

# Topological mixing and irreversibility in animal chromosome evolution

Darrin T. Schultz<sup>1,\*</sup>, Arno Blümel<sup>1</sup>, Dalila Destanović<sup>1,2</sup>, Fatih Sarigol<sup>1</sup>, Oleg Simakov<sup>1,\*</sup>

## Affiliations

<sup>1</sup>Department of Neuroscience and Developmental Biology, University of Vienna, Vienna 1010, Austria.

<sup>2</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA.

\*Correspondence to:

Darrin T. Schultz      darrin.schultz@univie.ac.at,  
Oleg Simakov          oleg.simakov@univie.ac.at

**Keywords:** Chromosome, Evolution

## ORCID (alphabetical):

Blümel,	Arno	<a href="https://orcid.org/0009-0000-8672-2319">https://orcid.org/0009-0000-8672-2319</a>
Destanović,	Dalila	<a href="https://orcid.org/0000-0002-3744-5789">https://orcid.org/0000-0002-3744-5789</a>
Sarigol,	Fatih	<a href="https://orcid.org/0000-0002-5842-6114">https://orcid.org/0000-0002-5842-6114</a>
Schultz,	Darrin T.	<a href="https://orcid.org/0000-0003-1190-1122">https://orcid.org/0000-0003-1190-1122</a>
Simakov,	Oleg	<a href="https://orcid.org/0000-0002-3585-4511">https://orcid.org/0000-0002-3585-4511</a>

## Emails (alphabetical):

Blümel,	Arno	arno.bluemel@proton.me
Destanović,	Dalila	dalila.destanovic@yale.edu
Sarigol,	Fatih	fatih.sarigoel@gmail.com
Schultz,	Darrin T.	darrin.schultz@univie.ac.at
Simakov,	Oleg	oleg.simakov@univie.ac.at

## Abstract/Summary Paragraph:

Animal chromosomes can persist with recognizable homology over hundreds of millions of years, in spite of homology-obfuscating processes such as chromosomal fusion and translocation. The frequency and pace of these major genome structural changes are unknown, and it remains unclear whether or how they impact long-term genome evolution. Here, we compare whole chromosomal sequences of 3,631 genomes from 2,291 species spanning all major animal clades and show that animal karyotypes evolve primarily via karyotype contraction, associated with increased rates of chromosomal fusion-with-mixing and dispersion that largely obey chromosomal algebra<sup>1</sup>, or karyotype expansion, which breaks up ancestral linkage groups and forms new chromosomal elements via non-algebraic changes. We show that chromosomal changes can be associated with major extinction events. Using a multi-scale encoding of pan-animal genome homology and a manifold representation of genomic changes, we find that genome evolution is not only driven by changes at the chromosomal level, but that subchromosomal mixing and irreversibility define clade-specific evolution. Using this ‘evolutionary genome topology’ approach, we calculate extrema of irreversible genomic configurations and identify species that occupy intermediate manifold positions, providing evidence for distinct macro-evolutionary trajectories. We propose that investigation of mixed state accumulation around important gene loci (such as Hox) will be crucial in capturing and further study of clade-specific regulatory innovations.

## Introduction

Animal chromosomes have remained remarkably conserved since the last common animal ancestor<sup>1-3</sup> that existed over 700 million years ago<sup>4</sup>. Some of the ancient animal chromosomal homologies date back even further to the ancestor of the Filozoa 804 million to 1.18 billion years ago<sup>4</sup>, as animal chromosomes share synteny with pieces of chromosomes<sup>1,3</sup>, or whole chromosome arms<sup>3</sup> in their unicellular relatives (e.g., the filasterean amoeba *Capsaspora*). These portions of chromosomes that are conserved over time across many species are called ancestral linkage groups<sup>5</sup> (ALGs), or chromosomal elements<sup>6</sup>. Many animal chromosomes can be explained by four algebraic<sup>1</sup> combinations of ancestral chromosomal elements. A key property of chromosomal algebra is fusion-with-mixing (FWM), a statistically irreversible process where genes on a fused chromosome mix, and can be used as strong phylogenetic characters<sup>1,3</sup>. However, the prevalence of chromosomal algebra and long-term (macro-)evolutionary effects caused by these chromosomal changes<sup>1</sup>, and the changes in the selective landscape they create remain poorly understood.

Inferring algebraic changes in chromosome evolution relies on detectable chromosomal homology, which is eroded over time through chromosomal rearrangements like translocations. Inter-chromosomal translocations are considered to be rare events as a portion of gametes produced after translocation are unbalanced and selected against<sup>7</sup>. Therefore, organisms that reproduce sexually or have not passed through substantial population bottlenecks are expected to retain their ancestral chromosomal complement, while asexually reproducing organisms may avoid this selection and more quickly fix translocations. Quantification of chromosomal translocations (**Supplementary Information 1, dispersal**) across different animal clades, however, has been hampered by the lack of chromosome-scale assemblies. The only metazoan baseline estimate for gene dispersal out of chromosomes is 1% per 40 million years<sup>1</sup>. However, whether and how this rate is different among clades or chromosomal element sizes has not been studied. Only now, with the increased availability of chromosome-scale genomes, are we in a position to study and quantify the evolutionary history of chromosomal homology by measuring dispersion in extant organisms’ genomes.

On top of these algebraic descriptions, several transitions in animal karyotypes have been reported that do not fit clearly within the framework of the four algebraic chromosomal changes. In this manuscript, we refer to these as non-algebraic changes, where fissions and/or large inter-chromosomal translocations break the chromosomal element homology. In particular, there are animal clades in which the ancestral animal chromosomal elements have partly or wholly dispersed or rearranged beyond recognizability. Clades with such rearrangements include the genome reshuffling in cephalopods<sup>8</sup>, the derived karyotype of ctenophores<sup>3</sup>, the ‘Merian elements’ of lepidopterans<sup>9</sup>, the ‘Nigon elements’ of nematodes<sup>10</sup>. Clades with rearrangements beyond recognition include the ‘Muller elements’ of drosophilids<sup>6,11</sup>, and clitellate annelids<sup>12</sup>. Whether and how these transitions can still be explained by the core algebraic processes involving complete ancestral chromosomal elements, or whether a non-algebraic breakage of chromosomes is required, remains to be investigated. In either case, these departures from ancestral animal chromosomal elements, some of which occur rapidly on short evolutionary branches<sup>12</sup>, can result in what we would recognize as new sets of clade-specific chromosomal elements. Whether and how these saltatory, rapid transitions are followed by stasis or further algebraic or non-algebraic changes remains unclear.

The presence of algebraic and non-algebraic processes as two major possibilities to change genomes poses the question of whether and how such chromosomal rearrangements have any long-term impact on the evolution of functions such as gene regulation. No global functional enrichment in any chromosomal element has been reported<sup>1,3</sup>. While this suggests that chromosome-level changes could be explained as a random process, chromosomal fusion-with-mixing may have a fundamental and barely studied outcome: by both expanding the chromosomal coordinate system and enabling new combinations of genes to enter regulatory interactions, it may drive phenotypic innovations<sup>13</sup> over the long-term.

There are known regulatory constraints on local genome structure, including co-expression<sup>14</sup> and co-regulation<sup>15</sup> and it has been suggested that genome-wide changes in chromosomal composition may help drive<sup>16</sup> or strengthen<sup>17</sup> novel genomic interactions. The functional importance of the conservation and change in the animal genome architecture among different animal clades remains a nascent research topic due to the limited gene regulatory information for many of these clades. However, it is clear that ancient local gene linkages can be conserved as functional regulatory units<sup>15,18</sup> and that chromosomal changes can<sup>13</sup> and do<sup>19</sup> facilitate the evolution of novel regulatory linkages and phenotypic novelties<sup>19–21</sup>.

Lastly, while theoretical models have suggested that large-scale genomic changes could be correlated with extinction and speciation<sup>22</sup>, there is little evidence connecting these patterns over macro-evolutionary time<sup>23</sup>. Furthermore, there is no explanation of how changes in species diversity may be reflected in the genomic record, including the alleged<sup>24</sup> 62 million year<sup>25,26</sup> periodicity of species extinction and origination. Understanding what type, if any, of chromosomal rearrangements are associated with speciation or extinction, and correlating these changes with patterns seen in the paleontological record, are key steps toward understanding the forces inherent in chromosome evolution. New methods are needed to resolve the outstanding question of the degree to which chromosomal rearrangements cause, or are consequences of, speciation and extinction.

Using a broad sampling of 3,631 metazoan chromosome-scale genomes, we address these questions by identifying general trends of algebraic and non-algebraic changes. Determining the time scale and relationships among large chromosomal changes and the evolution of new regulatory linkages is key to

understanding the evolutionary potential of animal chromosomes and the emergence of novelties (**Fig. 1**). In this manuscript, we quantify this time scale, and characterize the macro-evolutionary impact of large chromosomal changes. We find that these large chromosomal changes are correlated with species origination and extinction in different clades, and find that cycles in chromosomal rearrangement corroborate cycles in species diversity from the fossil record. We generalize the existing algebraic approach to involve both chromosomal and sub-chromosomal scales, governed by meiotic and regulatory constraints, by applying manifold (topological) approaches to study genome evolution. Our results show that each clade can be characterized by its own set of ‘mixed’ chromosomal and sub-chromosomal characters. This approach augments fusion-with-mixing, which is a process at the chromosomal level, and reveals a similar process of mixing and entanglement at the sub-chromosomal level. Our findings suggest that animal genome evolution to be a stepwise process of accumulation of irreversibly mixed states at different levels of genomic organization. We term this approach ‘evolutionary genome topology’, and propose that irreversible mixing is the driving force behind the observed macro-evolutionary changes, or modes, of genome evolution.

## Results and Discussion

### *Algebraic representation and completeness of metazoan genomes*

We first sought to construct an overview of animal genome representation in the context of the Bilaterian-Cnidarian-Sponge ancestral linkage groups (BCnS ALGs<sup>1</sup>, **Supplementary Information 1**). Using available chromosome-scale genomes for 2291 metazoan species (**Methods - Constructing a genome database**), we inferred orthology with BCnS chromosomal elements (**Methods - Identifying BCnS ALGs**), as well as the BCnS chromosomal element persistence, fusions, and fissions (**Methods - Inferring ALG evolutionary history**). Among 406 possible combinations of BCnS ALG fusion pairs, we were able to detect 397 as either fused (●) or fused-and-mixed (⊗) in at least one species.

First considering subclades of the Metazoa, the Cnidaria are known for their well-preserved BCnS ALG chromosomal complement<sup>1</sup>. For a set of 17 chromosome-scale cnidarian genomes (n=17 species), all 29 BCnS ALGs were detected in at least one species, totaling 107 out of 406 possible ALG fusion combinations, meaning that 299 ALG fusion combinations did not appear in any of the 17 cnidarian species. Of the 107 possible ALG fusion combinations, 46 (33% of 107) were species-specific. These estimates of fusion represent a lower limit, as these counts can be impacted by species with poorly assembled chromosomes, or chromosomes that have rearranged to the point where certain ALGs are dispersed across chromosomes. In nematodes, a clade known for its genome contraction and proposed rearrangement-rich history<sup>10</sup>, 28 ALGs were detected in at least one of 37 species. However, the ALG composition varied strongly between nematode species. The highest number of retained ALGs was 27 in a single species, the median was 19 ALGs per species, compared to 28 in cnidarians. Similarly, in strong contrast to cnidarians, 5 nematode species completely lacked recognizable ALGs. In nematodes, 101 out of total 378 possible fusions were present, 34 of which were species-specific. Of these 101 fusions in nematodes, 76 were fusions also seen in at least one cnidarian genome, predictably so given the convergence expected from the low karyotype number in nematodes. These data point to different levels of ALG retention across metazoans, but interestingly also highlight a similar level of algebraic combination space exploration, and thus algebraic representation, of nematodes and cnidarians.

Across our dataset we found only nine ALG combinations that have not been fused in any of the species. These unobserved fusions include the combinations A1a⊗●Qb, B1⊗●Qb, B1⊗●Qc, C2⊗●G, J1⊗●Qb, O1⊗●Qb, O1⊗●Qc, O2⊗●R. We note that the BCnS ALGs R, Qb, and Qc are among the smallest BCnS ALGs with 24, 12, and 14 constitutive orthologs. Our ability to detect such fusions is affected by dispersion of genes to non-homologous chromosomes.

### *The time scale of chromosomal element homology dispersal*

We next characterized dispersal, a process in which inter-chromosomal translocations cause genes to move from one ALG to other chromosomes. We investigated these properties by computing the dispersion state in the animal genomes (n=3,631 genomes) in relation to the BCnS ALGs (**Methods - Dispersal characterization**). The average dispersion of genes relative to the BCnS ALG complement is as low as 0.2% for chromosomally highly preserved genomes like the scallop *Mimachlamys varia* (**Fig. 2b**, **Extended Data Fig. 1**), to between 53%-73% for genomes in the nematode genus *Caenorhabditis*, reflecting partial BCnS ALG loss, and 100% for species like *Drosophila* or *Anopheles* whose genomes are highly rearranged and the BCnS ALGs are no longer detectable (**Fig. 2**).

The total dispersal scales with evolutionary time and the lineage in question. For example, a comparison of the genomes from two closely related bivalve species exhibit very little dispersal. Comparing genomes of animal or unicellular species with a common ancestor at the Choanozoa, Filozoa, or Holozoa nodes (**Extended Data Fig. 1**) reveals that very few chromosomal elements are conserved.

We find that the number of dispersed orthologs is independent of the chromosomal element size (**Fig. 2**). This has several important implications. First it enables predicting the life span of a given chromosomal homology. For example, because larger chromosomes have more positions available for homology detection, their homology can remain detectable longer than that of the smaller chromosomes. The smallest chromosome of the scallop *Pecten maximus* (18 BCnS orthogroups in 32.5 Mbp, BCnS ALG C2), for example, is unrecognizable in species that diverged over 550 million years ago, while the homology with larger *Pecten* chromosomes remain detectable (**Fig. 2**). For these reasons, while even the smaller chromosomes are still likely to persist within animals, without special analyses<sup>3</sup> the same chromosomal homologies are completely undetectable in animal-unicell comparisons. The second implication is that for smaller chromosomes, homology detection will rely either on a shorter evolutionary distance or, with larger evolutionary distance, will rely more on them having fused with other chromosomes. When a small chromosome fuses with another, the probability that its genes are dispersed is reduced by the ratio of summed lengths of the two original chromosomes versus the total genome size. One example is the bilaterian linkage group (BLG) Q element<sup>1,27</sup>, which became colocalized on the same chromosome in the ancestor of Chordates from four very small elements (Qa, Qb, Qc, Qd) that are also found colocalized in various combinations in other animal clades. Under this model, smaller ancestral elements are predictably undetectable in certain clades, a prominent example being BCnS ALG R, which is not present in chordate genomes<sup>1</sup>. Together, these data clarify the time decay of BCnS chromosomal element homology, and highlight the process of dispersal that happens along algebraic chromosome evolution.

### *Uneven rates of chromosomal element evolution*

Slow chromosomal evolution is a marked feature in animals<sup>1,27</sup>, and is characterized by the presence and preservation of the BCnS ALGs, each on a single or a small number of chromosomes. However, it is unclear how much the fusion and dispersal rates can vary. To assay the evolutionary history across the whole animal tree we developed a phylogeny-aware method of ALG fusion rate estimation (**Methods - ALG evolution simulations**, **Extended Data Fig. 2**) and applied it to our dataset (**Fig. 3a,b**). These results show similar fusion rates within each clade for any given BCNS ALG combination. We did not find a clear preference of smaller BCnS ALGs to undergo fusions. However, using a phylogenetically-aware method (**Methods**), we observed general discrepancies between fusion rates in various clades. For example, in sponges the fusion rate is 1.07 fusions/million years whereas in spiralian it is lower at 0.8825 fusions/million years (**Extended Data Fig. 2**).

Our results also show that chromosome evolution in animals did not occur at a steady rate, and that the processes of fusion and homology loss through dispersal vary in timing and intensity (**Fig. 3b**). For example, we find that the rate of BCnS ALG changes through both dispersal and fusion are highest in the branches leading to the last common ancestors of major extant clades (**Fig. 3b,c**). Notably, across all animals the averaged rate of BCnS ALG fusions increased during the end-Ediacaran 550 million years ago and peaked at the beginning of the Cambrian (approximately 1 fusion per 8 million years). The rate of chromosomal element fusions lowered by the end of the Cambrian period, and the rate varied (1 fusion per 32±16 million years) until the end of the Mesozoic, 66 million years ago. During this period, the rate of homology loss through dispersal was lower than the rate of fusions by a factor of two, with peaks of homology loss in the Cambrian (1 loss per 22 million years), Devonian (1 loss per 64 million years), and Permian (1 loss per 32 million years) periods.

Such cases of increased rates of chromosome rearrangements can result in observed saltatory changes along an evolutionary branch, in which the genomes of the extant species are highly rearranged relative to typical animal chromosomes with intact BCnS ALGs, and these changes are restricted to particular



epochs. In some clades, for example the Panarthropoda, after an initial burst of fusions and losses at the beginning of the clade diversification during the Cambrian approximately 580 million years ago<sup>28</sup> (1 fusion per 4 million years), these changes continued to happen at a rate twice as high as found averaged across animals (1 fusion and 1 loss per 16 My - **Extended Data Fig. 3d**). In other clades, like mosquitos or clitellate annelids, the changes appear to happen so rapidly that we did not observe remnants of transitional genomic states in closely related species, and therefore can only estimate lower bounds of rates of change along these branches. For example, clitellate annelid genomes are dramatically rearranged relative to the genomes of their close polychaete relatives<sup>12</sup>, and none of the 29 ancestral BCnS ALGs are recognizable.

With the chromosomal evolutionary rates across the history of animal evolution in hand, we tested for a correlation between the observed rate of chromosomal changes and species origination or extinction<sup>26</sup> (**Methods - Correlation inference with species diversity**). In one of the most species-rich groups in our dataset, the protostomes (n=1828 genomes), we found that chromosomal fusion rates are correlated with species extinction (one-tailed Spearman's rank correlation coefficient,  $r_s = 0.53, p < 1 \times 10^{-40}, n = 542$  million-year bins) and species origination rates (one-tailed Spearman's rank correlation coefficient,  $r_s = 0.29, p < 1 \times 10^{-10}, n = 542$  million-year bins). However, in vertebrates (n=1,598 genomes), both chromosomal fusion and loss rates are anti-correlated with species extinction (one-tailed Spearman's rank correlation coefficient, BCnS ALG fusions  $r_s = -0.29, p < 1 \times 10^{-11}$ , BCnS ALG losses  $r_s = -0.51, p < 1 \times 10^{-37}, n = 542$  million-year bins) and origination rates (one-tailed Spearman's rank correlation coefficient, BCnS ALG fusions  $r_s = -0.46, p < 1 \times 10^{-29}$ , BCnS ALG losses  $r_s = -0.35, p < 1 \times 10^{-16}, n = 542$  million-year bins) (**Extended Data Fig. 3e-n**). We find that several extinction events, such as the Permian–Triassic, Triassic–Jurassic, and the Cretaceous–Paleogene extinction events co-occur with spikes in both BCnS ALG fusions and losses across all animals (**Fig. 3b,c** and **Extended Data Fig. 3**).

Together these results suggest that protostome and vertebrate speciations, and the clades within them, had distinct underlying chromosomal evolution mechanisms that vary in intensity over time. In the case of protostomes, these results suggest that large-scale chromosomal changes that become fixed along an evolutionary branch may be the consequence of massive changes in selective constraints, like population bottlenecks caused by near-extinction events. In the case of vertebrates, on the contrary, our results suggest that higher rates of chromosomal change occurred in times of population stability. While counter-intuitive, this pattern could be expected for the rediploidization process that follows whole genome duplication, and therefore can happen in absence of any major extinction events<sup>29</sup>. Interestingly, we found evidence for periodicity in both BCnS ALG fusions and losses across animals (**Extended Data Fig. 4, Supplementary Information 2**), including a cycle of 62-64 million years in BCnS ALG fusions in Metazoa (Monte Carlo simulations, 93% support) similar to the proposed 62 million year cycle found in species origination and extinction events<sup>25,26</sup> (**Methods - Chromosome evolution periodicity analysis**). Regardless of the observed extinction event co-occurrence, our data clearly indicate that fusion and dispersal rates vary with time and clade, but happen to the same degree across all BCnS ALGs (**Extended Data Fig. 3**). With this in mind, the divergent, clade-specific trends of changing rates of chromosome evolution are a new orthogonal dataset to better interpret patterns of speciation seen in the paleontological record.

### *Algebraic and non-algebraic processes of chromosome evolution*

The balancing act between dynamic rates of chromosomal fusions and translocations has had a substantial role in generating distinct genomic organizations in animals<sup>1,30,31</sup>. Using our pan-metazoan dataset, we next asked whether the observed chromosomal evolutionary patterns could be categorized according to their evolutionary mechanisms (**Methods - Inferring modes of animal evolution**). Because the ancestral myriazoan<sup>3</sup> genome comprised between 20-30 chromosomes<sup>1</sup>, we reasoned that the chromosome evolution patterns of extant animals with fewer than 20 chromosomes would differ from the pattern found in extant animals with more than 30 chromosomes. In our dataset we find two mechanistically distinct modes of karyotype evolution (which we term consolidation and dissociation) that cause deviations from the ancestral myriazoan karyotype, and a ‘default’ mode of karyotypic stasis (**Fig. 3d**). In species with fewer than 10 chromosomes, we found that the predominant mechanism of karyotype reduction could be explained by extensive chromosomal fusions. Example clades with reduced karyotypes include: *Caenorhabditis* (n=28 genomes, 14 species), in which the BCnS ALGs have been fused and mixed, but are still recognizable on single chromosomes and can be expressed as algebraic BCnS ALG combinations; or *Drosophila* (n=216 genomes, 44 species) and the Culicidae (mosquitos, n=79 genomes, 36 species), in which the ancestral BCnS ALGs are completely dispersed across all chromosomes.

We found that the predominant pattern in animal genomes with few chromosomes is to have recognizable ALGs after fusions. This first mode of chromosome evolution, consolidation, is defined by the rate of chromosomal fusions being greater than the rate of chromosome fissions or translocations. We find that nematodes are archetypal of this evolutionary pattern. Despite nematodes having rapid sequence evolution rates<sup>32,33</sup>, and being suggested as having a new set of chromosomes<sup>34</sup>, nematode genomes (n=57 genomes, 37 species) can still be described by algebraic operations. In nematodes, BCnS ALGs are often retained as complete units, and only one BCnS homology loss (ALG C2) is shared among species in the clade (**Fig. 3d**). Similarly, we found that oyster (Ostreida, n=18 genomes, 9 species) chromosomes are the result of many fusions (43), but all nine species in our dataset still have all 29 BCnS ALGs recognizable and intact on single oyster chromosomes. As a consequence of the chromosomal fusions, the karyotype reduction during chromosome evolution causes genomes to be more dramatically affected by subsequent dispersal, resulting in the observed mode of evolution often associated with ‘faster evolving’ genomes.

The genomes that have substantially increased their chromosome counts since the ancestral myriazoan karyotype, on the other hand, could have done so through two processes. We call this second mode of evolution dissociation. In the first pattern of the dissociation mode, which is the non-algebraic type mentioned earlier, chromosomal fissions or translocations result in extant genomes in which BCnS ALGs are present on a large number of chromosomes. Furthermore, the BCnS ALGs are weakly or not dispersed, in contrast to the consolidation mode. The second pattern of the dissociation mode that could explain an expanded karyotype is whole genome duplications, which often result in secondary gene loss on the duplicated homeologous chromosomes, as seen in vertebrates<sup>27,35,36</sup>. While obeying different mechanisms, both dissociation scenarios result in similar syntenic patterns, stemming from the increase in karyotype number that can undergo additional fusions<sup>8</sup>. However, the nature of the non-algebraic fission/translocation process makes it impossible to represent these genomes in algebraic form, therefore a new ALG notation is required. Examples for these clades include coleoid cephalopods<sup>13</sup> (**Fig. 3d**) and clitellate annelids<sup>12</sup>. Vertebrates, due to duplication of complete ancestral chromosomes, can fully be represented by whole-genome duplications and algebraic combinations<sup>1,27</sup>.



The causes of the fissions could be due to their holocentric nature<sup>37</sup>, centric fissions, or possibly due to chromothripsis<sup>38</sup>. Given that chromosomal fissions separate ALGs into unique combinations of ancestral ALGs, and the improbability of those loci coming together again<sup>3</sup>, non-algebraic chromosome evolution is therefore a major contributor to saltatory changes over short evolutionary time and may thus be more strongly associated with periods of species origination and extinction than algebraic evolution.

Finally, our analysis also confirms the ‘default’ mode of animal chromosome evolution is a general karyotypic stasis with few intrachromosomal rearrangements, as previously described for several animal clades<sup>1</sup>. In the karyotypes of these species, centered around 20 chromosomes, there have been very few chromosomal fission or fusion events in the approximately 700 million years<sup>4</sup> of evolution from the myriazoan ancestor.

Together, our data point to three predominant modes of chromosome evolution in animal genomes that have shaped animal genome evolution since divergence from the ancestral myriazoan genome. Due to the different timings and degrees of fusion/fission and dispersal, the processes of karyotype consolidation via extensive fusion and dissociation are inherently distinct, reflecting different meiotic constraints and population dynamics. Any species that has a consolidation or dissociation of the ancestral chromosome count is therefore the result of a lineage in which the balance of stasis shifted, irrevocably changing the substrate of future chromosomal evolutionary events. Importantly, our study provides evidence that there is no intermediate state or trajectory between the dissociation and consolidation modes. While over longer evolutionary times one mode could follow after another, such cases are extremely rare.

### *Macro-evolutionary impact of chromosomal fusion processes*

Both algebraic and non-algebraic processes bring together genes that previously existed on different chromosomal elements. As the functional advantage of distal relationships on single chromosomes is unknown<sup>39</sup>, we instead investigated the impact of fusion-with-mixing, and more generally chromosomal inversions, on the emergence of local subchromosomal linkages, which are far better understood and can confer coregulation, for example, the provenance of the closely linked gene clusters in bilaterians, such as Hox<sup>17,40</sup> and many other local, or micro-syntenic, linkages<sup>15,41</sup>.

First, we sought to characterize the mixing time scales via a simple simulation model of inversions (**Methods - Simulation of fusion-with-mixing**) using beads-on-a-string genomes. We define  $\tau_s$  as the number of inversion cycles required to reach a fully mixed, high-entropy state (**Extended Data Fig. 5**). We compared the  $\tau_s$  of chromosomes of different sizes. We find that smaller chromosomes (fusion products of 100 genes) approach  $\tau_s$  very quickly already after ~50 inversion cycles. Expectedly, larger chromosomes (fusion products of 1000 genes) approach  $\tau_s$  at 1000 inversions. Direct measurements from cephalopod genomes (n=4 species, **Methods - Inversion breakpoint inference**) show that inversions in this clade occur at a rate of 4.29-29.2 inversions per million years<sup>42</sup>, suggesting a timeframe of 34.2-233.1 million years to reach the highly entropic state, for a chromosome of 1000 genes. These results suggest that achieving an irreversible mixed state after chromosome fusion depends on the chromosome size, but can be relatively quick on the macro-evolutionary time scale especially for the smaller chromosomes.

As a proxy for local regulatory linkages, we next sought to measure how many novel neighborhoods can be explored using this chromosome inversion model (**Extended Data Fig. 5, Methods - Simulation of fusion-with-mixing**). We define a measure  $\tau_{50}$  that quantifies how many inversion cycles are needed to explore 50% of all possible neighborhoods. These interaction neighborhoods are defined either as direct neighbors, or all genes within windows of 2, 5, or 10 genes. Such neighborhoods can reflect regulatory compartments of animal genomes, such as topological associating domains (TADs) in which multiple genes can be coregulated<sup>43,44</sup>. Similar to  $\tau_s$ ,  $\tau_{50}$  increases with chromosome sizes, yet the scale is vastly different. To reach 50% of direct neighbor interactions for a fusion product of 100 genes, approximately

1,700 inversion cycles are required on average (25 times more than for  $\tau_s$ ) (n=10 trials). Similarly, for a fusion product of 1,000 genes, 173,000 inversion cycles are needed (130 times larger than the respective  $\tau_s$ ) (n=10 trials, **Extended Data Fig. 5**). Through direct measurement (mixing value  $m^3$ ) of ancient fusion-with-mixing events in the scallop *Pecten maximus* we confirmed that entropic saturation through chromosomal inversions has not been reached yet. These results suggest that for an average metazoan chromosome that has been retained over 600 or more million years, the vast majority of interactions are still unexplored.

Persistent topological properties can substantially reduce  $\tau_{50}$  by allowing distal gene interactions to be maintained. When such neighborhoods are taken into account,  $\tau_{50}$  is substantially reduced. Exploring 50% of all possible neighbors in 5-gene neighborhoods reduces the  $\tau_{50}$  time by 93.5% (n=10 trials). This finding is important as it suggests that presence of persistent topologically associating domains (TADs) allows for faster evolutionary exploration of the regulatory environment given that genomes with interactions at the scale of TADs will explore the available combinations faster than genomes dominated by cis-interactions that can interact only with their immediate neighbors. Our expectation is therefore that non-bilaterian animal genomes that have been suggested to lack maintained loop extrusion through CTCF or similar insulator factors<sup>45,46</sup>, would be slow in the exploration of their regulatory potential, as opposed to genomes with the ability to establish and maintain more distal regulatory properties.

The long-term nature of continuous emergence of novel local neighborhoods on old or newly formed chromosomal elements describes an evolutionary process on the macro-evolutionary time scale (**Fig. 1**). However, whether and how local neighborhoods can be considered an irreversible FWM-like process, and the probabilities of a new sub-chromosomal FWM event given the macro-syntenic background, remain unexplored. The degree of irreversibility of local gene linkage mixing can therefore be seen as the cost of losing that regulatory constraint, as is the case for the Hox clusters<sup>17</sup> and many other micro-syntenic clusters<sup>15,41,47</sup>. While the functionalities of the majority of these clusters are still unexplored, their colocalization property implies a different regulatory constraint<sup>15,18</sup>. Therefore, these micro-syntenic clusters, if evolving under persistent regulatory constraint, will be ‘un-mixable’ into their original states. Our results on genome rearrangements (**Methods - Inversion Breakpoint Inference**) and previous work<sup>48–50</sup> highlight that inversion breakpoints tend to happen at insulator regions, i.e., TAD boundaries, suggesting high selective constraint to keep genes and their regulatory landscape within compartments together.

Different genomic compositions, such as repetitiveness or the chromosome size and composition after algebraic and non-algebraic changes, impact how much regulatory novelty can still be evolutionarily explored within each chromosome. The larger chromosomes are more likely to still harbor unexplored neighborhoods even after hundreds of millions of years after chromosomal element fusion, whereas small chromosomal elements are more likely to exhaust their combinatorial potential. To quantify how much modern genomes can be represented by both chromosomal and sub-chromosomal mixed states, we next describe a topological approach to this problem.

### *Topologically mixed states that define distinct animal genome evolutionary trajectories*

Comparative genomics analyses often focus on a single genomic scale at a time. To study FWM and other chromosomal changes simultaneously it is necessary to implement all genomic scales and features in one framework<sup>31</sup>: from single genes, to micro-synteny and small structural changes, to whole chromosome composition and organization. Each level evolves under certain meiotic, gene-regulatory, or other functional constraints. To study this, we implemented an ‘evolutionary genome topology’ approach, which, in mathematical terms, allows us to view genomes and FWM processes on chromosomes as topological operations.

In this manuscript, we calculate pairwise distances between BCnS orthologs, but this approach can be generalized to any homologous features in genomes, such as: genes, conserved non-coding elements (CNEs), or repeat families and their insertions (**Methods - Forms of genome topological analysis**). This allows us to implement two topological approaches. The first approach compares the topology of many genomes simultaneously, which we term ‘multi-genome topology’ (MGT) and allows the analyses of evolutionary trajectories across species. The second approach, which we call ‘multi-locus topology’ (MLT), allows the simultaneous analysis of the topology evolution of loci within and between clades of genomes. In the MLT analyses, the interaction values for pairs of homologous loci are phylogenetically averaged, while the MGT uses distance measures from individual genomes (**Methods - Evolutionary topology implementation**). Dimensionality reduction and projection methods, such as uniform manifold approximation and projection (UMAP)<sup>51</sup> or t-SNE<sup>52</sup>, are then used to visualize the interaction matrix of whole-chromosome differences between genomes, or the within-genome topology of single or multiple genomes (**Methods - Evolutionary topology implementation**). Eigenvector and singular value decomposition approaches are the key concepts for FWM generalization across genomic scales of organization, as they can be used as a measure of mixing (**Supplementary Information 3**). Raw orthology information and positional data can be stored and processed with graph database approaches (**Methods - Graph database integration, Extended Data Fig. 6**) for the downstream dimensionality and manifold projection approaches (**Fig. 4, Extended Data Fig. 7, Extended Data Fig. 8**).

The multi-genome topology UMAP (**Fig. 5a**) reflects key genomic groupings. The genomes that are least derived from the ancestral BCnS state, mainly mollusc and cnidarian genomes, form a cluster of species on the UMAP (**Extended Data Fig. 9a-d**). The central hub of genomes is composed of species that have some chromosome rearrangements (**Extended Data Fig. 9e**), but are not dramatically rearranged from the ancestral animal genome. The derived states where the ancestral chromosomal element composition has been most heavily modified are found in the species clouds located at the edges of the plot (**Extended Data Fig. 9f,g,i**). These distinct clouds are connected via intermediate genomic configurations, suggesting that some genomes’ configurations are in less-derived states and are still on the way to the advanced differentiation in terms of their BCnS ALG composition and mixing. For example, the genomes of a tick (*Dermacentor silvarum*), a beetle (*Tribolium castaneum*), mosquitoes and fruit flies can be linked by many intermediate genomes, forming a trajectory (**Extended Data Fig. 9h**). These analyses recapitulate the two major outcomes (modes) of chromosome evolution: consolidation and dissociation. However, these states can now be refined and seen as being composed of multiple very distinct evolutionary trajectories (**Fig. 5, Extended Data Fig. 10b**).

The different trajectories within even closely related species are exemplified by applying the multi-genome topology approach to individual animal clades (**Extended Data Fig. 10**). Closely related species, even after tens of millions of years of divergence, occupy close positions in the UMAP space (**Extended Data Fig. 10a**). True oysters (n=18 genomes, 9 species), which have a reduced karyotype of 1n=10, form a cluster distant from the other mollusc genomes (n=108 genomes, 80 species), other clades with unique paths to karyotype reduction (limpets: n=6 genomes, 5 species, karyotype 1n=9, scaphopods: n = 2 genomes, 2 species, karyotype 1n=9), additionally form distinct clusters separate from one another. Similarly, the major clades in Diptera (n=509 genomes, 275 species), form clusters in UMAP space (**Extended Data Fig. 10b**). We recently introduced the use of macrosyntenic FWM to resolve ancient, recalcitrant phylogenetic nodes<sup>3</sup>, and others have used microsynteny similarity to resolve more recent evolutionary relationships<sup>53</sup>. The MGT approach, because it simultaneously incorporates both chromosomal and sub-chromosomal localization data, has clear potential to be applied to resolve outstanding phylogenetic questions.

Our findings from the MGT approach suggests that irreversible chromosomal changes preclude a direct evolutionary path between trajectories, and more importantly, back to the ancestral state. We hypothesized that accumulation of irreversible topologically mixed states defines the directionality of this macro-evolutionary process.

To test the degree and stability of mixing at the sub-chromosomal level, we investigated the evolutionary genome topology within two or more genomes in specific clades using the MLT approach. To do this we averaged, with phylogenetic weighting (**Methods - Evolutionary topology implementation**), the linear distance matrices of species within various clades and computed the sub-chromosomal, whole-genome (**Fig. 5b**) and individual chromosomal UMAPs (**Fig. 5c**). The homologous chromosomes form clearly distinct regions, which are separate entities in the manifold projection. Within each chromosome, local feature clusters emerge as well. While the maximum resolution of our analysis is the 2367 BCnS ALGs present mapped to the genome, these features reflect local conserved micro-synteny, enhancer-promoter interactions, and other interactions which can be further explored at the limits of orthology detection.

The MLT manifold representation also recovers known FWM events, such as the A2 $\otimes$ N and A1b $\otimes$ B3 in cnidarians (**Fig. 5b** - Cnidaria) relative to clades without these fusions, such as non-cephalopod molluscs (**Fig. 5b** - Mollusca). Similarly, the manifold representation can be used to investigate the degree of FWM. For example, the dots comprising the cnidarian A2 $\otimes$ N ancestral FWM are highly mixed on the manifold representation (**Fig. 5b** - Cnidaria), while the poor mixing between B2 and Ea $\otimes$ Eb in echinoderms (n=21 genomes, 18 species) is also clear (**Fig. 5b** - Echinodermata). Similar to chromosome-level FWM events, we can also use local clustering of features on the manifold, or their connectivity, to estimate the degree of local FWM. For highly preserved micro-syntenic linkages, such as Hox, (**Fig. 5c**), their unlinking would imply substantially increasing the distances of their individual components on the manifold. For example, we find that the posterior Hox genes tend to be located in a separate topological cluster in protostomes (**Fig. 5d**).

The evolutionary topology approach also identifies the presence of loci in clade-specific linkages<sup>54</sup>. These linked loci remain similarly positioned, not necessarily adjacent, relative to one another in 2D or 3D space against the backdrop of inter- and intra-chromosomal changes. Our approach provides sets of mixed states that define each clade (**Methods**). Pairs of loci can define the archetypal genome for a clade, and are topologically linked, if (1) the pairs are unique to this clade - that is these two sequences are only present on the same chromosome within genomes in this clade, (2) the pairs are fixed at a stable distance in this clade, regardless of whether they are close or distant, or (3) the distances between the pairs are very small in genomes in this clade (*cis*- interactions). It is therefore possible that type 3 pairs (small inter-locus distance, or *cis*-) pairs are also type 2 pairs (stable inter-locus distance), but not vice versa. Type 3 pairs can generally be considered to reflect widely studied micro-syntenic regions, or constraints that act on very short (*cis*-) distances. Such regions can be easily identified via comparisons of both chromosome- and non chromosome-scale genomes<sup>15,41,47</sup>. Type 2 pairs, on the other hand, can only be uncovered by averaging dozens to hundreds of chromosome-scale genomes. Type 2 pairs therefore may reflect more distal entanglement (**Supplementary Information 1**). Extensive mixing, leading to entangled pairs, of many gene enhancer-promoter regulatory links within a defined genomic region (e.g., a topological domain) can be seen to make the domain ‘resistant’ to random translocations that would break one or more of such regulatory links.

There are generally few locus pairs unique to any given animal clade (type 1 locus pairs), reflecting the degree of dispersion and convergent colocalizations during metazoan evolution. On the contrary, there are many locus pairs that are characteristically stable (type 2) or close (type 3) in any given clade. For example, in Spiralia there are 2,398 pairs stable in the clade, and 2,471 especially close pairs. While the



test for type 1 pairs is binary, the stringency of our search for type 2 and type 3 pairs (one-sided  $\mu$ -2 $\sigma$  cutoffs,  $\geq 50\%$  of species with this pair, **Methods**) mean that these numbers are conservative estimates of the number of identifiable markers. We note that the most strongly stable pairs (type 2) are also close pairs (type 3), and these novel linkages appear to be rarely lost in genomes within the clade once the loci become linked (**Extended Data Fig. 11c,f**). These results suggest that within-chromosomal mixing and constraint, rather than karyotype-level changes alone, are dominant forces in creating new clade-specific chromosomal landscapes. Moreover, the presence of clade-specific stable pairs suggests that novel sub-chromosomal linkages have a degree of irreversibility, and changes in them appear to be selected against.

Our approach reveals emergent patterns in stable pair formation that reflect different evolutionary histories of chromosomes. For example, BCnS ALG B2 that contains the Hox cluster and BCnS ALG A1 that contains the pharyngeal cluster<sup>55</sup> as well as many homeobox genes<sup>27</sup> show contrasting patterns between deuterostomes and spiralian (**Extended Data Fig. 11b,c,e,f**). In spiralian, both chromosomes are characterized exclusively by the accumulation of within-chromosomal stable pairs (BCnS ALGs B2 or A1, **Extended Data Fig. 11b,e,f**). With A1 forming almost 5 times the number of novel, spiralian-specific, stable pairs. In contrast, in deuterostomes, also inter-ALG pairings can be observed to contribute to both B2 and A1, with BCnS ALG B2 particularly enriched in novel inter-ALG pairings. This highlights different levels of neighborhood exploration and evolution of putatively novel regulatory constraints within these ancient chromosomal elements.

