, 24 Copia clades have been identified in metazoans. Clades were defined using the same two criteria as in our previous studies [17, 21]: they must be shared by at least 3 species and have a bootstrap value of at least 70. A detailed description of the diversity of Copia elements observed in each of the different taxa groups represented in the eight trees is presented in the Additional file 7, and for each species a summary of the distribution of Copia elements by clade was compiled in Additional file 3. In most cases, the different clades are easily identifiable and well-supported. The Chelicerata taxa (without the mites) provide a good representation of Copia diversity, since they alone account for 13 clades (Fig. 2A). In total, seven new clades (CoMar, CoLand, CoProto, CoEcdy, CoEuar, CoEuar2, and CoArac) were characterized across multiple taxa. The CoMol and CoMar clades can sometimes be grouped together with a high bootstrap value, for example for the cnidaria or mandibulate trees (see Additional file 5, trees 1 and 6). However, this is not the case in the chelicerates (trees 5). As these two clades are always clearly individualized, we have chosen to define them as two distinct close sister clades. Although we only have the single Mtanga element from *Anopheles gambiae* as the initial reference, this clade is clearly distinguishable in numerous trees. Furthermore, six other new clades appear to be more specific to a single taxon (CoCnid of Cnidaria, CoRot of Rotifera, CoAcari1&2 of Acari, CoAran of Araneae, CoBran of Branchiopoda, and CoCol of Collembola). Two types of taxon-specific clades can be distinguished: those that appear in addition to classical clades, and those that completely differentiate Copia elements of a taxon such as CoRot or CoCol. Apart from the annotated elements, isolated sequences or those forming monospecific monophyletic groups (annotated as 'out of clade') remain few and mostly limited to particular taxa, especially Acari and Insecta.

In order to confirm the diversity established from all eight trees (Additional File 5), a distance tree based on all metazoan Copia sequences was constructed, with sequences labeled either by their clade or as out-of-clade (Fig. 3). The tree obtained does not substantially change the results. Major clades and taxon-specific clades remain well-supported, with only minor changes involving the integration of rare sequences previously annotated out-of-clade: a few Copia of a scorpion and an insect in the Coproto clade, and a few sequences of a myriapod and an insect in the CoEuar2 clade. The main contribution of this new tree, encompassing all metazoans, is the characterization of 4 minor clades (i.e., with fewer than 12 clusters observed across a maximum of 4 species) from the grouping of additional out-of-clade sequences: CoEcdy2 (2 nematodes and 2 mites), CoEuar3 (2 scorpions and 2 insects), CoEuar4 (2 mites and 1 insect), CoEuar5 (2 spiders and 2 insects). Apart from that, the majority of out-of-clade sequences, particularly from mites and insects, remained ungrouped.

Copia distribution

Seventeen metazoan clades show varied distributions among metazoans, which may partly be linked to each clade's representativeness in terms of the number of clusters within different taxa (Fig. 4). Overall, the clades can be divided into four categories: < 10 clusters in total (CoEcdy2, CoEuar3-5, 1731 and Copia), 10 to 20 clusters (CoProto, CoEuar & CoEuar 2, CoArac), several dozen clusters (GalEa, CoMol, CoMar, and CoLand), and >100 clusters (Hydra, Mtanga, CoEcdy). There is not necessarily a direct relationship between the number of clusters and the clade's distribution. Despite being represented by over 160 clusters, the CoEcdy clade has a distribution similar to CoEuar and less extensive than CoProto, which are ten times less diverse in term of clusters. Similarly, the Mtanga clade does not exhibit a significantly broader distribution despite having a record number of clusters, as 90% of them originate from squamates.

The distribution of Copia clades among metazoans does not seem to strictly follow species phylogeny. Indeed, the distribution of each clade is dispersed

