Conservation of animal genome structure is the exception not the rule

Thomas D. Lewin[1,*], Isabel Jiah-Yih Liao[1], and Yi-Jyun Luo[1,*]

Author affiliation:
[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

*To whom correspondence may be addressed. E-mail: tlewin@gate.sinica.edu.tw (T.D.L.), yjluo@gate.sinica.edu.tw (Y.-J.L.).

Abstract

Species from diverse animal lineages, including molluscs, hemichordates, and chordates, have retained groups of orthologous genes on the same chromosomes for over half a billion years since the last common ancestor of bilaterians. Though there are notable exceptions, it has been proposed that the conservation of chromosome-scale gene linkages is the norm among animals. Here, by examining interchromosomal rearrangements in 64 chromosome-level genomes across 15 bilaterian phyla and at least 52 classes, we show that cases of genome structure conservation are exceptionally rare. Large-scale genome restructuring events correlate with increased rates of protein sequence evolution and may represent a previously underappreciated contributor to adaptation and animal diversity.

Introduction

A remarkable feature of animal genomes is the retention of orthologous genes on the same chromosomes over long evolutionary distances. This concept of chromosomal colocalization is known as macrosynteny from the Greek for long (*makros*), together (*syn*), and ribbon (*taenia*) (1). The sequencing of genomes to chromosome level has revealed that macrosynteny has been conserved with high fidelity across animals since the bilaterian-cnidarian split (2) and, particularly, since the last common ancestor of bilaterians over 500 million years ago (3, 4). Notable examples of macrosynteny conservation from both protostomes and deuterostomes, the scallop *Pecten yessoensis* (5), the hemichordate *Ptychodera flava* (6), and the amphioxus *Branchiostoma floridae* (7), have deviated from this genome structure only through small numbers of chromosome fusions. Furthermore,

conserved macrosynteny can be identified not only within bilaterians but also between bilaterians and cnidarians, sponges, placozoans, ctenophores, and even the unicellular ancestors of animals (8). This evolutionarily deep emergence and subsequent stability suggests strong selective pressures have favored the conservation of genome structure uninterrupted for hundreds of millions of years.

Chromosome fusion, fission, and translocation events, together referred to as interchromosomal rearrangements, can degrade macrosynteny by separating genes from a single linkage group or uniting genes from different linkage groups. Indeed, clitellate annelids (9–11), nematodes (5), planarian flatworms (12), and bryozoans (13) have all lost the inherited bilaterian genome structure to varying extents. However, there remains an assumption that this is the exception, and that most bilaterian lineages retain the 24 linkage groups inherited from their common ancestor with relatively few changes, mostly chromosome fusions (4). A comprehensive assessment of macrosynteny conservation across bilaterians is urgently required to test this assumption and contextualize these individual discoveries.

Results and Discussion

We performed a quantitative profiling of the extent of macrosynteny conservation across bilaterians. We first assembled a dataset of 64 chromosome-level bilaterian genomes from 15 phyla and at least 52 classes, with the aim of maximizing phylogenetic coverage while minimizing biases caused by taxon overrepresentation. Our final dataset includes all three major bilaterian groups, with three deuterostome phyla, four ecdysozoan phyla, and seven spiralian phyla, as well as several genomes from the enigmatic Xenacoelomorpha. All bilaterians inherited 24 ancestral linkage groups (ALGs), sets of genes colocated on the same chromosome in their common ancestor (4). Within each assembly in our dataset, we assigned genes to one of the 24 bilaterian ALGs using orthology and applied our newly-developed rearrangement index (RI) (9) to quantify the conservation or loss of these ALGs. The RI distills the extent of ALG fusion and fission into a single number ranging from 0, exact conservation of bilaterian ALGs, to ≈1, complete loss of ALGs. We validated this index by applying it to species previously reported to have varying levels of interchromosomal rearrangements and found that RI scores increase in line with the extent of described rearrangements, suggesting that the index is a valid metric for measuring rearrangement (Dataset S1). Principal component analysis of ALG splitting and combining indices (see Extended Materials and Methods) supports a reported division of bilaterian genomes into high rearrangement (RI > 0.75) and low rearrangement (RI < 0.75) groups (9) (Fig. 1*A*). The difference between these groups is driven by an increased level of ALG splitting in high rearrangement genomes (Fig. 1*B*), although levels of ALG combining are also elevated (Fig. 1*C*).