In summary, our evolutionary topology approach, currently based on pre-computed BCnS ortholog distances, can be used to place genomes on the manifold, identify their macro-evolutionary trajectory, and to identify evolutionarily mixed chromosomal and sub-chromosomal linkages. This analysis supports the view that saltatory changes in genomes (such as accumulation of algebraic FWMs or non-algebraic transitions) put genomes on distinct and irreversible evolutionary trajectories. Our results describe genome evolution as an iterative process with each step consisting of one or several algebraic or non-algebraic events that lead to the irreversible, iterative erasure of ancestral states, with strong intra-chromosomal entanglement. We show that this process can be described in a mathematical topological context that reflects and predicts extensive evolutionary constraints that act at whole-chromosomal and regulatory linkage levels.

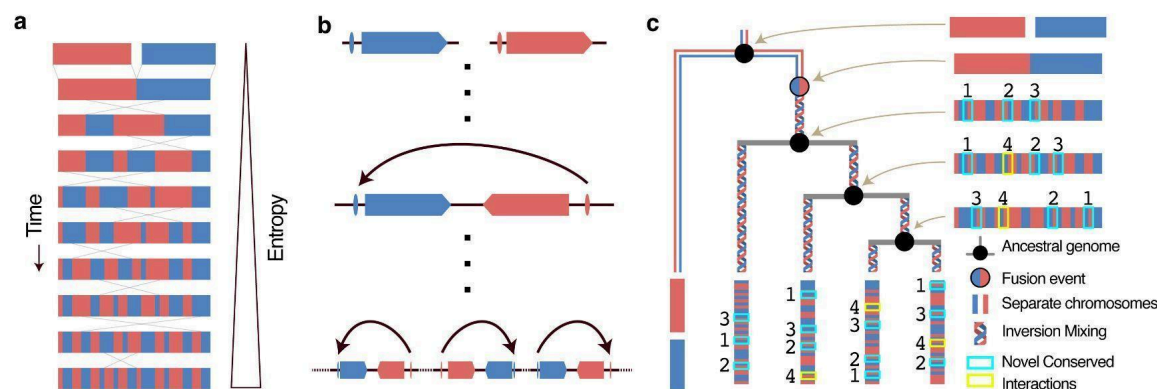
## Conclusions

In this paper, we describe and quantify evolutionary time scales of the processes that govern changes in chromosomal and sub-chromosomal homologies, and connect these patterns of genomic change to patterns of speciation and diversity seen in the paleontological record. As new chromosome-scale data is emerging for many phylogenetically informative clades, together with genome topological and regulatory data, we can determine the evolutionary trajectories that are set and defined by the present-day configurations of animal genomes. We propose that using the evolutionary genome topology approach, as described in this manuscript, will be crucial in understanding the evolution of mixing, loci linkage, and irreversibility across genomic scales. Our data reveals thousands of clade-specific stable sub-chromosomal configurations that can be used as synapomorphic characters in phylogenetic studies. We propose that the irreversibility of such mixed states is the main contributing factor to the observed genomic macro-evolutionary trends. Moreover, our data provides insights into the potential, at the macro-evolutionary time-scale, to evolve new regulatory linkages within available chromosomal elements. This approach also highlights the role of saltatory changes in chromosomal complement reorganization, via both consolidation and dissociation means, to increased innovation through exploration of novel

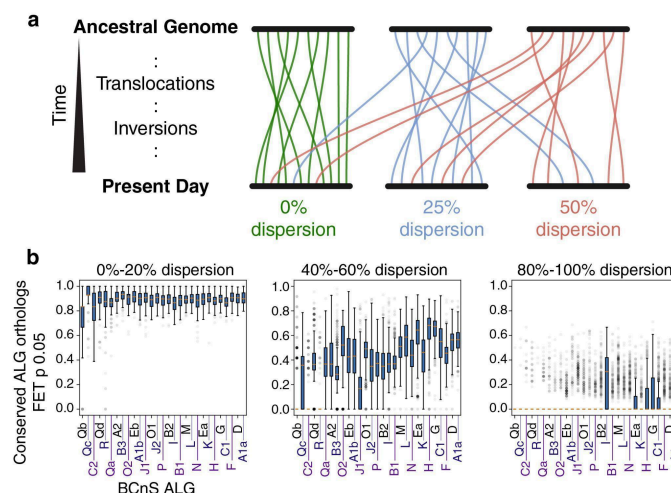


regulatory interactions that are predicted in this manuscript to be explored over the span of hundreds of millions of years. Understanding the degree of and the mechanisms underlying their predicted irreversibility with functional studies in various metazoan model organisms will provide further evaluation to this hypothesis.

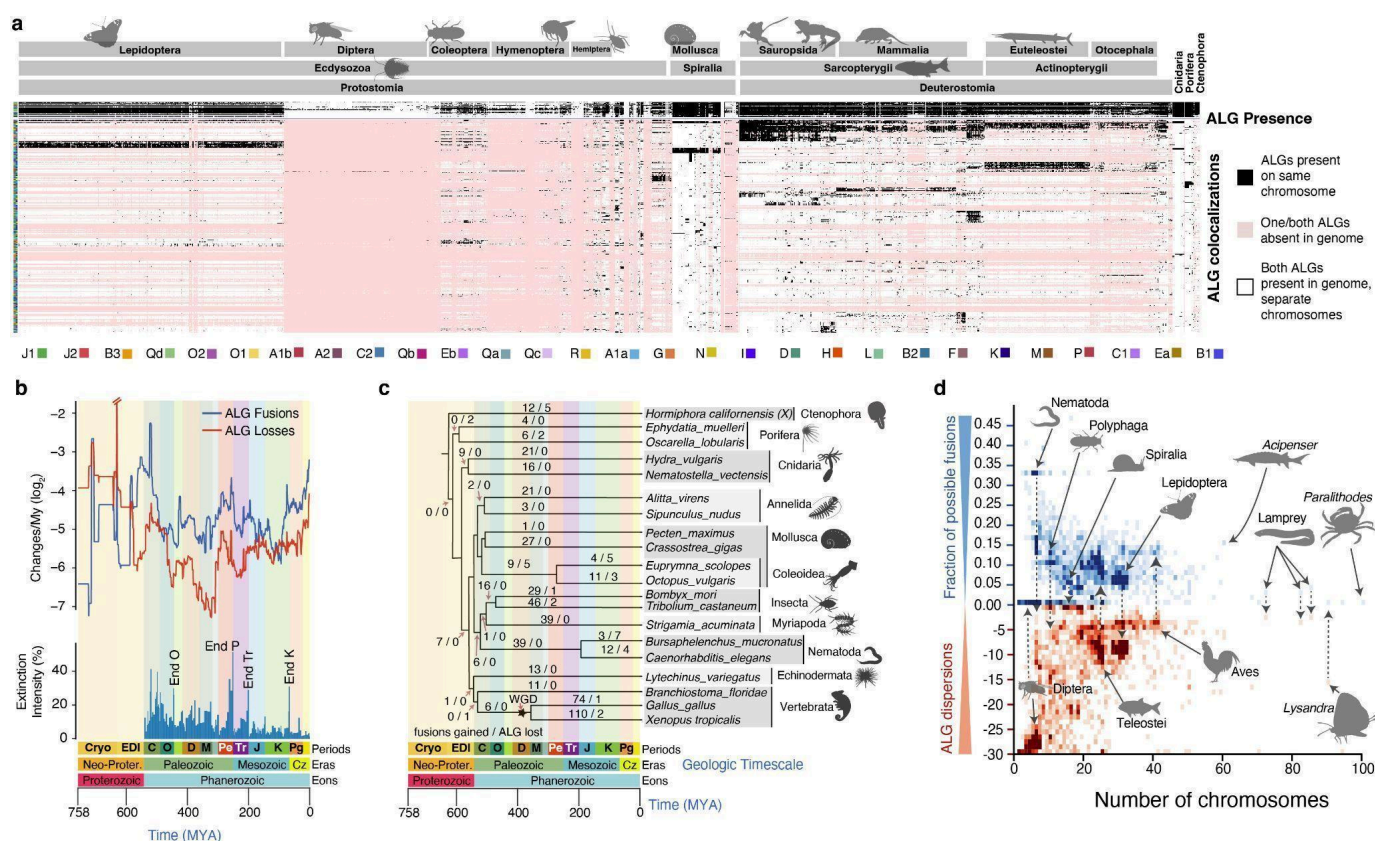
## Figures and Tables - Main Text



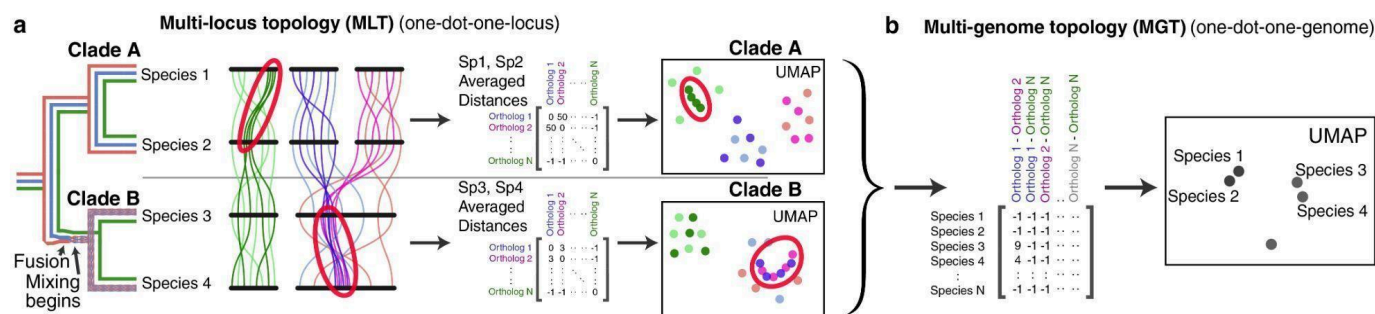
**Fig. 1 | Fusion with mixing (FWM) at chromosomal and sub-chromosomal levels.** **a**, Chromosome fusion-with-mixing (FWM) is a process that happens over long evolutionary time scales and leads to mixing of chromosomal regions previously on two separate chromosomes. The mixing over time can be measured as a form of entropy. **b**, Through FWM previously separate chromosome regions (top) can form new regulatory interactions not possible in the original separate chromosomes (middle). Over time many novel regulatory interactions can arise (bottom) and become constrained (which we refer to as entangled) through complex regulatory interactions. **c**, The FWM process is a sum of chromosomal-level mixing of gene sets and emergence of novel regulatory constrained linkages, novel regulatory ‘mixings’ can arise long after the original chromosomal fusions and therefore are synapomorphic local linkages.



**Fig. 2 | Dispersion of chromosomal elements is independent of their size.** **a**, Dispersion is the process of loci leaving their chromosome of origin through translocations. Inversions over time cause loci that were previously on separate chromosomes to mix. **b**, Chromosome dispersion in animal genomes ( $n=3,631$  genomes, 2,291 species) varies greatly depending on the evolutionary history of the species since the BCnS ancestor. Some species retain nearly 100% of BCnS ALGs as recognizable on one or more chromosomes, while some species retain no recognizable BCnS ALGs ( $p \leq 0.05$ , Bonferroni-corrected one-sided Fisher's exact test). Genomes are binned by the total percent of BCnS ALG dispersion across the entire genome. Individual BCnS ALG's dispersion values for each genome are plotted for the boxplots. Box centerlines are the median values, box limits span from the first quartile to the third quartile, fliers span to 1.5 times the interquartile range ( $Q3-Q1$ ), and outliers are plotted individually. See also **Extended Data Fig. 1** for *Pecten*-centered dispersion calculations.

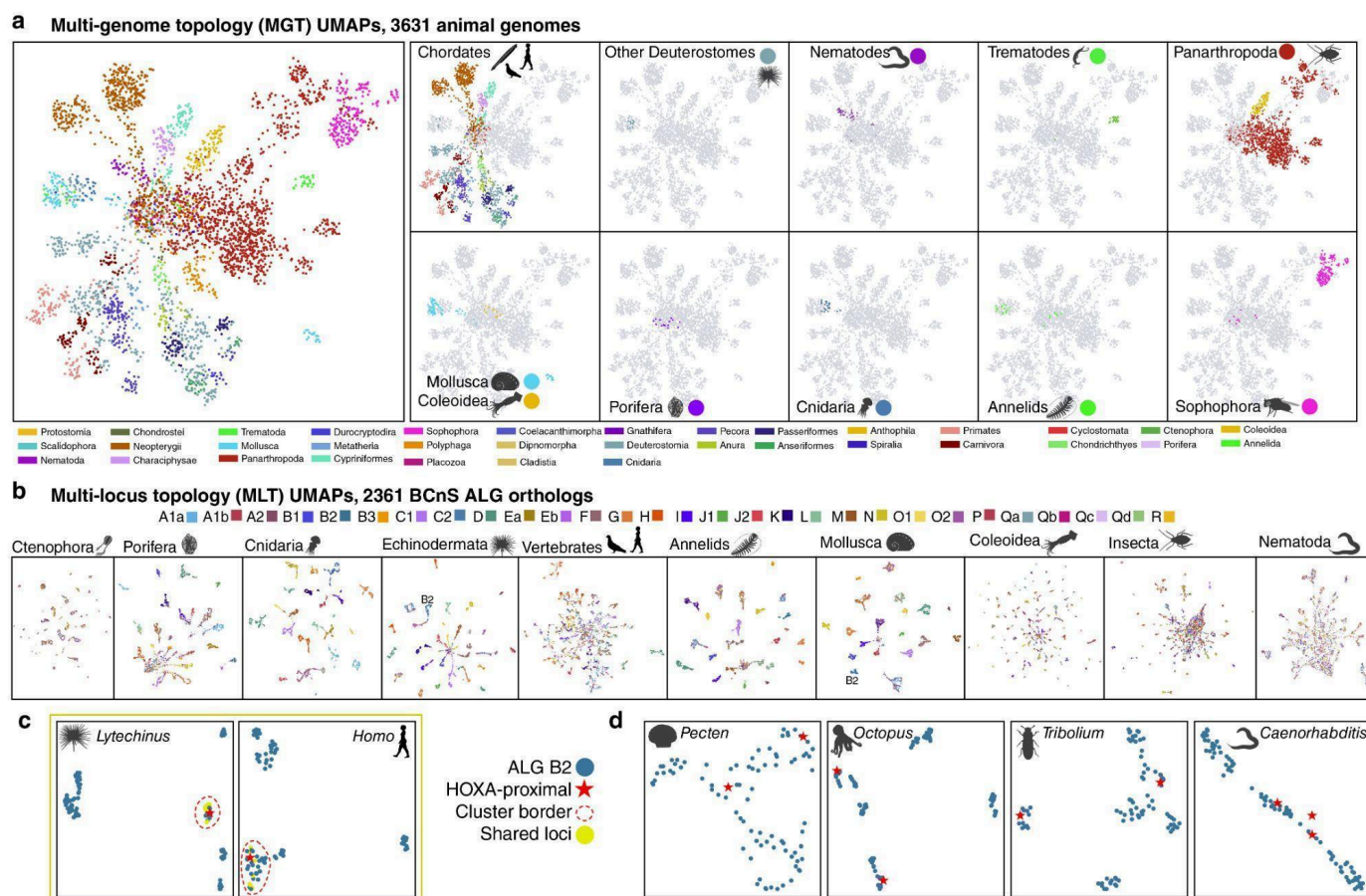


**Fig. 3 | Global patterns of chromosomal fusions and fissions across metazoans.** **a**, The grid of pink, black, and white pixels shows the BCnS ALG presence and colocalization in 2262 animal species. One column encodes one genome from each species. Black pixels in the top row, *ALG Presence*, shows BCnS ALG persistence in the extant genomes. The lower rows, labeled *ALG colocalizations*, depict pairs of ALG colocalizations on the same chromosome (black), or on separate chromosomes (white) are depicted. Pink indicates a lack of detectability due to a lost BCnS ALG homology. See **Extended Data Fig. 3** for a phylogenetic reconstruction of the extant genomes. **b**, Inferred rates of chromosomal changes on the animal tree in million-year bins. The period of most intense chromosomal changes occurred early during animal diversification. **c**, Counts of shared fusions and dispersions in major metazoan evolutionary transitions. **d**, Distribution of fusion and dispersion of BCnS linkage groups as a function of chromosome number (n=3,631 genomes).



**Fig. 4 | FWM representation on topological spaces.** **a**, Mixing dynamics and gene colocalization can be approximated by averaged pairwise distances between orthogroups in each genome for a given clade to produce manifold projects for each clade. This distance matrix is then parsed through a manifold reduction algorithm like UMAP to visualize orthologous regions that are more prone to mixing. Specifically, FWM of chromosomes will result in their orthogroups to be inseparable on the UMAP. Similar observations can be made for mixed sub-chromosomal regions (red oval in Clade A UMAP). **b**, Orthogroup pair distances across species can further be used to identify species with similar colocalization distributions.

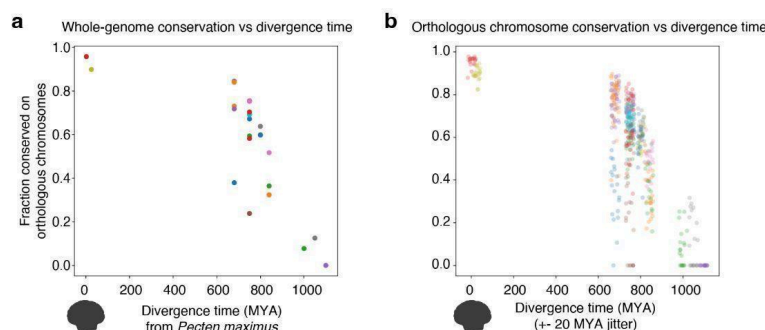




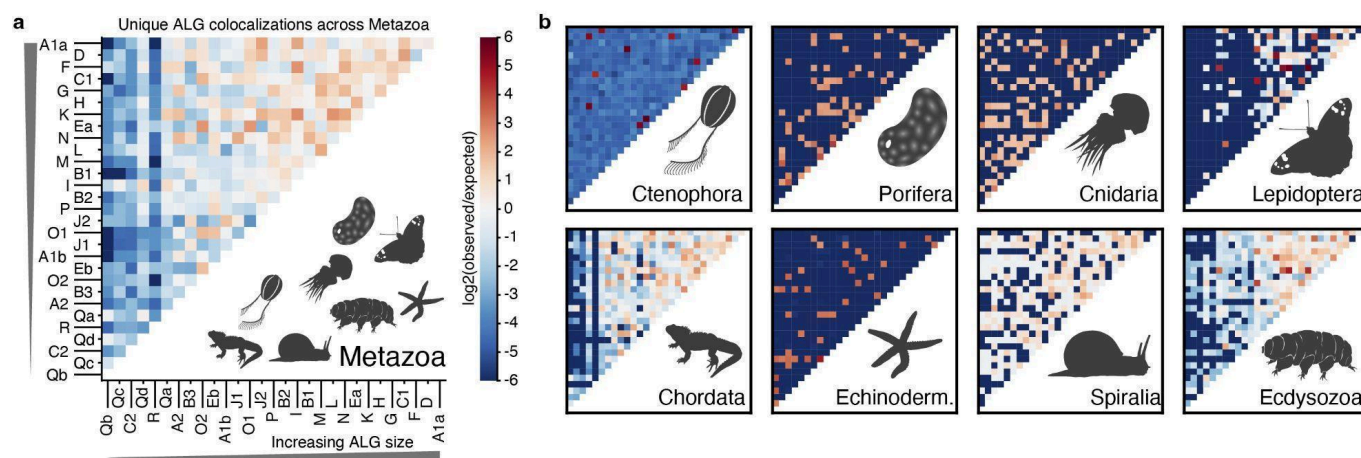
**Fig. 5 | Evolutionary topology of chromosomal elements.** **a**, A multi-genome topology (MGT) projection of 3631 chromosome-scale animal genomes ( $n=2,291$  species), in which each dot represents one genome. We use the genes from the previously-defined BCnS ALGs as anchors in the genomes of all chromosome-scale genomes, showing distinct colocalizations that are often clade specific. **b**, Multi-locus topology (MLT) projections for individual clades reflects evolutionary properties of chromosomes. One dot represents one orthologous protein shared across animals ( $n=2,361$  BCnS ALGs). **c**, Sub-chromosomal topology around the Hox cluster orthologs for BCnS ALG B2 genes allows the identification of a set of co-mixed orthologs. **d**, Chromosomal-level topology around the Hox cluster in several animal groups.



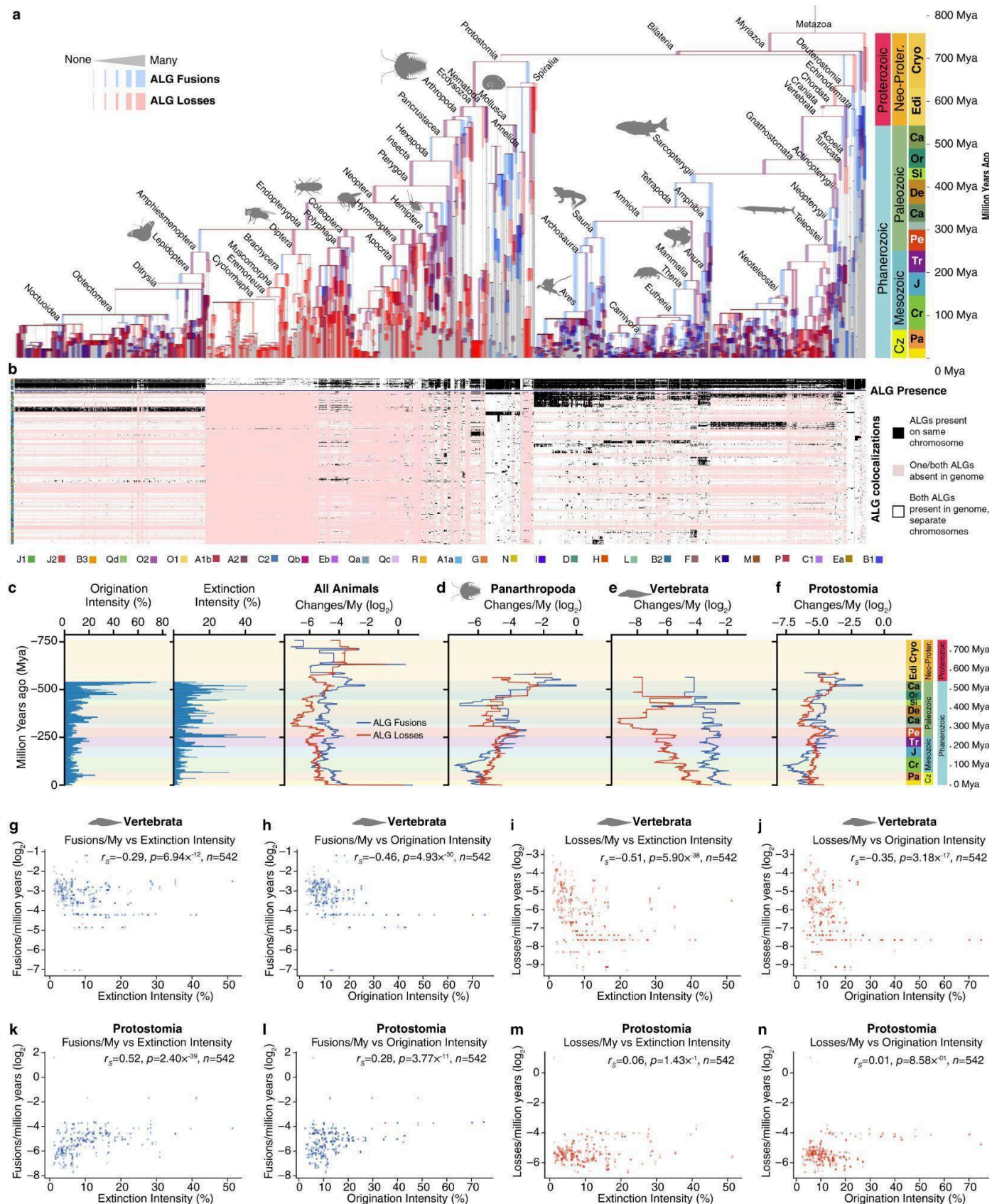
## Extended Data Figures



**Extended Data Fig. 1 | Time dependence of BCnS ALG dispersal.** **a**, We compared the chromosome conservation of the whole *Pecten maximus* genome to the genomes of other organisms (n=22 species), as distantly related as the filasterean amoeba *Capsaspora owczarzaki*. The y-axis shows the total fraction of orthologs conserved in the genome. The decay of homology is a time dependent process, with a rate of slowed change between present day and 600 million years ago. **b**, The same data, but the conservation value of individual *Pecten* chromosomes relative to the other species are shown.

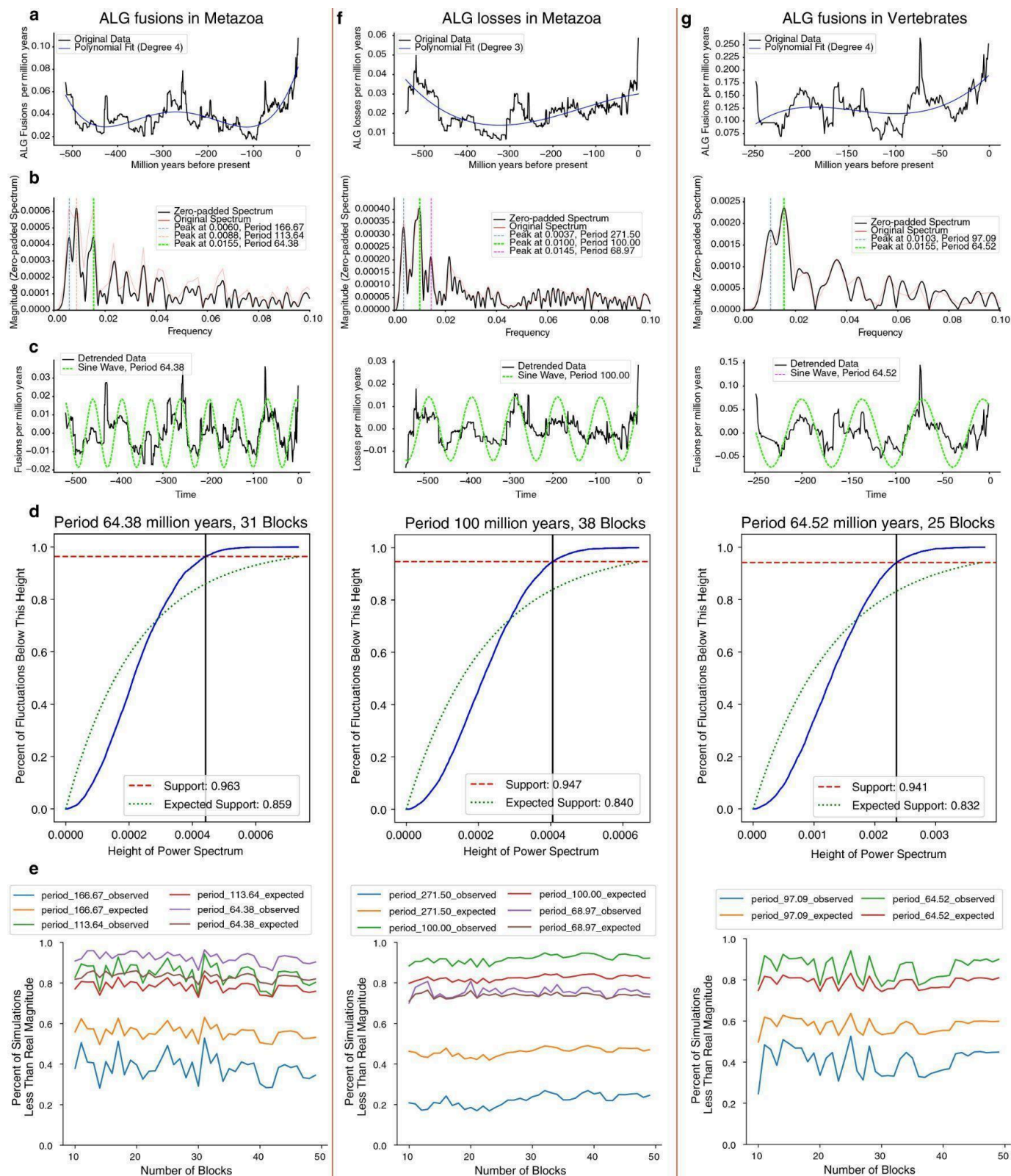


**Extended Data Fig. 2 | Phylogeny-aware counting of BCnS linkage group fusions.** **a**, A general overview across metazoan genomes (n=3,631 genomes) shows depletion of smaller BCnS ALG fusions. **b**, BCnS fusion rates are different among animal clades and is a function of both chromosomal number within each clade and dispersion rate (see also Fig. 3).



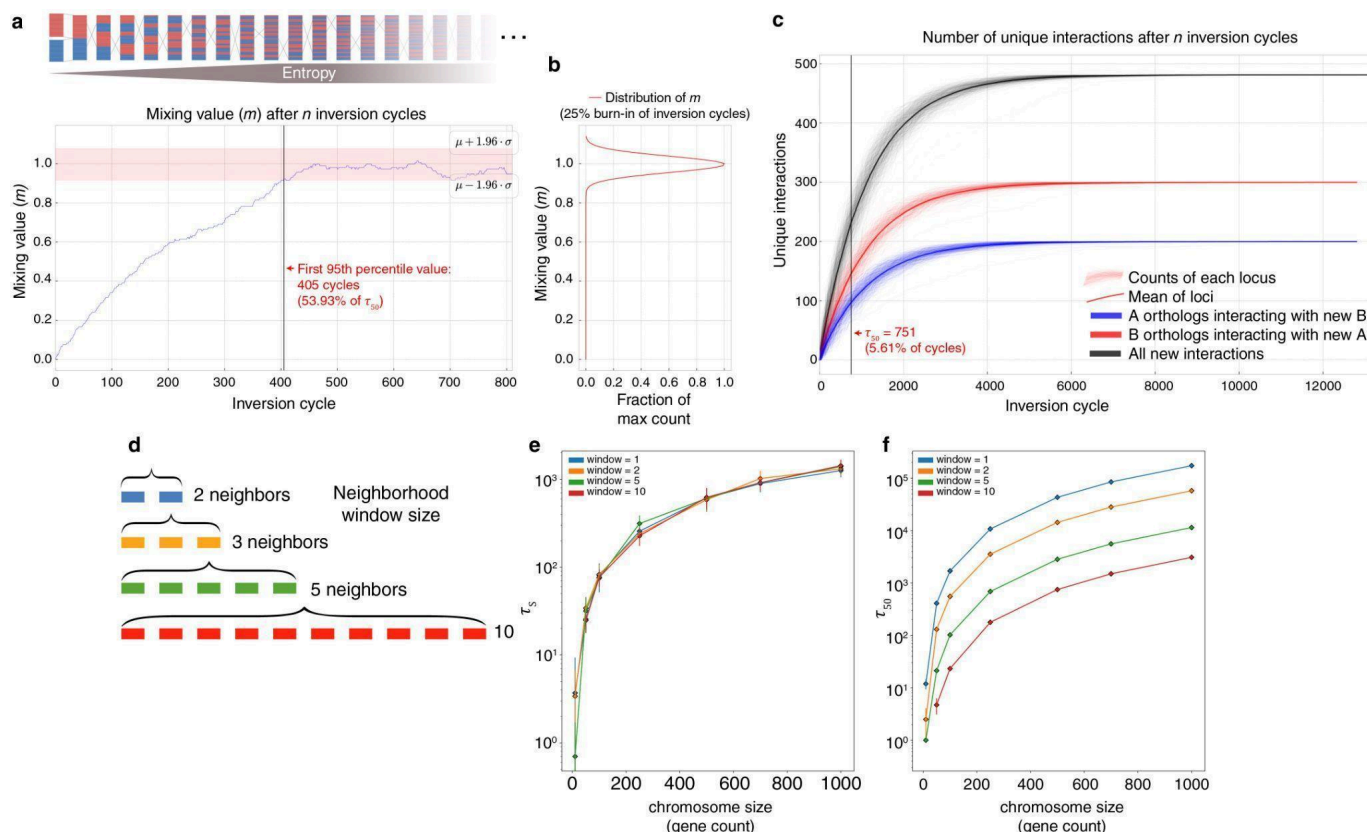
**Extended Data Fig. 3 | BCnS linkage profiling across metazoan chromosome-scale genomes. a**, Chromosome evolutionary history encoded along the phylogenetic tree of 2262 species. BCnS ALG losses are red, and BCnS ALG fusions are blue. **b**, The grid encoded in **Fig. 3a**, where each column corresponds to a tree tip in panel a. Black pixels in the top row, *ALG Presence*, shows BCnS ALG persistence in the extant genomes. The lower rows, labeled *ALG colocalizations*, depict pairs of ALG colocalizations on the same chromosome (black), or on separate chromosomes (white) are depicted. Pink indicates a lack of detectability due to a lost BCnS ALG homology. **c**, Unedited panel shown in **Fig. 3b**, which shows the changes per million years across the evolutionary tree shown in panel a. **d**, The changes per million years in the Panarthropoda, **e**, Vertebrata, and **f**, Protostomia. **g-j**, plots of changes per million years versus the species origination or extinction intensity for the changes within Vertebrata. Statistics reported in the panels are one-tailed Spearman's ranked correlation coefficients. **k-n**, the same plots as panels g-j for changes within Protostomia.



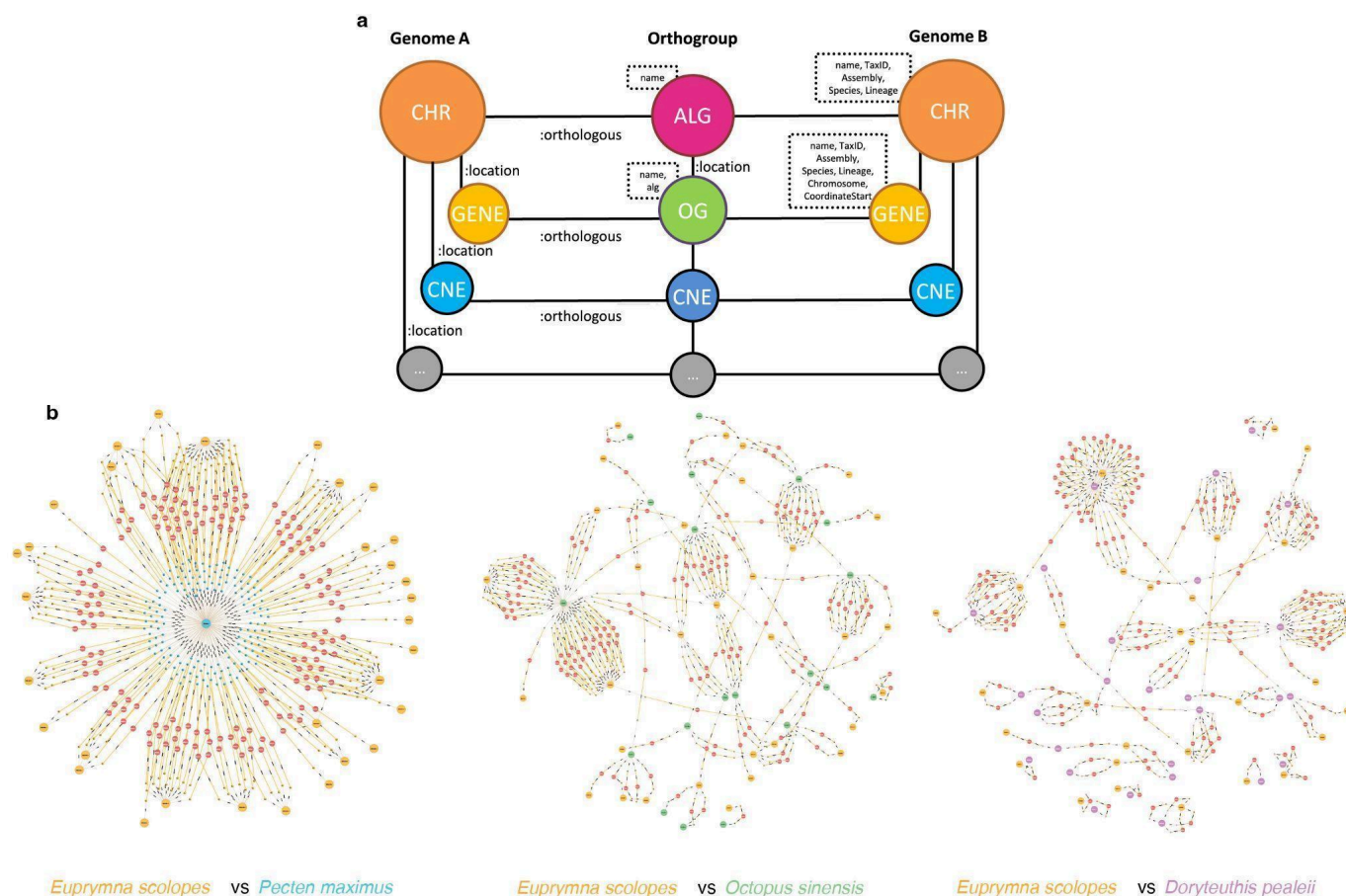


**Extended Data Fig. 4 | Periodicity analysis of BCnS ALG fusion or loss events.** **a**, Rates of BCnS ALG fusions across the Metazoa in million-year bins, fitted with a quartic function. **b**, The power spectrum of the zero-padded discrete Fourier transform of the detrended dataset (black line) and the transformed unpadded power spectrum (red line). The three strongest peaks are highlighted with vertical dotted lines (periods of 166.67 million years, 113.64 million years, and 64.38 million years). **c**, The detrended dataset overlaid with a sine wave of a 64.38 million-year period. **d**, Example results from one Monte Carlo simulation for the 64.38 million year period. The datasets transitions were split into 31 blocks and shuffled. The green dotted line represents the exponential expectation function based on the average magnitudes in the power spectrum from the Monte Carlo trials. The vertical black line indicates the observed peak magnitude. The red support value is the percentage of simulated Fourier transforms with magnitudes less than the observed peak's magnitude. **e**, The effect of the number of blocks on the observed support for the peak from the Monte Carlo simulations. **f**, Similar analysis as panels a-e for BCnS ALG losses in metazoans. **g**, Similar analysis as panels a-e for BCnS ALG fusions in Vertebrates.