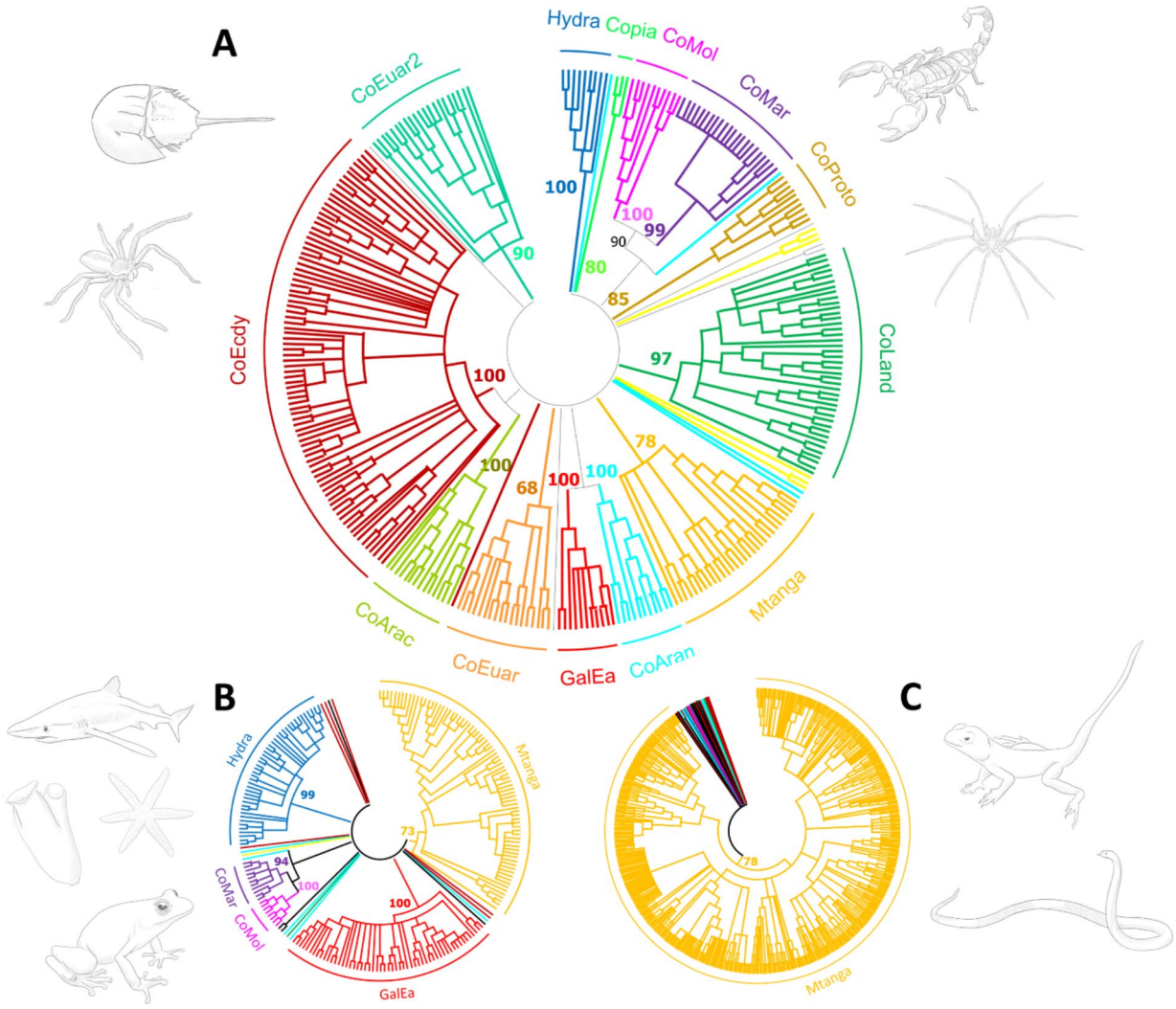

Fig. 2 Classification of Copia retrotransposons. Trees are based on Neighbor-Joining analysis of RT/RNaseH domain amino acid sequences. The Copia clades observed in a taxon are represented by different colors and labelled with their names. 'Out-of-clade' sequences are represented by yellow branches. When a clade contains only Copia reference elements and appears to be absent from the taxon, the resulting sub-tree has been collapsed into a single line, whose color corresponds to that used in the rest of the manuscript. However, taxon-specific clades are all color-coded in light blue. Refence Copia clades from plants are represented by grey lines. Extended representations of these trees are available in the supporting information text, with node statistical support values from non-parametric bootstrapping using 100 replicates. (**A**) Chelicerata (see Additional file 5, tree 5); (**B**) Deuterostomia except Amniota (see Additional file 5, tree 7); (**C**) Amniota (see Additional file 5, tree 8). Drawings were made by Laure Lamothe

throughout the host taxonomy. However, when examined taxon by taxon, a clear dichotomy is observed (Fig. 4): the GalEa, Hydra, CoMol and CoMar clades form a first set because they are similar in their distribution, since they are most often found together in given taxa; while the other clades are also grouped together in different taxa, forming a second set. To relate this clade division to an external factor, each taxon was annotated according to the environment most representative of its evolutionary history, i.e., the predominant environment since radiation (Fig. 4, see discussion). Indeed, the dichotomy is not linked to the habitat of each species, as for example the two *hydrophis* sea snakes appear to only possess Copia from the Mtanga clade like all related terrestrial squamates. However, their taxa can be generally affected to the terrestrial environment due to its evolutionary history. Even among taxonomically close taxa, such as within Chelicerata or Mandibulata, the dichotomy is consistent with this parameter, indicating that the distribution of Copia clades among metazoans is correlated to the environment. The GalEa, Hydra, CoMol, and CoMar clades are almost exclusively associated with marine taxa.