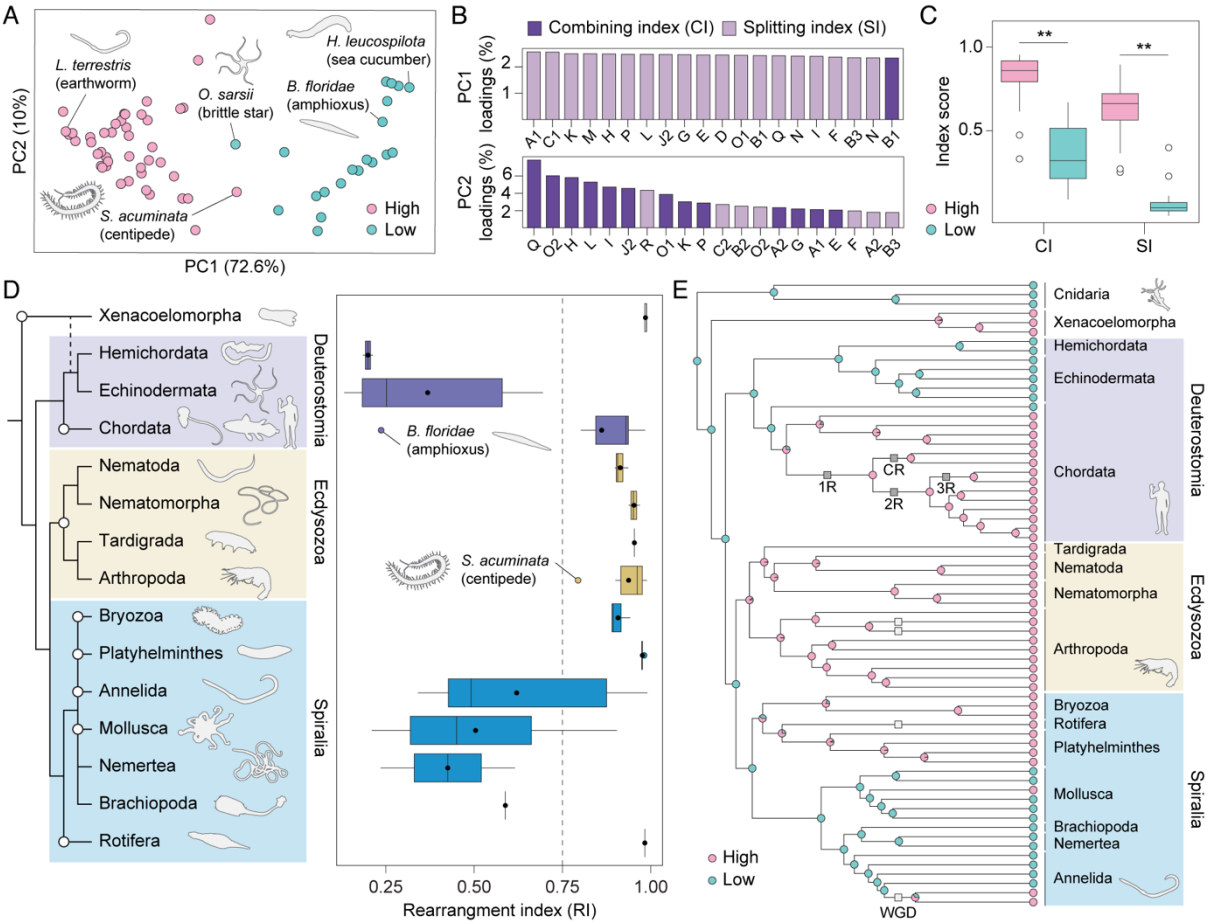

Fig. 1. Massive genome rearrangements are the norm in bilaterians. (*A*) PCA of combining and splitting indices for the 24 individual ALGs in each species. (*B*) Contribution of top 20 ALG variables to principal components. (*C*) High rearrangement species have significantly higher combining (*t*-test $p = 2.41 \times 10^{-11}$) and splitting (*t*-test $p = 2.07 \times 10^{-25}$) indices. (*D*) Boxplot of genome rearrangement index (RI) for 64 bilaterian species from 15 phyla. Blank circles on the tree indicate clades containing species with RI > 0.75. (*E*) Ancestral state reconstruction of genome rearrangement levels in a phylogenetic analysis of 67 animal species. Pie charts at inner nodes represent the proportion of stochastic character maps which produced each state. Maximum likelihood phylogeny (JTT+CAT) was built using an alignment of 148 concatenated orthologous protein sequences. Squares indicate whole genome duplication (WGD) events across vertebrate lineages (including one round, two rounds, three rounds, and a cyclostome-specific round) as well as lineage-specific WGD events in other taxa.