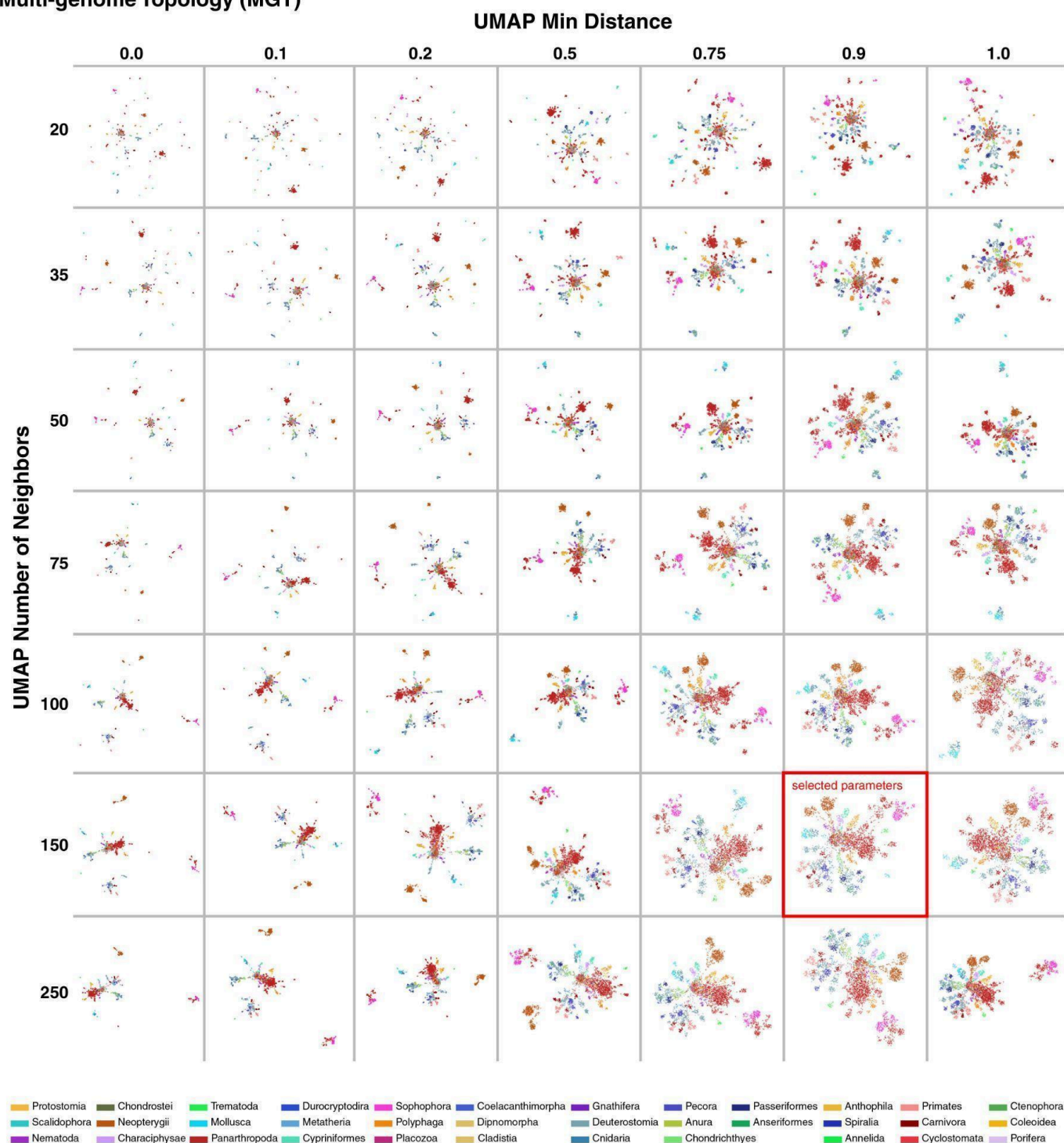


**Extended Data Fig. 5 | Time scale of mixing and genomic neighborhood exploration.** **a**, Trace of the mixing value  $m$  (Methods) for two beads-on-a-string chromosomes after fusion. This simulation is for 300 genes in group A fusing with 200 genes in group B. **b**, The mixing value  $m$  reaches the mass of the distribution ( $\mu - 1.96\sigma < m < \mu + 1.96\sigma$  of  $m$  after 25% burn-in) of a well mixed state after 405 inversion cycles. **c**, Counting the number of unique direct (nearest neighbor) interactions or pairs during the simulation. For the gene and window parameters of panel a, half of all possible neighborhoods have been explored only after 751 inversion cycles. **d**, The window size parameter changes the window size of genes that can be considered for new interactions. Small number of neighbors  $n$  may best model genomes with cis-regulation, while a large number of neighbors may best model genomes with gene regulation grouped by TADs. **e**, Window size has little effect on the amount of time it takes for genes to fully explore the number of possible combinations. Error bars are the standard deviations of 10 simulations. **f**, More practically, however, the time to explore half or more of the combinations is reduced exponentially as the window size increases.



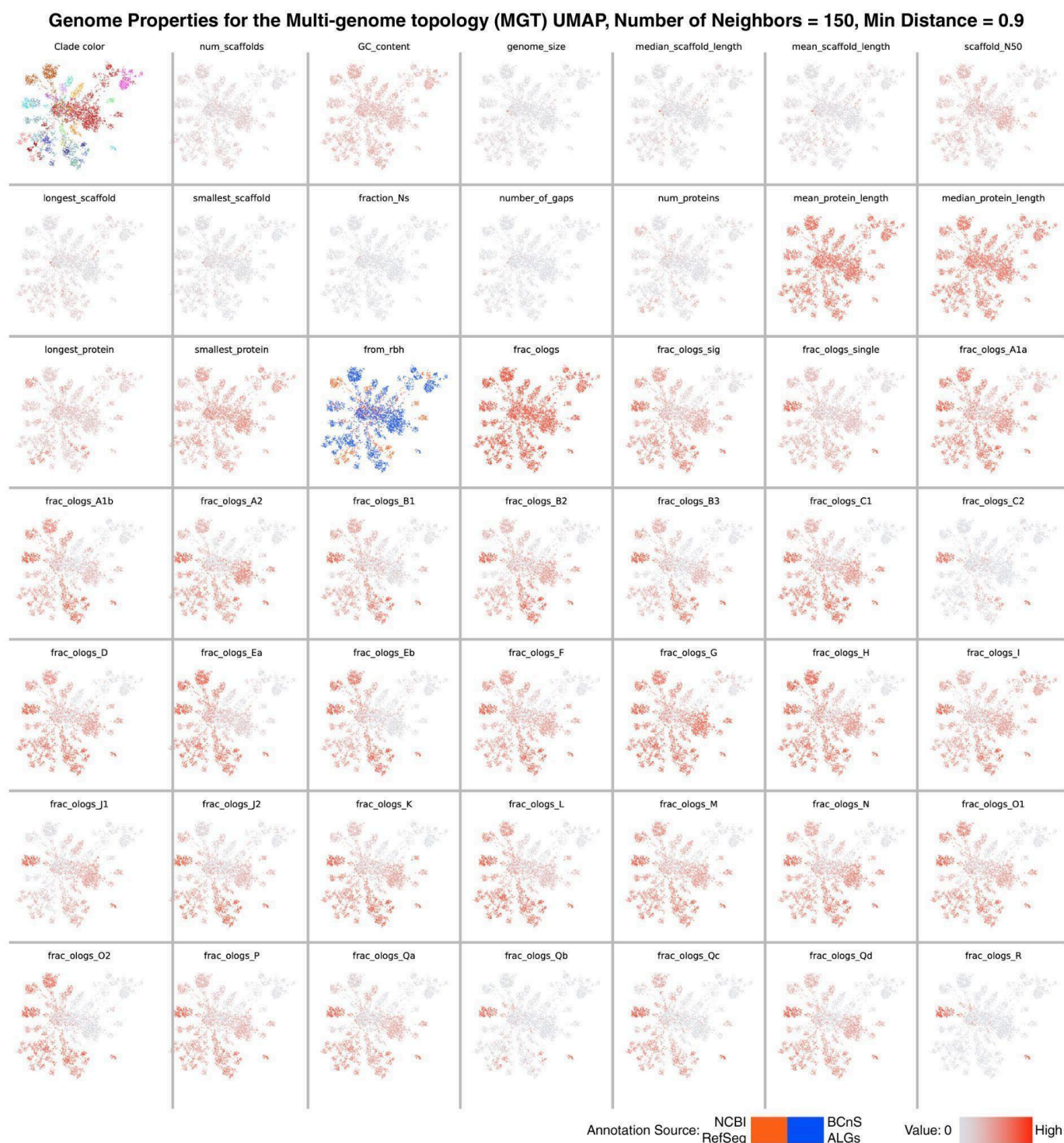
**Extended Data Fig. 6 | Graph database representation of orthologous connections between chromosomes, genes, and noncoding elements.** **a**, data structure as implemented in the Neo4j database. Properties of different node types are shown inside dashed lines. **b**, example networks for different species groups highlight simpler (one-to-one chromosomal, with some gene dispersion) relationships between two squid species (*Euprymna* and *Doryteuthis*), and fission of the ancestral linkage groups (as represented by a *Pecten* chromosome) in *Euprymna*. Orthogroup nodes belonging to BCnS ALG A1a are in red color in middle size, large nodes are chromosomes, small nodes are genes, and nodes of each species are colored as in the species names shown.

# Multi-genome Topology (MGT)



**Extended Data Fig. 7 | UMAP parameter space exploration.** While local distribution of species on UMAPs change according to neighborhood and min distance parameters, the overall clade assignment and composition remains stable.

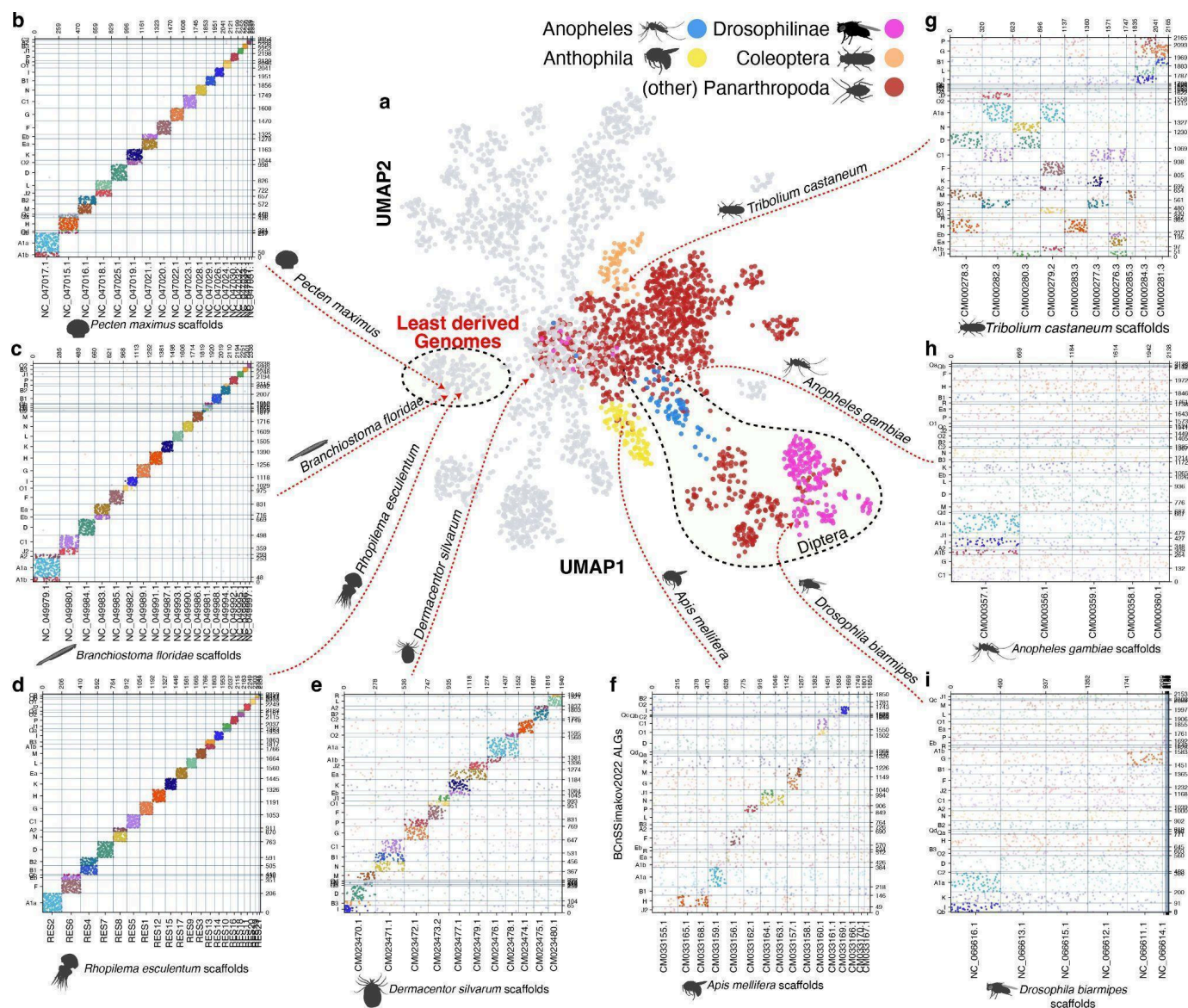




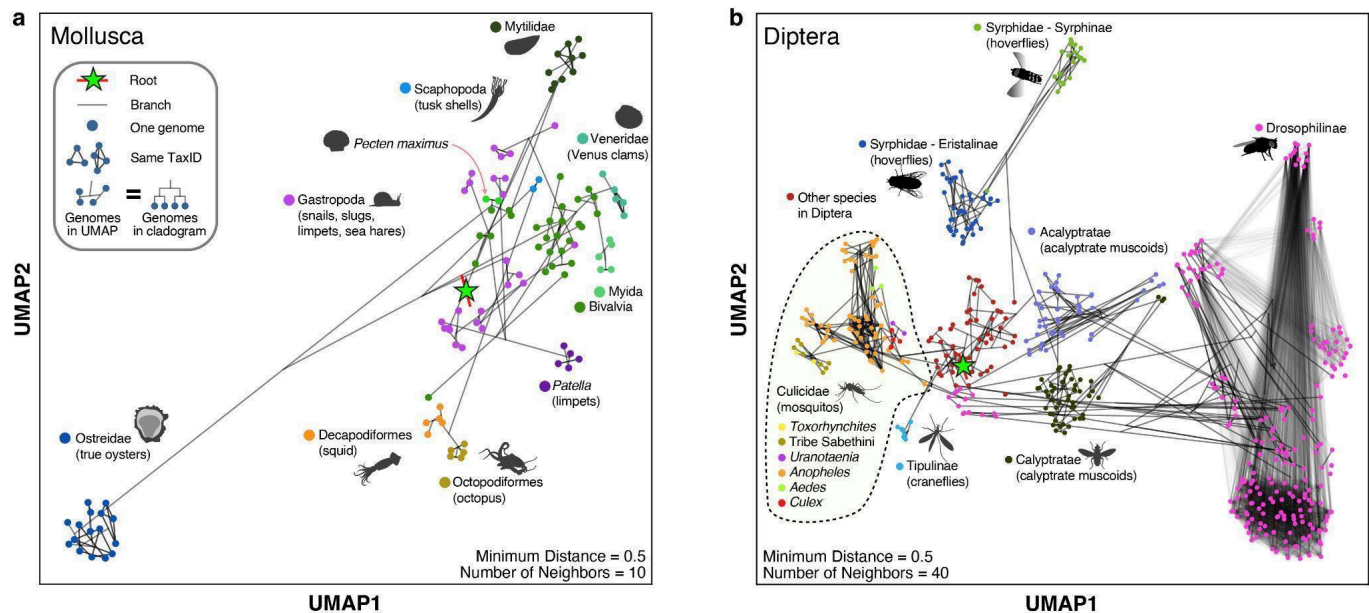
**Extended Data Fig. 8 | Genomic property localization on UMAP species embedding.** The UMAP visualization of genomes shows localized evolutionary history of basic genome properties and BCnS ALG changes. For each subplot, the scale is varied depending on the range of the values of all the genomes. The fields of the figure are: *clade\_color*, each dot is colored by the clade code from **Extended Data Fig. 7**; *num\_scaffolds*, the number of scaffolds in the genome assembly fasta file; *GC\_content*, the fraction of G or C nucleotides in the genome assembly fasta file; *genome\_size*, the number of basepairs in the genome

assembly fasta file; *median\_scaffold\_length*, the median length (basepairs) of all of the scaffolds in the genome assembly fasta file; *mean\_scaffold\_length*, same as the previous field, but the mean; *scaffold\_N50*, the N50 of all the scaffolds in the genome assembly fasta file; *longest\_scaffold*, the longest scaffold (basepairs) in the genome assembly fasta file; *fraction\_Ns*, the number of Ns in the genome assembly fasta file divided by the total number of bases in the fasta file; *number\_of\_gaps*, The number of gaps, or continuous stretch of at least 10 N characters, in the genome assembly fasta file; *num\_proteins*, the number of proteins present in the .chrom file for this genome; *mean\_protein\_length*, the mean length (amino acids) of the proteins present in the protein fasta file for this genome; *median\_protein\_length*, same as the previous field, but the median; *longest\_protein*, the longest protein (amino acids) present in the protein fasta file for this genome; *smallest\_protein*, the same as the last field, but the shortest protein; *from\_rbh*, blue indicates that the genome was annotated with the 2361 BCnS ALG sequences (**Methods**), and orange indicates that the genome annotation was taken from NCBI; *frac\_ologs*, the fraction of the 2361 BCnS ALG orthologs that were present on any scaffold in the genome; *frac\_ologs\_sig*, the fraction of the 2361 BCnS ALG orthologs that were present in a significant amount on any scaffold in the genome ( $p \leq 0.05$ , one-sided Bonferroni-corrected Fisher's exact test); *frac\_ologs\_single*, the fraction of the 2361 BCnS ALG orthologs significantly present on a single chromosome ( $p \leq 0.05$ , one-sided Bonferroni-corrected Fisher's exact test); *frac\_ologs\_{ALG name}*, similar to the previous two fields, this is the fraction of a single BCnS ALG that is significantly present ( $p \leq 0.05$ , one-sided Bonferroni-corrected Fisher's exact test) on one or more scaffolds in the genome.



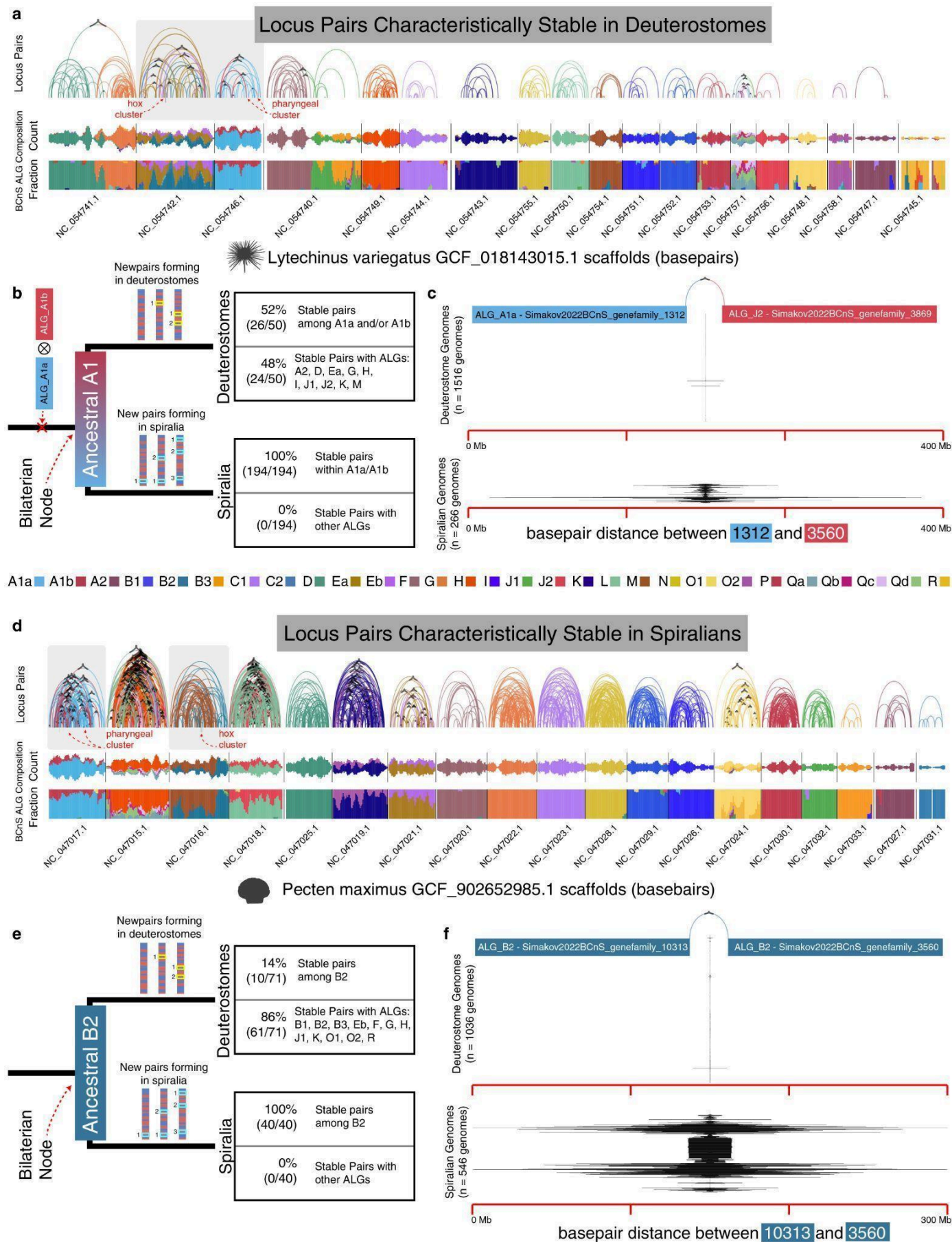


**Extended Data Fig. 9 | Multi-genome topology reflects genome evolutionary changes.** **a**, UMAP from Figure 5 ( $n=3,631$  genomes), with species within Insecta highlighted. The genomes with the fewest chromosomal tectonic changes since the ancestral myriazoan genome are circled. Genomes outside of this circle have varying degrees of tectonic changes from the ancestral BCnS ALGs. A clade with highly derived genomes, the Diptera, is also circled. Most chromosomes within the cluster of the least derived genomes contain single ALGs in **b**, *Pecten maximus*, **c**, *Branchiostoma floridae*, and **d**, *Rhopilema esculentum*. From the centroid of the UMAP mass there are both genomes with significant conservation of BCnS ALGs ( $p < 0.05$ , one-sided Bonferroni-corrected Fisher's exact test) **e**, in the tick *Dermacentor silvarum* and derived genomes with poor species sampling. Well-represented clades with lineage-specific genomic changes form clusters and lobes, such as **f**, bees (Anthophila, *Apis mellifera*), and **g**, beetles (Coleoptera, *Tribolium castaneum*). The multi-genome topology also demonstrates how genomes with similar highly reduced karyotypes can be discerned, as seen in **h**, mosquitos (*Anopheles gambiae*), and **i**, drosophilid flies (*Drosophila biarmipes*).



**Extended Data Fig. 10 | Multi-genome topology reflects genome phylogenetic changes.** Both panels are UMAP embeddings of clades within the all-animal dataset. Individual genomes (dots) are colored according to phylogeny, and a cladogram is drawn such that branches between genomes and ancestral nodes are gray lines. **a**, A reimbedding of mollusc genomes (n=108 genomes, 80 species) captures major phylogenetic transitions. Lineage-specific divergences from the ancestral spiralian genome, such as the karyotype reductions in oysters and limpets, can be seen as individual clusters of closely related genomes. Furthermore, lineage-specific changes in squids and octopus since the coleoid cephalopod ancestor are also captured by this approach. **b**, Reimbedding of Diptera genomes (n=509 genomes, 275 species), circled in **Extended Data Fig. 9**. Major dipteran clades colocalize in UMAP space, capturing lineage-specific genomic changes.





**Extended Data Fig. 11 | Evolutionary genome topology captures genome-defining features.** **a**, The chromosomes of the urchin *Lytechinus variegatus*. The bottom row is the fraction of BCnS ALG gene identities along the chromosome. The middle row is the absolute count of BCnS ALG genes along the chromosomes. The colorful arcs on the top row are the pairs of BCnS ALG that are characteristically stable in the Deuterostomes. Arcs with a black peak on top are stable pairs where the two loci are from different BCnS ALGs. **b**, After ALG A1a and A1b fused onto one chromosome before the ancestor of the Bilateria, the evolutionary history of the ALGs progressed differently in the deuterostomes versus spiralian. **c**, The distance between one locus pair that is characteristically stable in the deuterostomes. The locus is tightly linked in deuterostome genomes, but varies in genomic distance widely, in non-deuterostome genomes. **d**, Locus pairs that are characteristically stable in spiralian, plotted in the context of the scallop *Pecten maximus*. **e**, There are more inter-ALG stable pairs that have formed in deuterostomes with BCnS ALG B2, the ALG associated with the Hox cluster, versus the spiralian. **f**, New characteristically stable pairs can form in lineage-specific ways, even from loci that were present on the same chromosome in the ancestral organism. This is a B2-B2 pair that became characteristically stable in deuterostomes, but is highly variable in spiralian.



# References

1. Simakov, O. *et al.* Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).
2. Lv, J., Havlak, P. & Putnam, N. H. Constraints on genes shape long-term conservation of macro-synteny in metazoan genomes. *BMC Bioinform.* **12**, 1–12 (2011).
3. Schultz, D. T. *et al.* Ancient gene linkages support ctenophores as sister to other animals. *Nature* **618**, 110–117 (2023).
4. Strasser, J. F. H., Irisarri, I., Williams, T. A. & Burki, F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat. Commun.* **12**, 1879 (2021).
5. Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008).
6. Sturtevant, A. H. & Novitski, E. The Homologies of the Chromosome Elements in the Genus *Drosophila*. *Genetics* **26**, 517–541 (1941).
7. Wright, S. On the Probability of Fixation of Reciprocal Translocations. *Am. Nat.* **75**, 513–522 (1941).
8. Albertin, C. B. *et al.* Genome and transcriptome mechanisms driving cephalopod evolution. *Nat. Commun.* **13**, 2427 (2022).
9. Wright, C. J., Stevens, L., Mackintosh, A., Lawniczak, M. & Blaxter, M. Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol* **8**, 777–790 (2024).
10. Tandonnet, S. *et al.* Chromosome-Wide Evolution and Sex Determination in the Three-Sexed Nematode *Auanema rhodensis*. *G3* **9**, 1211–1230 (2019).
11. Drosophila 12 Genomes Consortium *et al.* Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
12. Schultz, D. T. *et al.* Acceleration of genome rearrangement in clitellate annelids. *bioRxiv* 2024.05.12.593736 (2024) doi:10.1101/2024.05.12.593736.
13. Schmidbaur, H. *et al.* Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. *Nat. Commun.* **13**, 2172 (2022).
14. Hurst, L. D., Williams, E. J. B. & Pál, C. Natural selection promotes the conservation of linkage of co-expressed genes. *Trends Genet.* **18**, 604–606 (2002).
15. Irimia, M. *et al.* Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* **22**, 2356–2367 (2012).
16. Yin, Y. *et al.* Molecular mechanisms and topological consequences of drastic chromosomal rearrangements of muntjac deer. *Nat. Commun.* **12**, 6858 (2021).
17. Duboule, D. The rise and fall of Hox gene clusters. *Development* **134**, 2549–2560 (2007).
18. Zimmermann, B., Robert, N. S. M., Technau, U. & Simakov, O. Ancient animal genome architecture reflects cell type identities. *Nat Ecol Evol* **3**, 1289–1293 (2019).
19. M Real, F. *et al.* The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. *Science* **370**, 208–214 (2020).
20. Marlétaz, F. *et al.* The little skate genome and the evolutionary emergence of wing-like fins. *Nature* **616**, 495–503 (2023).
21. Gilbertson, S. E. *et al.* Topologically associating domains are disrupted by evolutionary genome rearrangements forming species-specific enhancer connections in mice and humans. *Cell Rep.* **39**, 110769 (2022).
22. Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358 (2001).
23. Lucek, K. *et al.* The Impact of Chromosomal Rearrangements in Speciation: From Micro- to Macroevolution. *Cold Spring Harb. Perspect. Biol.* **15**, (2023).
24. Bambach, R. K. Phanerozoic Biodiversity Mass Extinctions. *Annu. Rev. Earth Planet. Sci.* **34**, 127–155 (2006).
25. Thomson, K. S. Explanation of large scale extinctions of lower vertebrates. *Nature* **261**, 578–580 (1976).
26. Rohde, R. A. & Muller, R. A. Cycles in fossil diversity. *Nature* **434**, 208–210 (2005).
27. Simakov, O. *et al.* Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* **4**,

- 820–830 (2020).
28. Rota-Stabelli, O., Daley, A. C. & Pisani, D. Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. *Curr. Biol.* **23**, 392–398 (2013).
29. Redmond, A. K., Casey, D., Gundappa, M. K., Macqueen, D. J. & McLysaght, A. Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. *Nat. Commun.* **14**, 2879 (2023).
30. Hane, J. K. *et al.* A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol.* **12**, R45 (2011).
31. Simakov, O. & Kawashima, T. Independent evolution of genomic characters during major metazoan transitions. *Dev. Biol.* **427**, 179–192 (2017).
32. Aguinaldo, A. M. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
33. Lartillot, N. & Philippe, H. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 1463–1472 (2008).
34. Wang, Y., Gasser, R. B., Charlesworth, D. & Zhou, Q. Evolution of sexual systems, sex chromosomes and sex-linked gene transcription in flatworms and roundworms. *Nat. Commun.* **13**, 3239 (2022).
35. Ohno, S. *Evolution by Gene Duplication*. (Springer Science & Business Media, 1970).
36. Dehal, P. & Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
37. Lukhtanov, V. A. *et al.* Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9610–E9619 (2018).
38. Leibowitz, M. L., Zhang, C.-Z. & Pellman, D. Chromothripsis: A New Mechanism for Rapid Karyotype Evolution. *Annu. Rev. Genet.* **49**, 183–211 (2015).
39. Clarence, T. *et al.* Robust 3D modeling reveals spatiotopic properties of animal genomes. *iScience* **26**, 106136 (2023).
40. Ryan, J. F. *et al.* Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*. *PLOS ONE* **2**, e153 (2007).
41. Robert, N. S. M. *et al.* Emergence of distinct syntenic density regimes is associated with early metazoan genomic transitions. *BMC Genomics* **23**, 143 (2022).
42. Destanović, D. Comparative structural genomics of Octopus. (University of Vienna, 2023). doi:10.25365/thesis.73889.
43. Acemel, R. D., Maeso, I. & Gómez-Skarmeta, J. L. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, (2017).
44. Rogers, T. F. & Simakov, O. Emerging questions on the mechanisms and dynamics of 3D genome evolution in spiralians. *Brief. Funct. Genomics* **22**, 533–542 (2023).
45. Zimmermann, B. *et al.* Topological structures and syntenic conservation in sea anemone genomes. *Nat. Commun.* **14**, 8270 (2023).
46. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17507–17512 (2012).
47. Engström, P. G., Ho Sui, S. J., Drivenes, O., Becker, T. S. & Lenhard, B. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17**, 1898–1908 (2007).
48. Krefting, J., Andrade-Navarro, M. A. & Ibn-Salem, J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol.* **16**, 87 (2018).
49. Lazar, N. H. *et al.* Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. *Genome Res.* **28**, 983–997 (2018).
50. Liao, Y., Zhang, X., Chakraborty, M. & Emerson, J. J. Topologically associating domains and their role in the evolution of genome structure and function in *Drosophila*. *Genome Res.* **31**, 397–410 (2021).
51. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* (2018).
52. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
53. Parey, E. *et al.* Genome structures resolve the early diversification of teleost fishes. *Science* **379**, 572–575 (2023).

54. Batut, P. J. *et al.* Genome organization controls transcriptional dynamics during development. *Science* **375**, 566–570 (2022).
55. Simakov, O. *et al.* Hemichordate genomes and deuterostome origins. *Nature* **527**, 459–465 (2015).

## Methods

### *Software requirements*

The software requirements for these analyses are: bokeh v.3.3.4<sup>83</sup>, ETE3 v.3.1.3<sup>57</sup>, fastapy v.1.0.3<sup>62</sup>, matplotlib v.3.8.0<sup>65</sup>, networkx v.3.1<sup>69</sup>, numpy v.1.26.4<sup>66</sup>, pandas v.2.1.4<sup>60,61</sup>, PIL v.10.2.0<sup>64</sup>, scikit-learn v.1.4.1<sup>67</sup>, scipy v.1.12.0<sup>68</sup>, snakemake v.7.32.3<sup>56</sup>, and umap v.0.5.4<sup>51</sup>.

### *Constructing a genome database*

We sought to analyze a comprehensive dataset of all chromosome-scale animal genomes available through NCBI. For unannotated genomes we mapped the BCnS proteins to the chromosomes with miniprot v.0.12-r237<sup>59</sup> with standard parameters.

### *Identifying BCnS ALGs*

We used hmmer v.3.4<sup>63</sup> to compare the ancestral BCnS ALGs<sup>1</sup> to the chromosome-scale genomes of extant animals. The HMM search was performed with hmmsearch from hmmer v.3.4<sup>63</sup> with a maximum e-value inclusion cutoff of  $1 \times 10^5$  (parameter `--incE 1E-5`). For the unannotated genomes, the HMM search was not performed, as the genomes were already annotated with the BCnS ALG orthologs using miniprot. For each genome, the *p*-values of conservation between each pair of BCnS ALGs and extant chromosomes were calculated with a Bonferroni-corrected one-sided Fisher's exact test<sup>3</sup>.

### *Inferring ALG evolutionary history*

First, we calculated basic statistics about the presence/absence of BCnS ALGs in extant genomes. One genome for each NCBI taxid was selected to represent each species in the dataset ( $n=2,291$  genomes). To identify whether a BCnS ALG was present on the chromosome of an extant animal, we selected significant BCnS ALG-chromosome combinations ( $p \leq 0.05$ , Bonferroni-corrected one-sided Fisher's exact test). Colocalized ALGs were scored as such simply if they were significant and present on the same scaffold.

Using the same program, we next inferred the history of animal chromosome evolution from the BCnS ancestral linkage groups using a lineage-aware algorithm ( $n=3631$  genomes). Briefly, the program works by reducing the information about BCnS ALG localization in extant genomes into a matrix that encodes whether two BCnS ALGs are colocalized on the same chromosome in that genome. After matrix construction, for each genome the program performs a depth-first search from tip (species) to the animal root and infers the branch along which specific chromosome fusion or loss events occurred. The algorithm pseudo code is:

Inputs:

- *d*: DataFrame with information about ALG presence and colocalization on single chromosomes.
- *t*: NCBI taxid string for a species in *d*.

- $n$ : Node representing the inference point for ALGs. Can be a species or ancestral node.
- $min_M$ : Threshold for determining missing ALGs. We use 0.8. This means that 80% of the genomes being considered must be missing this ALG to consider it lost from that species set. This value below 100% allows for the detection of ALGs that are still present, but difficult to detect, due to the evolutionary history of the chromosomes in that clade.
- $min_N$ : Threshold for determining non-colocalized ALG pairs. We use 0.5. This means that if 50% of the species in this dataset do not have an ALG pair colocalized on the same chromosome, we mark that the ALGs are not colocalized on the branch leading up to the most basal shared node for this clade. If ALGs are colocalized on the same chromosome in the branch leading up to a clade, we would expect to find the colocalization in 100% of species. However, due to dispersion or the limitations of Fisher's exact test in genomes with few chromosomes, it is possible that using a cutoff of 100% will produce false negatives for ALG colocalizations. Therefore a cutoff below 100% is necessary to find colocalizations. Because the algorithm below polarizes observed colocalizations in one species against a sister clade, it instead measures the percent of species in the sister clade that do not contain the ALG colocalization in question. If convergent fusions were not possible, we would optimally test that 0% of species in the sister clade contain the ALG fusion in question to locate the branch on which the fusion occurred. However, the algorithm must accommodate the possibility of convergent fusions, lineage-specific fusions, the effects of dispersion, and the unequal sampling across the phylogenetic tree. For these reasons, to mark two ALGs as not being colocalized in a given clade, we first check that the ALGs are detectable in that clade in at least 50% of species, and that at least 50% of those species do not have the two ALGs colocalized to the same chromosome.



Algorithm 1:

```

for each species  $S$  in  $d$ :
     $currentTaxID \leftarrow t$  of  $S$ 
     $parentTaxID \leftarrow S$ 's parent taxonomic unit
    Identify ALGs conserved, missing, or colocalized in the current species.
    while  $parentTaxID \neq 1$ : # 1 is the root of NCBI TaxID
         $parentTaxID \leftarrow t$  of  $currentTaxID$ 's parent taxonomic unit
         $sisterClade \leftarrow$  all species from  $d$  whose  $t$  contains  $parentTaxID$ , but not  $currentTaxID$ .
        if  $parentTaxID == 33208$ :
            Unaccounted ALG losses/fusions are on the  $parentTaxID$ - $currentTaxID$  branch.
            break
        if no species are present in  $sisterClade$ :
            goto the  $currentTaxID$  reassignment line below.
        if ALGs not present in the  $S$  but are present in the  $sisterClade$  species:
            # Note: we use cutoff value  $min_M$  to make the above determination.
            Mark ALG loss on the  $parentTaxID$ - $currentTaxID$  branch.
        if an ALG pair is colocalized in  $S$  but are not colocalized in  $sisterClade$  species:
            # Note: we use cutoff value  $min_N$  to make the above determination.
            Mark ALG fusion on the  $parentTaxID$ - $currentTaxID$  branch.
         $currentTaxID \leftarrow parentTaxID$ 
    Print out the inferred ALG losses and fusions for  $S$  with information on the events' branches.

```

The final result is a string representation of ALG changes for each species.

One calculated value for each change is a measure of uniqueness of this specific change to this evolutionary branch. In our previous work<sup>1,3,27</sup> we identified major algebraic chromosomal changes that happened on specific evolutionary branches, which became fixed before the crown group of that clade and consequently this evolutionary state can be observed in all extant species in that clade. For example, the fusion of BCnS ALGs A1b and A1a occurred in the ancestor of the Bilateria<sup>1</sup>, and 98.3% of the inferred A1a-A1b fusion events independently inferred from the viewpoint of individual extant species occurred on the branch leading up to Bilateria. Similarly, the known Ea-Eb fusion is captured in this analysis, and 91.1% of all inferred Ea-Eb fusions occurred on this branch. This field can be used as a cutoff to select the most confidently-called ALG change events.

Another value calculated is the measure of the number of species in this clade in which this change could be detected. For example, despite the fact that 98.3% of A1a-A1b events detected were in the branch leading up to Bilateria, this change was only detectable in 54.5% of the species in this clade. This value is reflective of species in which A1a or A1b are no longer detectable due to dispersal.

### Dispersal characterization

We sought to characterize the degree of BCnS ALG dispersal since the myriazoan ancestor, as well as the rate of dispersal over that time period. For each genome, we measured the total fraction of the 2367 BCnS ALG orthologs that were identifiable in the genome, the fraction of 2367 BCnS ALG orthologs that were significantly grouped on any scaffold ( $p \leq 0.05$ , Bonferroni-corrected one-sided Fisher's exact test), the fraction of 2367 BCnS ALG orthologs that were significantly grouped on the largest chromosome ( $p \leq 0.05$ , Bonferroni-corrected one-sided Fisher's exact test), and the fraction each BCnS ALG's orthologs that are significantly on any chromosome ( $p \leq 0.05$ , Bonferroni-corrected one-sided Fisher's exact test). The genomes were then binned into five categories of decay.

To show the relationship between evolutionary time elapsed and the effects of dispersal, we compared genomes in our dataset (n=22 genomes, 22 species) to the chromosomes of the scallop *Pecten maximus*, which have undergone few inter-chromosomal changes since the myriazoan ancestor. Comparisons between extant species allow for estimation of the amount of dispersal over a given divergence time. We collected estimated divergence times for species in our analyses with TimeTree v.5<sup>70</sup>, and for species with no estimates on TimeTree, we selected the most closely related node in the tree with a time estimate. The value of conservation between *Pecten maximus* and other species was the total fraction of genes on the *Pecten maximus* chromosomes that were significantly conserved on any chromosome in the other species ( $p \leq 0.05$ , Bonferroni-corrected one-sided Fisher's exact test).

### *ALG evolution simulations*

To estimate whether any given category of ALG fusion, or dispersal in any clade in the tree happened more often than could be expected by random chance in an identical evolutionary scenario, we developed an ALG evolution Monte Carlo simulation.

To estimate a background rate of each type of event on every branch (n=6,376 branches, 6364 taxid nodes), the Monte Carlo simulation was run as follows: for each iteration of the Monte Carlo simulation a random seed value was selected, and the 29 BCnS ALG labels were shuffled. From these labels, the previously described algorithm (**Methods** - *Inferring ALG evolutionary history*) was used to infer the branches on which changes to ALGs occurred. The topology of the cladogram on which the changes were inferred was identical to that used for previous analyses. The results for 10,000 Monte Carlo simulations were summed to estimate the background rate of any given ALG change event happening on any given branch.

### *Correlation inference with species diversity*

We then measured the one-tailed Spearman's ranked correlation coefficient between the ALG fusion and loss rates over time versus the species extinction and origination intensity<sup>26</sup>. Briefly, a vector was constructed in which one unit represented one million years. We recursed over the phylogenetic tree and recorded the mean rate of ALG fusions or losses that occurred in each million-year bin for all species in the clade in question.

### *Chromosome evolution periodicity analysis*

To determine whether changes in chromosomal evolutionary rates have significant underlying periods, we conducted Fourier analyses and Monte Carlo simulations, following the methods described by Rohde and Muller (2005)<sup>26</sup>. Our dataset included rates of ALG fusions and losses across vertebrates and protostomes. We truncated the oldest time for each clade to between 250 to 525 million years ago depending on the age of the clade, and removed the most recent million-year bin, to minimize the chance of over- or under-estimating rates of change. Cubic or quartic polynomials were fitted to the rates, and detrended datasets were created by subtracting these functions. Discrete Fourier transforms were then performed on both raw and zero-padded data to identify the strongest peaks in the spectral power distribution for further analysis.

To estimate false discovery rates for the identified peaks, we employed Monte Carlo simulations of the Fourier analysis through random walks through clustered transitions (type 'W'), wherein the original dataset was partitioned into vectors and shuffled. We performed simulations with varying block sizes (10 to 50 blocks). Each simulation comprised 5000 random trials, and the results were summarized by reporting the percentage of trials with peak magnitudes matching those observed in the real data. This approach allowed us to assess the periodicity and potential significance of chromosomal changes in evolutionary history.