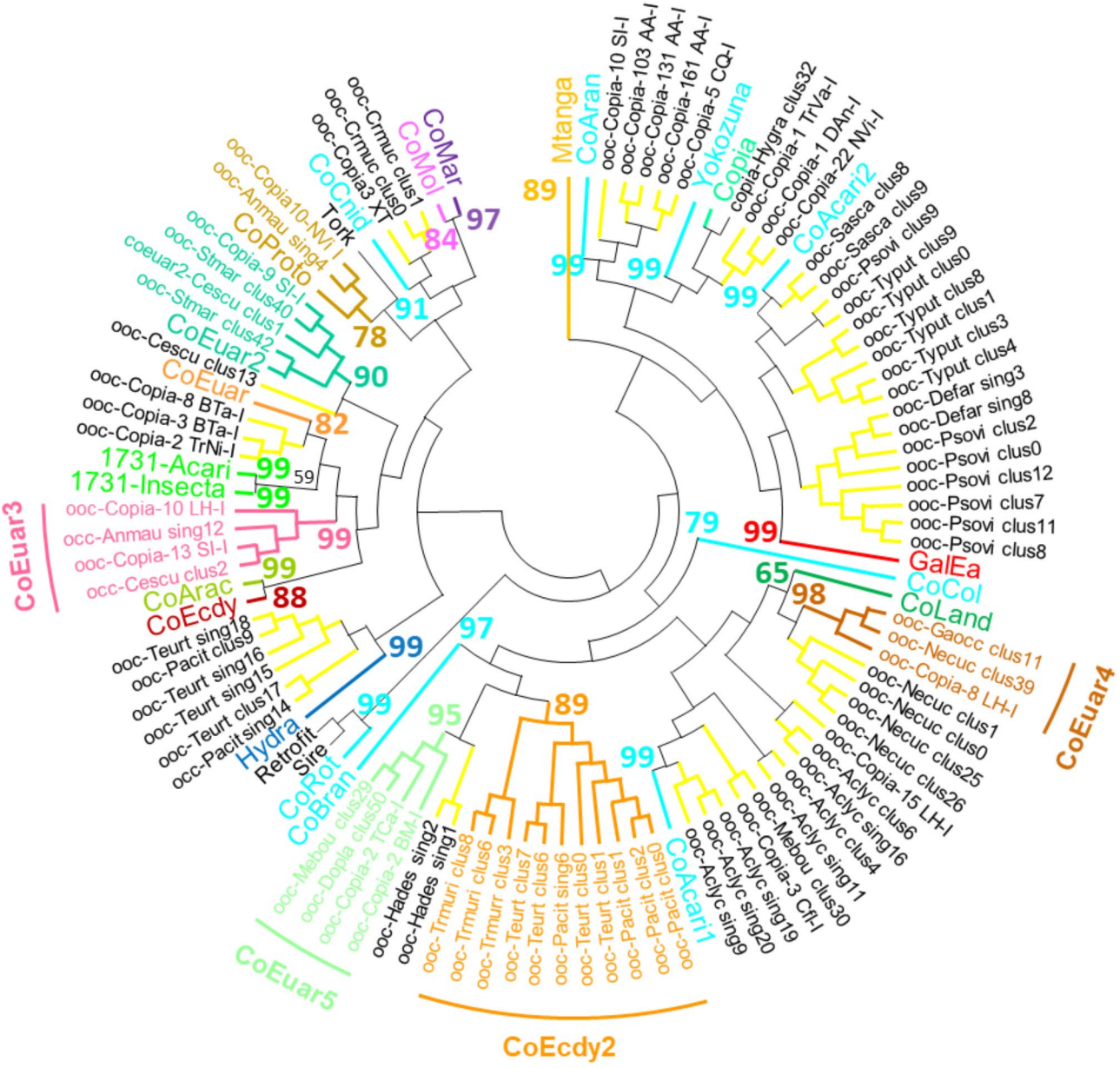

Fig. 3 Diversity of Copia retrotransposons among all metazoans. These distance tree highlight all the elements considered as 'out-of-clade' (tagged with 'ooc-') in previous distance trees (Addfile 3 and 5). Yellow branches represent sequences that are still 'out-of-clade'. For clades containing only correctly annotated Copia elements, the sub-tree has been collapsed into a single-colored line (with light blue representing all the different taxon-specific clades). The CoEuar3, CoEuar4, CoEuar5 and CoEcdy2 are four new minor Copia clades. The tree is based on Neighbor-Joining analysis of RT/RNaseH domain of 1,081 amino acid sequences. Node statistical support values (> 70%) come from non-parametric bootstrapping using 100 replicates

Conversely, the other clades are mainly linked to terrestrial taxa. Among the 40 taxa showing clade diversity, only freshwater ones exhibit a mixture of clades from both sets with the three leeches (Hirudinea), the bichir (Cladistia), and five out of eight frog species (Anura).

The overall representativeness of each major clade in terms of the number of taxa, species, or clusters was compared (Fig. 5). In the case of the marine/brackish environment, the relative importance of the four clades is roughly the same for all three estimators. The sister groups CoMol and CoMar, and the two clades GalEa and Hydra each represent a third of the taxa and species where they are observed. However, the GalEa and Hydra clades appear slightly more important and predominant in terms of the number of clusters. In the freshwater environment, despite the limited number of available species and taxa, GalEa, Hydra, and Mtanga hold comparable importance in terms of taxa representation. However, the importance of the CoLand and especially Mtanga clades increases in terms of the number of species and