Having validated the RI, we used it to quantify the level of rearrangement in each assembly in our dataset. RI scores varied from 0.132 in the sea cucumber *Holothuria leucospilota* (Echinodermata) to 0.990 in the earthworm *Lumbricus terrestris* (Annelida), with a mean of 0.760. The striking result is the abundance of species with highly rearranged genomes and the diversity of phyla to which they belong. Eight of the 15 phyla in our dataset contain exclusively high rearrangement genomes (Arthropoda, Bryozoa, Platyhelminthes, Nematoda, Nematomorpha, Rotifera, Tardigrada, and Xenacoelomorpha); three (Annelida,

Chordata, and Mollusca) contain both high and low rearrangement genomes; and only four (Brachiopoda, Echinodermata, Hemichordata, and Nemertea) contain exclusively low rearrangement genomes (Fig. 1*D*). Of the 64 genomes analyzed, 44 fall into the high rearrangement category (Fig. 1*E*). Fourteen of 15 phyla (all except Hemichordata) have at least one species with an RI over 0.55 (Fig. 2*A*). Overall, two-thirds of sampled species have highly rearranged genomes and over 70% of phyla contain highly rearranged species, demonstrating that loss of macrosynteny is much more common within bilaterians than previously thought.

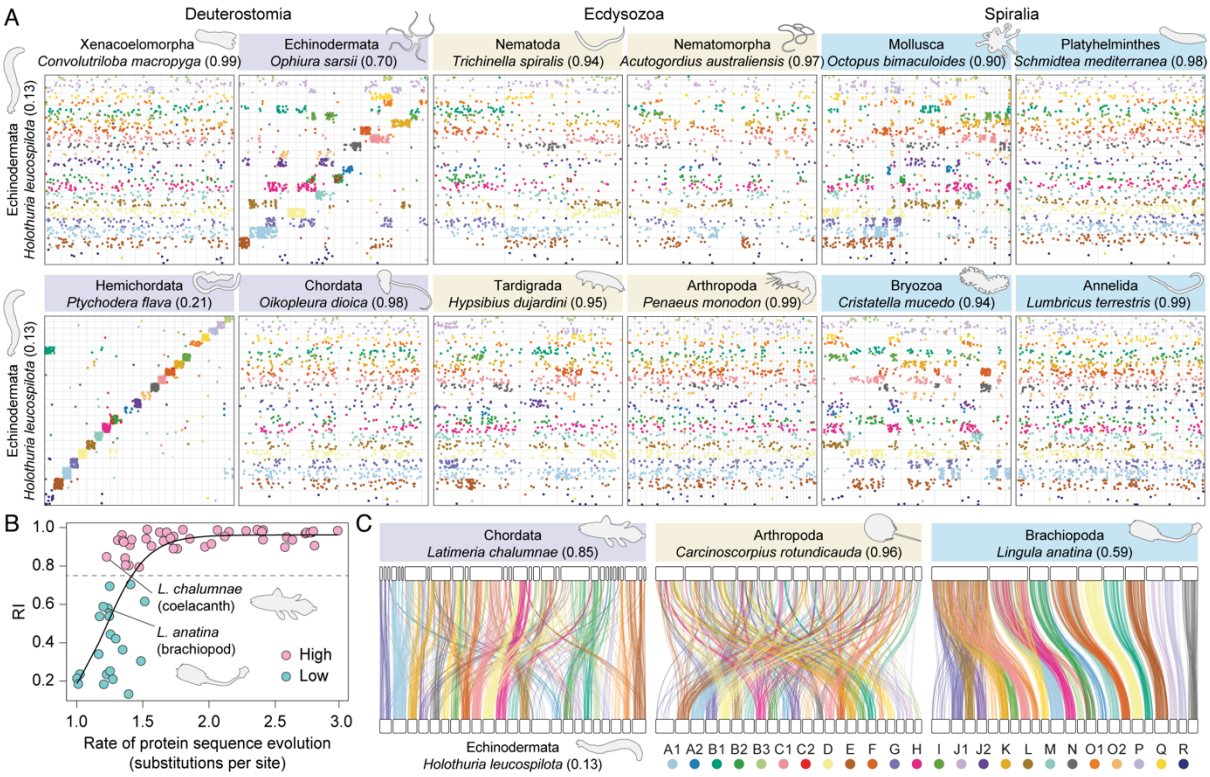

Fig. 2. Highly rearranged genomes are widespread across bilaterian phyla and are associated with higher rates of protein sequence evolution. (*A*) Dot plots for the genome with the highest RI value (shown in parentheses) from each phylum. Dots represent the position of orthologous genes, color-coded by their respective ALGs; chromosomes are separated by gray lines. (*B*) Rearrangement index versus the rate of protein sequence evolution for the studied genomes. Genomes with high rates of protein sequence evolution also have high levels of interchromosomal rearrangement. (*C*) Idiograms for the living fossils *Lingula anatina* (lamp shell), *Latimeria chalumnae* (coelacanth), and *Carcinoscorpius rotundicauda* (horseshoe crab) reveal highly rearranged genomes. White bars represent chromosomes. Vertical lines connect the genomic positions of orthologous genes, colored by the bilaterian ALG to which the genes belong.