### *Simulation of fusion-with-mixing*

We sought to simulate the mixing process of chromosomal fusion-with-mixing (FWM) events on single (simplified) chromosomes. During the simulations, we measured the rate and amount of gene mixing entropy over time on simplified chromosomes.

Such a chromosome consists of genes assigned to two linkage groups (LGs) A and B. Initially, the two groups are each present in contiguous blocks, only joined to each other at one A-to-B transition. This configuration emulates the state of a chromosome after a recent fusion or translocation event, and before any chromosome inversions have spanned into the genes that were originally on two separate chromosomes. Next, the chromosome is subjected to a process of stochastic inversions.

In each chromosome inversion cycle, both inversion breakpoints are chosen with a uniform distribution. We only allow inversions between genes, therefore we can treat the length of the chromosome as a vector of length  $L$ , one shorter than the number of genes on the chromosome,  $L = A + B - 1$ . Therefore, each value between 0 and  $L$  represents the cutpoint between a single gene. These two random variables range between  $[0, L]$ , and the distance between the two random variables,  $d$ , has a linear relationship because of the max value imposed by  $L$ . Given the range of possible values in  $[0, L]$ , the probability density function for  $d$  can be expressed as  $f_D(d) = \frac{2(L-d)}{L^2}$  for  $0 \leq d \leq L$ .

Therefore, the probability of yielding an inversion of length  $d$  linearly decreases.

At the beginning of the simulation, and after each inversion cycle, we calculate the mixing value  $m$  — a measurement of the entropy of the A and B genes that were previously on separate chromosomes. For a definition of  $m$ , see Supplementary Information Section 13.1.2, page 81, in Schultz et al. 2023<sup>3</sup>. Each simulation is run until the value  $m$  reaches a state in which further inversion events do not cause  $m$  to change dramatically — in other words after a burn-in period when  $m$  is exploring the main mass of the probability distribution of  $m$  in an ideally-mixed genome. After collecting this distribution, it is possible to estimate how many inversions it would take to reach a certain entropy value  $m$  observed in reality.

An additional feature of the simulation is that the novelty of local gene neighborhoods are measured after each inversion cycle. Every time there is an inversion, there is the possibility for two or more genes to come into close 2D proximity to one another for the first time. The simulation keeps track of newly encountered genes in windows of two to ten genes. This feature allows estimation of the number of inversions required to fully explore the space of possible gene combinations after the chromosome fusion event. The simulation offers the ability to collate the results for many runs and calculate averages or compare the mixing and novel locus encounter behavior under different parameter sets.

We ran chromosome simulations for four different window sizes  $W = 1, 2, 5, 10$  by executing the simulations with A, B gene counts:  $(A, B) = (4, 6), (20, 30), (45, 55), (100, 150), (200, 300), (300, 400), (450, 550)$ . Each combination of of chromosome size tuples  $(A, B)$  and window sizes  $W$  was simulated 10 times, unless  $2 \cdot W > A + B$ , (effectively, this means that  $(A, B) = (4, 6)$  was not run with  $W = 10$ ). Mean, standard deviation, median, 1<sup>st</sup> and 3<sup>rd</sup> quantiles, as well as minimum and maximum values were calculated from 11 runs with each parameter set.

These simulations can also be run using chromosome sizes, and A and B group sizes, inferred from chromosomal orthology. This mode of operation can infer how far along a chromosome is on the process of mixing through inversions by comparing the observed  $m$  to the distribution of  $m$  over the course of the simulation. We used *Pecten maximus* chromosomes that underwent ancient fusion-with-mixing events to

demonstrate the varied evolutionary history of these events. First, we identified which *Pecten maximus* fusion with mixing events. Consistent with previous results<sup>1,3</sup>, we found five chromosomes that are composed of two BCnS ancestral linkage groups that have fused and mixed. Among the five chromosomes there are two chromosomes with bilaterian synapomorphies ([NC\\_047017.1](#) - A1a⊗A1b, [NC\\_047021.1](#) - Ea⊗Eb), two chromosomes with spiralian synapomorphies ([NC\\_047018.1](#) - J2⊗L, [NC\\_047019.1](#) - K⊗O2), and one chromosome with a scallop synapomorphy ([NC\\_047016.1](#) - B2⊗M).

### *Inversion breakpoint inference*

We sought to infer rates at which chromosomal inversions become fixed by reconstructing the evolutionary history of homologous chromosomes in groups of closely related species.

Briefly, the analysis uses paf-format whole genome alignment files generated with minimap2 v.2.28<sup>72</sup>, subsetting the alignments to contain homologous chromosomes that have been identified using odp v.0.3.3<sup>3</sup> or D-GENIES v.1.5.0<sup>73</sup>. Only alignments of a minimal length of 100kb and quality score of 30 are considered. Collinearity breaks are then identified as discontinuities in the sorted pairwise alignment position table. Alignments are considered contiguous if there is a gap smaller than 5 million basepairs between them, and they share the same alignment orientation. This gap parameter can be adjusted for different species combinations to minimize falsely identified breakpoints.

To further study whether collinearity breakpoints overlap with insulation boundaries, breakpointer2 searches for relationships between breakpoints and regions between TADs and other 3D chromosomal topological features. The hypothesis that is tested is whether the breakpoints tend to occur between TADs or other 3D topological features. We define regions that occur between TADs or other 3D topological features as regions with low insulation scores. The null hypothesis is that breakpoints do not tend to occur between TADs and 3D topological features, and we model this distribution of randomly sampled insulation scores. To prepare the data for these analyses, the Hi-C reads were aligned to reference genomes with chromap v.0.2.3<sup>74</sup> with a quality cutoff of 0. Then pairix v.0.3.7<sup>75,76</sup> was used to bgzip compress the .pairs files from chromap. The program cooler v.0.9.1<sup>77</sup> was used to generate .cool matrices with bin sizes of 5000 bp, and 100,000 bp. The .cool files were then normalized with HiCNormalization from HiCExplorer v.3.7.2<sup>78</sup>. The normalized .cool files were balanced using cooler balance<sup>77</sup>. The normalized and balanced .cool files were converted into the .fance format using FAN-C v.0.9.26<sup>79</sup>, and insulation scores were generated with FAN-C for windows of 100 kbp, 250 kbp, 500 kbp, 750 kbp, and 1 Mbp. The resulting insulation scores, in .bed format, are used as the input for analyzing breakpoints with insulation score.

To complete the analysis of breakpoints and insulation scores, first the actual breakpoint coordinate is computed as the midpoint between the stop and the start of the previous and the next alignment region, respectively. The insulation scores at such locations and for each chromosome are then compared to the random sample across the same chromosome, with the sample size being the same as the number of observed breakpoints. The median insulation score for the random samples are calculated for 100,000 trials to generate a distribution of insulation score medians for a given chromosome. The medians of the observed breakpoint insulation scores are then tested against the random sampling median distribution using a one-tailed permutation test. The one-tailed permutation test yields a false discovery rate of finding an observed median insulation score smaller than the randomized background distribution. In other words, a false discovery rate of less than 0.05 represents breakpoints that occur in the lowest 5% of possible insulation scores for an identically sized sample, and likely represent regions between TADs or other topological features that have high insulation scores.



We characterized the breakpoints of the octopus species *Octopus vulgaris* ([GCA\\_951406725.2](#))<sup>71</sup>, *Octopus sinensis* ([GCA\\_006345805.1](#))<sup>80</sup>, *Octopus bimaculoides* ([GCA\\_001194135.2](#))<sup>8</sup>, and *Amphioctopus fangsiao* (<https://figshare.com/s/fa09f5dadcd966f020f3>)<sup>81</sup>. Using the inferred phylogenetic relationship for these four species<sup>81</sup>, and by comparing the .bed output of breakpoint2 using bedtools v.2.31.1<sup>82</sup>, we inferred shared and derived inversions along each evolutionary branch for these species. Inversion breakpoints found to exist on two sister species, but not in outgroup species, were assumed to be shared, derived inversions that occurred on the branch leading to the crown group of the two sister species.

### Forms of genome topological analysis

We implemented two forms of topological analysis in this study. In the first type, the topology of entire genomes is encoded for comparison. We term this analysis paradigm as ‘multi-genome topology’, or MGT. This allows not only for interspecies comparisons, but also for comparisons of multiple haplotypes or individuals within a given species. Each sample is a single genome, and when visualized each genome is represented by a single point. Each sample has a vector that contains all of the possible pairwise interactions for an orthology set. The values of the pairwise interactions could be basepair distance, basepair distance divided by chromosome length, Euclidean distance in 3D space, or other weight values derived from these or other sources. In this manuscript, we use the linear distance in basepairs for the distance measure. For measurements of linear distance along a chromosome, ortholog pairs on separate chromosomes should be represented as a number much larger than is possible on real chromosomes. Negative numbers should not be used to encode orthologs on separate chromosomes, as phylogenetically weighting those values could inadvertently bring the values close to a positive value that could be found by bona fide ortholog pairs on the same chromosome. We colloquially refer to visualizations and 2D projections of this matrix as ‘multi-genome topology’ (MGT) plots.

In the second form of topological analysis each sample is a single locus of a single genome, or of a phylogenetically weighted average from a collection of genomes. We refer to this analysis paradigm as ‘multi-locus topology’, or MLT. This analysis enables comparison of changes within genomes, such as colocalizations of loci, chromosomal tectonic changes, new or broken gene linkages. In other words, this analysis paradigm allows for simultaneous analysis of the concepts of colinearity, microsynteny, macrosynteny, and other chromosomal relationship concepts. Each sample is a single locus, and when visualized each locus is represented by a single point. In this manuscript, we use a 2367 x 2367 matrix of the BCnS ALG orthologs as the loci. Each sample, a single BCnS ALG ortholog, has a vector that contains the distance of this BCnS ALG to all other orthologs. This matrix is symmetric. The same rules for representing ortholog pairs on separate chromosomes described for multi-genome topology should also be followed for this approach, multi-locus topology. We refer to 2D projections of this matrix as One-Dot-One-Locus (ODOL) topology plots.

### Evolutionary topology implementation

The multi-locus topology (MLT) matrix creation and visualization procedure differs slightly from that of the multi-genome topology matrices. Each multi-locus topology matrix is a symmetric  $N \times N$  matrix, where  $N$  is the total number of orthologous loci across the entire dataset. The goal of this matrix is to express the topology of a genome for a single genome or collection of species, and to be able to compare topologies between species or clades. For this reason, it is necessary to average the distances between ortholog pairs for all of the samples in the comparison.

To perform phylogenetic weighting of the matrix, we first implement a directed, acyclic graph (DAG) reflecting the structure of the evolutionary relationships between species. In the phylogenetic weighting algorithm, the  $N \times N$  distance matrix of each node is calculated by normalizing the contribution of

different species depending on their relatedness. This prevents clades with many representative genomes, such as the many sequenced genomes of drosophilids or vertebrates, from biasing the averaged weights. In this approach, ortholog pairs that exist on separate chromosomes and are encoded as large values (999,999,999,999) are not treated differently in the weighting algorithm.

A more detailed explanation of the algorithm is below:

1. We construct the DAG, and its inverse graph, from NCBI taxid lineage information, and the branch lengths between taxids in the lineages were encoded as 1s. In this graph, nodes are extant species or ancestors of one or more species. Edges represent branches on the phylogenetic tree, and the value of the edge is the length of the branch. The root is determined and is defined as the node that has no parents. Below, we will refer to a node of the DAG as  $DAG[node1]$ , and the edge weight between two nodes as  $DAG[node1][node2]$ .
2. After constructing the DAG, the branch lengths are normalized such that the total distance from the root to any leaf is 1. The reason this is done is to enable matrix multiplication to be used to perform the phylogenetic averaging. Therefore, both cladograms and phylograms can be used as input.

The algorithm finds the longest path from the root and scales the branch lengths accordingly. Let  $P_{longest}$  be the longest path from the root to a leaf with  $\ell$  nodes (hence  $\ell - 1$  edges). We define

$$\begin{aligned} remaining\_length &= 1 \\ edge\_length &= \frac{remaining\_length}{\ell - 1} \end{aligned}$$

For each child  $c$  of node  $n$ :

$$dag[c][n] = \begin{cases} x(n) & \text{if } c \text{ is a leaf} \\ x(n - 1) & \text{otherwise} \end{cases}$$

The algorithm then recursively adjusts the branch lengths for the subtree of each child node. This is all performed in the function `adjacency_dag.normalize_branch_lengths()`.

3. A distance matrix is generated for each pair of tips  $(t_i, t_j)$ , where each tip  $t_i$  is one of the species in the tree that is being averaged. The distance between two tips  $(t_i, t_j)$  is the sum of the weights along the path from  $t_i$  to the last common ancestor node shared with  $t_j$ . The distance matrix,  $D$ , is then normalized to be between the maximum weight recorded in the matrix.

$$D_{normalized} = \frac{D}{\max(D)}$$

4. The phylogenetic weight calculation is as follows:

$$T_i = \sum_j D_{normalized}[i][j]$$

The total distance for each sample:

$$w_i = \frac{T_i}{\sum T_i}$$

The weight vector for the samples:

5. The phylogenetic weight matrix is calculated by calculating the dot product of the transposed multi-genome topology (MGT)  $matrix^T$  and the weight vector  $w$ .

$$matrix_{\text{phylogenetic average}} = matrix^T \cdot w$$

6. The final vector is transformed into the  $N \times N$  distance matrix based on the predetermined mapping of MGT matrix indices to ortholog IDs.

### Graph database integration

We used the Neo4j Community Edition server v.5.20.0 (<https://neo4j.com>) to build a graph database of orthology at multiple levels across the species utilized in this study, which consists of relationships, nodes, and node properties, all of which can be used for various types of semantic queries. The BCnS ALGs and orthogroups (separate ALG and OG nodes), the orthologous proteins from each species (gene nodes), and the scaffolds for each species (chromosome nodes) were all loaded as different node types into the graph. The *orthologous* relationship in our graph database connects the chromosome nodes with the ALG nodes, and the gene nodes with the OG nodes. The *location* relationship connects the gene nodes with the chromosome nodes, and the OG nodes with the ALG nodes. Both gene and chromosome nodes have the following properties: *TaxID*, which is the NCBI taxid of the species; *Assembly*, which is the version of the genome assembly; *Species*, which is the scientific name of the species; *Lineage*, which is the complete set of NCBI taxids forming the taxonomic lineage of the organism starting from the species ID; and *name*, which is the name of the gene or the chromosome. In cases when a chromosome node is actually not a real chromosome but a smaller scaffold that may still be present in a chromosome level genome assembly, a node is still labeled as a chromosome. However, we note that we only downloaded the chromosomal scaffolds for the genomes we collected from NCBI. In addition, the gene nodes have the properties *Chromosome* and *CoordinateStart*, which define the chromosome the gene is located on and the genomic coordinate where the gene starts. Both gene and chromosome nodes are labeled with a string with the first field as the *CHR* or *GENE*, plus the NCBI taxid of the source organism. For instance, nodes labeled as CHR9606 and GENE9606 belong to *Homo sapiens*. This taxid identifier can be used to query nodes with the same taxonomic source. A schematic of the graph database and an example illustration comparing the orthology of genes from an ALG between three different pairs of species are shown in **Extended Data Fig. 6a,b**.

The graph was visualized with Neo4j Browser v.5.15.0. We further designed a tool using the Neo4j Cypher language to generate queries to retrieve subgraphs that satisfy different topological questions. The query forms provide the user with options such as species and ALG names, to generate Cypher codes that can be copied or downloaded and run on our browser to explore the subgraph of interest. The graph data structure is also available as a Neo4j image for loading it locally. We have also included the graph as a GraphML file which is the common format for exchanging graph structure data that can be used in other graph database softwares, as a JSON file, and as a CSV file, all of which may also be loaded into other graph data structures for exploration.

## Methods References

56. Köster, J. & Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
57. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
58. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **49**, D10–D17 (2021).
59. Li, H. Protein-to-genome alignment with minimap2. *Bioinformatics* **39**, (2023).
60. The pandas development team. *Pandas-Dev/pandas: Pandas*. doi:10.5281/zenodo.10426137.
61. McKinney, W. Data Structures for Statistical Computing in Python. 56–61 (2010).
62. Zieleszinski, A., Michalczyk, M. & Dudeczak, K. *Azule/fastapy: Fastapy 1.0.3*. doi:10.5281/zenodo.10462111.
63. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
64. Murray, A. *et al.* *Python-pillow/Pillow: 10.2.0*. (2024). doi:10.5281/zenodo.10450403.
65. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
66. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
67. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **abs/1201.0490**, (2011).
68. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
69. Hagberg, A., Swart, P. J. & Schult, D. A. *Exploring Network Structure, Dynamics, and Function Using NetworkX*. <https://www.osti.gov/biblio/960616> (2008).
70. Kumar, S. *et al.* TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol. Biol. Evol.* **39**, (2022).
71. Destanović, D. *et al.* A chromosome-level reference genome for the common octopus, *Octopus vulgaris* (Cuvier, 1797). *G3 (Bethesda)* **13**, (2023).
72. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
73. Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).
74. Zhang, H. *et al.* Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat. Commun.* **12**, 1–6 (2021).
75. *Pairix: 1D/2D Indexing and Querying on Bgzipped Text File with a Pair of Genomic Coordinates*. (Github).
76. Lee, S., Bakker, C. R., Vitzthum, C., Alver, B. H. & Park, P. J. Pairs and Pairix: a file format and a tool for efficient storage and retrieval for Hi-C read pairs. *Bioinformatics* **38**, 1729–1731 (2022).
77. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
78. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
79. Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.* **21**, 303 (2020).
80. Li, F. *et al.* Chromosome-level genome assembly of the East Asian common octopus (*Octopus sinensis*) using PacBio sequencing and Hi-C technology. *Mol. Ecol. Resour.* **20**, 1572–1582 (2020).
81. Jiang, D. *et al.* The gold-ringed octopus (*Amphioctopus fangsiao*) genome and cerebral single-nucleus transcriptomes provide insights into the evolution of karyotype and neural novelties. *BMC Biol.* **20**, 289 (2022).
82. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
83. Bokeh Development Team. *Bokeh: Python Library for Interactive Visualization*. (2018).



# Acknowledgements

D.T.S., D.D., and O.S. were supported by the European Research Council's Horizon 2020: European Union Research and Innovation Programme, grant No. 945026. D.D. was also supported by the Vienna Haus des Meeres Verein: Rupert Riedl Prize. F.S. was supported by the Austrian Science Fund grant P32190 to O.S. The authors thank Nils Carqueville and Jacob Winnikoff for scientific discussions. The computation for this work was performed using the Life Science Compute Cluster (LiSC) of the University of Vienna.

# Author Contributions

O.S. and D.T.S. designed the scientific objectives of the study. Various parts of the code were written by D.T.S., A.B., F.S., and D.D. D.T.S., A.B., and F.S. created figures and tables. D.T.S. and O.S. wrote the manuscript under input from all other authors. All of the authors contributed to the interpretation, presentation and writing of the Article and the Supplementary Information.

# Competing Interests

The authors declare that they have no competing interests.

# Supplementary information

The online version contains supplementary information.

# Materials & Correspondence

Correspondence to: D.T. Schultz (darrin.schultz@univie.ac.at), O. Simakov (oleg.simakov@univie.ac.at)

# Data Availability

The genomes used are available through the INSDC database (ENA, NCBI, DDBJ) accession numbers: *Abramis brama* (GCA\_022829085.1), *Abrostola tripartita* (GCA\_905340225.1), *Abrostola triplasia* (GCA\_946251915.1), *Acanthisitta chloris* (GCA\_016904835.1), *Acanthochromis polyacanthus* (GCA\_021347895.1, GCF\_021347895.1), *Acanthococcus lagerstroemiae* (GCA\_031841125.1), *Acanthopagrus latus* (GCA\_013436515.1, GCA\_020080195.1, GCA\_904848185.1, GCF\_904848185.1), *Acanthosoma haemorrhoidale* (GCA\_930367205.1), *Accipiter gentilis* (GCA\_929443795.1, GCA\_929443795.2, GCF\_929443795.1), *Acentria ephemerella* (GCA\_943193645.1), *Achatina immaculata* (GCA\_009760885.1), *Achlya flavicornis* (GCA\_947623365.1), *Acholoe squamosa* (GCA\_949317995.1), *Acinonyx jubatus* (GCF\_027475565.1, GCA\_027475565.2), *Acipenser oxyrinchus oxyrinchus* (GCA\_030684275.1), *Acipenser ruthenus* (GCA\_902713425.2, GCF\_902713425.1, GCA\_010645085.2, GCF\_010645085.2, GCA\_902713435.2), *Acleris cristana* (GCA\_948252455.1), *Acleris emargana* (GCA\_927399475.2), *Acleris holmiana* (GCA\_949316455.1), *Acleris literana* (GCA\_946894065.1), *Acleris sparsana* (GCA\_923062465.1), *Acomys cahirinus* (GCA\_029890205.1), *Acomys dimidiatus* (GCA\_907164435.1), *Acomys kempfi* (GCA\_907164505.1), *Acomys percivali* (GCA\_907169655.1), *Acomys russatus* (GCF\_903995435.1, GCA\_903995435.1), *Acridotheres tristis* (GCA\_027559615.1), *Acrobasis consociella* (GCA\_963555685.1), *Acrobasis repandana* (GCA\_963576875.1), *Acrobasis suavella* (GCA\_943193695.1), *Acrocephalus scirpaceus scirpaceus* (GCA\_910950805.1), *Acrocera orbiculus* (GCA\_947359355.1), *Acronicta aceris* (GCA\_910591435.1), *Acronicta leporina* (GCA\_947256265.1), *Acronicta psi* (GCA\_946251955.1),

*Acropora hyacinthus* (GCA\_020536085.1), *Acropora millepora* (GCF\_013753865.1, GCA\_013753865.1), *Acrossocheilus wenchowensis* (GCA\_029620275.1), *Acutogordius australiensis* (GCA\_030555515.1), *Acyrtosiphon pisum* (GCF\_005508785.2, GCA\_005508785.2), *Adalia bipunctata* (GCA\_910592335.1), *Adalia decempunctata* (GCA\_951802165.1), *Adelphocoris suturalis* (GCA\_030762985.1), *Adineta vaga* (GCA\_021613535.1), *Aedes aegypti* (GCA\_002204515.1, GCF\_002204515.2), *Aedes aegypti aegypti* (GCA\_025407655.1), *Aelia acuminata* (GCA\_911387785.2), *Agapetus fuscipes* (GCA\_951799405.1), *Agapostemon virescens* (GCA\_028453745.1), *Agelaius phoeniceus* (GCA\_020745825.3, GCF\_020745825.1), *Agelas oroides* (GCA\_949130485.1), *Agelastica alni* (GCA\_950111635.1), *Agonopterix arenella* (GCA\_927399405.1), *Agonopterix heracliensis* (GCA\_963693445.1), *Agonopterix subpropinquella* (GCA\_922987775.1), *Agonum fuliginosum* (GCA\_947534325.1), *Agrilus cyanescens* (GCA\_947389935.1), *Agrilus mali* (GCA\_029378335.1), *Agriopsis aurantiaria* (GCA\_914767915.1), *Agriopsis leucophaearia* (GCA\_949125355.1), *Agriopsis marginaria* (GCA\_932305915.1), *Agriotes lineatus* (GCA\_940337035.1), *Agriphila genicula* (GCA\_943789525.1), *Agriphila straminella* (GCA\_950108535.1), *Agriphila tristella* (GCA\_928269145.1), *Agrochola circellaris* (GCA\_914767755.1), *Agrochola litura* (GCA\_949152395.1), *Agrochola lota* (GCA\_947364205.1), *Agrochola lychnidis* (GCA\_963680765.1), *Agrochola macilenta* (GCA\_916701695.1), *Agrotis clavis* (GCA\_954870645.1), *Agrotis exclamationis* (GCA\_950005045.1), *Agrotis ipsilon* (GCA\_028554685.1), *Agrotis puta* (GCA\_943136025.2), *Agrypnus murinus* (GCA\_929113105.1), *Ahaetulla prasina* (GCA\_028640845.1, GCF\_028640845.1), *Ailuropoda melanoleuca* (GCA\_002007445.3, GCA\_029963865.1, GCF\_002007445.2), *Albula goreensis* (GCA\_022829145.1), *Alca torda* (GCA\_008658365.1), *Alcis repandata* (GCA\_949125135.1), *Aldabrachelys gigantea* (GCA\_026122505.1), *Alectoris magna* (GCA\_030625065.1), *Aleiodes testaceus* (GCA\_963565655.1), *Alentia gelatinosa* (GCA\_950022655.1), *Alitta virens* (GCA\_932294295.1), *Allacma fusca* (GCA\_947179485.1), *Alligator mississippiensis* (GCA\_030867095.1, GCF\_030867095.1), *Allophyes oxyacanthae* (GCA\_932294325.1), *Alloplasta piceator* (GCA\_946863875.1), *Allygus modestus* (GCA\_963675035.1), *Alosa alosa* (GCF\_017589495.1, GCA\_017589495.2), *Alosa sapidissima* (GCF\_018492685.1, GCA\_018492685.1, GCA\_019202745.1), *Alsophila aescularia* (GCA\_946251855.1), *Alviniconcha strummeri* (GCA\_963584105.1), *Amaurobius ferox* (GCA\_951213105.1), *Amazona aestiva* (GCA\_017639355.1), *Amblyjoppa fuscipennis* (GCA\_963575735.1), *Amblyraja radiata* (GCA\_010909765.2, GCF\_010909765.2), *Amblyteles armatorius* (GCA\_933228735.2), *Ambystoma mexicanum* (GCA\_002915635.3), *Ameiurus melas* (GCA\_012411365.1), *Amia calva* (GCA\_017591415.1), *Ammodytes dubius* (GCA\_026122265.1), *Ammodytes marinus* (GCA\_949987685.1), *Ammospiza caudacuta* (GCA\_027887145.1, GCF\_027887145.1), *Ammospiza nelsoni* (GCF\_027579445.1, GCA\_027579445.1), *Amorpha juglandis* (GCA\_949126905.1), *Amphiduros pacificus* (GCA\_949316495.1), *Amphilophus citrinellus* (GCA\_013435755.1), *Amphimallon solstitiale* (GCA\_963170755.1), *Amphipoea lucens* (GCA\_947508005.1), *Amphipoea oculatea* (GCA\_945859645.1), *Amphiprion clarkii* (GCA\_027123335.1), *Amphiprion ocellaris* (GCA\_022539595.1, GCF\_022539595.1), *Amphiprion percula* (GCA\_003047355.2), *Amphipyra berbera* (GCA\_910594945.2), *Amphipyra tragopoginis* (GCA\_905220435.1), *Anabas testudineus* (GCA\_900324465.2, GCF\_900324465.2, GCA\_900324465.3), *Anableps anableps* (GCA\_014839685.1), *Anadara kagoshimensis* (GCA\_021292105.1), *Anania crocealis* (GCA\_949315895.1), *Anania fuscalis* (GCA\_950371115.1), *Anania hortulata* (GCA\_963576865.1), *Anarsia innoxia* (GCA\_947563765.1), *Anas platyrhynchos* (GCF\_003850225.1, GCF\_015476345.1, GCA\_015476345.1, GCA\_008746955.3, GCA\_003850225.1, GCA\_900411745.1, GCA\_030704485.1), *Anas platyrhynchos platyrhynchos* (GCA\_002743455.1), *Anaspis maculata* (GCA\_949128115.1), *Anastrepha ludens* (GCF\_028408465.1, GCA\_028408465.1), *Anastrepha obliqua* (GCA\_027943255.1, GCF\_027943255.1), *Ancistrocerus nigricornis* (GCA\_916049575.1), *Andrena bicolor* (GCA\_960531205.1), *Andrena bucephala* (GCA\_947577245.1), *Andrena camellia* (GCA\_029168295.1, GCA\_029448645.1), *Andrena chrysosceles* (GCA\_963855975.1), *Andrena dorsata* (GCA\_929108735.1), *Andrena fulva* (GCA\_946251845.1), *Andrena haemorrhoa* (GCA\_910592295.1), *Andrena hattorfiana* (GCA\_944738655.2), *Andrena minutula* (GCA\_929113495.1), *Andrena trimmerana* (GCA\_951215215.1, GCA\_952773225.1), *Anguilla anguilla* (GCA\_018320845.1, GCF\_013347855.1, GCA\_013347855.1), *Anguilla japonica* (GCA\_025169545.1), *Anguilla rostrata* (GCA\_018555375.2), *Anisus vortex* (GCA\_949126835.1), *Anolis carolinensis* (GCF\_000090745.2, GCA\_000090745.2), *Anomalospiza imberbis* (GCA\_031753505.1), *Anomoia purmunda* (GCA\_951828415.1), *Anopheles albimanus* (GCF\_013758885.1, GCA\_000349125.2, GCA\_015501965.1, GCA\_013758885.1),

*Anopheles aquasalis* (GCA\_943734665.1, GCF\_943734665.1), *Anopheles arabiensis* (GCF\_016920715.1, GCA\_016920715.1), *Anopheles atroparvus* (GCA\_015501955.1), *Anopheles bellator* (GCF\_943735745.2, GCA\_943735745.2), *Anopheles coluzzii* (GCA\_016097195.1, GCA\_016508615.1, GCA\_943734685.1, GCA\_016920705.1, GCA\_016097205.1, GCA\_016097185.1, GCA\_016097225.1, GCF\_016920705.1, GCA\_016097175.1, GCA\_016097095.1, GCF\_943734685.1), *Anopheles coustani* (GCA\_943734705.2, GCF\_943734705.1), *Anopheles cruzii* (GCA\_943734635.1, GCF\_943734635.1), *Anopheles darlingi* (GCF\_943734745.1, GCA\_943734745.1), *Anopheles funestus* (GCA\_943734845.1, GCF\_943734845.2, GCA\_003951495.1), *Anopheles funestus-like sensu Spillings et al. (2009)* (GCA\_016170035.1), *Anopheles gambiae* (GCA\_943734735.2), *Anopheles gambiae str. PEST* (GCF\_000005575.2, GCA\_000005575.1), *Anopheles longipalpis* (GCA\_016170015.1), *Anopheles maculipalpis* (GCA\_943734695.1, GCF\_943734695.1), *Anopheles marshallii* (GCA\_943734725.1, GCF\_943734725.1), *Anopheles merus* (GCF\_017562075.2, GCA\_017562075.2), *Anopheles moucheti* (GCA\_943734755.1, GCF\_943734755.1), *Anopheles nili* (GCA\_943737925.1, GCF\_943737925.1), *Anopheles parensis* (GCA\_016254315.1), *Anopheles rivulorum* (GCA\_018908115.1), *Anopheles sp. NFL-2015* (GCA\_016171315.1), *Anopheles stephensi* (GCF\_013141755.1, GCA\_023078585.1, GCA\_003448955.2, GCA\_013141755.1, GCA\_017562265.1, GCA\_020882675.1, GCA\_003448975.2), *Anopheles vaneedeni* (GCA\_016170025.1), *Anopheles ziemanni* (GCF\_943734765.1, GCA\_943734765.2), *Anoplius nigerrimus* (GCA\_914767735.1), *Anoplopoma fimbria* (GCA\_027596085.2, GCF\_027596085.1), *Anorthoa munda* (GCA\_945859665.1), *Anser cygnoides* (GCA\_026259575.1, GCA\_025388735.1), *Anser cygnoides domesticus* (GCA\_030515125.1), *Anser indicus* (GCA\_025583725.1), *Antechinus flavipes* (GCF\_016432865.1, GCA\_016432865.2), *Antedon bifida* (GCA\_963402885.1), *Antennarius maculatus* (GCA\_013358685.1), *Anthidium xuezhongii* (GCA\_022405125.1), *Anthocharis cardamines* (GCA\_905404175.1), *Anthonomus grandis grandis* (GCA\_022605725.3, GCF\_022605725.1), *Anthonomus grandis thurberiae* (GCA\_030068095.1), *Anthophora plumipes* (GCA\_951804975.1), *Anticlea derivata* (GCA\_947579855.1), *Antitype chi* (GCA\_947359405.1), *Antrozous pallidus* (GCA\_027563665.1), *Apalone spinifera* (GCA\_030068395.1), *Apamea anceps* (GCA\_951799955.1), *Apamea crenata* (GCA\_949629185.1), *Apamea epomidion* (GCA\_947507525.1), *Apamea monoglypha* (GCA\_911387795.2), *Apamea sordens* (GCA\_945859715.1), *Apeira syringaria* (GCA\_934044485.1), *Apeltes quadracus* (GCA\_021346845.1), *Aphelinus atriplicis* (GCA\_020882685.1), *Aphelinus certus* (GCA\_020882725.1), *Aphidius gifuensis* (GCA\_014905175.1, GCF\_014905175.1), *Aphidoletes aphidimyza* (GCA\_030463065.1), *Aphis gossypii* (GCA\_917880025.4, GCA\_020184165.1, GCA\_020184175.2, GCF\_020184175.1), *Aphrocallistes beatrix* (GCA\_963281255.1), *Aphrophora alni* (GCA\_963513935.1), *Apis cerana* (GCA\_020141525.2, GCA\_029531695.1, GCA\_029169275.1), *Apis cerana cerana* (GCA\_011100585.1), *Apis mellifera* (GCF\_000002195.4, GCF\_003254395.2, GCA\_000002195.1, GCA\_003254395.2), *Apis mellifera carnica* (GCA\_013841245.2), *Apis mellifera ligustica* (GCA\_019321825.1), *Apis mellifera mellifera* (GCA\_003314205.2), *Aplidium turbinatum* (GCA\_918807975.1), *Aplocera efformata* (GCA\_921293045.1), *Aplochiton taeniatus* (GCA\_017639675.1), *Aplysina aerophoba* (GCA\_949841015.1), *Apocheima hispidaria* (GCA\_947579745.1), *Apoda limacodes* (GCA\_946406115.1), *Apodemus sylvaticus* (GCF\_947179515.1, GCA\_947179515.1), *Apoderus coryli* (GCA\_911728435.2), *Apolygus lucorum* (GCA\_009739505.2), *Apomyeloides bistriatella* (GCA\_947044815.1), *Aporia crataegi* (GCA\_912999735.1), *Aporophyla lueneburgensis* (GCA\_932294355.1), *Aporophyla nigra* (GCA\_947507805.1), *Aporrectodea icterica* (GCA\_963556255.1), *Apotomis betuletana* (GCA\_932273695.1), *Apotomis capreana* (GCA\_947623375.1), *Apotomis turbidana* (GCA\_905147355.2), *Aproaerema taeniolella* (GCA\_949987775.1), *Apus apus* (GCA\_020740795.1, GCF\_020740795.1), *Aquila chrysaetos chrysaetos* (GCF\_900496995.4, GCA\_900496995.4), *Ara ararauna* (GCA\_028858755.1, GCA\_028858555.1), *Aradus depressus* (GCA\_963662175.1), *Archilochus colubris* (GCA\_023079485.1), *Archips crataeganus* (GCA\_947859365.1), *Archips xylosteana* (GCA\_947563465.1), *Archivesica marissinica* (GCA\_014843695.1), *Archocentrus centrarchus* (GCA\_007364275.2, GCF\_007364275.1), *Arctocephalus townsendi* (GCA\_028646355.1), *Argentina silus* (GCA\_951799395.1), *Argiope bruennichi* (GCF\_947563725.1, GCA\_947563725.1, GCA\_015342795.1), *Argynnis bischoffii washingtonia* (GCA\_033884665.1), *Argyresthia albistria* (GCA\_963457715.1), *Argyresthia goedartella* (GCA\_949825045.1), *Argyrosomus regius* (GCA\_946959425.1), *Aricia agestis* (GCA\_944452695.1, GCA\_905147365.1, GCF\_905147365.1), *Aricia artaxerxes* (GCA\_937612035.1), *Arion vulgaris* (GCA\_020796225.1), *Artemia franciscana* (GCA\_032884065.1), *Artemia sinica* (GCA\_027921565.1), *Arvicanthus niloticus* (GCF\_011762505.1,



GCA\_011762505.1), *Arvicola amphibius* (GCA\_903992535.2, GCF\_903992535.2), *Ascaris suum* (GCA\_013433145.1), *Ascidia mentula* (GCA\_947561715.1), *Aselliscus stoliczkanus* (GCA\_033961575.1), *Aspidapion aeneum* (GCA\_963576565.1), *Aspiorhynchus laticeps* (GCA\_023376895.1), *Astatotilapia calliptera* (GCA\_900246225.3, GCF\_900246225.1, GCA\_900246225.5), *Asterias amurensis* (GCA\_032118995.1), *Asterias rubens* (GCF\_902459465.1, GCA\_902459465.3), *Asteroscopus sphinx* (GCA\_949699075.1), *Astyanax mexicanus* (GCA\_023375845.1, GCA\_023375835.1, GCA\_000372685.2, GCF\_023375975.1, GCA\_019721115.1, GCA\_023375975.1, GCF\_000372685.2), *Atethmia centrigo* (GCA\_905333075.3), *Athalia rosae* (GCF\_917208135.1, GCA\_917208135.1), *Atherix ibis* (GCA\_958298945.1), *Athetis lepigone* (GCA\_033675125.1), *Athrips mouffetella* (GCA\_947532105.1), *Athripsodes cinereus* (GCA\_947579605.1), *Auanema freiburgensis* (GCA\_030370435.1), *Augochlora pura* (GCA\_028453695.1), *Augochlorella aurata* (GCA\_028455555.1), *Autographa gamma* (GCA\_905146925.1), *Autographa pulchrina* (GCA\_905475315.1), *Aythya baeri* (GCA\_026413565.1), *Aythya fuligula* (GCA\_009819795.1, GCF\_009819795.1), *Aythya marila* (GCA\_029042245.1), *Azorinus chamasolen* (GCA\_963576725.1), *Babyrousa celebensis* (GCA\_028533215.1), *Baccha elongata* (GCA\_951217065.1), *Bacillus rossius redtenbacheri* (GCA\_032445375.1), *Bactericera cockerelli* (GCA\_024516035.1), *Bactrocera correcta* (GCA\_027475135.1), *Bactrocera dorsalis* (GCF\_023373825.1, GCA\_023373825.1, GCA\_030710565.1), *Bactrocera neohumeralis* (GCA\_024586455.2, GCF\_024586455.1), *Bactrocera tryoni* (GCA\_016617805.2, GCF\_016617805.1), *Balaenoptera acutorostrata* (GCA\_949987535.1, GCF\_949987535.1), *Balaenoptera musculus* (GCF\_009873245.2, GCA\_009873245.3), *Balaenoptera ricei* (GCF\_028023285.1, GCA\_028023285.1), *Balanococcus diminutus* (GCA\_959613365.1), *Balearica regulorum gibbericeps* (GCA\_011004875.1), *Barbatula barbatula* (GCA\_947034865.1), *Barbus barbus* (GCA\_023566175.1, GCA\_936440315.1), *Bathymodiolus brooksi* (GCA\_963680875.1), *Bathymodiolus septemdierum* (GCA\_963383655.1), *Beaufortia kweichowensis* (GCA\_019155185.1), *Bedellia somnulentella* (GCA\_963576735.1), *Bellardia bayeri* (GCA\_950370525.1), *Bellardia pandia* (GCA\_916048285.2), *Belonocnema kinseyi* (GCA\_010883055.1, GCF\_010883055.1), *Bembecia ichneumoniformis* (GCA\_910589475.2), *Bemisio tabaci* (GCA\_918797505.1), *Beris chalybata* (GCA\_949128065.1), *Beris morrisii* (GCA\_951812415.1), *Betta splendens* (GCA\_024678985.1, GCA\_013403625.1, GCF\_900634795.4, GCA\_900634795.4), *Bibio marci* (GCA\_910594885.2), *Bicyclus anynana* (GCA\_947172395.1, GCF\_947172395.1), *Bimastos eiseni* (GCA\_959347315.1), *Biomphalaria glabrata* (GCF\_947242115.1, GCA\_947242115.1), *Biston betularia* (GCA\_905404145.2), *Biston stratarius* (GCA\_950106695.1), *Blastobasis adustella* (GCA\_907269095.1), *Blastobasis laticolella* (GCA\_905147135.1), *Blastomussa wellsi* (GCA\_947652115.1), *Blennius ocellaris* (GCA\_963422515.1), *Blera fallax* (GCA\_946965025.1), *Blomia tropicalis* (GCA\_029204025.1), *Bolinopsis microptera* (GCA\_026151205.1), *Boloria euphrosyne* (GCA\_951802675.2), *Boloria selene* (GCA\_905231865.2), *Bombina bombina* (GCA\_027579735.1, GCF\_027579735.1), *Bombus affinis* (GCA\_024516045.2, GCF\_024516045.1), *Bombus breviceps* (GCA\_014825925.1), *Bombus campestris* (GCA\_905333015.3), *Bombus haemorrhoidalis* (GCA\_014825975.1), *Bombus hortorum* (GCA\_905332935.1), *Bombus huntii* (GCF\_024542735.1, GCA\_024542735.1), *Bombus hypnorum* (GCA\_911387925.2), *Bombus ignitus* (GCA\_014825875.1), *Bombus pascuorum* (GCF\_905332965.1, GCA\_905332965.1), *Bombus pratorum* (GCA\_930367275.1), *Bombus pyrosoma* (GCA\_014825855.1, GCF\_014825855.1), *Bombus sylvestris* (GCA\_911622165.2), *Bombus terrestris* (GCA\_910591885.2, GCA\_000214255.1, GCF\_910591885.1, GCF\_000214255.1), *Bombus turneri* (GCA\_014825825.1), *Bombus vestalis* (GCA\_963556215.1), *Bombylius discolor* (GCA\_939192795.1), *Bombylius major* (GCA\_932526495.1), *Bombyx mandarina* (GCA\_030267445.2), *Bombyx mori* (GCA\_014905235.2, GCF\_014905235.1, GCA\_027497135.1, GCA\_027497115.1, GCA\_026075555.1, GCA\_030269925.2), *Borostomias antarcticus* (GCA\_949987555.1), *Bos gaurus* (GCA\_014182915.2), *Bos gaurus x Bos taurus* (GCA\_946052875.1), *Bos grunniens* (GCA\_027580245.1, GCA\_005887515.3), *Bos grunniens x Bos taurus* (GCA\_009493645.1, GCA\_009493655.1), *Bos indicus* (GCA\_030271805.1, GCA\_026262375.1, GCA\_002933975.1, GCA\_000247795.2, GCA\_030270995.1, GCA\_030271795.1, GCA\_030267355.1, GCA\_030272505.1, GCA\_030270855.1, GCF\_000247795.1, GCA\_030270935.1, GCA\_030270945.1, GCA\_026262465.1, GCA\_030272475.1, GCA\_030272235.1, GCA\_030270865.1, GCA\_030272135.1, GCA\_026262515.1, GCA\_030270805.1, GCA\_029378735.1, GCA\_030270725.1, GCA\_030270715.1, GCA\_030270705.1, GCA\_030270685.1, GCA\_030270955.1, GCA\_030269815.1, GCA\_030269505.1, GCA\_030267375.1, GCA\_026262545.1, GCA\_030272495.1,