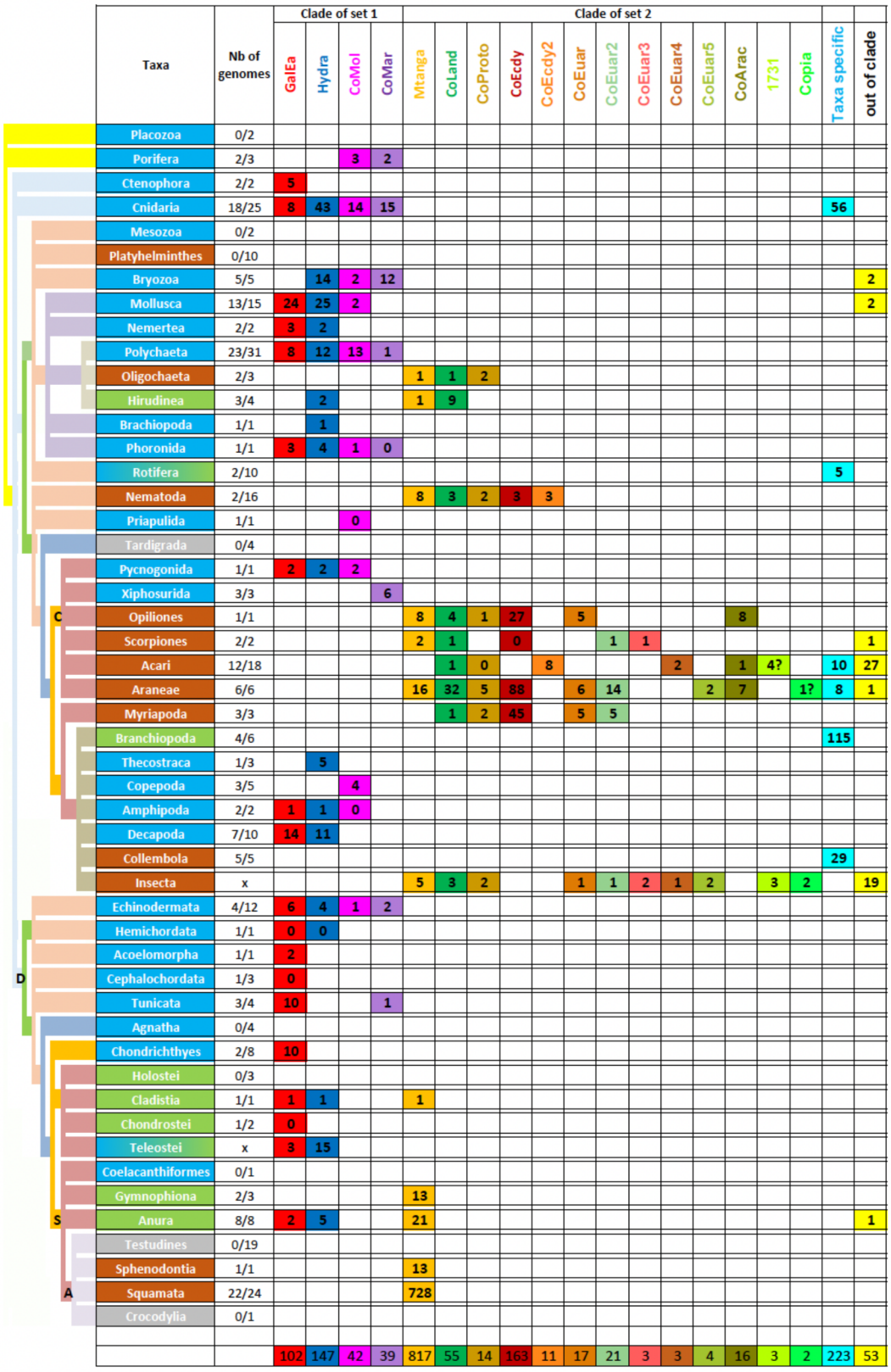

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Dispatching of Copia clusters in clades among metazoan taxa and their environment. Taxa are arranged according to a simplified classification shown on the left (C = Chelicerata, D = Deuterostomia, S = Sarcopterygii, A = Amniota) and their global environment is indicated by colour (marine/brackish in blue, freshwater in green, and terrestrial in brown). The number of genomes represents those with at least one Copia RT/RNaseH sequence identifiable in distance trees out of the total number of species studied. The clades are organized into 2 sets according to their comparable distribution among taxa. Clades represented only by orphan sequences are illustrated with a 0

clusters. These results appear entirely different for the terrestrial environment with much more clade diversity. In terms of number of taxa and species, the main clades are Mtanga, followed by CoLand, CoProto and CoEcdy. However, there is a wide variation in relative abundances depending on the criterion studied, with an explosion of the number of clusters for the Mtanga and CoEcdy clades. This important variability in the terrestrial environment is also reflected in two other results. First, most taxon-specific clades are found in terrestrial taxa such as in mites, spiders, and collembolans; while only the cnidarian taxon-specific clade was found in the marine environment. Second, the out-of-clade sequences are almost exclusively found in terrestrial taxa, with nearly 95% of the clusters (49/50) and orphan sequences (42/51) concerned, particularly for mites and insects (Fig. 4).

To better understand the diversity and the relative importance of each clade in terms of species, their distribution within each taxon was mapped, considering the three environments while maintaining the taxon organization (Fig. 1, right). As expected, the contrast in the repartition of the different clades between terrestrial and marine taxa appears clearly, with freshwater taxa being somewhat intermediate. The noticeable difference between the high diversity observed in most terrestrial taxa and the more homogeneous appearance of the marine taxa is still obvious. When the number of species is considered as a whole (Fig. 5), the total values obtained may only result from the effect of certain taxa that are particularly rich in a type of clade. This is clearly not the case for the marine environment, which shows a certain balance as all marine clades are distributed well within the different taxa. For freshwater taxa, except those represented by very few species or taxon-specific clades, the remaining three taxa are highly mixed. Finally, seven of the terrestrial taxa show huge diversity without any truly dominant clades, which is particularly striking for spiders (Araneae). However, there is a sharp change in amniotes, where diversity seems to decrease significantly since there are no Copia observed in the 20 Testudines or the Chinese alligator; and among the Lepidosauria, only Mtanga is found.

Discussion

Copia clades in metazoans

One of the crucial points when studying the diversity of a type of TEs within a large group like metazoans is the selection of genomes. This choice depends on a balance between the number of genomes studied, their assembly quality, and a sampling that is representative of various taxa; the latter being our main criterion. We compiled a list of taxa based on "The Tree of Life - A Phylogenetic Classification" [33] and selected the most complete assemblies available on NCBI while limiting the number of species to about ten to maintain a balanced distribution among taxa. In cases where a large number of well-assembled genomes were available, we attempted to account for the phylogenetic diversity within the taxon. For some emblematic large taxa that have been poorly studied regarding Copia elements (e.g., Cnidaria, Echinodermata), we increased the number of analyzed species. Three types of taxa were considered separately: (i) Birds and mammals were not included as they do not have Copia elements [6]. (ii) Mollusks and polychaete annelids genomes have recently been deeply studied for LTR-retrotransposons [5, 16]. So, we focused our analyses on phylogenetic groups that have been poorly analyzed (e.g., cephalopods for Mollusca). (iii) Insects and fish have been extensively studied regarding TEs [34–38] and hundreds of genomes are available. Therefore, they will be the subject of two independent studies at a later date and are instead represented in this study with selected Copia sequences from Repbase. For insects, the 41 chosen elements span five major terrestrial clades, and five minor ones, making it the second most diverse taxon in our study after the Araneae, probably covering a greater diversity than would have been possible with a limited sample of 5 or 6 species.

This analysis of LTR-retrotransposon diversity focuses on elements whose sequences are not excessively damaged indicating recent transposition activity. An approach based on LTR recognition seems well-suited, and LTRharvest is known to yield very good results [39, 40]. The number of copies obtained with LTRharvest is an estimate and should only be considered as an indication of the order of magnitude of the number of Copia copies. To maximize the clades detection, we also included orphan sequences, even though many of them appeared very close to cluster sequences, and can therefore provide redundant information. Thus, while orphan sequences may provide important information at the species level, our data show that their usefulness may be much more limited in a broader study. It may therefore be far more interesting to increase the number of genomes analyzed rather than to use orphan sequences for identifying different clades within taxa.

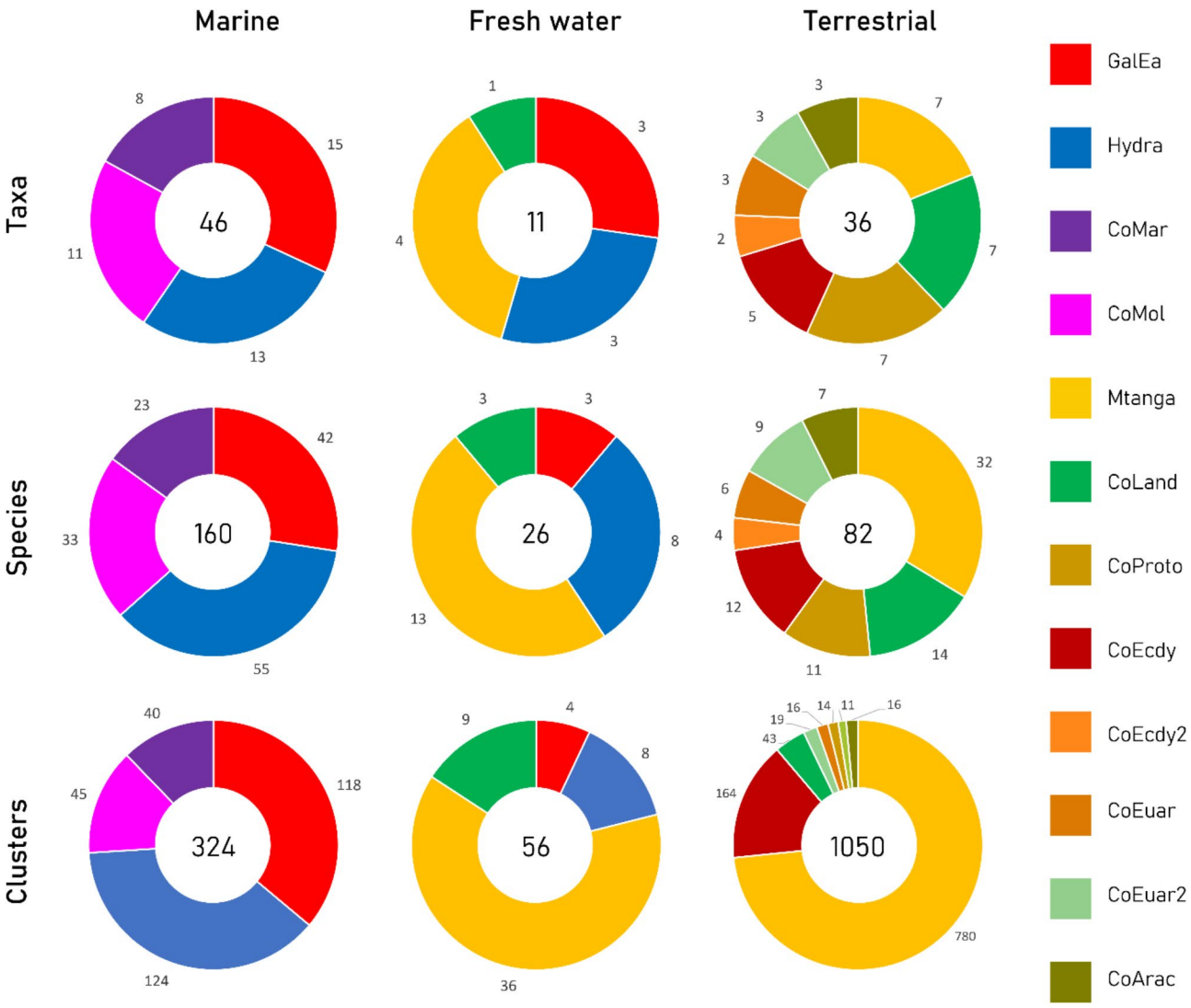

Fig. 5 Relative proportion of major Copia clades among metazoans. The figure presents the number of taxa and species in which each clade occurs for three types of environments, along with the distribution of clusters in different clades. Taxon-specific clades were not considered

The characterization of most of the major Copia clades in metazoans provides a valuable tool for refining the annotation of Copia in eukaryotes, thanks to more than 1,200 annotated reference sequences (provided in Additional file 8). We have doubled the number of widely distributed clades described within the metazoans. It is therefore likely that most of the clades that remain to be defined are either taxon-specific, or correspond to minor clades with a reduced distribution that remain undefinable at our sampling scale, being represented only by out-of-clade sequences. This underscores the information provided, especially by terrestrial taxa for which the diversity of clades defined until now outside insects was very limited [6]. It seems puzzling that the Mtanga clade has been so far underestimated outside of insects, even if Mtanga-like elements were previously described in the reptilian tuatara [41].

It is notable that among Deuterostomia, the Mtanga elements remain the only terrestrial Copia clade (Fig. 2B and C). There is thus a drastic drop in Copia diversity especially among Sarcopterygii, which seems to accompany the disappearance of Copia elements in many taxa (Additional file 9). In addition to birds and mammals, Copia seem also absent in the coelacanth, turtles, and crocodilians (no Copia elements in data from three crocodiles in Repbase). This absence remains confirmed, and also includes Dipneusti (lungfish), following a BLAST search on the nr/nt and tsa NCBI databases (RT/RNaseH sequences of the Mtanga, GalEa1 and Hydra1-1 elements were mapped against the databases). The evolutionary history of this disappearance is not easy to imagine, because the majority of other deuterostome taxa are marine. Based on the study of complete vertebrate mobilomes a strong reduction of TE diversity

was previously described in mammals and birds, even if some other vertebrate lineages contain many TE superfamilies. These results suggest a reduction of TE diversity through elimination of TE superfamilies in the sarcopterygian lineages having led to mammals and birds [14]. The loss of Copia among Sarcopterygii could be linked to the decrease in their diversity, while their maintenance would result only from the remarkable success of Mtanga elements within Lissamphibia and Lepidosauria (Additional file 9). This success could be linked to one or several episodes of strong amplifications, either following the reactivation of rare still active copies or horizontal transfers. So far, among Deuterostomia, only one Mtanga element has been detected outside of Sarcopterygii in the Cladistia *Polypterus senegalus*. This single occurrence is insufficient to conclude that the clade was present in the common ancestor of Deuterostomia. An alternative hypothesis is that Copia elements were acquired horizontally by Sarcopterygii after they had already diverged from a common ancestor. In this scenario, Copia HT would not have occurred in crocodilians turtles, or coelacanths.

Environmental impact

The major outcome of our study is based on the characterization of the overall environment of each taxon, which was conducted a priori and independently of our findings. The difficulty lies in not just making a simple tally of the habitats of the studied species but in considering the evolutionary history of the taxon to which they belong, i.e., the predominant environment since radiation, which ultimately corresponds to the majority of species or sub-taxa. Sequencing project choices may focus on uncommon species and be biased for comparative purposes, meaning that the habitats of sequenced species may not necessarily reflect the environment of the taxon. For example, one of the two sponges (Porphira) we analyzed has a freshwater habitat, whereas only one of twenty-nine orders in this phylum includes a few freshwater species.