GCA\_030270825.1, GCA\_029378745.1, GCA\_030267345.1, GCA\_030270845.1), *Bos indicus* x *Bos taurus* (GCA\_003369685.2, GCA\_003369695.2, GCF\_003369695.1), *Bos javanicus* (GCF\_032452875.1, GCA\_032452875.1), *Bos mutus* (GCA\_027580195.1, GCA\_002968435.1), *Bos taurus* (GCA\_028973685.2, GCA\_034097375.1, GCF\_002263795.3, GCF\_000003205.7, GCA\_905123515.1, GCA\_030378505.1, GCA\_002263795.4, GCA\_021234555.1, GCA\_000003055.5, GCA\_000003205.6, GCA\_905123885.1, GCF\_000003055.6, GCA\_947034695.1, GCA\_021347905.1), *Bos taurus* x *Bison bison* (GCA\_018282465.1, GCA\_018282365.1, GCA\_030254855.1), *Bostrychus sinensis* (GCA\_020698575.1), *Brachylomia viminalis* (GCA\_937001585.2), *Brachymystax lenok tsinlingensis* (GCA\_031305455.1, GCA\_030435695.1), *Brachyopa scutellaris* (GCA\_949775065.1), *Brachypalpus laphriformis* (GCA\_945910035.1), *Brachyptera putata* (GCA\_907164805.1), *Brachypterus glaber* (GCA\_958510825.1), *Bradysia coprophila* (GCA\_014529535.1, GCA\_014529535.2, GCF\_014529535.1), *Branchellion lobata* (GCA\_947562095.1), *Branchiostoma floridae* (GCA\_000003815.2, GCF\_000003815.2), *Branchiostoma floridae* x *Branchiostoma belcheri* (GCA\_022788385.1, GCA\_019207075.1, GCA\_019207045.1), *Branchiostoma floridae* x *Branchiostoma japonicum* (GCA\_013266295.2, GCA\_015852565.1), *Branchiostoma lanceolatum* (GCA\_927797965.1), *Branchipolynoe longqiensis* (GCA\_030323885.1), *Brassicogethes aeneus* (GCA\_921294245.1), *Brenthis daphne* (GCA\_947093115.1), *Brenthis hecate* (GCA\_951806755.1), *Brenthis ino* (GCA\_921882275.1), *Brienomyrus brachyistius* (GCF\_023856365.1, GCA\_023856365.1), *Bruchidius siliquastri* (GCA\_949316355.1), *Brugia pahangi* (GCA\_012070555.1), *Buathra laborator* (GCA\_934046635.1), *Bubalus bubalis* (GCF\_003121395.1, GCF\_019923925.1, GCA\_003121395.1, GCA\_019923925.1), *Bubalus carabanensis* (GCA\_029407905.2, GCF\_029407905.1), *Bucorvus abyssinicus* (GCA\_009769605.1), *Budorcas taxicolor* (GCF\_023091745.1, GCA\_023091745.2), *Bufo bufo* (GCA\_905171765.1, GCF\_905171765.1), *Bufo gargarizans* (GCF\_014858855.1, GCA\_014858855.1), *Bufotes viridis* (GCA\_033119425.1), *Bugulina stolonifera* (GCA\_935421135.1), *Bungarus multicinctus* (GCA\_023653725.1), *Bursaphelenchus mucronatus* (GCA\_025436335.1), *Cabera pusaria* (GCA\_954871355.1), *Caenorhabditis afra* (GCA\_963570955.1), *Caenorhabditis briggsae* (GCA\_021491975.1, GCA\_022453885.1, GCA\_029581135.1, GCF\_000004555.2, GCA\_000004555.3), *Caenorhabditis doughertyi* (GCA\_963572265.1), *Caenorhabditis drosophilae* (GCA\_963572285.1), *Caenorhabditis elegans* (GCA\_020450165.1, GCA\_004526295.1, GCA\_029748435.1, GCA\_000975215.1, GCA\_022984815.1, GCA\_028201515.1), *Caenorhabditis imperialis* (GCA\_963572205.1), *Caenorhabditis inopinata* (GCA\_003052745.1), *Caenorhabditis japonica* (GCA\_963572235.1), *Caenorhabditis latens* (GCA\_002259235.3), *Caenorhabditis nigoni* (GCA\_002742825.1, GCA\_027920645.1, GCA\_001643685.2), *Caenorhabditis remanei* (GCF\_010183535.1, GCA\_001643735.4, GCA\_002259225.3, GCA\_010183535.1), *Caenorhabditis sp. 8 KK-2011* (GCA\_963572245.1), *Caenorhabditis tropicalis* (GCA\_016735795.1), *Caenorhabditis uteleia* (GCA\_963573275.1), *Cairina moschata* (GCA\_018104995.1), *Cairina moschata domestica* (GCA\_009194515.1), *Calamotropha paludella* (GCA\_927399485.1), *Caligus rogercresseyi* (GCA\_013387185.1), *Callimorpha dominula* (GCA\_949752705.1), *Callinectes sapidus* (GCA\_020233015.1), *Calliopum simillimum* (GCA\_951812925.1), *Calliphora vicina* (GCA\_958450345.1), *Calliphora vomitoria* (GCA\_942486065.2), *Callitettix versicolor* (GCA\_022606455.1), *Callithrix jacchus* (GCF\_011100555.1, GCF\_009663435.1, GCA\_011078405.1, GCA\_013138085.1, GCA\_011100555.2, GCA\_009663435.2, GCA\_000004665.1, GCF\_000004665.1, GCA\_011100535.2), *Calypste anna* (GCA\_003957555.2, GCF\_003957555.1), *Camarhynchus parvulus* (GCA\_902806625.1, GCA\_901933205.1, GCF\_901933205.1), *Camelus dromedarius* (GCA\_000803125.3, GCF\_000803125.2), *Camelus ferus* (GCF\_009834535.1, GCA\_009834535.1), *Campaea margaritaria* (GCA\_912999815.1), *Campoletis raptor* (GCA\_948107755.1), *Camptogramma bilineatum* (GCA\_958496255.1), *Canis lupus* (GCA\_905319855.2), *Canis lupus dingo* (GCF\_003254725.2, GCA\_012295265.2, GCA\_003254725.2), *Canis lupus familiaris* (GCA\_000002285.4, GCA\_004886185.2, GCF\_014441545.1, GCF\_011100685.1, GCF\_005444595.1, GCF\_000002285.5, GCA\_008641055.3, GCA\_031010295.1, GCA\_011100685.1, GCA\_014441545.1, GCA\_031010635.1, GCA\_012044875.1, GCA\_005444595.1, GCA\_012045015.1, GCA\_031010765.1, GCA\_031165255.1, GCA\_013276365.2), *Cantharis flavilabris* (GCA\_949152465.1), *Cantharis lateralis* (GCA\_963170105.1), *Cantharis nigra* (GCA\_958510845.1), *Cantharis rufa* (GCA\_947369205.1), *Cantharis rustica* (GCA\_911387805.1), *Capitulum mitella* (GCA\_030062745.1), *Capra aegagrus* (GCA\_000978405.1), *Capra hircus* (GCF\_001704415.2, GCA\_000317765.2, GCA\_026652205.1, GCA\_001704415.2, GCA\_004361675.1, GCA\_015443085.1), *Capricornis sumatraensis* (GCA\_032405125.1),

*Caprimulgus europaeus* (GCA\_907165065.1), *Carabus problematicus* (GCA\_963422195.1), *Caradrina clavipalpis* (GCA\_932526535.1), *Caradrina kadenii* (GCA\_947462355.1), *Carassius auratus* (GCA\_003368295.1, GCA\_014332655.1, GCA\_019720715.2, GCF\_003368295.1), *Carassius carassius* (GCA\_963082965.1, GCF\_963082965.1), *Carassius gibelio* (GCA\_023724075.1, GCA\_019843895.2), *Carcharodon carcharias* (GCA\_017639515.1, GCF\_017639515.1), *Carcina quercana* (GCA\_910589575.2), *Carcinoscorpius rotundicauda* (GCA\_011833715.1), *Caretta caretta* (GCF\_023653815.1, GCA\_023653815.1), *Carettochelys insculpta* (GCA\_033958435.1), *Cariama cristata* (GCA\_009819825.1), *Carpatolechia fugitivella* (GCA\_951230895.1), *Carposina sasakii* (GCA\_014607495.2), *Carterocephalus palaemon* (GCA\_944567765.1), *Cataglyphis hispanica* (GCA\_021464445.1, GCA\_021464435.1, GCF\_021464435.1), *Catalaphyllia jardinei* (GCA\_022496165.2), *Catharus ustulatus* (GCA\_009819885.2, GCF\_009819885.2), *Catocala fraxini* (GCA\_930367265.1), *Catocala nupta* (GCA\_963675205.1), *Catocala sponsa* (GCA\_963564715.1), *Catoptria pinella* (GCA\_963556745.1), *Celastrina argiolus* (GCA\_905187575.2), *Centropristis striata* (GCA\_030273125.1, GCF\_030273125.1), *Ceraclea dissimilis* (GCA\_963576895.1), *Cerastis leucographa* (GCA\_963082945.1), *Cerastis rubricosa* (GCA\_949152445.1), *Cerastoderma edule* (GCA\_947846245.1), *Ceratotherium simum cottoni* (GCA\_021442165.1), *Ceratotherium simum simum* (GCA\_023653735.1), *Cerceris rybyensis* (GCA\_910591515.1), *Cercopithecus mitis* (GCA\_028627265.1), *Ceriagrion tenellum* (GCA\_963169105.1), *Certhia americana* (GCA\_018697195.1), *Cervus canadensis* (GCF\_019320065.1, GCA\_019320065.1), *Cervus elaphus* (GCA\_910594005.1, GCF\_910594005.1), *Cervus elaphus hippelaphus* (GCA\_002197005.1), *Cervus hanglu yarkandensis* (GCA\_010411085.1), *Cervus nippon* (GCA\_034195675.1), *Cetonia aurata* (GCA\_949128085.1), *Ceutorhynchus assimilis* (GCA\_917834065.1), *Chaenocephalus aceratus* (GCA\_023974075.1), *Chaetogeoica ovagalla* (GCA\_032441825.1), *Chalcis sispes* (GCA\_949987625.1), *Chalcosyrphus nemorum* (GCA\_949716465.1), *Channa argus* (GCA\_018997905.1, GCA\_033026475.1, GCA\_004786185.1), *Channa maculata* (GCA\_020496755.1), *Channa striata* (GCA\_033026295.1), *Channallabes apus* (GCA\_030522415.1), *Chanodichthys erythropterus* (GCA\_024489055.1), *Chanos chanos* (GCA\_902362185.1, GCF\_902362185.1), *Charanyca ferruginea* (GCA\_947361185.1), *Cheilinus undulatus* (GCA\_018320785.1, GCF\_018320785.1), *Cheilusia grossa* (GCA\_963082955.1), *Cheilusia impressa* (GCA\_948293265.1), *Cheilusia pagana* (GCA\_936431705.1), *Cheilusia scutellata* (GCA\_955612985.1), *Cheilusia soror* (GCA\_949372485.1, GCA\_948107745.1), *Cheilusia urbana* (GCA\_946477585.1), *Cheilusia variabilis* (GCA\_951230905.1), *Cheilusia vernalis* (GCA\_949126925.1), *Cheilusia vulpina* (GCA\_916610125.1), *Chelidonichthys spinosus* (GCA\_029853015.1), *Chelmon rostratus* (GCA\_017976325.1, GCF\_017976325.1), *Chelon labrosus* (GCA\_963514085.1), *Chelonia mydas* (GCF\_015237465.2, GCA\_015237465.2), *Chelonus formosanus* (GCA\_028641665.1), *Cherax quadricarinatus* (GCA\_026875155.2), *Chesias legatella* (GCA\_947359385.1), *Cheumatopsyche charites* (GCA\_024721215.1), *Chilo suppressalis* (GCA\_902850365.2), *Chiloscyllium plagiosum* (GCF\_004010195.1, GCA\_004010195.1), *Chionomys nivalis* (GCF\_950005125.1, GCA\_950005125.1), *Chironomus riparius* (GCA\_917627325.4), *Chironomus striatipennis* (GCA\_026123125.1), *Chironomus tentans* (GCA\_963573255.1), *Chiroxiphia lanceolata* (GCA\_009829145.1, GCF\_009829145.1), *Chlorocebus sabaeus* (GCA\_000409795.2, GCF\_000409795.2), *Chloroclysta siterata* (GCA\_932294275.1), *Chloroclystis v-ata* (GCA\_963691955.1), *Chlorops oryzae* (GCA\_020466095.1), *Choloepus didactylus* (GCF\_015220235.1, GCA\_015220235.1), *Chondrosia reniformis* (GCA\_947172415.1), *Choreutis nemorana* (GCA\_949316135.1), *Chorisops tibialis* (GCA\_963669355.1), *Choristoneura fumiferana* (GCA\_025370935.1), *Chroicocephalus ridibundus* (GCA\_030820635.1), *Chrysemys picta bellii* (GCA\_000241765.5, GCA\_011386835.2, GCF\_000241765.5), *Chrysocyon brachyurus* (GCA\_028533335.1), *Chrysodeixis includens* (GCA\_903961255.1, GCA\_941860345.1), *Chrysolina americana* (GCA\_958502065.1), *Chrysolina haemoptera* (GCA\_958298965.1), *Chrysolina oricalcia* (GCA\_944452925.2), *Chrysomallon squamiferum* (GCA\_012295275.1), *Chrysomela aeneicollis* (GCA\_027562985.1), *Chrysopa pallens* (GCA\_020423425.1), *Chrysoperla carnea* (GCA\_905475395.1, GCF\_905475395.1), *Chrysoteuchia culmella* (GCA\_910589605.1), *Chrysotoxum bicinctum* (GCA\_911387755.1), *Chydorus sphaericus* (GCA\_030141595.1), *Chymomyza fuscimana* (GCA\_949987675.1), *Ciconia maguari* (GCA\_017639555.1), *Cinclus cinclus* (GCA\_963662255.1), *Ciona intestinalis* (GCF\_000224145.3, GCA\_000224145.2), *Cirrhinus molitorella* (GCA\_033026305.1), *Cistogaster globosa* (GCA\_937654795.1), *Clarias fuscus* (GCA\_030347435.1), *Clarias gariepinus* (GCA\_024256465.1, GCF\_024256425.1, GCA\_024256435.1, GCA\_024256425.2), *Clavelina lepadiformis* (GCA\_947623445.1), *Cleipsis dumicolana*

(GCA\_963691665.1), *Clistopyga incitator* (GCA\_947507545.1), *Cloeon dipterum* (GCA\_949628265.1), *Clonorchis sinensis* (GCA\_003604175.2), *Clostera curtula* (GCA\_905475355.2), *Clupea harengus* (GCA\_900700415.2, GCF\_900700415.2), *Clusia tigrina* (GCA\_920105625.2), *Cnaphalocrocis medinalis* (GCA\_014851415.1), *Coccinella septempunctata* (GCA\_907165205.1, GCF\_907165205.1), *Coelioxys conoideus* (GCA\_947623535.1), *Coelopa frigida* (GCA\_017309665.1), *Coelopa pilipes* (GCA\_947389925.1), *Coendou prehensilis* (GCA\_028534165.1), *Coenobita brevipennis* (GCA\_032717465.1), *Coenonympha glycerion* (GCA\_963855885.1), *Coilia nasus* (GCA\_027475355.1, GCA\_007927625.1), *Colaptes auratus* (GCA\_015227895.2), *Coleophora deauratella* (GCA\_958295455.1), *Coleophora flavipennella* (GCA\_947284805.1), *Colias croceus* (GCA\_905220415.1, GCF\_905220415.1), *Colius striatus* (GCA\_028858725.1, GCA\_028858625.1), *Collichthys lucidus* (GCA\_004119915.2), *Colobus guereza* (GCA\_021498455.1, GCA\_030247045.1), *Coloeus monedula* (GCA\_013407035.1), *Cololabis saira* (GCA\_033807715.1), *Colostygia pectinataria* (GCA\_951394395.1), *Columba livia* (GCA\_001887795.1, GCA\_028654425.1), *Conchocele bisecta* (GCA\_029237695.1), *Conger conger* (GCF\_963514075.1, GCA\_029692045.1, GCA\_963514075.1), *Congerius kuscieri* (GCA\_027627225.1), *Conistra vaccinii* (GCA\_948150665.1), *Conogethes punctiferalis* (GCA\_031163375.1), *Conomelus anceps* (GCA\_948455865.1), *Conops quadrifasciatus* (GCA\_949752815.1), *Conus canariensis* (GCA\_033310375.1), *Conus ventricosus* (GCA\_018398815.1), *Cordilura impudica* (GCA\_963682025.1), *Cordylocheres scorpioides* (GCA\_030710605.1), *Coregonus clupeaformis* (GCA\_020615455.1, GCA\_018398675.1, GCF\_018398675.1, GCF\_020615455.1), *Coregonus sp. 'balchen'* (GCA\_902810595.1), *Coregonus ussuriensis* (GCA\_031479575.1), *Corella eumyota* (GCA\_963082875.1), *Coremacera marginata* (GCA\_914767935.1), *Coreoperca whiteheadi* (GCA\_011952105.1), *Coronaproctus castanopsis* (GCA\_032883995.1), *Corticium candelabrum* (GCA\_963422355.1), *Corvus cornix* (GCA\_000738735.6, GCF\_000738735.6), *Corvus hawaiiensis* (GCF\_020740725.1, GCA\_020740725.1), *Corvus moneduloides* (GCA\_009650955.1, GCF\_009650955.1), *Corythoichthys intestinalis* (GCF\_030265065.1, GCA\_030265065.1), *Cosmia pyralina* (GCA\_946251885.1), *Cosmia trapezina* (GCA\_905163495.3), *Cosmorhoe ocellata* (GCA\_963675405.1), *Cotesia glomerata* (GCF\_020080835.1, GCA\_020080835.1), *Cottoperca gobio* (GCA\_900634415.1, GCF\_900634415.1), *Cottus gobio* (GCA\_023566465.1), *Coturnix japonica* (GCF\_001577835.2, GCA\_001577835.2), *Crambus lathoniellus* (GCA\_949710035.1), *Craniophora ligustri* (GCA\_905163465.1), *Cranoglanis boudierius* (GCA\_026119655.1, GCA\_023630585.1), *Crassostrea angulata* (GCA\_025612915.2, GCF\_025612915.1, GCA\_025765675.3), *Crassostrea ariakensis* (GCA\_020458035.1, GCA\_020567875.1), *Crassostrea gigas* (GCA\_011032805.1, GCF\_902806645.1, GCA\_902806645.1, GCA\_025765685.3), *Crassostrea hongkongensis* (GCA\_015776775.1), *Crassostrea nippona* (GCA\_033439105.1), *Crataerina pallida* (GCA\_949710015.1), *Crepidodera aurea* (GCA\_949320105.2), *Cricetulus griseus* (GCA\_003668045.2, GCF\_003668045.3), *Crioceris asparagi* (GCA\_958507055.1), *Criorhina berberina* (GCA\_917880715.2), *Criorhina ranunculi* (GCA\_951813785.1), *Crocallis elinguarina* (GCA\_907269065.1), *Cromileptes altivelis* (GCA\_013133815.1, GCA\_019925165.1), *Crossocerus cetratus* (GCA\_963675795.1), *Crotalus viridis viridis* (GCA\_003400415.2), *Cryptocephalus moraei* (GCA\_946251905.1), *Cryptocephalus primarius* (GCA\_963576515.1), *Cryptophagus acutangulus* (GCA\_963556235.1), *Cryptoprocta ferox* (GCA\_028646485.1), *Cryptosula pallasiana* (GCA\_945261195.1), *Ctenicera cuprea* (GCA\_958336395.1), *Ctenopharyngodon idella* (GCA\_019924925.1, GCF\_019924925.1), *Cuculus canorus* (GCF\_017976375.1, GCA\_017976375.1), *Culex pipiens molestus* (GCA\_024516115.1), *Culex pipiens pallens* (GCF\_016801865.2, GCA\_016801865.2), *Culex quinquefasciatus* (GCA\_015732765.1, GCF\_015732765.1), *Cyaniris semiargus* (GCA\_905187585.1), *Cybosia mesomella* (GCA\_946251805.1), *Cyclophora albipunctata* (GCA\_963082685.1), *Cyclophora punctaria* (GCA\_951394245.1), *Cyclopterus lumpus* (GCA\_963457625.1, GCF\_009769545.1, GCA\_009769545.1), *Cyclura pinguis* (GCA\_030412105.1), *Cydalima perspectalis* (GCA\_951394215.1), *Cydia amplana* (GCA\_948474715.1), *Cydia fagiglandana* (GCA\_963556715.1), *Cydia pomonella* (GCA\_033807575.1), *Cydia splendana* (GCA\_910591565.2), *Cydia strobilella* (GCA\_947568885.1), *Cygnus atratus* (GCA\_013377495.2, GCF\_013377495.2), *Cygnus olor* (GCF\_009769625.2, GCA\_009769625.2), *Cynaetus angustus* (GCA\_030157275.1), *Cynocephalus volans* (GCA\_027409185.1), *Cynoglossus semilaevis* (GCA\_000523025.1, GCF\_000523025.1), *Cynopterus sphinx* (GCA\_030015415.1), *Cyprinodon diabolis* (GCA\_030533445.1), *Cyprinodon nevadensis mionectes* (GCA\_030533455.1), *Cyprinus carpio* (GCA\_018340385.1, GCA\_000951615.2, GCF\_018340385.1, GCF\_000951615.1, GCA\_027406505.1),



*Cyrtorhinus lividipennis* (GCA\_019603395.1), *Dallia pectoralis* (GCA\_029465975.1), *Dama dama* (GCA\_033118175.1, GCF\_033118175.1), *Danaus plexippus* (GCA\_018135715.1, GCF\_018135715.1), *Danaus plexippus plexippus* (GCF\_009731565.1, GCA\_009731565.1), *Danio aesculapii* (GCF\_903798145.1, GCA\_903798145.1), *Danio kyathit* (GCA\_903798195.1), *Danio rerio* (GCA\_903798175.1, GCA\_000002035.4, GCA\_903798165.1, GCA\_020184715.1, GCA\_018400075.1, GCA\_903684855.2, GCA\_008692375.1, GCA\_903684865.1, GCF\_000002035.6, GCA\_903798185.1, GCA\_944039275.1), *Daphnia carinata* (GCF\_022539665.2, GCA\_022539665.4), *Daphnia galeata* (GCA\_030770115.1), *Daphnia magna* (GCA\_030254905.1, GCF\_020631705.1, GCA\_003990815.1, GCF\_003990815.1, GCA\_020631705.2), *Daphnia pulex* (GCA\_021134715.1, GCA\_028752575.1, GCA\_023526725.1, GCA\_028752595.1, GCF\_021134715.1), *Daphnia pulicaria* (GCA\_021234035.2, GCF\_021234035.1, GCA\_017493165.1, GCA\_021234015.2), *Daphnia sinensis* (GCA\_013167095.2), *Dascillus cervinus* (GCA\_949768715.1), *Dascyllus trimaculatus* (GCA\_024666655.1), *Dastarcus helophoroides* (GCA\_028583605.1), *Dasypolia templi* (GCA\_963555695.1), *Dasypus novemcinctus* (GCA\_030445055.1, GCA\_030445035.1, GCF\_030445035.1), *Dasysyrphus albostrigatus* (GCA\_946251815.1), *Dasyurus viverrinus* (GCA\_028643805.1, GCA\_020854095.1), *Decapterus maruadsi* (GCA\_030347415.2), *Deilephila elpenor* (GCA\_949752805.1), *Deilephila porcellus* (GCA\_905220455.2), *Delia radicum* (GCA\_021234595.1), *Delphinus delphis* (GCA\_949987515.1, GCF\_949987515.1, GCA\_030062865.1), *Dendrolimus kikuchii* (GCA\_019925095.2), *Dendrolimus pini* (GCA\_949752895.1), *Dendrolimus punctatus* (GCA\_012273795.1), *Dendropsophus ebraccatus* (GCA\_027789765.1), *Denticeps clupeoides* (GCA\_900700375.1, GCF\_900700375.1, GCA\_900700375.2), *Dermacentor silvarum* (GCA\_013339745.2, GCF\_013339745.2), *Dermatophagoides farinae* (GCA\_024713945.1), *Dermochelys coriacea* (GCA\_009764565.4, GCF\_009764565.3), *Deroplatys truncata* (GCA\_030765065.1), *Desmodus rotundus* (GCA\_022682495.1, GCF\_022682495.1, GCA\_022682395.1), *Diabrotica balteata* (GCA\_918026665.1), *Diabrotica virgifera virgifera* (GCF\_917563875.1, GCA\_917563875.2), *Diachrysia chrysitis* (GCA\_932294365.1), *Diadema setosum* (GCA\_033980235.1), *Diadumene lineata* (GCA\_918843875.1), *Diaperis boleti* (GCA\_963583935.1), *Diaphora mendica* (GCA\_949125395.1), *Diaphorina citri* (GCA\_030643865.1), *Diarsia brunnea* (GCA\_949774965.1), *Diarsia dahlii* (GCA\_949775195.1), *Diarsia mendica* (GCA\_949316265.1), *Diarsia rubi* (GCA\_932274075.1), *Diatraea saccharalis* (GCA\_918026875.4), *Diceros bicornis* (GCA\_028533375.1), *Diceros bicornis minor* (GCA\_020826845.1, GCF\_020826845.1), *Dicrocoelium dendriticum* (GCA\_944474145.2), *Dicycla oo* (GCA\_948252095.1), *Dicyrtomina minuta* (GCA\_949802685.1), *Diglossa brunneiventris* (GCA\_019023105.1), *Diloba caeruleocephala* (GCA\_947459985.1), *Dilophus febrilis* (GCA\_958336335.1), *Diogma glabrata* (GCA\_963693315.1), *Diorhabda carinata* (GCA\_029229535.1), *Diorhabda carinulata* (GCA\_026250575.1, GCF\_026250575.1), *Diorhabda elongata* (GCA\_026230145.1), *Diorhabda sublineata* (GCA\_026230105.1, GCF\_026230105.1), *Diplazon laetatorius* (GCA\_963662125.1), *Diplosoma virens* (GCA\_963680785.1), *Diprion similis* (GCA\_021155765.1, GCF\_021155765.1), *Dircenna loreta* (GCA\_963555665.1), *Discoglossus pictus* (GCA\_027410445.1), *Dissostichus eleginoides* (GCA\_031216635.1), *Ditula angustiorana* (GCA\_963691745.1), *Dolichopus griseipennis* (GCA\_963082915.1), *Dolichovespula media* (GCA\_911387685.1), *Dolichovespula saxonica* (GCA\_911387935.1), *Dolichovespula sylvestris* (GCA\_918808275.2), *Dolomedes plantarius* (GCA\_907164885.2), *Dorcus hopei* (GCA\_033060865.1), *Dorcus parallelipipedus* (GCA\_958336345.1), *Doryrhamphus excisus* (GCF\_030265055.1, GCA\_030265055.1), *Doryteuthis pealeii* (GCA\_023376005.1), *Dreissena polymorpha* (GCA\_020536995.1, GCF\_020536995.1), *Drepana falcatoria* (GCA\_945859725.1), *Drepanosiphum platanoidis* (GCA\_948098885.1), *Dromaius novaehollandiae* (GCA\_016128335.2), *Dromiciops gliroides* (GCA\_019393635.1, GCF\_019393635.1), *Drosophila albomicans* (GCF\_009650485.2, GCA\_009650485.2), *Drosophila americana* (GCA\_030788265.1), *Drosophila ananassae* (GCA\_017639315.2, GCA\_030555595.1, GCF\_017639315.1), *Drosophila athabasca* (GCA\_003185025.1, GCA\_008121215.1, GCA\_008121225.1), *Drosophila biarmipes* (GCA\_025231255.1, GCF\_025231255.1), *Drosophila bifasciata* (GCA\_009664405.1), *Drosophila bipectinata* (GCA\_030179905.1), *Drosophila busckii* (GCF\_001277935.1, GCA\_011750605.1, GCA\_001277935.1, GCF\_011750605.1), *Drosophila elegans* (GCA\_011057505.1), *Drosophila funebris* (GCA\_958295475.1), *Drosophila gunungcola* (GCA\_025200985.1, GCF\_025200985.1), *Drosophila histrio* (GCA\_958299025.2), *Drosophila immigrans* (GCA\_963583835.1, GCA\_963584065.1), *Drosophila innubila* (GCA\_004354385.2, GCF\_004354385.1, GCA\_004354385.1), *Drosophila kepulauaniana* (GCA\_023558455.1), *Drosophila kikkawai* (GCA\_030179895.1), *Drosophila lowei* (GCA\_008121275.1),



*Drosophila mauritiana* (GCA\_004382145.1, GCF\_004382145.1), *Drosophila melanogaster* (GCA\_029775775.1, GCA\_020141795.1, GCA\_029775175.1, GCA\_029774715.1, GCA\_029774615.1, GCA\_029774635.1, GCA\_029775755.1, GCA\_029774655.1, GCA\_029774675.1, GCA\_029774695.1, GCA\_029775815.1, GCA\_020142005.1, GCA\_029775195.1, GCF\_000001215.4, GCA\_020141985.1, GCA\_020141835.1, GCA\_029775795.1, GCA\_020142055.1, GCA\_020142045.1, GCA\_020141815.1, GCA\_020142025.1, GCA\_020141765.1, GCA\_020141845.1, GCA\_020141935.1, GCA\_024500395.1, GCA\_029774375.1, GCA\_029774395.1, GCA\_020142105.1, GCA\_020142085.1, GCA\_029774415.1, GCA\_029774435.1, GCA\_029774455.1, GCA\_029775475.1, GCA\_029774475.1, GCA\_002300595.1, GCA\_002310755.1, GCA\_002310775.1, GCA\_003397115.2, GCA\_003401685.1, GCA\_003401735.1, GCA\_003401745.1, GCA\_003401795.1, GCA\_000001215.4, GCA\_003401805.1, GCA\_020169495.1, GCA\_029775115.1, GCA\_029775135.1, GCA\_020141485.1, GCA\_020141745.1, GCA\_003401925.1, GCA\_003401975.1, GCA\_003402005.1, GCA\_020141585.1, GCA\_020141575.1, GCA\_020141625.1, GCA\_020141515.1, GCA\_015832445.1, GCA\_029775155.1, GCA\_015852585.1, GCA\_020141505.1, GCA\_020141495.1, GCA\_004798055.1, GCA\_004798075.2, GCA\_020141925.1, GCA\_003402015.1, GCA\_003401855.1, GCA\_020141595.1, GCA\_003402055.1, GCA\_029775675.1, GCA\_029775275.1, GCA\_029775655.1, GCA\_029775295.1, GCA\_029775635.1, GCA\_029775315.1, GCA\_029775615.1, GCA\_029775335.1, GCA\_020141665.1, GCA\_029775355.1, GCA\_020141855.1, GCA\_029775375.1, GCA\_029775555.1, GCA\_029775535.1, GCA\_029775515.1, GCA\_029775395.1, GCA\_020141875.1, GCA\_029775415.1, GCA\_029775495.1, GCA\_029775215.1, GCA\_029775715.1, GCA\_029775735.1, GCA\_020141675.1, GCA\_020141705.1, GCA\_003401885.1, GCA\_029775255.1, GCA\_029775695.1, GCA\_029775575.1, GCA\_020141655.1, GCA\_029775095.1, GCA\_020141735.1, GCA\_029775595.1, GCA\_003401915.1, GCA\_020141955.1, GCA\_029775235.1, GCA\_029775455.1), *Drosophila miranda* (GCF\_003369915.1, GCA\_003369915.2, GCF\_000269505.1, GCA\_000269505.2), *Drosophila nasuta* (GCA\_023558535.1, GCF\_023558535.1), *Drosophila nebulosa* (GCA\_024703675.1), *Drosophila novamexicana* (GCA\_030788195.1), *Drosophila obscura* (GCA\_963584055.1, GCA\_963583975.1), *Drosophila pallidifrons* (GCA\_023558445.1), *Drosophila persimilis* (GCA\_020698555.1), *Drosophila phalerata* (GCA\_951394115.2), *Drosophila pseudoobscura* (GCF\_009870125.1, GCA\_004329205.1, GCA\_009870125.2), *Drosophila pseudoobscura pseudoobscura* (GCA\_000001765.3, GCA\_020698565.1), *Drosophila santomea* (GCA\_016746245.2, GCF\_016746245.2), *Drosophila sechellia* (GCF\_004382195.2, GCA\_004382195.2), *Drosophila simulans* (GCA\_029774515.1, GCA\_029774535.1, GCA\_016746395.2, GCA\_029774575.1, GCA\_029774995.1, GCA\_029775075.1, GCA\_029774955.1, GCA\_029775035.1, GCA\_029775015.1, GCA\_029774895.1, GCA\_029775055.1, GCA\_029774915.1, GCA\_000259055.1, GCF\_000259055.1, GCA\_029774975.1, GCA\_029774595.1, GCA\_000259045.1, GCF\_016746395.2, GCA\_029774875.1, GCF\_000754195.2, GCA\_029774495.1, GCA\_029774935.1, GCA\_029774555.1, GCA\_029774735.1, GCA\_004382185.1, GCA\_000754195.3, GCA\_029774835.1, GCA\_029774815.1, GCA\_029774795.1, GCA\_000820565.1, GCA\_029774755.1, GCA\_029774855.1, GCA\_029774775.1), *Drosophila subobscura* (GCF\_008121235.1, GCA\_008121235.1, GCA\_963583985.1, GCA\_963584095.1), *Drosophila subpulchrella* (GCA\_014743375.2, GCF\_014743375.2), *Drosophila sulfurigaster albostrigata* (GCA\_023558435.1), *Drosophila sulfurigaster bilimbata* (GCA\_023558465.1), *Drosophila sulfurigaster sulfurigaster* (GCA\_023558475.1), *Drosophila suzukii* (GCF\_013340165.1, GCA\_013340165.1), *Drosophila takahashii* (GCA\_030179915.1), *Drosophila teissieri* (GCA\_016746235.2, GCF\_016746235.2), *Drosophila triauraria* (GCA\_014170255.2), *Drosophila virilis* (GCA\_007989325.2, GCA\_030788295.1, GCA\_000004125.1), *Drosophila willistoni* (GCF\_018902025.1, GCA\_018902025.2), *Drosophila yakuba* (GCA\_000005975.1, GCF\_016746365.2, GCA\_016746335.2, GCF\_000005975.2, GCA\_016746365.2), *Dryas iulia moderata* (GCA\_019049465.1), *Drymonia ruficornis* (GCA\_947859195.1), *Dryobates pubescens* (GCA\_014839835.1, GCF\_014839835.1), *Dryobates labecula* (GCA\_947523025.1), *Dryobotodes eremita* (GCA\_917490735.1), *Dryococelus australis* (GCA\_029891345.1), *Dryomyza anilis* (GCA\_951804985.1), *Dugong dugon* (GCA\_030035585.1), *Dunckerocampus dactyliophorus* (GCA\_027744805.2, GCF\_027744805.1), *Dysmachus trigonus* (GCA\_949715965.1), *Dyspetes luteomarginatus* (GCA\_963669185.1), *Ecdyonurus torrentis* (GCA\_949318235.1), *Echeneis naucrates* (GCA\_900963305.2, GCF\_900963305.1, GCA\_031770045.1, GCA\_900963305.1), *Echiichthys vipera* (GCA\_963691815.1), *Echinococcus granulosus* (GCA\_021556725.1), *Echinometra lucunter* (GCA\_029962745.1), *Ecliptopera silaceata*