For numerous taxa, there is no ambiguity about the environment, for which the information is commonly known. Some taxa are either marine/brackish (e.g., Ctenophora, Brachiopoda, Priapulida, Pycnogonida, Xiphosurida, Echinodermata) or terrestrial (Scorpiones, Opiliones). Other taxa are easily assignable to an environment: many taxa are almost exclusively marine, such as Cnidaria, Polychaeta or Chondrichthyes; the vast majority of leeches (Hirudinea) are freshwater, even though there are terrestrial species and other parasites of marine fish; Arachnids, hexapods, Sphenodontia or Squamata are largely terrestrial, although rare species inhabit coasts or intertidal zones. Dominance of the environment can also be reflected at a higher taxonomic level, with some taxa having almost exclusively marine families, such as Thecostraca (12 marine families out of 13), Acoelomorpha (16 out of 18), or Tunicata (34 out of 35). Defining a global environment can sometimes seem more difficult because there is a large number of species that have colonized different habitats. If we take the example of decapods, many terrestrial and freshwater species are described. However, the number of terrestrial species remains marginal, and most freshwater species are restricted to crayfish (2 superfamilies mainly freshwater), crabs (but 81% of Brachyura are marine [42]), or shrimp (but 78% of Caridea are marine [43]). Therefore, decapods can be labeled as predominantly marine.

The impact of sharing the same environment on the distribution of Copia elements can be revealed at two levels: the marked separation between clades of marine and terrestrial taxa, and twice as many clades identified in terrestrial taxa. This high diversity can be explained by the fact that "land environments tend to incorporate a wider range of environmental heterogeneity (e.g., a wider range of microclimates and microenvironmental conditions; varying levels of restrictions in water supply)" [44]. Additionally, it has been estimated that more than 70% of the planet's species are terrestrial [45], a prevalence that persists even without counting arthropods. In contrast, due to the likely marine origin of life, marine environments have a deeper taxonomic diversity, with many taxa that have not transitioned to terrestrial environment but this does not seem to increase the diversity of Copia clades.

The idea of an influence of sharing the same environment on the presence of a clade has already been mentioned. For example, among all the features of the GalEa-like elements, one of the most striking features was their complex distribution apparently limited to marine environments. This suggested that the "aquatic environment may facilitate horizontal transfers or that some specific evolutionary forces may act in such ecosystems" [9]. By creating genetic homogeneity within these two physically distinct environments, the influence of horizontal Copia transfers could partly explain the dichotomy between the marine and terrestrial clades. This is even more evident given that there does not seem to be any link with the phylogeny of the elements, as the marine clades GalEa and Hydra belong to branches 1 and 2 of the Copia, respectively [6].

Horizontal transfer of transposable elements (HTT) is an important process shaping eukaryote genomes [46–48]. In arthropods, numerous HTT of Copia elements have been detected [49, 50]. Ray-finned fishes (Actinopterygii) contribute to the majority of transfer events in vertebrates (~94%), but there is little evidence that aquatic environments favor HTT [51]. It has been proposed that horizontal transfer of transposons in aquatic

taxa contributes to the distribution of intron-generating transposable elements, known as Introners, which are disproportionately common in the genomes of aquatic organisms [52, 53]. Indeed, Transposable elements and marine taxa are associated with high rates of horizontal gene transfer [54, 55]. The marine environment appears more conducive with lateral mixing processes of water masses, possibly facilitated by ocean currents [51]. For example, it has been demonstrated that the transmission of cancer cells in the marine environment has originated multiple times within and across various species [56]. HTT could thus help in maintaining the four marine Copia clades across metazoans. The presence of HTT in terrestrial environments could likewise support the maintenance and distribution of other clades, potentially with less homogeneity, as observed in collembolans or sarcopterygians, where diversity loss may be apparent.

The effect of horizontal transfers could be reinforced by the dynamic model attributed to Copia elements, the 'domino days spreading' branching process, in which successive amplifications may interact positively. Following this model, even if they are affected by "arms race", Copia elements have dynamics where only a few clades are maintained due to amplification bursts in specific taxonomic groups [16, 18]. The possible role of horizontal transfers in the domino days model has already been suggested as helping to maintain and ensure the success of a certain elements and clades, acting as rescue paths when classical transmission of Copia elements fails due to their low number of copies and diversity.

These results provide a solid basis for studying Copia diversity in metazoans. However, further studies are necessary to elucidate the relation between clades presence, host phylogeny and environment. It would be interesting to analyze additional taxa or delving deeper into others through complementary approaches. In particular, it would be useful to have more freshwater taxa, more diagnostic groups with bi-environments comprising marine and terrestrial taxa, and potentially to target specific taxa. For example, analyzing new taxa phylogenetically close or internal to Hexapoda (e.g., Remipedia, Diplura) could provide a better understanding of the significant difference in clade diversity between insects and collembolans, despite being phylogenetically and ecologically similar. To reinforce our hypotheses, it will also be necessary to integrate the results obtained from ongoing analyses on the diversity of other LTR-retrotransposons, knowing that they are not expected to follow the same dynamics as Copia. Already, the diversity of Copia clades described here appears to be greater than that of BEL/Pao and equal to that of Gypsy (9 and 23 clades described at present, respectively [16]),. However, the diversity of clades described for these superfamilies has only been established in a limited number of taxa and not yet across all metazoans. More anecdotally, it is interesting to note that the nature of Copia elements could provide insights into the evolutionary history of certain taxa, such as Tardigrada, which do not currently display a dominant type of environment.

Conclusion

This research presents the first large-scale study of Copia LTR-retrotransposon diversity in metazoans. By identifying 18 distinct new Copia clades, we likely characterized the majority of the large Copia clades in metazoans, offering a valuable tool for refining Copia annotation in eukaryotes. We revealed significant variability in clade distribution among different taxa, particularly across environments. The overall distribution of each clade is strongly related to the taxa environment, with a clear contrast between terrestrial and marine taxa, while freshwater taxa exhibit intermediate characteristics. This dichotomy between marine and terrestrial clades may suggest a strong influence of horizontal transfer of TEs, with HTT being more prevalent within environments than between them. Our findings enhance the understanding of transposable element evolution and underscore the impact of ecological contexts on genomic diversity.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13100-025-00346-z.

Supplementary Material 1: Additional file 1: List of genomes included in the study. Genomes are listed with their accession numbers, combined sequence sizes (in Mb), the number of sequences included in the analysis, and various metrics. Species classification are color-coded, with bold taxa corresponding to those used in the article. (.xls)

Supplementary Material 2: Additional file 2: Copia elements used as references in distance trees. List of all the reference Copia elements used for phylogenetic analysis, together with the species and taxon name from which they were derived. (.xls)

Supplementary Material 3: Additional file 3: Presence of different Copia clades within species. For each species where Copia sequences were detected, details include the number of Copia sequences, the presence of clades indicated by colored rectangles, and the number of clusters/orphans annotated. Clades are color-coded, and clades unique to a single taxon are shown in turquoise. Sequences that could not be annotated from classification analyses are designated in yellow as 'out of clade'. The table also references the distance tree number and indicates the species most common habitat: marine (brackish) in blue, fresh water in green and terrestrial in brown. (.xls)

Supplementary Material 4: Additional file 4: Distribution of species according to their Copia copy number. The x-axis scale shows steps of 10, 100, and 1000 copies (e.g., 10: species that contain between 1 and 10 copies). (.pdf)

Supplementary Material 5: Additional file 5: Classification relationships of Copia retrotransposons among metazoans. Trees are based on Neighbor-Joining analysis of RT/RNaseH domain amino acid sequences. Reference Copia elements are indicated in black, with their clade specified at the beginning of their name. Branch lines and brackets of different colors represent major clades, while turquoise indicates clades appearing in a single taxon. Yellow branch lines represent out-of-clade sequences. When

a branch is compressed, it means that it contains only Copia of reference. Node statistical support values (> 70%) come from non-parametric bootstrapping using 100 replicates. Taxa are arranged in height different trees according to taxonomy. (.pdf)

Supplementary Material 6: Additional file 6: Classification relationships of Copia retrotransposons within 7 species. Specific trees are provided for species with a large number of Copia clusters, similar to those in Additional file 5. (.pdf)

Supplementary Material 7: Additional file 7: Brief description of the diversity of Copia clades among metazoans according to the eight distance trees (showed in Additional file 5) analyzed and discussed individually. (.pdf)

Supplementary Material 8: Additional file 8: RT/RNaseH amino acid sequences of newly characterized and some reference Copia LTR-retrotransposons. Only the consensus sequences obtained from clusters are given, annotated at the level of the Copia clade according to classification analyses. (.fasta)

Supplementary Material 9: Additional file 9: Loss of Copia diversity in Deuterostomia. Classification of Deuterostomia showing Taxa with (green) or without (red) Copia elements. For each taxon with Copia, the corresponding clades are shown in a colored circle. (.pdf)

Acknowledgements
The authors would particularly like to thank Hervé Le Guyader, François Lallier, and Jean-Yves Toullec for their valuable help in classifying and identifying taxa environments. They also thank the ABIMS bioinformatic platform for providing the necessary support for the analyses. We would like to extend our sincere thanks to Emmanuelle Lerat and Laure Teysset for their valuable help with the revised version of the manuscript. This study was partly carried out with the support of the Brittany Region.

Author contributions
SF, DH, and EB: designed research; KK, and SF: performed research; KK., LL., and EB: analyzed data; KK., SF., LL., and EB: wrote and reviewed the paper; EB: Coordinated the study.

Funding
Sorbonne Université and CNRS provided searchers financial supports. The region of Brittany partly funded KK's salary as part of the SAD BivET project. These funding sources had no role in the design of this study and will not have any role during its execution, analyses, interpretation of the data, or decision to submit results.

Data availability
All data analyzed during this study, along with full references, are provided in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 9 September 2024 / Accepted: 21 February 2025
Published online: 08 March 2025

References

1. Kidwell MG, Lisch DR. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution. 2001;55:1–24.
2. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82.
3. Eickbush TH, Jamburuthugoda VK. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res. 2008;134:221–34.
4. Gao X, Havecker ER, Baranov PV, Atkins JF, Voytas DF. Translational recoding signals between gag and pol in diverse LTR retrotransposons. RNA. 2003;9:1422–30.
5. Donnart T, Piednoël M, Higuet D, Bonnivard É. Filamentous ascomycete genomes provide insights into Copia retrotransposon diversity in fungi. BMC Genomics, 2017;18.
6. Llorens C, Muñoz-Pomer A, Bernad L, Botella H, Moya A. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biol Direct. 2009;4:41.
7. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 1990;9:3353–62.
8. McClure MA. Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol Biol Evol. 1991;8:835–56.
9. Piednoël M, Donnart T, Esnault C, Graça P, Higuet D, Bonnivard E. LTR-Retrotransposons in rimicaris exoculata and other crustaceans: the outstanding success of GalEa-Like Copia elements. PLoS ONE. 2013;8:e57675.
10. Llorens C, Marin I. A mammalian gene evolved from the integrase domain of an LTR retrotransposon. Mol Biol Evol. 2001;18:1597–600.
11. Brandt J, Veith AM, Volff JN. A family of neofunctionalized Ty3/gypsy retrotransposon genes in mammalian genomes. Cytogenet Genome Res. 2005;110:307–17.
12. Lynch C, Tristem M. A co-opted gypsy-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats. Curr Biol. 2003;13:1518–23.
13. Suh L, Smeds H, Ellegren. Abundant recent activity of retrovirus-like retrotransposons.
14. Chalopin D, Volff JN, Galiana D, Anderson JL, Schartl M. Transposable elements and early evolution of sex chromosomes in fish. Chromosome Res. 2015;23:545–60.
15. Dodonova SO, Prinz S, Bilanchone V, Sandmeyer S, Briggs JA. Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. Proc Natl Acad Sci USA. 2019;116(20):10048–57.
16. De la Chaux N, Wagner A. BEL/Pao retrotransposons in metazoan genomes. BMC Evol Biol. 2011;11:154.
17. Thomas Bulle C, Piednoël M, Donnart T, Filée J, Jollivet D, Bonnivard É. Mollusc genomes reveal variability in patterns of LTR retrotransposons dynamics. BMC Genomics. 2018;19:821.
18. Terrat Y, Bonnivard E, Higuet D. GalEa retrotransposons from galatheid squat lobsters (Decapoda, Anomura) define a new clade of Ty1/copia-like elements restricted to aquatic species. Mol Gen Genomics. 2008;279:63–73.
19. Auvinet J, Graça P, Ghigliotti L, Pisano E, Dettaï A, Ozouf-Costaz C, Higuet D. Insertion hot spots of DIRS1 retrotransposon and chromosomal diversifications among the Antarctic teleosts Nototheniidae. Int J Mol Sci. 2019;20(3):701.
20. Grau JH, Poustka AJ, Meixner M, Plötner J. LTR retroelements are intrinsic components of transcriptional networks in frogs. BMC Genomics. 2014;15:1–16.
21. Filée J, Farhat S, Higuet D, Teysset L, Marie D, Thomas-Bulle C, et al. Comparative genomic and transcriptomic analyses of transposable elements in polychaetous annelids highlight LTR retrotransposon diversity and evolution. Mob DNA. 2021;12:1–24.
22. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 2000;16(9):418–20.
23. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de Novo detection of LTR retrotransposons. BMC Bioinformatics. 2008;9:18.
24. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):1–9.
25. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460–1.
26. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
27. Piednoël M, Gonçalves IR, Higuet D, Bonnivard E. Eukaryote DIRS1-like retrotransposons: an overview. BMC Genom. 2011;12:621.
28. Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, et al. The Gypsy database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. 2011;39:D70–4.
29. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.
30. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7.

31. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39:783.
32. Lee MS, Sanders KL, King B, Palci A. Diversification rates and phenotypic evolution in venomous snakes (Elapidae). R Soc Open Sci. 2016;1:150277.
33. Lecointre G, Le Guyader H. The tree of life: a phylogenetic classification. Harvard University Press; 2006.
34. Gilbert C, Peccoud J, Cordaux R. Transposable elements and the evolution of insects. Annu Rev Entomol. 2021;66:355–72.
35. Klai K, Chenais B, Zidi M, Djebbi S, Caruso A, Denis F, Khemakhem M. Screening of Helicoverpa armigera mobilome revealed transposable element insertions in insecticide resistance genes. Insects. 2020;11(12):879.
36. Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, et al. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Ecol Evol. 2019;19:1–15.
37. Carducci F, Barucca M, Canapa A, Carotti E, Biscotti MA. Mobile elements in ray-finned fish genomes. Life. 2020;10(10):221.
38. Shao F, Han M, Peng Z. Evolution and diversity of transposable elements in fish genomes. Sci Rep. 2019;9(1):15399.
39. Lerat E. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity. 2010;104(6):520–33.
40. Ou S, Jiang N. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. Mob DNA. 2019;10:1.
41. Gemmell NJ, Rutherford K, Prost S, Tollis M, Winter D, Macey JR, et al. The Tuatara genome reveals ancient features of amniote evolution. Nature. 2020;584(7821):403–9.
42. Yeo DC, Ng PK, Cumberlidge N, Magalhaes C, Daniels SR, Campos MR. Global diversity of crabs (Crustacea: decapoda: Brachyura) in freshwater. Freshw Anim Divers Assess. 2008;595(1):275–86.
43. De Grave S, Smith KG, Adeler NA, Allen DJ, Alvarez F, Anker A et al. Dead shrimp blues: a global assessment of extinction risk in freshwater shrimps (Crustacea: decapoda: Caridea). PLoS ONE 10(3), e0120198 (2015).
44. Díaz S, Malhi Y, Biodiversity. Concepts, patterns, trends, and perspectives. Annu Rev Environ Resour. 2022;47:31–63.
45. Ruggiero MA, Gordon DP, Orrell TM, Bailly N, Bourgoin T, Brusca RC, Kirk… A higher level classification of all living organisms. PLoS ONE. 2015;10(4):e0119248.
46. Panaud O. Horizontal transfers of transposable elements in eukaryotes: the flying genes. C R Biol. 2016;339:296–9.
47. Schaack S, Gilbert C, Feschotte C, Promiscuous DNA. Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 2010;25:537–46.
48. Wallau GL, Vieira C, Loreto ELS. Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome. Mob DNA. 2018;9:6.
49. Zhang HH, Peccoud J, Xu MRX, Zhang XG, Gilbert C. Horizontal transfer and evolution of transposable elements in vertebrates. Nat Commun. 2020;11(1):1362.
50. Gozashti L, Roy SW, Thornlow B, Kramer A, Ares M Jr, Corbett-Detig R. Transposable elements drive intron gain in diverse eukaryotes. Proceedings of the National Academy of Sciences. 119(48), e2209766119 (2022).
51. McDaniel LD, et al. High frequency of horizontal gene transfer in the oceans. Science. 2010;330:50.
52. Farhat S, Le P, Kayal E, Noel B, Bigeard E, Corre E, Guillou L. Rapid protein evolution, organellar reductions, and invasive intronic elements in the marine aerobic parasite dinoflagellate amoebophrya spp. BMC Biol. 2021;19:1–21.
53. Venner S, et al. Ecological networks to unravel the routes to horizontal transposon transfers. PLoS Biol. 2017;15:e2001536.
54. Metzger MJ, Paynter AN, Siddall ME, Goff SP. Horizontal transfer of retrotransposons between bivalves and other aquatic species of multiple phyla. Proc Natl Acad Sci. 2018;115(18):E4227–35.
55. Peccoud J, Loiseau V, Cordaux R, Gilbert C. Massive horizontal transfer of transposable elements in insects. Proc Natl Acad Sci. 2017;114(18):4721–6.
56. Reiss D, Mialdea G, Miele V, de Vienne DM, Peccoud J, Gilbert C, Charlat S. Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot. PLoS Genet. 2019;15(2):e1007965.

Publisher's note