(GCA\_932527185.1), *Ectemnius continuus* (GCA\_910591665.1), *Ectemnius lituratus* (GCA\_910593735.2), *Ectropis crepuscularia* (GCA\_963693475.1), *Ectropis grisescens* (GCA\_017562165.1), *Eidolon dupreanum* (GCA\_028627145.1), *Eilema caniola* (GCA\_949126895.1), *Eilema depressum* (GCA\_914767945.1), *Eilema griseolum* (GCA\_963662185.1), *Eilema sororculum* (GCA\_914829495.1), *Electrona antarctica* (GCA\_951216825.1), *Electrophaes corylata* (GCA\_947095575.1), *Electrophorus electricus* (GCA\_013358815.1, GCF\_013358815.1), *Elegia similella* (GCA\_947532085.1), *Elephas maximus* (GCA\_033060105.1, GCA\_032718585.1, GCA\_032718755.1), *Elephas maximus indicus* (GCF\_024166365.1, GCA\_024166365.1), *Eleutherodactylus coqui* (GCA\_019857665.1), *Elmis aenea* (GCA\_947652605.1), *Elophila nymphaeata* (GCA\_955850985.1), *Elysia crispata* (GCA\_963854125.1), *Emmelina monodactyla* (GCA\_916618145.1), *Empis stercorea* (GCA\_949752835.1), *Emys orbicularis* (GCA\_028017835.1), *Endomychus coccineus* (GCA\_958510875.1), *Endotricha flammealis* (GCA\_905163395.2), *Engystomops pustulosus* (GCA\_019512145.1), *Ennomos erosarius* (GCA\_963580505.1), *Ennomos fuscantarius* (GCA\_905220475.3), *Ennomos quercinarius* (GCA\_910589525.2), *Entelurus aequoreus* (GCA\_033978785.1), *Entomobrya proxima* (GCA\_029691765.1), *Entosphenus tridentatus* (GCA\_014621495.2), *Ephestia elutella* (GCA\_018467065.1), *Epiblema foenella* (GCA\_963556455.1), *Epicampocera succincta* (GCA\_932526305.1), *Epinephelus cyanopodus* (GCA\_026686955.1), *Epinephelus fuscoguttatus* (GCF\_011397635.1, GCA\_011397635.1), *Epinephelus lanceolatus* (GCA\_005281545.1, GCF\_005281545.1, GCA\_017165655.1), *Epinephelus moara* (GCA\_006386435.1, GCF\_006386435.1), *Epinotia bilunana* (GCA\_947049275.1), *Epinotia brunnichana* (GCA\_963854355.1), *Epinotia demarniana* (GCA\_945867215.1), *Epinotia nisella* (GCA\_932294315.1), *Epinotia ramella* (GCA\_947578815.1), *Epirrhoe alternata* (GCA\_963565295.1), *Epirrhoe tristata* (GCA\_951394285.1), *Epirrita christyi* (GCA\_951392215.1), *Epistrophe eligans* (GCA\_951394125.1), *Epistrophe grossulariae* (GCA\_929447395.1), *Epistrophehella euchroma* (GCA\_947049315.1), *Episyrphus balteatus* (GCA\_945859705.1, GCF\_945859705.1), *Erinaceus europaeus* (GCA\_950295315.1, GCF\_950295315.1), *Eriocheir sinensis* (GCA\_013436485.1, GCA\_024679095.1, GCA\_023348025.1, GCF\_024679095.1, GCA\_033459325.1), *Eriosoma lanigerum* (GCA\_013282895.1), *Eristalinus aeneus* (GCA\_955652365.1), *Eristalinus sepulchralis* (GCA\_944738645.1), *Eristalis arbustorum* (GCA\_916610145.1), *Eristalis pertinax* (GCA\_907269125.1), *Eristalis tenax* (GCA\_905231855.2), *Erithacus rubecula* (GCA\_903797595.2), *Erpetoichthys calabaricus* (GCA\_900747795.4, GCF\_900747795.2), *Erynnis tages* (GCA\_905147235.1), *Erythrolamprus reginae* (GCA\_031021105.1), *Eschrichtius robustus* (GCA\_028021215.1), *Esox lucius* (GCF\_000721915.3, GCA\_011004845.1, GCF\_004634155.1, GCF\_011004845.1, GCA\_004634155.1, GCA\_000721915.3), *Esperia sulphurella* (GCA\_947086405.1), *Etheostoma cragini* (GCF\_013103735.1, GCA\_013103735.1), *Etheostoma nigrum* (GCA\_029931835.1), *Etheostoma perlongum* (GCA\_026937815.1), *Etheostoma spectabile* (GCA\_008692095.1, GCF\_008692095.1), *Ethmia dodecea* (GCA\_963855545.1), *Eubalaena glacialis* (GCA\_028564815.2, GCF\_028564815.1, GCA\_028571275.2), *Eubasilissa splendida* (GCA\_031772225.1), *Eublepharis macularius* (GCF\_028583425.1, GCA\_028583425.1), *Euclidia mi* (GCA\_944739405.2), *Eucosma cana* (GCA\_951800055.1), *Eudemis profundana* (GCA\_947034925.1), *Eudonia lacustrata* (GCA\_947562085.1), *Eudonia mercurella* (GCA\_963082485.1), *Eudonia truncicolella* (GCA\_949315975.1), *Eueides isabella* (GCA\_019049475.1), *Eugnorisma glareosa* (GCA\_947578425.1), *Eulemur mongoz* (GCA\_028534055.1), *Euleptes europaea* (GCF\_029931775.1, GCA\_029931775.1), *Eulithis prunata* (GCA\_918843925.1), *Eulithis testata* (GCA\_947507515.1), *Eumerus sabulonum* (GCA\_951905685.1), *Eupeodes corollae* (GCA\_945859685.1, GCF\_945859685.1), *Eupeodes latifasciatus* (GCA\_920104205.1), *Eupeodes luniger* (GCA\_951509635.1), *Eupithecia abbreviata* (GCA\_943735965.1), *Eupithecia centaureata* (GCA\_944547425.1), *Eupithecia dodoneata* (GCA\_947044415.1), *Eupithecia exigua* (GCA\_947086465.1), *Eupithecia insigniata* (GCA\_947859395.1), *Eupithecia inturbata* (GCA\_963662085.1), *Eupithecia subfuscata* (GCA\_963564075.1), *Eupithecia subumbrata* (GCA\_949316285.1), *Eupithecia tripunctaria* (GCA\_955876795.1), *Eupithecia vulgata* (GCA\_946478455.1), *Euplexia lucipara* (GCA\_921972225.1), *Euproctis similis* (GCA\_905147225.2), *Euprymna scolopes* (GCA\_024364805.1), *Eupsilia transversa* (GCA\_914767815.1), *Eurois occulta* (GCA\_950022335.1), *Euspilapteryx auroguttella* (GCA\_951802225.1), *Eustalomyia histrio* (GCA\_949748255.1), *Euthynnus affinis* (GCA\_029490765.1), *Eutrigla gurnardus* (GCA\_963514095.1), *Euzophera pinguis* (GCA\_947363495.1), *Exephanes ischioxanthus* (GCA\_958510785.1), *Fabriciana adippe* (GCA\_905404265.1), *Falcaria lacertinaria* (GCA\_951449985.1), *Falco biarmicus* (GCA\_023638135.1, GCF\_023638135.1), *Falco cherrug*





*Harmonia axyridis* (GCA\_914767665.1, GCF\_914767665.1, GCA\_011033045.2), *Harmothoe impar* (GCA\_947462335.1), *Harpalus rubripes* (GCA\_963082715.1), *Harpalus rufipes* (GCA\_951394225.1), *Harpia harpyja* (GCA\_026419915.1, GCF\_026419915.1), *Hecatera dysodea* (GCA\_905332915.2), *Hedya salicella* (GCA\_905404275.2), *Helarctos malayanus* (GCA\_028533245.1), *Heliconius charithonia* (GCA\_030704555.1), *Heliconius sara* (GCA\_917862395.2), *Helicoverpa armigera* (GCF\_023701775.1, GCA\_023701775.1, GCA\_026262555.1), *Helicoverpa assulta* (GCA\_029618815.1), *Helicoverpa zea* (GCA\_022343045.1, GCF\_022581195.2, GCA\_022581195.1), *Heligmosomoides bakeri* (GCA\_947359475.1), *Heligmosomoides polygyrus polygyrus* (GCA\_947396885.1), *Heliocidaris erythrogramma* (GCA\_025617745.1), *Heliocidaris tuberculata* (GCA\_025618425.1), *Heliothis peltigera* (GCA\_958496145.1), *Hemaris fuciformis* (GCA\_907164795.1), *Hemibagrus wyckioideus* (GCA\_019097595.1, GCF\_019097595.1), *Hemicordylus capensis* (GCF\_027244095.1, GCA\_027244095.1), *Hemicrepidius niger* (GCA\_963082805.1), *Hemiprocne comata* (GCA\_020745705.1), *Hemiscyllium ocellatum* (GCA\_020745735.1, GCF\_020745735.1), *Hemistola chrysoprasaria* (GCA\_947063395.1), *Hemitea aestivaria* (GCA\_947507615.1), *Hepsetus odoe* (GCA\_017165825.1), *Hermaphysa mercurialis* (GCA\_951812935.1), *Hermetia illucens* (GCF\_905115235.1, GCA\_905115235.1), *Herminia tarsipennalis* (GCA\_945859575.2), *Hesperia comma* (GCA\_905404135.1), *Hestina assimilis* (GCA\_023373885.1), *Heterobilharzia americana* (GCA\_944470555.2, GCA\_944470545.2), *Heterocephalus glaber* (GCA\_944319715.1, GCA\_944319725.1), *Heterodera glycines* (GCA\_004148225.2), *Heterohyrax brucei* (GCA\_028571685.1), *Heteromyza rotundicornis* (GCA\_951394025.1), *Heteronetta atricapilla* (GCA\_011075105.1), *Heteronotia binoei* (GCA\_032191835.1, GCF\_032191835.1), *Heteropelma amictum* (GCA\_959613375.1), *Heterotis niloticus* (GCA\_018136845.1), *Hexactinellida sp. DTS-2022* (GCA\_028753945.1), *Hexaprotodon liberiensis* (GCA\_023065765.1), *Himalopsyche anomala* (GCA\_031772345.1), *Hipparchia semele* (GCA\_933228805.2), *Hippocampus abdominalis* (GCA\_018466805.1), *Hippocampus zosterae* (GCA\_026261805.1, GCF\_025434085.1, GCA\_025434085.3), *Hippoglossus hippoglossus* (GCF\_009819705.1, GCA\_009819705.1), *Hippoglossus stenolepis* (GCF\_022539355.2, GCA\_022539355.2), *Hippopotamus amphibius* (GCA\_023065835.1), *Hippopotamus amphibius kiboko* (GCF\_030028045.1, GCA\_030028045.1), *Hippopus hippopus* (GCA\_946811185.1), *Hipposideros larvatus* (GCA\_031876335.1), *Hirtodrosophila cameraria* (GCA\_949708635.1), *Hirundo rustica* (GCA\_015227805.3, GCF\_015227805.2), *Hofmannophila pseudospretella* (GCA\_947369225.1), *Hololepta plana* (GCA\_963695495.1), *Holothuria leucospilota* (GCA\_029531755.1), *Holotrichia oblita* (GCA\_023690525.1), *Homalodisca vitripennis* (GCF\_021130785.1, GCA\_021130785.2), *Homo sapiens* (GCA\_016695395.2, GCA\_948473225.1, GCA\_000002125.2, GCA\_948474725.1, GCA\_021951015.1, GCA\_016700455.2, GCA\_001292825.2, GCA\_003634875.1, GCA\_000365445.1, GCA\_948473255.1, GCA\_022833125.2, GCA\_001712695.1, GCA\_024586135.1, GCA\_002180035.3, GCA\_000002115.2, GCA\_000001405.29, GCA\_948473215.1, GCA\_020497115.1, GCA\_001524155.4, GCA\_000306695.2, GCA\_021950905.1, GCA\_000002135.3, GCA\_948473235.1, GCA\_000212995.1, GCA\_000252825.1, GCA\_948474675.1, GCA\_000442335.2, GCA\_948474735.1, GCA\_015476435.1, GCA\_016894425.1, GCA\_002077035.3, GCF\_000001405.40, GCA\_018852605.2, GCF\_000002125.1, GCF\_000306695.2, GCA\_018852615.2, GCA\_018873775.2, GCA\_020497085.1, GCA\_020881995.2, GCA\_011064465.2, GCA\_014905855.1), *Hopl意思 malabaricus* (GCA\_029633875.1), *Hoplocladia ambigua* (GCA\_949774945.1), *Hoplocladia blanda* (GCA\_949316365.1), *Horismia vitalbata* (GCA\_951804965.1), *Hormaphysa cornu* (GCA\_017140985.1), *Hormiphora californensis* (GCA\_020137815.1), *Hyalomma asiaticum* (GCA\_013339685.2), *Hydra vulgaris* (GCA\_022113875.1, GCA\_024232925.1, GCF\_022113875.1), *Hydractinia symbiolongicarpus* (GCF\_029227915.1, GCA\_029227915.2), *Hydraecia micacea* (GCA\_914767645.1), *Hydriomena furcata* (GCA\_912999785.1), *Hydrochoerus hydrochaeris* (GCA\_015741225.1), *Hydrophis curtus* (GCA\_019472885.1), *Hydrophis cyanocinctus* (GCA\_019473425.1), *Hydrophis major* (GCA\_033807585.1), *Hydropotes inermis* (GCA\_020226075.1), *Hydrotaea cyrtoneurina* (GCA\_958296145.1), *Hydrotaea diabolus* (GCA\_963513945.1), *Hyla sarda* (GCA\_029499605.1, GCF\_029499605.1), *Hylaea fasciaria* (GCA\_905147375.1), *Hylaeus volcanicus* (GCA\_026283585.1, GCF\_026283585.1), *Hyles euphorbiae* (GCA\_023078785.2), *Hyles vespertilio* (GCA\_009982885.2), *Hylobates pileatus* (GCA\_021498465.1), *Hylyphantes graminicola* (GCA\_023701765.1), *Hymenochirus boettgeri* (GCA\_019447015.1), *Hymenolepis microstoma* (GCA\_000469805.3), *Hymenopus coronatus* (GCA\_030762935.1), *Hypanus sabinus* (GCF\_030144855.1, GCA\_030144855.1, GCA\_030144785.1), *Hypena proboscidalis* (GCA\_905147285.1), *Hypercompe scribonia* (GCA\_949316085.1), *Hyperoodon ampullatus*



(GCA\_949752795.1), *Hyperoplus immaculatus* (GCA\_949357725.1), *Hyperoplus lanceolatus* (GCA\_026929865.2), *Hypomecis punctinalis* (GCA\_949316475.1), *Hypomesus transpacificus* (GCF\_021917145.1, GCA\_021917145.1, GCA\_021870715.1), *Hypophthalmichthys molitrix* (GCA\_022817975.1), *Hypophthalmichthys nobilis* (GCA\_019925145.1), *Hyppa rectilinea* (GCA\_951799385.1), *Hypsopygia costalis* (GCA\_937001555.2), *Icerya purchasi* (GCA\_952773005.1), *Ichneumon xanthorius* (GCA\_917499995.1), *Ichthyophis bannanicus* (GCA\_033557465.1), *Ictalurus furcatus* (GCA\_023701845.2, GCF\_023375685.1, GCA\_023375685.2), *Ictalurus punctatus* (GCF\_001660625.3, GCA\_001660625.3, GCA\_004006655.3), *Ictidomys tridecemlineatus* (GCA\_016881025.1, GCF\_016881025.1), *Idaea aversata* (GCA\_907269075.1), *Idaea biselata* (GCA\_958496205.1), *Idaea dimidiata* (GCA\_949358125.1), *Idaea straminata* (GCA\_951213275.1), *Ilyophis sp. 1 JC-2022* (GCA\_022702515.1), *Ilyophis sp. 2 JC-2022* (GCA\_022702525.1), *Ilyophis sp. JC\_2022\_1* (GCA\_022702415.1), *Incurvaria masculella* (GCA\_946894095.1), *Indicator indicator* (GCA\_027791375.2, GCF\_027791375.1), *Iphiclides podalirius* (GCA\_933534255.1), *Ips nitidus* (GCA\_018691245.2), *Ischnura elegans* (GCF\_921293095.1, GCA\_921293095.2, GCA\_921293095.1), *Isoperla grammatica* (GCA\_945910005.1), *Istiophorus platypterus* (GCA\_016859345.1), *Ixodes scapularis* (GCA\_031841145.1), *Jaculus jaculus* (GCA\_020740685.1, GCF\_020740685.1), *Junco hyemalis* (GCA\_003829775.2), *Karalla daura* (GCA\_029224185.1), *Kogia breviceps* (GCF\_026419965.1, GCA\_026419965.1, GCA\_026419985.1), *Kryptolebias hermaphroditus* (GCA\_007896545.1), *Kryptolebias marmoratus* (GCF\_001649575.2, GCA\_001649575.2), *Labeo rohita* (GCA\_022985175.1, GCF\_022985175.1), *Labia minor* (GCA\_963082975.1), *Labroides dimidiatus* (GCA\_030710495.1), *Labrus mixtus* (GCA\_963584025.1, GCF\_963584025.1), *Lacanobia oleracea* (GCA\_950371165.1), *Lacanobia wlatinum* (GCA\_947578705.1), *Lacerta agilis* (GCF\_009819535.1, GCA\_009819535.1), *Lagenorhynchus albirostris* (GCF\_949774975.1, GCA\_949774975.1), *Lagopus muta* (GCA\_023343835.1, GCF\_023343835.1), *Lagorchestes hirsutus* (GCA\_028533205.1), *Lagria hirta* (GCA\_947359425.1), *Lama glama* (GCA\_028534125.1), *Lamellibrachia columna* (GCA\_963662155.1), *Lampris incognitus* (GCA\_029633865.1, GCF\_029633865.1, GCA\_022059245.1), *Lampropteryx suffumata* (GCA\_948098915.1), *Lamprotornis superbus* (GCA\_015883425.2), *Laodelphax striatellus* (GCA\_014465815.1, GCA\_017141395.1), *Laothoe populi* (GCA\_905220505.1), *Lapara coniferarum* (GCA\_949316025.1), *Larimichthys crocea* (GCA\_000972845.2, GCA\_004352675.2, GCA\_003845795.1, GCA\_003711585.2, GCF\_000972845.2, GCA\_900246015.1), *Lasioglossum calceatum* (GCA\_028455575.1), *Lasioglossum figueresi* (GCA\_028455805.1), *Lasioglossum lativentre* (GCA\_916610255.1), *Lasioglossum leucozonium* (GCA\_028454225.1), *Lasioglossum malachurum* (GCA\_028455605.1), *Lasioglossum marginatum* (GCA\_028455795.1), *Lasioglossum morio* (GCA\_916610235.2), *Lasioglossum oenotherae* (GCA\_028455765.1), *Lasioglossum pauxillum* (GCA\_028455745.1, GCA\_933228785.1), *Lasioglossum vierecki* (GCA\_028455595.1), *Lasioglossum zephyrum* (GCA\_028455615.1), *Lasiommata megera* (GCA\_928268935.1), *Lasius fuliginosus* (GCA\_949152495.1), *Laspeyria flexula* (GCA\_905147015.1), *Lateolabrax maculatus* (GCA\_004023545.1, GCA\_004028665.1, GCA\_031216445.1), *Lates calcarifer* (GCA\_001640805.2, GCF\_001640805.2), *Latheticus oryzae* (GCA\_030157265.1), *Lathronympha strigana* (GCA\_949128165.1), *Latrodectus elegans* (GCA\_030067965.1), *Leguminivora glycinivorella* (GCA\_023078275.1, GCF\_023078275.1), *Leistus spinibarbis* (GCA\_933228885.1), *Lemur catta* (GCA\_020740605.1, GCF\_020740605.2), *Leontopithecus rosalia* (GCA\_028533165.1), *Leopardus geoffroyi* (GCF\_018350155.1, GCA\_018350155.1), *Lepeophtheirus nordmannii* (GCA\_963680855.1), *Lepeophtheirus salmonis* (GCA\_016086655.3, GCA\_905330665.1, GCF\_016086655.3, GCA\_902860245.1, GCA\_016086655.4), *Lepidonotus clava* (GCA\_936440205.1), *Lepisosteus oculatus* (GCA\_000242695.1, GCF\_000242695.1), *Leptidea sinapis* (GCA\_905404315.1), *Leptinotarsa decemlineata* (GCA\_024712935.1), *Leptobrachium ailaonicum* (GCA\_032062155.1, GCA\_018994145.1), *Leptobrachium leishanense* (GCA\_009667805.1), *Leptodactylus fuscus* (GCA\_031893055.1, GCA\_031893025.1), *Leptodirus hochenwartii* (GCA\_947310635.1), *Leptogaster cylindrica* (GCA\_963082835.1), *Leptophobia aripa* (GCA\_951799465.1), *Leptopilina boulandi* (GCA\_032872485.1), *Leptopilina drosophilae* (GCA\_032873175.1), *Leptopilina heterotoma* (GCA\_032872495.1), *Leptopilina sp. ZJUHH005* (GCA\_032872475.1), *Leptopilina syphax* (GCA\_032872505.1), *Leptopterna dolabrata* (GCA\_954871275.1), *Leptura quadrifasciata* (GCA\_963675555.1), *Lepturacanthus savala* (GCA\_030544235.1), *Lepus europaeus* (GCA\_033115175.1), *Lepus timidus* (GCA\_009760805.1), *Lethenteron reissneri* (GCF\_015708825.1, GCA\_015708825.1), *Leucania comma* (GCA\_958295575.1), *Leucaspius delineatus* (GCA\_023718395.1), *Leuciscus idus* (GCA\_021554675.1), *Leucoma*

*salicis* (GCA\_948253155.1), *Leucophora obtusa* (GCA\_949987735.1), *Leucoraja erinacea* (GCA\_028641065.1, GCF\_028641065.1), *Leucozona laternaria* (GCA\_932273885.2), *Leuctra nigra* (GCA\_934045905.1), *Liasis olivaceus* (GCA\_030867105.1, GCA\_030867085.1), *Lichenostomus cassidix* (GCA\_008360975.2), *Ligdia adustata* (GCA\_947049295.1), *Limanda limanda* (GCF\_963576545.1, GCA\_963576545.1), *Limenitis camilla* (GCA\_905147385.1), *Limnephilus auricula* (GCA\_951813805.1), *Limnephilus lunatus* (GCA\_917563855.2), *Limnephilus marmoratus* (GCA\_917880885.1), *Limnephilus rhombicus* (GCA\_929108145.2), *Limnoperna fortunei* (GCA\_944474755.1), *Lineus longissimus* (GCA\_910592395.2), *Linnaemya tessellans* (GCA\_951800035.1), *Linnaemya vulpina* (GCA\_963675445.1), *Lipophrys pholis* (GCA\_963383615.1), *Liposcelis brunnea* (GCA\_023512825.1), *Lithobates sylvaticus* (GCA\_028564925.1), *Lithophane leautieri* (GCA\_949152455.1), *Lithophane ornitopus* (GCA\_948473465.1), *Lithophane semibrunnea* (GCA\_947363905.1), *Lithophane socia* (GCA\_947522985.1), *Lithosia quadra* (GCA\_963576445.1), *Litosomoides sigmodontis* (GCA\_963070105.1), *Lobophora halterata* (GCA\_932526365.1), *Lochmaea capreae* (GCA\_949126875.1), *Lochmaea crataegi* (GCA\_947563755.1), *Locusta migratoria* (GCA\_026315105.1), *Lomographa bimaculata* (GCA\_948107665.1), *Lonchura striata domestica* (GCF\_005870125.1, GCA\_005870125.1), *Lophura swinhoii* (GCA\_030408155.1), *Loxodonta africana* (GCA\_033060095.1, GCA\_032717415.1, GCA\_030014295.1, GCA\_032717405.1), *Lucilia cuprina* (GCA\_022045245.1, GCF\_022045245.1), *Lucinisca nassula* (GCA\_963580285.1), *Luffia ferchaultella* (GCA\_949709985.1), *Luidia sarsii* (GCA\_949987565.1), *Lumbricus rubellus* (GCA\_945859605.1), *Lumbricus terrestris* (GCA\_949752735.1), *Luperina nickerlii* (GCA\_963855955.1), *Luperina testacea* (GCA\_927399505.1), *Lutjanus erythropterus* (GCA\_020091685.1), *Lutra lutra* (GCF\_902655055.1, GCA\_902655055.2), *Lutzomyia longipalpis* (GCA\_024334085.1, GCF\_024334085.1), *Lycaena phlaeas* (GCA\_905333005.2), *Lycaon pictus* (GCA\_001883655.1, GCA\_001887905.1), *Lycia hirtaria* (GCA\_947563715.1), *Lycodopsis pacificus* (GCA\_028022725.1), *Lycophotia porphyrea* (GCA\_950005105.1), *Lymantria dispar* (GCA\_963576585.1, GCA\_032191425.1), *Lymantria monacha* (GCA\_905163515.2), *Lynx canadensis* (GCA\_007474595.2, GCF\_007474595.2), *Lypha dubia* (GCA\_947311025.1), *Lysandra bellargus* (GCA\_905333045.1), *Lysandra coridon* (GCA\_905220515.1), *Lytechinus pictus* (GCF\_015342785.2, GCA\_015342785.2), *Lytechinus variegatus* (GCA\_018143015.1, GCF\_018143015.1), *Macaca cyclopis* (GCA\_026956025.1), *Macaca fascicularis* (GCF\_000364345.1, GCA\_011100615.1, GCF\_012559485.2, GCA\_000222185.1, GCA\_000230815.1, GCA\_012559485.3, GCA\_000364345.1), *Macaca mulatta* (GCA\_008058575.1, GCA\_003339765.3, GCF\_003339765.1, GCA\_001270425.1, GCA\_000230795.1, GCA\_000002255.2, GCF\_000772875.2, GCF\_000002255.3, GCA\_000772875.3), *Macaca thibetana thibetana* (GCA\_024542745.1, GCF\_024542745.1), *Macaria notata* (GCA\_927399415.1), *Machimus atricapillus* (GCA\_933228815.1), *Machimus rusticus* (GCA\_951509405.1), *Macrobrachium nipponense* (GCA\_015104395.1, GCA\_015110555.1), *Macrochelys suwanniensis* (GCA\_033349115.1, GCA\_033296515.1), *Macrophya alboannulata* (GCA\_949628255.1), *Macropis europaea* (GCA\_916610135.2), *Macropodus opercularis* (GCA\_030770545.1), *Macropus fuliginosus* (GCA\_028583105.1), *Macropus giganteus* (GCA\_028627215.1), *Mactra quadrangularis* (GCA\_025267735.1), *Magallana gigas* (GCA\_963853765.1), *Malachius bipustulatus* (GCA\_910589415.1), *Malaclemys terrapin pileata* (GCA\_027887205.1, GCF\_027887155.1, GCA\_027887155.1), *Malacosoma neustria* (GCA\_963693495.1), *Malaya genurostris* (GCF\_030247185.1, GCA\_030247185.2), *Malthinus flaveolus* (GCA\_950108345.1), *Malurus cyaneus samueli* (GCA\_009741485.1), *Mamestra brassicae* (GCA\_905163435.1), *Manduca sexta* (GCF\_014839805.1, GCA\_014839805.1), *Maniola hyperantus* (GCF\_902806685.1, GCA\_902806685.1, GCA\_963576595.1, GCA\_902806685.2), *Maniola jurtina* (GCA\_905333055.1, GCF\_905333055.1), *Manis pentadactyla* (GCA\_030020395.1, GCF\_030020395.1), *Mantis religiosa* (GCA\_030765055.1), *Marasmarcha lunaedactyla* (GCA\_923062675.1), *Marasmia exigua* (GCA\_019059595.1), *Martes flavigula* (GCA\_029410595.1), *Martes martes* (GCA\_963455335.1), *Marthasterias glacialis* (GCA\_911728455.2), *Mastacembelus armatus* (GCA\_900324485.3, GCA\_019455535.1, GCA\_019455525.1, GCA\_900324485.2, GCF\_900324485.2), *Mastomys coucha* (GCF\_008632895.1, GCA\_008632895.1), *Matsumurasca onukii* (GCA\_018831715.1), *Mauremys mutica* (GCF\_020497125.1, GCA\_020497125.1, GCA\_032357905.1), *Mauremys reevesii* (GCA\_016161935.1, GCF\_016161935.1), *Maylandia zebra* (GCF\_000238955.4, GCA\_000238955.5), *Meandrina meandrites* (GCA\_963693305.1), *Mechanitis mazaeus* (GCA\_959347395.1), *Mechanitis messenoides* (GCA\_959347415.1), *Meconema thalassinum* (GCA\_946902985.2), *Mecyna flavalis* (GCA\_949319885.1), *Meda fulgida* (GCA\_030578275.1), *Megachile leachella* (GCA\_963576845.1), *Megachile ligniseca* (GCA\_945859555.1),

*Megachile willughbiella* (GCA\_945859595.2), *Megaleporinus macrocephalus* (GCA\_021613375.1), *Megalobrama amblycephala* (GCF\_018812025.1, GCA\_018812025.1), *Megalops atlanticus* (GCA\_019176425.1), *Megalops cyprinoides* (GCF\_013368585.1, GCA\_013368585.1), *Megalurothrips usitatus* (GCA\_026979955.1), *Megamerina dolium* (GCA\_963854835.1), *Meganola albula* (GCA\_936450015.1), *Meghimatium bilineatum* (GCA\_034231615.1), *Meiosimyza decipiens* (GCA\_963680825.1), *Meiosimyza platycephala* (GCA\_963662105.1), *Melanargia galathea* (GCA\_920104075.1), *Melanchra persicariae* (GCA\_947386135.1), *Melangyna quadrimaculata* (GCA\_949320155.1), *Melanophora roralis* (GCA\_963583895.1), *Melanostigma gelatinosum* (GCA\_949748355.1), *Melanostoma mellinum* (GCA\_914767635.1), *Melanostoma scalare* (GCA\_949752695.1), *Melanotaenia boesemani* (GCA\_017639745.1, GCF\_017639745.1), *Melanotaenia duboulayi* (GCA\_026261665.1), *Melanotus villosus* (GCA\_963082815.1), *Meleagris gallopavo* (GCA\_000146605.4, GCA\_943295565.1, GCF\_000146605.3, GCA\_943294205.1), *Meles meles* (GCF\_922984935.1, GCA\_922984935.2, GCA\_922984935.1), *Meliera crassipennis* (GCA\_963668005.1), *Melinaea marsaeus rileyi* (GCA\_918358865.1), *Melinaea menophilus n. ssp. AW-2005* (GCA\_918358695.1), *Meliscaeva auricollis* (GCA\_948107695.1), *Melitaea cinxia* (GCF\_905220565.1, GCA\_905220565.1), *Mellicta athalia* (GCA\_905220545.2), *Melolontha melolontha* (GCA\_935421215.2), *Melopsittacus undulatus* (GCA\_012275295.1, GCF\_012275295.1), *Melospiza georgiana* (GCA\_028018845.1, GCF\_028018845.1), *Membranipora membranacea* (GCA\_914767715.1), *Mercenaria mercenaria* (GCA\_021730395.1, GCA\_014805675.2, GCF\_021730395.1, GCF\_014805675.1), *Meriones unguiculatus* (GCA\_030254825.1, GCF\_030254825.1), *Merodon equestris* (GCA\_958301585.1), *Merops nubicus* (GCA\_009819595.1), *Merzomyia westermanni* (GCA\_949987695.1), *Mesoligia furuncula* (GCA\_916614155.1), *Mesoplodon densirostris* (GCF\_025265405.1, GCA\_025265405.1), *Mesopsocus fuscifrons* (GCA\_950004255.1), *Meta bourneti* (GCA\_933210815.1), *Metalampra italica* (GCA\_949699065.1), *Metallyticus violacea* (GCA\_030762175.1), *Metaphire vulgaris* (GCA\_018105865.1), *Metellina segmentata* (GCA\_947359465.1), *Metopia argyrocephala* (GCA\_963576795.1), *Metopolophium dirhodum* (GCF\_019925205.1, GCA\_019925205.1), *Metridium senile* (GCA\_949775045.1), *Microcaecilia unicolor* (GCA\_901765095.2, GCF\_901765095.1, GCA\_901765095.1), *Microcebus murinus* (GCF\_000165445.2, GCA\_000165445.3), *Microchirus variegatus* (GCA\_963457635.1), *Microchrysa polita* (GCA\_949715475.1), *Microctonus brassicae* (GCA\_940306215.1, GCA\_918226075.1), *Microplitis demolitor* (GCF\_026212275.2, GCA\_026212275.2), *Microplitis manilae* (GCA\_029641195.1, GCA\_030273425.1), *Microplitis mediator* (GCF\_029852145.1, GCA\_029852145.1), *Micropterix aruncella* (GCA\_944548615.1), *Micropterus dolomieu* (GCA\_021292245.1, GCF\_021292245.1), *Micropterus salmoides* (GCA\_022435785.1, GCA\_019677235.1), *Microtus ochrogaster* (GCF\_000317375.1, GCA\_000317375.1), *Mitochrista miniata* (GCA\_933228765.1), *Mimachlamys varia* (GCA\_947623455.1), *Mimas tiliae* (GCA\_905332985.1), *Mimumesa dahlbomi* (GCA\_917499265.3), *Miopithecus talapoin* (GCA\_028551445.1), *Mirounga angustirostris* (GCF\_021288785.2, GCA\_021288785.3), *Misgurnus anguillicaudatus* (GCF\_027580225.1, GCA\_027580225.1), *Mobula birostris* (GCA\_030028105.1, GCA\_030035685.1), *Molanna angustata* (GCA\_963576475.1), *Molothrus ater* (GCF\_012460135.2, GCA\_012460135.3), *Monochamus saltuarius* (GCA\_025584915.1), *Monodelphis domestica* (GCF\_000002295.2, GCF\_027887165.1, GCA\_000002295.1, GCA\_027887165.1), *Monomorium pharaonis* (GCA\_013373865.2, GCF\_013373865.1), *Monopis laevigella* (GCA\_947458855.1, GCA\_947359445.1), *Monoplex corrugatus* (GCA\_030674185.1), *Montipora capitata* (GCA\_949126865.1), *Morinoia aosen* (GCA\_030386875.1), *Morus bassanus* (GCA\_031468805.1, GCA\_031468815.1), *Motacilla alba alba* (GCF\_015832195.1, GCA\_015832195.1), *Mugilogobius chulae* (GCA\_016735935.1), *Mungos mungo* (GCA\_028533875.1), *Muntiacus crinifrons* (GCA\_020276665.1, GCA\_020226055.1), *Muntiacus muntjak* (GCA\_008782695.1), *Muntiacus reevesi* (GCA\_020226045.1, GCA\_008787405.2), *Muraenolepis orangiensis* (GCA\_027704905.1), *Muricea muricata* (GCA\_963855995.1), *Mus caroli* (GCA\_900094665.2, GCF\_900094665.2), *Mus musculus* (GCA\_921999865.2, GCA\_031761455.1, GCA\_001624675.1, GCA\_001624535.1, GCA\_947599735.1, GCA\_001632575.1, GCA\_000001635.9, GCA\_921998905.2, GCA\_001624215.1, GCA\_921998355.2, GCA\_921998635.2, GCA\_921998315.2, GCA\_001632525.1, GCA\_921998325.2, GCA\_001624505.1, GCA\_001632555.1, GCA\_921998555.2, GCA\_001624475.1, GCA\_921997125.2, GCA\_030265425.1, GCA\_949316315.1, GCA\_922000895.2, GCA\_001632615.1, GCA\_001624185.1, GCF\_000001635.27, GCA\_001624295.1, GCA\_921997145.2, GCA\_947593165.1, GCA\_031763685.1, GCA\_000002165.1, GCA\_001624745.1), *Mus musculus castaneus* (GCA\_921999005.2, GCA\_001624445.1), *Mus musculus domesticus* (GCA\_921998345.2, GCA\_001624835.1),



*Mus musculus molossinus* (GCA\_921999095.2), *Mus musculus musculus* (GCA\_921998335.2, GCA\_001624775.1), *Mus pahari* (GCF\_900095145.1, GCA\_900095145.2), *Mus spretus* (GCA\_921997135.2, GCA\_001624865.1), *Muscardinus avellanarius* (GCA\_963383645.1), *Muscidifurax raptorellus* (GCA\_020010945.3), *Musotima nitidalis* (GCA\_949126915.1), *Mustela erminea* (GCF\_009829155.1, GCA\_009829155.1), *Mustela lutreola* (GCF\_030435805.1, GCA\_030435805.1), *Mustela nigripes* (GCF\_022355385.1, GCA\_022355385.1), *Mya arenaria* (GCF\_026914265.1, GCA\_026914265.1), *Myathropa florea* (GCA\_930367185.1), *Myiopsitta monachus* (GCA\_017639245.1), *Myopa tessellatipennis* (GCA\_943737955.2), *Myopa testacea* (GCA\_949629155.1), *Myotis daubentonii* (GCF\_963259705.1, GCA\_963259705.1), *Myripristis murdjan* (GCF\_902150065.1, GCA\_902150065.1), *Myrrha octodecimguttata* (GCA\_958510865.1), *Mystacides longicornis* (GCA\_963576905.1), *Mythimna albipuncta* (GCA\_929112965.1), *Mythimna ferrago* (GCA\_910589285.2), *Mythimna impura* (GCA\_905147345.3), *Mythimna lalium* (GCA\_949319445.1), *Mythimna loreyi* (GCA\_029852875.1), *Mythimna pallens* (GCA\_961205895.1), *Mythimna separata* (GCA\_026898235.1, GCA\_030763345.1, GCA\_020882275.1, GCA\_029852925.1), *Mythimna vitellina* (GCA\_949316375.1), *Mytilisepta virgata* (GCA\_028015205.1), *Mytilus coruscus* (GCA\_017311375.1), *Mytilus edulis* (GCA\_963676685.1, GCA\_019925275.1, GCA\_025215535.1, GCA\_025276775.1), *Mytilus galloprovincialis* (GCA\_025277285.1), *Myxocyprinus asiaticus* (GCA\_019703515.2, GCF\_019703515.2), *Myzus persicae* (GCA\_029232275.1), *Naja naja* (GCA\_009733165.1), *Napeogenes inachia* (GCA\_959347355.1), *Napeogenes sylphis* (GCA\_959347405.1), *Nasalis larvatus* (GCA\_000772465.1), *Nasonia vitripennis* (GCF\_000002325.3, GCA\_009193385.2, GCF\_009193385.2, GCA\_000002325.2), *Nasua narica* (GCA\_028533885.1), *Nebria brevicollis* (GCA\_944738965.1), *Nebria salina* (GCA\_944039245.1), *Necator americanus* (GCA\_031761385.1), *Nematolebias whitei* (GCF\_014905685.2, GCA\_014905685.2), *Nematopogon swammerdamellus* (GCA\_946902875.1), *Nematostella vectensis* (GCA\_932526225.1, GCF\_932526225.1, GCA\_033964005.1), *Nemotelus nigrinus* (GCA\_947369275.1), *Nemoura dubitans* (GCA\_921293005.1), *Nemurella pictetii* (GCA\_921293315.2), *Neoaliturus tenellus* (GCA\_030545055.2), *Neoaricus graeffei* (GCF\_027579695.1, GCA\_027579695.1), *Neoscasia interrupta* (GCA\_947623515.1), *Neoceratitis asiatica* (GCA\_030068015.2), *Neoceratodus forsteri* (GCA\_016271365.2), *Neocrepidodera transversa* (GCA\_963243735.1), *Neodiprion fabricii* (GCF\_021155785.1, GCA\_021155785.1), *Neodiprion lecontei* (GCA\_021901455.1, GCA\_001263575.2, GCF\_021901455.1), *Neodiprion pinetum* (GCF\_021155775.1, GCA\_021155775.1), *Neodiprion virginianus* (GCA\_021901495.1, GCF\_021901495.1), *Neofelis nebulosa* (GCF\_028018385.1, GCA\_028018385.1, GCA\_030324275.1), *Neogale vison* (GCF\_020171115.1, GCA\_020171115.1), *Neotamus cyanurus* (GCA\_947538895.1), *Neolamprologus multifasciatus* (GCA\_963576455.1), *Neomonachus schauinslandi* (GCA\_002201575.2, GCF\_002201575.2), *Neoneuromus ignobilis* (GCA\_024320075.1, GCA\_029203775.1), *Neosalanx taihuensis* (GCA\_030340665.1), *Neostethus bicornis* (GCA\_902685375.1), *Neotoxoptera formosana* (GCA\_022818045.1), *Nephrotettix cincticeps* (GCA\_023375725.1), *Nephrocera scutellatus* (GCA\_947095585.1), *Nephrotoma appendiculata* (GCA\_947310385.1), *Nephrotoma flavescens* (GCA\_932526605.1), *Nephrotoma guestfalica* (GCA\_963691935.1), *Neria commutata* (GCA\_963457695.1), *Nerophis lumbriciformis* (GCA\_033978685.1), *Nerophis ophidion* (GCA\_033978795.1), *Nesoenas mayeri* (GCA\_963082525.1), *Netelia dilatata* (GCA\_946811545.1), *Netelia virgata* (GCA\_951802705.1), *Nettapus auritus* (GCA\_011076525.1), *Nezara viridula* (GCA\_928085145.1), *Nibea albiflora* (GCA\_902410095.1, GCA\_014281875.1), *Nibea coibor* (GCA\_023373845.1), *Nicrophorus investigator* (GCA\_963457615.1), *Nilaparvata lugens* (GCA\_014356525.1, GCA\_015708395.1, GCF\_014356525.2), *Nipponacmea schrenckii* (GCA\_030562195.1), *Nippostrongylus brasiliensis* (GCA\_030553155.1), *Noctua comes* (GCA\_963082995.1), *Noctua fimbriata* (GCA\_905163415.1), *Noctua janthe* (GCA\_910589295.1), *Noctua janthina* (GCA\_963576755.1), *Noctua pronuba* (GCA\_905220335.1), *Nomada fabriciana* (GCA\_907165295.1), *Nomada ferruginata* (GCA\_963583965.1), *Nomada flava* (GCA\_951802745.1), *Nomada fucata* (GCA\_948146005.1), *Nomada hirtipes* (GCA\_951802735.1), *Nomada lathburiana* (GCA\_963667235.1), *Nomada panzeri* (GCA\_951802685.1), *Nomada ruficornis* (GCA\_951802695.1), *Nomascus leucogenys* (GCF\_000146795.2, GCA\_000146795.3, GCA\_006542625.1, GCF\_006542625.1), *Nomophila noctuella* (GCA\_958496325.1), *Notamacropus eugenii* (GCA\_028372415.1), *Nothobranchius furzeri* (GCA\_014300015.1, GCF\_001465895.1, GCA\_001465895.2), *Notocelia uddmanniana* (GCA\_905163555.1), *Notodonta dromedarius* (GCA\_905147325.1), *Notodonta ziczac* (GCA\_918843915.1), *Notolabrus celidotus* (GCA\_009762535.1, GCF\_009762535.1), *Notothenia rossii* (GCA\_949606895.1), *Nowickia ferox*



(GCA\_936439885.1), *Nudaria mundana* (GCA\_963556515.1), *Numida meleagris* (GCA\_002078875.2, GCF\_002078875.1), *Nyctea revayana* (GCA\_947037095.2), *Nyctibius grandis* (GCA\_013368605.1), *Nycticebus bengalensis* (GCA\_023898255.1), *Nycticebus coucang* (GCF\_027406575.1, GCA\_027406575.1), *Nylanderia fulva* (GCA\_024268025.1), *Nymphalis c-album* (GCA\_963151315.1), *Nymphalis io* (GCA\_905147045.1, GCF\_905147045.1), *Nymphalis polychloros* (GCA\_905220585.2), *Nymphalis urticae* (GCA\_905147175.2), *Nymphula nitidulata* (GCA\_947347705.1), *Nyssus spinosus* (GCA\_910591585.2), *Obolodiplosis robiniae* (GCA\_028476595.1), *Ochloides sylvanus* (GCA\_905404295.2), *Ochotona princeps* (GCA\_030435715.1, GCA\_014633375.1, GCF\_014633375.1, GCF\_030435755.1, GCA\_030435755.1), *Ochropacha duplaris* (GCA\_951361185.1), *Ochropleura leucogaster* (GCA\_958449745.1), *Ochropleura plecta* (GCA\_905475445.1), *Octodonta nipae* (GCA\_034190945.1), *Octopus bimaculoides* (GCA\_001194135.2, GCF\_001194135.2), *Octopus sinensis* (GCF\_006345805.1, GCA\_006345805.1), *Octopus vulgaris* (GCA\_951406725.2), *Ocyrops olens* (GCA\_910593695.2), *Odocoileus hemionus* (GCA\_020976825.1), *Odocoileus virginianus* (GCA\_023699985.2), *Odontamblyopus lacepedii* (GCA\_032888595.1), *Odontamblyopus rebecca* (GCA\_030686955.1), *Odontesthes bonariensis* (GCA\_027942865.1), *Odontoceros albicorne* (GCA\_949825065.1), *Oedothorax gibbosus* (GCA\_019343175.1), *Oegoconia quadripuncta* (GCA\_949316235.1), *Oenanthe melanoleuca* (GCF\_029582105.1, GCA\_029582105.1), *Oikopleura dioica* (GCA\_907165135.1), *Okapia johnstoni* (GCA\_024291935.2), *Oligia fasciuncula* (GCA\_963082905.1), *Oligia latruncula* (GCA\_948474745.1), *Oligia strigilis* (GCA\_951800025.1), *Omphaloscelis lunosa* (GCA\_916610215.1), *Oncorhynchus gorbuscha* (GCA\_021184085.1, GCA\_017355495.1, GCF\_021184085.1), *Oncorhynchus keta* (GCA\_012931545.1, GCF\_012931545.1, GCF\_023373465.1, GCA\_023373465.1), *Oncorhynchus kisutch* (GCF\_002021735.2, GCA\_002021735.2), *Oncorhynchus mykiss* (GCA\_013265735.3, GCA\_025558465.1, GCF\_013265735.2, GCA\_029834435.1, GCF\_002163495.1, GCA\_002163495.1, GCA\_900005705.1), *Oncorhynchus nerka* (GCF\_006149115.2, GCA\_006149115.2), *Oncorhynchus tshawytscha* (GCA\_002872995.1, GCF\_002872995.1, GCA\_002831465.1, GCA\_018296145.1, GCF\_018296145.1), *Onychomys arenicola* (GCA\_949786405.1, GCA\_949786425.1), *Onychomys leucogaster* (GCA\_949786395.1, GCA\_949786385.1), *Onychomys torridus* (GCA\_949787125.1, GCF\_903995425.1, GCA\_903995425.1), *Onychostoma macrolepis* (GCA\_012432095.1, GCF\_012432095.1), *Ooceraea biroii* (GCA\_003672135.1, GCF\_003672135.1), *Operophtera brumata* (GCA\_932527175.1), *Ophion costatus* (GCA\_951751655.1), *Ophion ellenae* (GCA\_963210295.1), *Ophion slaviceki* (GCA\_944452715.1), *Ophonus ardosiacus* (GCA\_943142095.1), *Opisthocomus hoazin* (GCA\_030867145.1, GCA\_030867165.1), *Opisthograptis luteolata* (GCA\_931315375.2), *Orchestes rusci* (GCA\_958502075.1), *Orcinus orca* (GCA\_937001465.1, GCF\_937001465.1), *Oreochromis aureus* (GCA\_016625535.1, GCF\_013358895.1, GCA\_013358895.1), *Oreochromis niloticus* (GCF\_001858045.2, GCA\_000188235.2, GCF\_000188235.2, GCA\_001858045.3, GCA\_013350305.1), *Orgyia antiqua* (GCA\_916999025.1), *Ormia ochracea* (GCA\_963402855.1), *Ornithorhynchus anatinus* (GCA\_004115215.4, GCA\_000002275.2, GCF\_004115215.2), *Orthonevra nobilis* (GCA\_963555765.1), *Orthosia gothica* (GCA\_949775005.1), *Orthosia gracilis* (GCA\_947562075.1), *Orthosia incerta* (GCA\_948252205.1), *Oryctolagus cuniculus* (GCF\_000003625.3, GCF\_009806435.1, GCA\_009806435.2, GCA\_000003625.1), *Oryctolagus cuniculus cuniculus* (GCA\_013371645.1), *Oryzias curvinotus* (GCA\_023969325.1), *Oryzias javanicus* (GCA\_003999625.1), *Oryzias latipes* (GCA\_004348095.1, GCA\_002234715.1, GCA\_004347445.1, GCF\_000313675.1, GCA\_002234695.1, GCA\_002234675.1, GCF\_002234675.1, GCA\_000313675.1, GCA\_004348135.1), *Oryzias melastigma* (GCF\_002922805.2, GCA\_002922805.2), *Oscarella lobularis* (GCA\_947507565.1), *Oscheius dolichura* (GCA\_932521035.1), *Oscheius onirici* (GCA\_932521025.1), *Oscheius sp. DF5120* (GCA\_932521415.1), *Oscheius sp. JU1382* (GCA\_932521405.1), *Oscheius tipulae* (GCA\_013425905.1), *Osmerus eperlanus* (GCA\_963692335.1), *Osmia bicornis bicornis* (GCA\_907164935.1, GCF\_907164935.1), *Ostrea edulis* (GCA\_024362745.1, GCA\_032173915.1, GCF\_947568905.1, GCA\_947568905.1), *Ostrinia nubilalis* (GCA\_963855985.1), *Othius punctulatus* (GCA\_951805005.1), *Otis tarda* (GCA\_026413225.1), *Otocolobus manul* (GCA\_028564725.2), *Ovis ammon polii* (GCA\_028583565.1), *Ovis ammon polii x Ovis aries* (GCA\_023701675.1), *Ovis aries* (GCF\_002742125.1, GCF\_016772045.2, GCA\_022432835.1, GCA\_022416775.1, GCA\_002742125.1, GCA\_011170295.1, GCA\_016772045.2, GCA\_017524585.1, GCA\_018804185.1, GCA\_019145175.1, GCF\_000298735.2, GCA\_024222265.1, GCA\_030512445.1, GCA\_024222175.1, GCA\_022244695.1, GCA\_022432845.1, GCA\_022416725.1, GCA\_022432825.1, GCA\_022244705.1, GCA\_022416685.1, GCA\_022416695.1,

GCA\_022416745.1, GCA\_022416755.1, GCA\_022416785.1, GCA\_022416915.1, GCA\_033439445.1, GCA\_022538005.1, GCA\_000298735.2, GCA\_000005525.1), *Ovis canadensis canadensis* (GCA\_001039535.1), *Ovis orientalis* (GCA\_014523465.1), *Oxytorus armatus* (GCA\_958009045.1), *Oxyura jamaicensis* (GCA\_011077185.1, GCF\_011077185.1), *Pachypeltis micranthus* (GCA\_020466155.1), *Pachypsilla venusta* (GCA\_012654025.1), *Pachyrhynchus sulphureomaculatus* (GCA\_019049505.1), *Pachyuromys duprasi* (GCA\_028658305.1), *Paedocypris micromegethes* (GCA\_028031935.1), *Paguma larvata* (GCA\_030068075.1), *Palloptera scutellata* (GCA\_958295655.1), *Pammene aurita* (GCA\_947086415.1), *Pammene fasciana* (GCA\_911728535.1), *Pan paniscus* (GCA\_028858845.1, GCA\_028858825.1, GCA\_000258655.2, GCF\_029289425.1, GCF\_000258655.2, GCA\_029289425.1, GCA\_013052645.3), *Pan troglodytes* (GCF\_000001515.7, GCA\_028858805.1, GCF\_028858775.1, GCA\_000090855.1, GCA\_000001515.5, GCA\_002880755.3, GCA\_028858775.1, GCF\_002880755.1), *Pan troglodytes verus* (GCA\_000002175.2, GCF\_000002175.1), *Pandemis cinnamomeana* (GCA\_932294345.1), *Pandemis corylana* (GCA\_949127965.1), *Pandemis heparana* (GCA\_963854515.1), *Pangasianodon gigas* (GCA\_022758105.1), *Pangasianodon hypophthalmus* (GCA\_027358585.1, GCF\_009078355.1, GCA\_009078355.1, GCA\_016801045.1, GCF\_027358585.1), *Pangasius djambal* (GCA\_022985145.1), *Panopea generosa* (GCA\_029582155.1), *Panorpa germanica* (GCA\_963678705.1), *Pantala flavescens* (GCA\_020796165.1), *Panthera leo* (GCA\_008795835.1, GCF\_018350215.1, GCA\_018350215.1), *Panthera onca* (GCF\_028533385.1, GCA\_028533385.1), *Panthera tigris* (GCA\_024034525.1, GCA\_018350195.2, GCF\_018350195.1), *Panthera tigris tigris* (GCA\_021131075.1, GCA\_021130815.1), *Panthera uncia* (GCF\_023721935.1, GCA\_023721935.1, GCA\_028646445.1), *Panzeria rudis* (GCA\_956483635.1), *Pao palembangensis* (GCA\_015343265.1), *Papilio elwesi* (GCA\_029849275.1, GCA\_029641285.1), *Papilio machaon* (GCF\_912999745.1, GCA\_912999745.1), *Papio anubis* (GCA\_008728515.2, GCF\_000264685.3, GCA\_008728515.1, GCF\_008728515.1, GCA\_000264685.2), *Papio papio* (GCA\_028645565.1), *Paracanthobrama guichenoti* (GCA\_018749465.1), *Paraescarpia echinospica* (GCA\_020002185.1), *Paralichthys olivaceus* (GCA\_001904815.2), *Paralithodes platypus* (GCA\_013283005.1, GCA\_032716605.1), *Parambassis ranga* (GCA\_900634625.1, GCA\_900634625.2, GCF\_900634625.1), *Paraponyx stratiotata* (GCA\_910589355.1), *Pararge aegeria* (GCA\_905163445.1, GCF\_905163445.1), *Parascotia fuliginaria* (GCA\_963082885.1), *Parasteatoda lunata* (GCA\_949128135.1), *Pardosa pseudoannulata* (GCA\_032207245.1), *Parus major* (GCA\_001522545.3, GCF\_001522545.3), *Pasiphila rectangulata* (GCA\_963082625.1), *Passer domesticus* (GCA\_001700915.1), *Passerculus sandwichensis* (GCA\_031885435.1), *Patella depressa* (GCA\_948474765.1), *Patella pellucida* (GCA\_917208275.1), *Patella ulyssiponensis* (GCA\_963678685.1), *Patella vulgata* (GCA\_932274485.2, GCF\_932274485.2), *Patiria pectinifera* (GCA\_029964075.1), *Pecten maximus* (GCA\_902652985.1, GCF\_902652985.1), *Pectinophora gossypiella* (GCF\_024362695.1, GCA\_024362695.1), *Pedicia rivosa* (GCA\_963082725.1), *Pelecanus crispus* (GCA\_030463565.1), *Pelobates cultripes* (GCA\_933207985.1), *Pelochelys cantorii* (GCA\_032595735.1), *Pelosia muscerda* (GCA\_963691645.1), *Pemphredon lugubris* (GCA\_933228935.1), *Penaeus chinensis* (GCA\_019202785.2, GCA\_016920825.1), *Penaeus monodon* (GCF\_015228065.2, GCA\_015228065.1), *Perca flavescens* (GCF\_004354835.1, GCA\_004354835.1), *Perca fluviatilis* (GCF\_010015445.1, GCA\_010015445.1), *Perca schrenkii* (GCA\_025617405.1), *Percottus glenii* (GCA\_024416835.2), *Peribatodes rhomboidaria* (GCA\_911728515.1), *Periophthalmus magnuspinnatus* (GCA\_026225835.1, GCF\_009829125.3, GCA\_009829125.3), *Periparus ater* (GCA\_032357785.1), *Periphyllus acericola* (GCA\_949715065.1), *Perizoma affinitatum* (GCA\_961405105.1), *Perizoma flavofasciatum* (GCA\_958496245.1), *Perognathus longimembris pacificus* (GCA\_023159225.1, GCF\_023159225.1), *Peromyscus eremicus* (GCA\_949786415.1, GCA\_949786435.1, GCF\_949786415.1), *Peromyscus leucopus* (GCF\_004664715.2, GCA\_004664715.2), *Peromyscus maniculatus bairdii* (GCF\_003704035.1, GCA\_003704035.3), *Peromyscus polionotus subgriseus* (GCA\_003704135.2), *Petromyzon marinus* (GCA\_010993605.1, GCF\_010993605.1), *Petrophora chlorosata* (GCA\_951640565.1), *Petrosia ficiformis* (GCA\_947044365.1), *Phacochoerus africanus* (GCF\_016906955.1, GCA\_016906955.1), *Phaedon cochleariae* (GCA\_918026855.4), *Phaenicophaeus curvirostris* (GCA\_032191515.2), *Phalacrocorax aristotelis* (GCA\_949628215.1), *Phalanger gymnotis* (GCA\_028646595.1), *Phalera bucephala* (GCA\_905147815.2), *Phasia obesa* (GCA\_949628195.1), *Phenacoccus solenopsis* (GCA\_009761765.1), *Phengaris arion* (GCA\_963565745.1), *Pheosia gnoma* (GCA\_905404115.1), *Pheosia tremula* (GCA\_905333125.1), *Pherbina coryleti* (GCA\_943735915.1), *Philereme vetulata* (GCA\_918857605.2),

*Philonthus cognatus* (GCA\_932526585.2), *Philonthus spinipes* (GCA\_963082785.1), *Phlebotomus papatasi* (GCF\_024763615.1, GCA\_024763615.2), *Phlogophora meticulosa* (GCA\_905147745.2), *Phocoena sinus* (GCA\_008692025.1, GCF\_008692025.1), *Phodopus sungorus* (GCA\_023856395.1), *Phoenicopterus ruber ruber* (GCA\_009819775.1), *Pholidichthys leucotaenia* (GCA\_020510985.1), *Pholis gunnellus* (GCA\_910591455.2), *Phorcus lineatus* (GCA\_921293015.1), *Phosphuga atrata* (GCA\_944588485.1), *Phoxinus phoxinus* (GCA\_949152265.1), *Phragmatobia fuliginosa* (GCA\_932526445.1), *Phrynosoma platyrhinos* (GCA\_020142125.1), *Phycodurus eques* (GCA\_024500275.1), *Phyllopteryx taeniolatus* (GCA\_024500385.1, GCA\_019802525.1, GCA\_019802545.1), *Phyllostomus discolor* (GCA\_004126475.3, GCF\_004126475.2), *Phyllotreta cruciferae* (GCA\_917563865.1), *Phyllotreta striolata* (GCA\_918026865.1), *Phymatocera aterrima* (GCA\_963170745.1), *Phymorhynchus buccinoides* (GCA\_017654935.2), *Physeter catodon* (GCF\_002837175.3, GCA\_002837175.5), *Phyto melanocephala* (GCA\_941918925.1), *Pictodentalium vernedei* (GCA\_031216915.1), *Picus viridis* (GCA\_033816785.1), *Pieris brassicae* (GCF\_905147105.1, GCA\_905147105.1), *Pieris mannii* (GCA\_028984075.1, GCA\_029001895.1), *Pieris napi* (GCA\_905231885.1, GCF\_905475465.1, GCA\_905475465.2), *Pieris rapae* (GCF\_905147795.1, GCA\_905147795.1), *Piezodorus guildinii* (GCA\_023052935.1), *Piliocolobus tephrosceles* (GCF\_002776525.5, GCA\_002776525.5), *Pilophorus perplexus* (GCA\_955831195.1), *Pinctada fucata* (GCA\_028142955.1, GCA\_028253585.1), *Pinctada imbricata* (GCA\_002216045.1), *Pipistrellus pipistrellus* (GCA\_903992545.1), *Pipistrellus pygmaeus* (GCA\_949987585.1), *Pisaster ochraceus* (GCA\_010994315.2), *Piscicola geometra* (GCA\_943735955.1), *Pithecia pithecia* (GCA\_028551515.1), *Plagioderia versicolora* (GCA\_963584125.1), *Plagadis dolabraria* (GCA\_963854805.1), *Planococcus citri* (GCA\_950023065.1), *Platichthys flesus* (GCA\_949316205.1), *Platichthys stellatus* (GCA\_016801935.1), *Platycheirus albimanus* (GCA\_916050605.2), *Platycnemis pennipes* (GCA\_933228895.1), *Platygaster robiniae* (GCA\_028476575.1), *Platypus cylindrus* (GCA\_949748235.1), *Plazaster borealis* (GCA\_021014325.1), *Plebejus argus* (GCA\_905404155.3), *Plecotus auritus* (GCA\_963455305.1), *Plectropomus leopardus* (GCF\_008729295.1, GCA\_026936395.1, GCA\_008729295.2, GCA\_011397275.1), *Plemyria rubiginata* (GCA\_963576535.1), *Pleurodeles waltl* (GCA\_026652325.1), *Pleuronectes platessa* (GCF\_947347685.1, GCA\_947347685.1), *Plodia interpunctella* (GCA\_027563975.1, GCF\_027563975.1), *Plotosus lineatus* (GCA\_024760905.1), *Plusia festucae* (GCA\_950381575.1), *Plutella xylostella* (GCA\_932276165.1, GCF\_932276165.1, GCA\_019096205.1), *Pluvialis apricaria* (GCA\_017639485.1), *Pocota personata* (GCA\_963082735.1), *Podabrus alpinus* (GCA\_932274525.1), *Podarcis cretensis* (GCA\_951804945.1), *Podarcis lilfordi* (GCA\_947686815.1), *Podarcis muralis* (GCA\_004329235.1, GCF\_004329235.1), *Podarcis muralis nigriventris* (GCA\_014706415.1), *Podarcis raffonei* (GCA\_027172205.1, GCF\_027172205.1), *Podargus strigoides* (GCA\_028020825.1), *Poecile atricapillus* (GCA\_030490855.1, GCA\_030490865.1, GCF\_030490865.1), *Poecilia formosa* (GCA\_013036135.2, GCA\_013036085.2), *Poecilia picta* (GCA\_033032245.1), *Poecilia reticulata* (GCF\_000633615.1, GCA\_000633615.2, GCA\_904066995.1), *Poecilobothrus nobilitatus* (GCA\_947095535.1), *Pogoniulus pusillus* (GCA\_015220805.1), *Pogonophryne albipinna* (GCA\_028583405.1), *Pogonus chalceus* (GCA\_002278615.1), *Polia nebulosa* (GCA\_951329385.1), *Poliates domitor* (GCA\_947397865.1), *Pollachius pollachius* (GCA\_949987615.1), *Pollenia amentaria* (GCA\_943735925.1), *Pollenia angustigena* (GCA\_930367215.1), *Pollenia labialis* (GCA\_949318255.1), *Polydactylus sextarius* (GCA\_016801845.1), *Polydrusus cervinus* (GCA\_935413205.1), *Polydrusus tereticollis* (GCA\_963576705.1), *Polymixis flavicincta* (GCA\_949987655.1), *Polymixis lichenea* (GCA\_949091785.1), *Polyodon spathula* (GCF\_017654505.1, GCA\_017654505.1), *Polyommatus icarus* (GCA\_937595015.1), *Polyommatus iphigenia* (GCA\_963422495.1), *Polypedilum vanderplanki* (GCA\_018290095.1), *Polyploca ridens* (GCA\_951394255.1), *Polypterus senegalus* (GCA\_016835505.1, GCF\_016835505.1), *Pomacea canaliculata* (GCA\_004794335.1, GCA\_003073045.1, GCF\_003073045.1), *Pongo abelii* (GCF\_028885655.1, GCA\_028885685.1, GCA\_000001545.3, GCF\_002880775.1, GCA\_028885655.1, GCA\_002880775.3), *Pongo pygmaeus* (GCA\_028885625.1, GCF\_028885625.1, GCA\_900086635.1, GCA\_028885525.1), *Porites lutea* (GCA\_958299795.1), *Porphyrio hochstetteri* (GCA\_020800305.1), *Portevinia maculata* (GCA\_949715645.1), *Portunus trituberculatus* (GCA\_032715055.1, GCA\_017591435.1), *Potorous gilbertii* (GCA\_028658325.1), *Potos flavus* (GCA\_028534135.1), *Prinia subflava* (GCA\_021018805.1), *Prionailurus bengalensis* (GCA\_016509475.2, GCF\_016509475.1), *Prionailurus viverrinus* (GCF\_022837055.1, GCA\_022837055.1, GCA\_028551425.1), *Pristionchus pacificus* (GCA\_000180635.4), *Pristis pectinata* (GCF\_009764475.1, GCA\_009764475.2),



*Procambarus clarkii* (GCA\_020424385.2), *Procyon lotor* (GCA\_028646535.1), *Prosilocerus akamusi* (GCA\_018397935.1), *Propylea japonica* (GCA\_013421045.1), *Proterorhinus semilunaris* (GCA\_021464625.1), *Protocalliphora azurea* (GCA\_932274085.1), *Protodeltote pygarga* (GCA\_936450705.2), *Protonemura montana* (GCA\_947568835.1), *Protonibea diacanthus* (GCA\_028641955.1), *Protophormia terraenovae* (GCA\_951394005.1), *Protopterus annectens* (GCA\_019279795.1, GCF\_019279795.1), *Protosalinx chinensis* (GCA\_030340685.1), *Protula sp. h* YS-2021 (GCA\_949752745.1), *Pseudochaenichthys georgianus* (GCA\_902827115.2, GCF\_902827115.1, GCA\_902827115.1), *Pseudocheirus occidentalis* (GCA\_028646575.1), *Pseudochirops corinnae* (GCA\_028646515.1), *Pseudochirops cupreus* (GCA\_028627135.1), *Pseudoips prasinana* (GCA\_951640165.1), *Pseudoliparis swirei* (GCF\_029220125.1, GCA\_029220125.1), *Pseudolycoriella hygida* (GCA\_029228625.1), *Pseudophryne corroborae* (GCA\_028390025.1), *Pseudorashora parva* (GCA\_024679245.1), *Psittacula echo* (GCA\_963264785.1), *Psylliodes chrysocephala* (GCA\_927349885.1), *Pterocles gutturalis* (GCA\_009769525.1), *Pteromalus puparum* (GCA\_012977825.3), *Pteropus rufus* (GCA\_028533765.1), *Pterostichus madidus* (GCA\_911728475.2), *Pterostichus niger* (GCA\_947425015.1), *Ptilodon capucinus* (GCA\_914767695.1), *Ptychobarbus kaznakovi* (GCA\_027422405.1), *Ptychoptera albiglana* (GCA\_961205885.1), *Puma concolor* (GCA\_028749985.3, GCA\_028749965.3), *Pungitius pungitius* (GCA\_949316345.1), *Puntigrus tetrazona* (GCA\_018831695.1, GCF\_018831695.1), *Pusa sibirica* (GCA\_028975605.1), *Pycnopodia helianthoides* (GCA\_032158295.1), *Pyemotes zhonghuajia* (GCA\_025170145.1), *Pygocentrus nattereri* (GCF\_015220715.1, GCA\_015220715.1), *Pyralis farinalis* (GCA\_947507595.1), *Pyrausta aurata* (GCA\_963584085.1), *Pyrausta nigrata* (GCA\_949316185.1), *Pyrgus malvae* (GCA\_911387765.1), *Pyrochroa serraticornis* (GCA\_905333025.2), *Pyrrhosoma nymphula* (GCA\_963573305.1), *Pyxicephalus adspersus* (GCA\_004786255.1, GCA\_032062135.1), *Rafetus swinhoei* (GCA\_019425775.1), *Raja brachyura* (GCA\_963514005.1), *Rana kukunoris* (GCA\_029574335.1), *Rana muscosa* (GCA\_029206835.1), *Rana temporaria* (GCF\_905171775.1, GCA\_905171775.1), *Rangifer tarandus caribou* (GCA\_019903745.2), *Rangifer tarandus platyrhynchus* (GCA\_949782905.1, GCA\_951394145.1), *Ranitomeya imitator* (GCA\_032444005.1), *Rapana venosa* (GCA\_028751875.1), *Rattus norvegicus* (GCA\_000001895.4, GCA\_015227675.2, GCF\_015227675.2, GCF\_000002265.2, GCF\_000001895.5, GCA\_023515785.1, GCA\_000002265.1, GCA\_023515805.1), *Rattus rattus* (GCA\_011064425.1, GCA\_011800105.1, GCF\_011064425.1), *Reinhardtius hippoglossoides* (GCA\_006182925.3), *Rhagium mordax* (GCA\_963680705.1), *Rhagonycha fulva* (GCA\_905340355.1), *Rhagonycha lutea* (GCA\_958510855.1), *Rhaphigaster nebulosa* (GCA\_948150685.1), *Rhea pennata* (GCA\_028389875.1), *Rhegmatorhina hoffmannsi* (GCA\_013398505.2), *Rhinatrema bivittatum* (GCA\_901001135.1, GCF\_901001135.1, GCA\_901001135.2), *Rhincodon typus* (GCF\_021869965.1, GCA\_021869965.1), *Rhineura floridana* (GCA\_030035675.1, GCF\_030035675.1), *Rhingia campestris* (GCA\_932526625.1), *Rhingia rostrata* (GCA\_949824845.1), *Rhinoceros unicornis* (GCA\_028646465.1), *Rhinolophus ferrumequinum* (GCA\_004115265.3, GCF\_004115265.2), *Rhinopithecus roxellana* (GCF\_007565055.1, GCA\_007565055.1), *Rhipicephalus microplus* (GCA\_013339725.1, GCF\_013339725.1), *Rhipicephalus sanguineus* (GCA\_013339695.2, GCF\_013339695.2), *Rhodophaea formosa* (GCA\_963082605.1), *Rhogogaster chlorosoma* (GCA\_944452935.1), *Rhopalosiphum maidis* (GCF\_003676215.2, GCA\_003676215.3), *Rhopalosiphum padi* (GCF\_020882245.1, GCA\_020882245.1), *Rhorus exstirpatorius* (GCA\_963564615.1), *Rhynchoetos jubatus* (GCA\_027574665.1), *Ricordea florida* (GCA\_949710005.1), *Riptortus pedestris* (GCA\_019009955.1), *Rissa tridactyla* (GCA\_028500815.1, GCF\_028500815.1, GCA\_028501385.1), *Romanogobio albipinnatus* (GCA\_023566225.1), *Rousettus madagascariensis* (GCA\_028533395.1), *Ruditapes philippinarum* (GCA\_009026015.1), *Rutilus rutilus* (GCA\_951802725.1), *Rutpela maculata* (GCA\_936432065.2), *Sabethes cyaneus* (GCF\_943734655.1, GCA\_943734655.2), *Sacculina carcini* (GCA\_916048095.2), *Saguinus midas* (GCA\_021498475.1), *Saguinus oedipus* (GCA\_031835075.1), *Salarias fasciatus* (GCA\_902148845.1, GCF\_902148845.1), *Salminus brasiliensis* (GCA\_030463535.1), *Salmo salar* (GCA\_000233375.4, GCA\_923944775.2, GCA\_905237065.2, GCA\_931346935.2, GCF\_000233375.1, GCF\_905237065.1, GCA\_021399835.1), *Salmo trutta* (GCA\_901001165.1, GCA\_901001165.2, GCF\_901001165.1), *Salpingus planirostris* (GCA\_949788335.1), *Salvelinus fontinalis* (GCF\_029448725.1, GCA\_029448725.1), *Salvelinus namaycush* (GCF\_016432855.1, GCA\_016432855.1), *Salvelinus sp.* IW2-2015 (GCA\_002910315.2, GCF\_002910315.2), *Samia ricini* (GCA\_014132275.2), *Sander lucioperca* (GCA\_008315115.2, GCF\_008315115.2, GCA\_023718385.1), *Sander vitreus* (GCA\_031162955.1), *Sarcophaga caerulescens*



(GCA\_927399465.1), *Sarcophaga carnaria* (GCA\_958299815.1), *Sarcophaga peregrina* (GCA\_014635995.1), *Sarcophaga rosellei* (GCA\_930367235.1), *Sarcophaga subvicina* (GCA\_936449025.2), *Sarcophaga variegata* (GCA\_932273835.1), *Sarcophilus harrisii* (GCF\_902635505.1, GCA\_902635505.1), *Sarcoptes scabiei* (GCA\_020844145.1), *Sardina pilchardus* (GCA\_963854185.1), *Saturnia japonica* (GCA\_033032175.1), *Saturnia pavonia* (GCA\_947532125.1), *Scaeva pyrastris* (GCA\_905146935.1), *Scardinius erythrophthalmus* (GCA\_024453875.1), *Scatophagus argus* (GCF\_020382885.2, GCA\_020382885.1), *Sceloporus tristichus* (GCA\_016801065.1, GCA\_016801415.1), *Sceloporus undulatus* (GCA\_019175285.1, GCF\_019175285.1), *Schistocerca americana* (GCA\_021461395.2, GCF\_021461395.2), *Schistocerca cancellata* (GCA\_023864275.2, GCF\_023864275.1), *Schistocerca gregaria* (GCA\_023897955.2, GCF\_023897955.1), *Schistocerca nitens* (GCA\_023898315.2, GCF\_023898315.1), *Schistocerca piceifrons* (GCF\_021461385.2, GCA\_021461385.2), *Schistocerca serialis cubense* (GCA\_023864345.3, GCF\_023864345.2), *Schistosoma bovis* (GCA\_944470425.2, GCA\_944470445.2), *Schistosoma curassoni* (GCA\_944474815.3), *Schistosoma guineensis* (GCA\_944470375.2), *Schistosoma haematobium* (GCA\_000699445.3, GCA\_944470465.2, GCF\_000699445.3, GCA\_944470455.2), *Schistosoma intercalatum* (GCA\_944470365.2, GCA\_944470385.2), *Schistosoma japonicum* (GCA\_021461655.1, GCA\_025215515.1), *Schistosoma mansoni* (GCA\_000237925.2, GCA\_000237925.5), *Schistosoma margrebowiei* (GCA\_944470205.2), *Schistosoma mattheei* (GCA\_944470405.2), *Schistosoma rodhaini* (GCA\_944470415.2, GCA\_944470435.2), *Schistosoma spindale* (GCA\_946903255.1), *Schistosoma turkestanicum* (GCA\_944470395.2), *Schizaphis graminum* (GCA\_020882235.1), *Schizopygopsis malacanthus* (GCA\_027474115.1), *Schizopygopsis pylzovi* (GCA\_027422435.1), *Schizothorax lantsangensis* (GCA\_027422415.1), *Schizotus pectinicornis* (GCA\_951805265.1), *Schlechtendalia chinensis* (GCA\_019022885.1), *Schmidtea mediterranea* (GCA\_022537955.1), *Schrankia costaestrigalis* (GCA\_905475405.1), *Sciaenops ocellatus* (GCA\_014183145.1, GCA\_033000465.1), *Sciurus carolinensis* (GCF\_902686445.1, GCA\_902686445.2, GCA\_028643795.1), *Sciurus vulgaris* (GCA\_902686455.2), *Scleropages formosus* (GCA\_023634345.2, GCF\_900964775.1, GCA\_023634425.1, GCA\_900964775.1), *Scolanthus callimorphus* (GCA\_033964015.1), *Scomber japonicus* (GCA\_027409825.1, GCF\_027409825.1), *Scomber scombrus* (GCA\_963691925.1), *Scophthalmus maximus* (GCF\_013347765.1, GCA\_003186165.1, GCF\_022379125.1, GCA\_013347765.1, GCA\_963854745.1, GCA\_022379125.1), *Scortum barcoo* (GCA\_023238725.1), *Scotopteryx bipunctaria* (GCA\_949320045.1), *Scrobipalpa costella* (GCA\_949820665.1), *Scyliorhinus canicula* (GCA\_902713615.2, GCF\_902713615.1, GCA\_902713615.1), *Sebastes schlegelii* (GCA\_014673565.1), *Sebastes umbrosus* (GCF\_015220745.1, GCA\_015220745.1), *Seladonia tumulorum* (GCA\_913789895.3), *Selenia dentaria* (GCA\_917880725.1), *Sepiolo atlantica* (GCA\_963556195.1), *Sepioteuthis lessoniana* (GCA\_963585895.1), *Serinus canaria* (GCA\_022539315.2, GCF\_022539315.1), *Seriola aureovittata* (GCA\_021018895.1, GCF\_021018895.1, GCA\_023856345.1), *Sesia apiformis* (GCA\_914767545.1), *Sesia bembeciformis* (GCA\_943735995.1), *Setipinna tenuifilis* (GCA\_030347295.1), *Setophaga coronata coronata* (GCA\_001746935.2), *Shargacucullia verbasci* (GCA\_947562105.1), *Shinisaurus crocodilurus* (GCA\_021292165.1), *Sialis fuliginosa* (GCA\_961205875.1), *Sialis lutaria* (GCA\_949319165.1), *Sicus ferrugineus* (GCA\_922984085.1), *Silurus aristotelis* (GCA\_946808225.1), *Silurus meridionalis* (GCF\_014805685.1, GCA\_023972475.1, GCA\_014805685.1), *Sinella curviseta* (GCA\_004115045.4), *Siniperca chuatsi* (GCA\_011952085.1, GCF\_020085105.1, GCA\_027580155.1, GCA\_020085105.1), *Siniperca kneri* (GCA\_011952075.1), *Siniperca scherzeri* (GCA\_011952095.1, GCA\_027580175.1), *Sinohyriopsis cumingii* (GCA\_028554795.1), *Sinonovacula constricta* (GCA\_009762815.1, GCA\_007844125.1), *Sinotaia purificata* (GCA\_028829895.1), *Siphamia tubifer* (GCA\_020466265.1), *Siphonurus alternatus* (GCA\_949825025.1), *Siphonodentalium dalli* (GCA\_032622095.1), *Sipunculus nudus* (GCA\_026874595.1), *Sisyr nigr*a (GCA\_958496155.1), *Sisyr terminalis* (GCA\_958496175.1), *Sitobion miscanthi* (GCA\_008086715.1), *Sitodiplosis mosellana* (GCA\_021018905.1), *Sitta carolinensis* (GCA\_032173965.1), *Slavum lentiscoides* (GCA\_032441835.1), *Smittia aterrima* (GCA\_033063855.1), *Smittia sp. YW1114-2* (GCA\_033064975.1), *Sogatella furcifera* (GCA\_017141385.1, GCA\_014356515.1), *Solea senegalensis* (GCA\_919967415.2, GCA\_019176455.1, GCF\_019176455.1), *Solea solea* (GCF\_958295425.1, GCA\_958295425.1), *Solen grandis* (GCA\_021229015.1), *Solenopsis invicta* (GCA\_018691235.1, GCF\_016802725.1, GCA\_009650705.1, GCA\_009299965.1, GCA\_009299975.1, GCA\_010367695.1, GCA\_016802725.1), *Somateria mollissima* (GCA\_030142145.1, GCA\_951416345.1, GCA\_951411735.1), *Sorex araneus* (GCA\_027595985.1, GCF\_027595985.1), *Sparus aurata* (GCA\_003309015.1, GCA\_900880675.1, GCF\_900880675.1, GCA\_900880675.2), *Spea bombifrons*

(GCA\_027358695.2, GCF\_027358695.1), *Sphaeramia orbicularis* (GCF\_902148855.1, GCA\_902148855.1), *Sphaerodactylus townsendi* (GCA\_021028975.2, GCF\_021028975.2), *Sphaerophoria taeniata* (GCA\_943590905.1), *Sphecodes monilicornis* (GCA\_913789915.3), *Sphenella marginata* (GCA\_951509765.1), *Spheniscus humboldti* (GCA\_027474245.1), *Sphinx pinastri* (GCA\_947568825.1), *Spilarctia lutea* (GCA\_916048165.1), *Spilosoma lubricipeda* (GCA\_905220595.1), *Spisula solida* (GCA\_947247005.1), *Spisula subtruncata* (GCA\_963678985.1), *Spodoptera exigua* (GCA\_011316535.1, GCA\_022315195.1, GCA\_902829305.4), *Spodoptera frugiperda* (GCA\_012979215.2, GCA\_011064685.2, GCF\_023101765.2, GCA\_023101765.3, GCF\_011064685.2), *Spodoptera littoralis* (GCA\_902850265.1), *Spodoptera litura* (GCF\_002706865.2, GCA\_002706865.3), *Spongilla lacustris* (GCA\_949361645.1), *Sprattus sprattus* (GCA\_963457725.1), *Squalius cephalus* (GCA\_022829025.1, GCA\_949319135.1), *Squalus acanthias* (GCA\_030390025.1), *Stegostoma tigrinum* (GCA\_022316705.1, GCA\_030684315.1, GCF\_022316705.1, GCF\_030684315.1), *Steinernema carpocapsae* (GCA\_000757645.3), *Steinernema hermaphroditum* (GCA\_030435675.1), *Stelis phaeoptera* (GCA\_943735885.1), *Stenchaetothrips biformis* (GCA\_030522485.1), *Stenella coeruleoalba* (GCA\_951394435.1), *Steno bredanensis* (GCA\_028646385.1), *Stenoptilia bipunctidactyla* (GCA\_944452665.1), *Stenurella melanura* (GCA\_963583905.1), *Sterna hirundo* (GCA\_009819605.1), *Steromphala cineraria* (GCA\_916613615.1), *Sthenelais limicola* (GCA\_942159475.1), *Stichopus chloronotus* (GCA\_021234535.1), *Stictonetta naevosa* (GCA\_011074415.1), *Stomorhina lunata* (GCA\_933228675.1), *Stomoxys calcitrans* (GCA\_963082655.1, GCF\_963082655.1), *Stramonita haemastoma* (GCA\_030674155.1), *Stratiomys singularior* (GCA\_954870665.1), *Streblospio benedicti* (GCA\_019095985.1), *Streptopelia turtur* (GCA\_901699155.2), *Strigamia acuminata* (GCA\_949358305.1), *Strigops habroptila* (GCA\_004027225.2, GCF\_004027225.2), *Strix aluco* (GCA\_031877795.1, GCA\_031877785.1), *Strongyloides ratti* (GCF\_001040885.1, GCA\_001040885.1), *Strongyloides stercoralis* (GCA\_029582065.1), *Sturmia bella* (GCA\_963662145.1), *Sturnus vulgaris* (GCA\_023376015.1), *Subacronicta megacephala* (GCA\_958496365.1), *Suillia variegata* (GCA\_949127995.1), *Suncus etruscus* (GCA\_024139225.1, GCF\_024139225.1), *Suricata suricatta* (GCA\_006229205.1, GCF\_006229205.1), *Sus scrofa* (GCA\_015776825.1, GCA\_900119615.2, GCA\_000003025.6, GCA\_023065335.1, GCA\_007644095.1, GCA\_030704935.1, GCF\_000003025.6, GCA\_023065355.1, GCA\_002844635.1, GCA\_031225015.1, GCA\_031306245.1, GCA\_024718415.1), *Sus scrofa domesticus* (GCA\_020567905.1, GCA\_017957985.1), *Sus scrofa scrofa* (GCA\_006511355.2), *Sylvia atricapilla* (GCA\_009819655.1), *Sylvia borin* (GCA\_014839755.1), *Sylvicola cinctus* (GCA\_963854165.1), *Sylvilagus bachmani* (GCA\_015711505.1), *Sylvilagus floridanus* (GCA\_949820135.1), *Sympetrum striolatum* (GCA\_947579665.1), *Symphalangus syndactylus* (GCA\_028878055.1, GCA\_028878085.1, GCA\_028642525.1, GCF\_028878055.1), *Symphodus melops* (GCA\_947650265.1), *Symsagittifera roscoffensis* (GCA\_963678635.1), *Synanceia verrucosa* (GCA\_029721515.1), *Synanthedon andrenaeformis* (GCA\_936446665.2), *Synanthedon formicaeformis* (GCA\_945859745.1), *Synanthedon myopaeformis* (GCA\_944738685.1), *Synanthedon tipuliformis* (GCA\_947623395.1), *Synanthedon vespiformis* (GCA\_918317495.1), *Synphobranchus kaupii* (GCA\_029718625.1), *Synchiropus splendidus* (GCA\_027744825.1, GCF\_027744825.2), *Syngnathoides biaculeatus* (GCA\_019802595.1), *Syngnathus acus* (GCA\_948146105.1, GCF\_901709675.1, GCA\_901709675.2), *Syngnathus scovelli* (GCA\_024217435.4), *Syngnathus typhle* (GCA\_033458585.1, GCF\_033458585.1), *Syrpitta pipiens* (GCA\_905187475.1), *Syrphus vitripennis* (GCA\_958431115.1), *Tachina fera* (GCA\_905220375.1), *Tachina grossa* (GCA\_949987645.1), *Tachina lurida* (GCA\_944452675.1), *Tachyglossus aculeatus* (GCA\_015852505.1, GCA\_015598185.1, GCF\_015852505.1), *Tachypleus gigas* (GCA\_014155125.1), *Tachypleus tridentatus* (GCA\_004210375.1), *Tachystola acroxantha* (GCA\_963506565.1), *Tachysurus fulvidraco* (GCA\_022655615.1, GCF\_022655615.1, GCA\_023638525.1), *Tachysurus vachellii* (GCF\_030014155.1, GCA\_030014155.1, GCA\_033026395.1), *Tadarida brasiliensis* (GCA\_030848825.1), *Taenia multiceps* (GCA\_001923025.3), *Taenia pisiformis* (GCA\_023968675.1), *Taenioides sp. WSHXM2023* (GCA\_029959045.1), *Taeniopygia guttata* (GCA\_008822105.2, GCA\_008822125.1, GCA\_008822115.3, GCA\_000151805.2, GCA\_003957565.4, GCF\_008822105.2, GCF\_003957565.2, GCA\_009859065.2, GCF\_000151805.1), *Takifugu bimaculatus* (GCA\_004026145.2), *Takifugu flavidus* (GCA\_003711565.2, GCF\_003711565.1), *Takifugu rubripes* (GCF\_901000725.2, GCA\_000180615.2, GCA\_901000725.3, GCF\_000180615.1, GCA\_901000725.2), *Tamandua tetradactyla* (GCA\_023851605.1), *Tamias sibiricus* (GCA\_025594165.1), *Tanypteryx hageni* (GCA\_028673005.1), *Taphrorychus bicolor* (GCA\_951812265.1), *Tapirus indicus* (GCA\_031878705.1, GCA\_031878655.1), *Tapirus*

*terrestris* (GCA\_028533255.1), *Tauraco erythrolophus* (GCA\_009769465.1), *Taurulus bubalis* (GCA\_910589615.1), *Tautogolabrus adspersus* (GCA\_020745685.1), *Tegillarca granosa* (GCA\_029721355.1), *Teleiodes luculella* (GCA\_948473455.1), *Telenomus remus* (GCA\_020615435.1), *Teleopsis dalmanni* (GCA\_002237135.2, GCA\_002237135.5, GCF\_002237135.1), *Telmatherina bonti* (GCA\_933228915.1), *Tenodera sinensis* (GCA\_030324105.1, GCA\_030765045.1), *Tenthredo distinguenda* (GCA\_947538915.1), *Tenthredo mesomela* (GCA\_943736025.1), *Tenthredo notha* (GCA\_914767705.2), *Terebella lapidaria* (GCA\_949152475.1), *Tetanocera ferruginea* (GCA\_958299015.1), *Tethea ocularis* (GCA\_963555595.1), *Tetheella fluctuosa* (GCA\_951216915.1), *Tetragnatha montana* (GCA\_963680715.1), *Tetramorium bicarinatum* (GCA\_928718305.1), *Tetrao urogallus* (GCA\_951394365.1), *Thalassophryne amazonica* (GCA\_902500255.1, GCF\_902500255.1), *Thaleichthys pacificus* (GCA\_023658055.1), *Thalpophila matura* (GCA\_948465475.1), *Thamnaconus septentrionalis* (GCA\_009823395.1), *Thamnophis elegans* (GCA\_009769535.1, GCF\_009769535.1), *Thecocarcelia acutangulata* (GCA\_914767995.1), *Thecophora atra* (GCA\_937620795.1), *Thelaira solivaga* (GCA\_947397855.1), *Theocolax elegans* (GCA\_026168455.1), *Thera britannica* (GCA\_939531255.2), *Thera obeliscata* (GCA\_947578465.1), *Thereira japonica* (GCA\_033459515.1), *Thereva nobilitata* (GCA\_963855945.1), *Thereva unica* (GCA\_949987705.1), *Therioaphis trifolii* (GCA\_027580255.1), *Theristicus caerulescens* (GCA\_020745775.1), *Theropithecus gelada* (GCA\_028533075.1, GCF\_003255815.1, GCA\_003255815.1), *Tholera decimalis* (GCA\_943138885.2), *Thomomys bottae* (GCA\_031878675.1, GCA\_031878665.1), *Thumatha senex* (GCA\_948477245.1), *Thunnus albacares* (GCA\_914725855.1, GCA\_914725855.2, GCF\_914725855.1), *Thunnus maccoyii* (GCA\_910596095.1, GCF\_910596095.1), *Thyatira batis* (GCA\_905147785.2), *Thylacinus cynocephalus* (GCA\_007646695.3), *Thymallus thymallus* (GCA\_023634145.1, GCA\_004348285.1), *Thymelicus acteon* (GCA\_951805285.1), *Thymelicus sylvestris* (GCA\_911387775.1), *Thysanoteuthis rhombus* (GCA\_963457665.1), *Tiaroga cobitis* (GCA\_030578255.1), *Tigriopus californicus* (GCA\_007210705.1, GCF\_007210705.1), *Tigriopus japonicus* (GCA\_010645155.1), *Tiliacea aurago* (GCA\_948098905.1), *Timandra comae* (GCA\_958496195.1), *Timema cristinae* (GCA\_002009905.3, GCA\_002928295.1), *Tinea pellionella* (GCA\_948150575.1), *Tinea semifulvella* (GCA\_910589645.1), *Tinea trinotella* (GCA\_905220615.1), *Tiphia femorata* (GCA\_944319695.1), *Tipula confusa* (GCA\_963556175.1), *Tipula helvola* (GCA\_963556165.1), *Tipula unca* (GCA\_951394425.1), *Tipula vernalis* (GCA\_958295665.1), *Tolmerus cingulatus* (GCA\_959613345.1), *Tomocerus qinae* (GCA\_020055645.1), *Topomyia yanbarensis* (GCF\_030247195.1, GCA\_030247195.1), *Tortricodes alternella* (GCA\_947859335.1), *Tortrix viridana* (GCA\_963241965.1), *Toxonevra muliebris* (GCA\_963691655.1), *Toxorhynchites rutilus septentrionalis* (GCA\_029784135.1, GCF\_029784135.1), *Toxotes chatareus* (GCA\_016801885.1), *Toxotes jaculatrix* (GCF\_017976425.1, GCA\_017976425.1), *Trachemys scripta elegans* (GCA\_013100865.1, GCF\_013100865.1), *Trachops cirrhosus* (GCA\_028533065.1), *Trachurus trachurus* (GCA\_905171665.2), *Tremarctos ornatus* (GCA\_028551375.1), *Tribolium castaneum* (GCA\_031307605.1, GCA\_000002335.3, GCF\_000002335.3), *Tribolium confusum* (GCA\_019155225.1), *Tribolium freemani* (GCA\_939628115.1, GCA\_022388455.1), *Trichinella murrelli* (GCA\_002221485.1), *Trichinella spiralis* (GCA\_008807755.1, GCA\_008807815.1, GCA\_000181795.3, GCA\_008807795.1, GCA\_008807775.1), *Trichobilharzia regenti* (GCA\_944472135.2), *Trichobilharzia szidati* (GCA\_944472155.2), *Tricholauxania praeusta* (GCA\_949775025.1), *Trichomycterus rosablanca* (GCA\_030014385.1), *Trichoplusia ni* (GCA\_003590095.1, GCF\_003590095.1), *Trichopria drosophilae* (GCA\_030407085.1), *Trichosurus vulpecula* (GCF\_011100635.1, GCA\_011100635.1), *Tridacna crocea* (GCA\_032873355.1, GCA\_943736015.1), *Tridacna derasa* (GCA\_963210305.1), *Tridacna gigas* (GCA\_945859785.2), *Trididemnum clinides* (GCA\_963675345.1), *Trilocha varians* (GCA\_030269945.2), *Triplophysa bombifrons* (GCA\_029783895.1), *Triplophysa dalaica* (GCA\_015846415.1, GCF\_015846415.1), *Triplophysa rosa* (GCF\_024868665.1, GCA\_024868665.2), *Triplophysa tibetana* (GCA\_008369825.1), *Triplophysa yarkandensis* (GCA\_033220385.1), *Trisateles emortualis* (GCA\_947095525.1), *Trogon surrucura* (GCA\_020746105.1), *Troides aeacus* (GCA\_033220335.2), *Tromatobia lineatoria* (GCA\_949699805.1), *Trypoxylus dichotomus* (GCA\_023509865.1), *Tuberolachnus salignus* (GCA\_956483605.1), *Tupaia chinensis* (GCA\_033439345.1), *Tursiops truncatus* (GCF\_011762595.1, GCA\_011762595.1), *Tuta absoluta* (GCA\_029230345.1), *Tyria jacobaeae* (GCA\_947561695.1), *Udea ferrugalis* (GCA\_950022985.1), *Udea olivalis* (GCA\_947369235.1), *Uloborus diversus* (GCA\_026930045.1, GCF\_026930045.1), *Uranotaenia lowii* (GCF\_029784155.1, GCA\_029784155.1), *Urbanus simplicius* (GCA\_949699795.1), *Urocyon cinereoargenteus* (GCA\_032313775.1), *Uromys caudimaculatus*



(GCA\_028551405.1), *Urophora cardui* (GCA\_960531455.1), *Ursus arctos* (GCA\_023065955.2, GCF\_023065955.2), *Valencia hispanica* (GCA\_963556495.2), *Vandiemena viatica* (GCA\_031002005.1), *Vanessa atalanta* (GCF\_905147765.1, GCA\_905147765.2), *Vanessa cardui* (GCA\_905220365.1, GCA\_905220365.2, GCF\_905220365.1), *Varecia variegata* (GCA\_028533085.1), *Venturia canescens* (GCF\_019457755.1, GCA\_019457755.1), *Venusia cambrica* (GCA\_951394065.1), *Verasper variegatus* (GCA\_013332515.1, GCA\_012393435.1, GCA\_026259375.1), *Vespa crabro* (GCF\_910589235.1, GCA\_910589235.2), *Vespa velutina* (GCF\_912470025.1, GCA\_912470025.1), *Vespula germanica* (GCA\_905340365.1, GCA\_014466195.1), *Vespula pensylvanica* (GCA\_014466175.1, GCF\_014466175.1), *Vespula vulgaris* (GCA\_905475345.1, GCA\_014466185.1, GCF\_905475345.1), *Vidua chalybeata* (GCF\_026979565.1, GCA\_026979565.1), *Vidua macroura* (GCF\_024509145.1, GCA\_024509145.1), *Villa cingulata* (GCA\_951394055.1), *Vimba vimba* (GCA\_022828995.1), *Vipera latastei* (GCA\_024294585.1), *Vipera ursinii* (GCA\_947247035.1), *Volucella bombylans* (GCA\_949129095.1), *Volucella inanis* (GCA\_907269105.1), *Volucella inflata* (GCA\_928272305.1), *Vombatus ursinus* (GCA\_028626985.1), *Vulpes corsac* (GCA\_030463245.1), *Vulpes ferrilata* (GCA\_024500485.1), *Vulpes lagopus* (GCA\_018345385.1, GCF\_018345385.1), *Watersipora subatra* (GCA\_963576615.1), *Watsonella binaria* (GCA\_929442735.1), *Wyeomyia smithii* (GCA\_029784165.1, GCF\_029784165.1), *Xanthia icteritia* (GCA\_949128155.2), *Xanthia togata* (GCA\_963853775.1), *Xanthogramma pedissequum* (GCA\_910595825.1), *Xanthorhoe designata* (GCA\_963582015.1), *Xanthorhoe spadicearia* (GCA\_947086425.1), *Xanthostigma xanthostigma* (GCA\_963575645.1), *Xenentodon cancila* (GCA\_014839995.1), *Xenopus borealis* (GCA\_024363595.1), *Xenopus laevis* (GCF\_017654675.1, GCA\_017654675.1, GCF\_001663975.1, GCA\_001663975.1), *Xenopus tropicalis* (GCA\_013368275.1, GCF\_000004195.4, GCA\_000004195.4), *Xerus rutilus* (GCA\_028644305.1), *Xestia ashworthii* (GCA\_950022955.1), *Xestia c-nigrum* (GCA\_916618015.1), *Xestia rhomboidea* (GCA\_963853795.1), *Xestia sexstrigata* (GCA\_941918905.2), *Xestia xanthographa* (GCA\_905147715.2), *Xestospongia muta* (GCA\_963693285.1), *Xiphias gladius* (GCF\_016859285.1, GCA\_016859285.1), *Xiphophorus couchianus* (GCF\_001444195.1, GCA\_001444195.3), *Xiphophorus hellerii* (GCA\_003331165.2, GCF\_003331165.1), *Xiphophorus maculatus* (GCA\_002775205.2, GCF\_002775205.1), *Xiphidria camelus* (GCA\_963678675.1), *Xylocampa areola* (GCA\_935421205.1), *Xylophagus ater* (GCA\_963422695.1), *Xylota segnis* (GCA\_963583995.1), *Xylota sylvarum* (GCA\_905220385.1), *Xyrauchen texanus* (GCA\_025860055.1, GCF\_025860055.1), *Xyrichtys novacula* (GCA\_962446985.1), *Yoshiicerus persimilis* (GCA\_020352615.2), *Yponomeuta cagnagella* (GCA\_947311075.1, GCA\_947310995.1), *Yponomeuta malinellus* (GCA\_947308005.1), *Yponomeuta plumbellus* (GCA\_947310845.1), *Yponomeuta sedellus* (GCA\_934045075.1), *Ypsolopha scabrella* (GCA\_910592155.1), *Ypsolopha sequella* (GCA\_934047225.1), *Zalophus californianus* (GCA\_009762305.2, GCF\_009762305.2), *Zele albiditarsus* (GCA\_958496275.1), *Zelleria hepariella* (GCA\_949319315.1), *Zerene cesonia* (GCA\_012273895.2, GCF\_012273895.1), *Zeugodacus cucurbitae* (GCF\_028554725.1, GCA\_028554725.2), *Zeugodacus tau* (GCA\_031772095.1), *Zeus faber* (GCA\_960531495.1), *Zeuzera pyrina* (GCA\_907165235.1), *Zingel zingel* (GCA\_023566125.1), *Zonotrichia leucophrys* (GCA\_028769735.1), *Zootoca vivipara* (GCF\_963506605.1, GCF\_011800845.1, GCA\_011800845.1, GCA\_963506605.1), *Zophobas atratus* (GCA\_022388445.1), *Zoroaster* sp. YZ-2022 (GCA\_029582265.1), *Zygaena filipendulae* (GCA\_907165275.2), *allo-octoploid hybrid of Carassius auratus x Cyprinus carpio* (GCA\_024542945.1), and *unclassified Cladorhizidae* (GCA\_028752895.1, GCA\_028752905.1). Genomes downloaded from other sources were: *Amphioctopus fangsiao* (<https://figshare.com/s/fa09f5dadcd966f020f3>), *Ephydatia muelleri* (<https://spaces.facscl.ualberta.ca/ephybase/>), *Rhopilema esculentum* ([http://gigadb.org/dataset/view/id/100720/Sample\\_page/2/File\\_page/2](http://gigadb.org/dataset/view/id/100720/Sample_page/2/File_page/2)).



# Supplementary Information Guide

## Topological mixing and irreversibility in animal chromosome evolution

Darrin T. Schultz<sup>1,\*</sup>, Fatih Sarigol<sup>1</sup>, Arno Blümel<sup>1</sup>, Dalila Destanović<sup>1,2</sup>, Oleg Simakov<sup>1,\*</sup>

### Affiliations

<sup>1</sup>Department for Neurosciences and Developmental Biology, University of Vienna, Vienna 1010, Austria.

<sup>2</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA.

\*Correspondence to:

Darrin T. Schultz      [darrin.schultz@univie.ac.at](mailto:darrin.schultz@univie.ac.at),  
Oleg Simakov          [oleg.simakov@univie.ac.at](mailto:oleg.simakov@univie.ac.at)

### ORCID (alphabetical):

Blümel,	Arno	<a href="https://orcid.org/0009-0000-8672-2319">https://orcid.org/0009-0000-8672-2319</a>
Destanović,	Dalila	<a href="https://orcid.org/0000-0002-3744-5789">https://orcid.org/0000-0002-3744-5789</a>
Sarigol,	Fatih	<a href="https://orcid.org/0000-0002-5842-6114">https://orcid.org/0000-0002-5842-6114</a>
Schultz,	Darrin T.	<a href="https://orcid.org/0000-0003-1190-1122">https://orcid.org/0000-0003-1190-1122</a>
Simakov,	Oleg	<a href="https://orcid.org/0000-0002-3585-4511">https://orcid.org/0000-0002-3585-4511</a>

### emails (alphabetical):

Arno Blümel	<a href="mailto:arno.bluemel@proton.me">arno.bluemel@proton.me</a>
Dalila Destanović	<a href="mailto:dalila.destanovic@yale.edu">dalila.destanovic@yale.edu</a>
Fatih Sarigol	<a href="mailto:fatih.sarigoel@gmail.com">fatih.sarigoel@gmail.com</a>
Darrin T. Schultz	<a href="mailto:darrin.schultz@univie.ac.at">darrin.schultz@univie.ac.at</a>
Oleg Simakov	<a href="mailto:oleg.simakov@univie.ac.at">oleg.simakov@univie.ac.at</a>

### This PDF file includes:

- Supplementary Information, containing:
  - 1. Glossary
  - 2. Periodicity in genomic changes
  - 3. State formulation of the (sub)chromosomal mixing process

# Supplementary Information 1

## Glossary

In an effort to clarify the intended meanings of our text, we include below a glossary of terms defined in this manuscript, or existing terms and the meaning of these terms based on previous publications.

<b>Algebraic changes</b>	<p>The <i>four</i> elementary operations that underlie metazoan chromosome dynamics. The types of changes are (A) <b>Syntenic equivalence</b>, (B) <b>Robertsonian translocation</b>, (C) <b>Centric insertion</b>, (D) and an extension of the above two processes, <b>Fusion-with-mixing</b>. These algebraic changes only are composed of <b>building</b> new chromosomes from ancestrally separate pieces of DNA, and does not account for chromosomes splitting apart, or a translocation. See <b>genome tectonics</b></p> <ul style="list-style-type: none"> <li>- Term introduced in<sup>1</sup>.</li> <li>- This term is also used in<sup>12</sup>.</li> <li>- This term is not used in<sup>8</sup>.</li> </ul>
<b>Ancestral linkage group</b>	<p>(ALGs) Refers to groups of genes that are conserved together over macro-evolutionary time, usually together on the same chromosome. <u>The term ALG does not require that the loci are one contiguous block. Therefore, ALGs can be an extension of the concept of <b>Macrosyteny</b>.</u> In Putnam et al 2008, the paper mostly refers to chordate linkage groups, then use the term ‘ancestral linkage group’ once.</p> <ul style="list-style-type: none"> <li>- Term introduced in<sup>5</sup></li> <li>- This term is used in<sup>1,3,8,27,84,85</sup></li> <li>- This term is not used in<sup>13</sup></li> </ul>
<b>Big bang</b>	<p>Chromosomal “big bangs” is the term used to date only in Albertin et al. 2022<sup>8</sup> to refer to saltational chromosomal changes, including what we call <b>non-algebraic changes</b> in this paper. See the definition of that term for a more complete description of the phenomena in this category.</p> <ul style="list-style-type: none"> <li>- This term is not used in<sup>1,3</sup></li> </ul>
<b>Centric insertion</b>	<p>One chromosome is inserted into another, denoted symbolically by <math>\searrow</math>. This is an irreversible change<sup>1</sup>.</p>
<b>Chromosomal element</b>	<p>In our usage, similar to <b>Ancestral linkage group</b>. Refers to groups of genes or loci that are conserved together as units over time. In the original usage by Sturtevant and Novitski<sup>6</sup>, it refers to <u>contiguous homologous regions</u> that still persist in closely related species, <i>Drosophila</i> in their case. Earlier, McClintock<sup>86</sup> used this term, but it only referred to parts of the genome that shape the nucleolus organization.</p> <ul style="list-style-type: none"> <li>- Term introduced in<sup>6</sup></li> <li>- Term used in<sup>1,27</sup></li> <li>- Term not used in<sup>3,5,8,84,85,87</sup></li> </ul>
<b>Dispers(al/ed)</b>	<p>The process of units being spread out across some area. In this case, See also <b>Dispersion</b>.</p>

- Earlier use in <sup>88</sup> to talk about homeobox gene clusters splitting apart from an ancestral clustered state.
- Dispersed introduced in <sup>27</sup>
- Dispersal introduced in <sup>1</sup>
- Term not in <sup>8,13</sup>

## Dispersion

The pattern of units being across some area. In this case, we would say “there is dispersion of the homeobox genes” across multiple chromosomes. Or, “From the dispersion of the ALG orthologs, we inferred the dispersal pattern along the evolutionary branches leading to these species. See also **dispers(al/ed)**.  
- Term introduced in this paper.

## Entanglement

When two or more loci come under the same regulatory or topological constraint, and the mixing of their regulatory interactions (such as enhancer-promoter links) results in a locus in which reverting the locus back to the ancestral pre-entangled state by translocation or other means, is selected against.

## Fragmentation and mixing

When a chromosome fragments into multiple pieces (whether by translocation or another mechanism), becomes fused with another chromosome, then mixes. In this manuscript, we will refer to this as one of the **non-algebraic changes**. See also **big bang**.  
- Term introduced in <sup>8</sup>.  
- Term not used in <sup>1,3</sup>.

## Fusion-with-mixing

The genes of two chromosomes are brought together (by either Robertsonian or end-to-end translocation or centric insertion) and, through a series of intrachromosomal rearrangements, become interspersed, denoted symbolically by the outer product symbol:  $\otimes$ . This is an irreversible change. Can be used to determine phylogenetics, in the case where shared, derived fusion-with mixing events are found.  
- Term introduced in <sup>1</sup>  
- Term also used in <sup>3</sup>, or in a sense in <sup>8,13</sup>

## Genome/chromosome tectonics

The basic mechanisms and the systematic construction of contemporary chromosomes from ancestral units. See **algebraic changes**.  
- Term introduced in <sup>1</sup>.  
- Term not used in <sup>8,13</sup>

## Linked

Linked refers to loci that are not only **mixed**, but also are conserved together over evolutionary time.

## Mixed

Mixed refers to the state of loci that were previously on separate chromosomes, but are now both on the same chromosome and the loci have become intermingled through inversions. For example, two separate chromosomes, [AAAAA] and [BBBBB] fused into a single chromosome [AAAAABBBBB]. Then, after a series of inversions the A and B loci are in rearranged, stochastic state [BAABABBAAB]. It is not necessary that A and B have the same number of loci being considered.

It is also not necessary that the A and B loci are alternating in a one-after-the-other pattern, like [ABABAB...]

## Modality

See *mode*.

## Mode

See also *modality*. We define a “mode of genome evolution”, or “mode of chromosome evolution” as the way in which chromosomes change over time. The mode of evolution is defined by the sum of the effects of translocations, inversions, and chromosome fissions.

The development of this term to the meaning above comes from a long line of research dating back to early comparative genomics studies predating modern DNA sequencing. One of the first usages of this term based on our searches was in reference to whether DNA evolution is ‘ignorant’ or ‘selfish’ (1980)<sup>89–91</sup>. Later, “Mode and tempo in primate chromosome evolution” was used to discuss the divergence in karyotype between primate species (1983)<sup>92</sup>. The term “Modes of Chromosomal Evolution” was used to reference the discovery of paracentric inversions in potato chromosomes (1988)<sup>93</sup>, and later for karyotype evolution in carcinomas (2003)<sup>94</sup>. The first usage in the context of “mode of genome evolution” or “mode of chromosome evolution” was in reference to macrosynteny shared between ascomycete fungi species, which they called “mesosynteny” (2011)<sup>30</sup>. The term “patterns of metazoan evolution” was used in 2017<sup>31</sup>, and later the term “mode of chromosome change” was used in the sense of our definition above, in reference to chromosome FWM events.

## Multi-genome topology

Acronym: MGT. A topological comparison of entire genomes. Each sample (dot on a plot) is a single genome. Each sample has a vector of comparison that contains all of the possible pairwise interactions for an orthology set. The values of the pairwise interactions could be basepair distance, basepair distance divided by chromosome length, Euclidean distance in 3D space, or other weight values derived from these or other sources. 2D projections of this matrix are called One-Dot-One-Genome (ODOG) topology plots.

## Multi-locus topology

Acronym: MLT. A topological comparison of the loci in a single genome or a phylogenetically weighted representation from a clade. Each sample (dot on a plot) is a single locus. In this manuscript, we use a 2367 x 2367 matrix of the BCnS ALG orthologs as the loci. Each sample, a single BCnS ALG ortholog, has a vector that contains the distance of this BCnS ALG to all other orthologs. This matrix is symmetric. The values of the pairwise interactions could be basepair distance, basepair distance divided by chromosome length, Euclidean distance in 3D space, or other weight values derived from these or other sources. 2D projections of this matrix are called One-Dot-One-Locus (ODOL) topology plots.

## Non-algebraic changes

When ALGs break apart. But often/always followed by 'normal' evolution, i.e. algebraic changes. This idea/term was alluded to in Albertin et al 2022<sup>8</sup>, in the passage where they refer to the evolutionary events leading



up to the cephalopods as “**big bang**” events. This encompasses major ALG fragmentations after whole genome duplications (explicitly mentioned in Albertin et al 2022), and through other processes like translocations (not mentioned explicitly in Albertin et al 2022).

- Term introduced in this manuscript.

- Terms not used in<sup>1</sup>.

<b>Robertsonian translocation</b>	Two chromosomes fuse without mixing, with a stable boundary denoted symbolically by • or ●. Robertsonian fusions are reversible if no mixing has occurred. May also describe end-to-end fusions <sup>1</sup> .
<b>Saltation</b>	The process or result of <b>saltatory changes</b> occurring.
<b>Saltatory changes</b>	Dramatic changes of chromosomes, and consequently of a whole genome, from the ancestral state. Usually in this process there are not only <b>algebraic changes</b> , as a complete fusion of chromosomes would result in a very small karyotype. Usually there are also <b>non-algebraic changes</b> that occur as well (chromosomal fissions).
<b>Syntenic equivalence</b>	One of the <b>algebraic changes</b> and part of <b>genome tectonics</b> . This is conserved syntenicity without regard to colinearity, denoted symbolically by ≡. - Term introduced <sup>1</sup> .

## **Supplementary Information 2**

### **Periodicity in genomic changes**

One persistent theme in paleobiology is the search for factors associated with, and the patterns of, speciation and extinction. One major pattern evident from the fossil record is that speciation and extinction events happen in bursts<sup>25,95</sup> in spite of an apparent constant probability of extinction<sup>96</sup>. Moreover, there appears to be a periodicity of these bursts<sup>25,26</sup>, most significantly in cycles of 62<sup>25,26</sup> and 140 million years<sup>26</sup>. The potential biologic and abiotic factors related to this periodicity are discussed in the literature cited above.

### **Methods**

With our dataset of chromosomal evolutionary rates across the tree of life, we sought to determine whether the changes in rates had significant underlying periods. To achieve this, we performed Fourier analyses and Monte Carlo simulations to estimate false discovery rates on the rates of chromosomal changes in vertebrates and protostomes, following the method described by Rohde and Muller (2005)<sup>26</sup>. In this manuscript, the authors analyzed the periodicity of the species origination and extinction rates through geological time by analyzing Sapkoski's compendium of fossil genera<sup>97</sup>.

Our input data were the rates of ALG fusions or ALG losses across the vertebrates or protostomes, generated earlier in the manuscript. To process the input data, we first limited the oldest time for each clade to between 500 and 525 million years ago, given that there are very few branches at the base of each clade, and therefore a higher chance to over- or under-estimate the rates of change. We also removed the bin representing the most recent million years, as this bin often contained outliers of evolutionary rates compared to the rest of the dataset. From the truncated datasets, we then fit cubic or quartic polynomials to the rates of change, and created a detrended dataset by subtracting this function from the data. We then performed discrete Fourier transforms on the raw data, and on zero-padded data. The results measuring periods taken from both padded and non-padded data are reported in this manuscript. We identified the strongest two or three peaks in each spectral power distribution, and selected these frequencies to analyze further with Monte Carlo false discovery rate simulations.

In Rohde and Muller (2005)<sup>26</sup>, two types of Monte Carlo simulations were performed to infer the false discovery rates ( $\alpha$ ) for the strengths of magnitude for the 62 and 140 million year peaks in the Fourier power spectrum. The first type, 'R', was a random walk through the observed transitions between timepoints in the species extinction or origination measurements. The second type, 'W', was a random walk through clusters of transitions, wherein the original dataset of transitions was partitioned into 20 equally-sized vectors, then shuffled. In both of these methods the starting rate was kept the same as the observed data, and therefore the rate of the final bin was also the same. The benefit of the 'W' type is that short-term correlations are preserved, whereas they are completely lost in the 'R' type. We propose that the 'R' type simulation is an extension of the 'W' type, in which the block size is the same size as the input vector of observations.

We performed both types, 'R' and 'W', of Monte Carlo simulations on the chromosome change rate data. From these simulations we noticed that contrary to the reported observation that changing the number of blocks for type 'W' having little effect on the background power spectra, that changing the block size for the chromosome change rate data caused variations of 20-30% in the percent of spectra with

magnitudes matching those in the real data. Given that “the choice of 20 blocks is somewhat arbitrary”<sup>26</sup> and the high variance of the background rate, we opted to perform the ‘W’ type simulations for a rank of block sizes ranging from 10 to 50 blocks. Each simulation for each peak contained 5000 random trials, and the results were summarized reporting the percent of trials whose target peaks were above a given magnitude (See Supplemental Figure S1 in Rohde and Muller (2005)<sup>26</sup>). For each simulation, we also calculated the function for the assumption of exponential form.

## Results

As we mentioned, Monte Carlo simulations were performed for block quantities between [10-50] for each condition of clade, chromosomal change type, padding treatment, and period. For this reason there is a range of support values for each simulation — one for each block quantity. We will report the median support value in text.

The two conditions with the highest support were BCnS ALG losses in protosomes with a period of 102.8-107.53 million years (support of 97.96% for unpadded, 98.04% for zero-padded), BCnS ALG losses in vertebrates with a period of 49.8 million years (support of 94.01% for unpadded, 93.45% for zero-padded).

Notably, we find that for our dataset of chromosome-scale animal genomes the most highly supported condition for Metazoans is fusions with a period of 62-64.38 million years (support of 93.8% for unpadded, 93.00% for zero-padded). This is the same period that was first noticed from a fossil diversity dataset<sup>25</sup>, and was later found to be the strongest period underlying the species extinction and origination patterns of more than 30 thousand metazoan marine genera<sup>26</sup>.

## Discussion

Our findings echo previous observations of periodicity in biological evolution from the fossil record, albeit our findings are derived independently from genomic changes. This suggests that similar underlying mechanisms may drive periodicity in both fossil records and genomic evolution. The consistency of the 62-million-year cycle across different data types and methods emphasizes its potential significance in understanding the temporal dynamics of evolution. We note, and others have noted<sup>26</sup>, that the number of 62 million years is a consequence of the sampling rate and number of observations. Future work in this area might focus on a refinement of node ages across the tree of life, and finer modeling of the periodicity given the uncertainties of node ages, to better ascertain the variances in the periodicity of genomic changes over time.

## **Supplementary Information 3**

### **State formulation of the (sub)chromosomal mixing process**

#### *Probability vector and matrix multiplication of genomic states*

This study suggests the presence of distinct irreversible evolutionary trajectories in animal genomes, with different degrees of irreversibility, depending on the number of components that are being mixed and the strength of their coupling (for example, chromosomal meiotic constraints or regulatory linkages at the sub-chromosomal level). The irreversibility is a product of mixing of genes that are located on distinct evolutionary units. In our previous studies<sup>1,27</sup>, these units were simply the chromosomal elements (or chromosomes) themselves. In this study, we generalize this mixing to include mixing of previously separated sub-chromosomal elements.

This degree of irreversibility depends on the size of the mixing units (chromosomal vs sub-chromosomal) and mechanistic reasons (regulatory linkage, meiotic constraints) behind the fusion process. An important property of the mixing is the loss of the information that the constituent elements can be separated back into the more ancestral and distinct units. For this reason, we can only identify the existence of these ancestral states through evolutionary outgroup comparisons. To begin to study these processes within a more rigorous and quantitative framework, we describe a mathematical approach to quantify the degree of irreversibility and its application in the manifold approximation context presented in this paper.

#### *(Sub-)chromosomal fusion-with-mixing representation*

We treat each chromosome as a coordinate system assigning a probability of observation for a given genomic feature. For example, the probability of finding a gene within a chromosome can be defined as a probability density vector of a size of total chromosomal positions. To find two genes to be colocalized at the same coordinate is simply the inner product of the two corresponding gene vectors ( $\mathbf{a} \cdot \mathbf{b}$ , or  $\langle \mathbf{a} | \mathbf{b} \rangle$ ). Genomic features move according to their propagation mechanism, such as translocations. An ‘evolution operator’  $U$  multiplied by a given state vector ( $U | \mathbf{a} \rangle$ ) gives the new probability of observing feature  $\mathbf{a}$  at the next evolutionary step. The system of two elements  $\mathbf{a}$  and  $\mathbf{b}$  can be described as a tensor product of these two,  $\mathbf{a} \otimes \mathbf{b}$ , as that (two-dimensional) matrix will peak at a point where both densities are highest in relation to the coordinate systems of  $\mathbf{a}$  and  $\mathbf{b}$ . For systems of many genes, this tensor product is simply generalized in a longer string of products ( $\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} \otimes \mathbf{d} \dots$ ).

In our previous studies, we represented the fusion-with-mixing process as a tensor product of two chromosomes. However, this is not a completely accurate representation, as the tensor (outer) product simply defines a combination of two coordinate systems and correspondences of each element between the two coordinate systems. If a gene ‘ $\mathbf{a}$ ’ on chromosome  $\mathbf{A}$  has probability density of  $\mathbf{a}=(1,0,0)$  and gene ‘ $\mathbf{b}$ ’ on chromosome  $\mathbf{B}$  has probability of  $\mathbf{b}=(0,0,1,0)$ , then  $\mathbf{a} \otimes \mathbf{b} = (0,0,1,0 ; 0,0,0,0 ; 0;0;0;0)$  (or an expanded vector). This describes the system state according to the corresponding coordinate systems (e.g., chromosomal position) of  $\mathbf{a}$  and  $\mathbf{b}$ , but does not provide a framework to qualitatively or quantitatively describe how and how well the chromosomes  $\mathbf{a}$  and  $\mathbf{b}$  have become mixed since their initial fusion event

To represent the mixing property one has to measure the ‘entanglement’ of the two vector spaces, where, as an initial proxy, correlative measures (such as Pearson correlation) can be used. An initial, unmixed, state consisting of two fused chromosomes (via a Robertsonian fusion), each of size two, and



each containing a single gene that can be found with equal probability at either of the two positions, can be represented as  $\mathbf{a} = (0.5, 0.5, 0, 0)$  and  $\mathbf{b} = (0, 0, 0.5, 0.5)$  and  $\mathbf{a} \otimes \mathbf{b} = (0, 0, 0.25, 0.25; 0, 0, 0.25, 0.25; 0, 0, 0, 0; 0, 0, 0, 0)$ . Similar to the eigenvector decomposition, one can then use singular value decomposition of  $\mathbf{a} \otimes \mathbf{b}$  to find that this matrix' 1st singular value is explaining 100% of the matrix, suggesting that  $\mathbf{a}$  and  $\mathbf{b}$  can be decomposed and are therefore not yet mixed. However, if the two genes translocate within the newly formed fused chromosome, one possible resulting matrix is  $(0, 0, 0, 0; 0.25, 0, 0.25, 0; 0.25, 0, 0, 0; 0.25, 0, 0, 0)$ , and the 1st singular value will explain less than 100% of variance in the matrix indicating that  $\mathbf{a}$  and  $\mathbf{b}$  coordinates cannot be decomposed clearly any longer, reflecting mixing of the two ancestral states. Together, this suggests that fusion (without yet mixing) is represented by the tensor (or outer) product ( $\otimes$  or  $|\mathbf{a}\rangle\langle\mathbf{b}|$ ), whereas the mixing itself is moving the elements within this state matrix. Therefore, braiding ( $\times$ , as defined in category theory) is mathematically a more appropriate representation of the mixing process. In our current implementation in this manuscript, we are using distance between orthologous genes as input for dimensionality reduction approaches and manifold projection. This allows us to compare evolutionarily very distant genomes, and focus on the variation in the genomic distance for the profiled set of orthogroups ( $n=2,361$ ). In this context, the positional information of the example above is encoded by the orthogroup pair identifier. In the future, instead of profiling orthogroups, this approach will also allow to generalize to other genomic features, yet always requiring a high confidence underlying homology analysis.

In a multi-dimensional system represented by the tensor product of its components' state vectors, a single mixing event (analogous to a chromosomal inversion or translocation) of the components along their coordinate systems is achieved by swapping values within the matrix. Each swap alone, without subsequent changes, is reversible (single braid). Irreversibility comes in when the topology-driven mixing results in a state where  $\mathbf{a}$  and  $\mathbf{b}$  identities cannot be distinguished anymore, without the knowledge of their separate identities from the outgroups. Thus, correlative approaches such as singular value decomposition, principal component analysis, and manifold approximation/projection are useful for the empirical identification of mixed states in genomic spaces. Additionally, singular value decomposition (or any other topological entropy measurements) can provide important insight into the 'strength' of mixing, with stronger mixing of the individual states being less likely to reverse.

## **Supplementary Information References**

84. Srivastava, M. *et al.* The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).
85. Srivastava, M. *et al.* The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010).
86. McClintock, B. The relation of a particular chromosomal element to the development of the nucleoli in *Zea mays*. *Z. Zellforsch. Mikrosk. Anat.* **21**, 294–326 (1934).
87. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
88. Luke, G. N. *et al.* Dispersal of NK homeobox gene clusters in amphioxus and humans. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5292–5295 (2003).
89. Doolittle, W. F. & Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603 (1980).
90. Orgel, L. E. & Crick, F. H. Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607 (1980).
91. Dover, G. & Doolittle, W. F. Modes of genome evolution. *Nature* **288**, 646–647 (1980).
92. Stanyon, R. & Chiarelli, B. Mode and tempo in primate chromosome evolution: Implications for hylobatid phylogeny. *J. Hum. Evol.* **12**, 305–315 (1983).
93. Bonierbale, M. W., Plaisted, R. L. & Tanksley, S. D. RFLP Maps Based on a Common Set of Clones Reveal Modes of Chromosomal Evolution in Potato and Tomato. *Genetics* **120**, 1095–1103 (1988).
94. Höglund, M., Gisselsson, D., Hansen, G. B., Säll, T. & Mitelman, F. Ovarian carcinoma develops through multiple modes of chromosomal evolution. *Cancer Res.* **63**, 3378–3385 (2003).
95. Thomson, K. S. Chapter 12 The Pattern of Diversification Among Fishes. in *Developments in Palaeontology and Stratigraphy* (ed. Hallam, A.) vol. 5 377–404 (Elsevier, 1977).
96. Van Valen, L. A new evolutionary law. *Evolutionary theory* **1**, 1 (1973).
97. Sepkoski, J. J. A compendium of fossil marine animal genera. *Bull. Am. Paleontol.* 1–560 (2002).