We next questioned whether high rearrangement genomes evolved just once or independently many times. Ancestral state reconstruction reveals that transitions to a high rearrangement state occurred within all three major bilaterian groups (Spiralia, Ecdysozoa,

116 and Deuterostomia) and on a minimum of seven occasions (within Chordata, Annelida, and
117 Mollusca; in stem Xenaceolomorpha and Ecdysozoa; and at least twice in Bryozoa, Rotifera,
118 and Platyhelminthes) (Fig. 1*E*). Massive genome rearrangements have occurred at varying
119 taxonomic levels. All species within Ecdysozoa (Arthropoda, Nematoda, Nematomorpha, and
120 Tardigrada) possess highly rearranged genomes, with ancestral state reconstruction
121 supporting a loss of macrosynteny common to many ecdysozoan lineages (Fig. 1*D* and *E*). In
122 contrast, large losses of conserved macrosynteny have also occurred within specific classes
123 (Annelida: Clitellata; Mollusca: Cephalopoda) while most lineages in the same phylum retain
124 highly conserved genome structures. Overall, the dispersal of highly rearranged genomes
125 around the bilaterian tree of life demonstrates that this state has evolved repeatedly and
126 independently on many occasions, in diverse lineages and at varied taxonomic levels (Fig.
127 1*E*).

128 We then tested whether any genomic characters are predictive of RI scores. Notably,
129 we found a strong positive correlation of RI with the rate of protein sequence evolution (Fig.
130 2*B*). This demonstrates that lineages with fast-evolving protein-coding sequences also have
131 high rates of genomic shuffling by interchromosomal rearrangement, suggesting that rapid
132 evolution of coding sequences and genome rearrangements may be driven by similar selective
133 pressures. In contrast, we found no correlation between RI and genome size or repetitive
134 element content. This suggests that loss of macrosynteny in bilaterians is not driven by
135 changes to genome size or transposable element coverage, consistent with recent
136 observations in the brittle star genome (14). However, it is noteworthy that lineages with whole
137 genome duplications have typically lost the conserved bilaterian macrosynteny (Fig.1*E*).
138 Additionally, several species proposed to be 'living fossils' possess highly shuffled genomes
139 (Fig. 2*C*). This observation supports previous reports of living fossils with high rates of genome
140 sequence evolution (15) and suggests a decoupling of the rate of genome evolution from that
141 of morphological change.

142 Though we aimed to maximize phylogenetic coverage given current genome
143 availability, our dataset represents a small snapshot of bilaterian diversity. Over half of phyla
144 remain unrepresented due to the absence of chromosome-level genomes, and Brachiopoda,
145 Rotifera, and Tardigrada are represented by only one species. Although all currently sampled
146 ecdysozoan genomes are highly rearranged, chromosome-level genomes from additional
147 phyla (particularly Kinorhyncha, Priapulida, and Loricifera) are necessary to understand the
148 extent to which rearrangement had begun in the last common ancestor of ecdysozoans. As it
149 stands, the lowest level of rearrangement in this clade is found in the genome of the centipede
150 *Strigamia acuminata* (Arthropoda, RI = 0.794). Myriapod genomes may therefore be key to
151 elucidating the process by which macrosynteny was lost within ecdysozoans.

Overall, our findings reveal that extensive rearrangement to bilaterian chromosomes and subsequent erosion of macrosynteny has occurred many times independently across the animal tree of life. We argue that the high-fidelity macrosynteny conservation reported in some phyla is the exception rather than the rule, and that large-scale interchromosomal rearrangements have been pervasive throughout bilaterian evolution. Whether these events are selectively neutral or potentially deleterious but widely tolerated, or represent a previously underappreciated route to adaptation, is therefore a pressing question in genome biology.

Materials and Methods

BRAKER3 (16) was used for gene prediction. SyntenyFinder (13) was utilized for macrosynteny analysis. The rearrangement index was calculated as described previously (9). See details in *SI Appendix*.

Data, Materials, and Software Availability

The dataset for this study is available as Dataset S1. Gene models for the annotated assemblies are deposited on Zenodo (https://doi.org/10.5281/zenodo.13148652). Custom R scripts are available in our GitHub repository (https://github.com/symgenoevolab/animal_genome_structure).

Acknowledgments

This work was supported by an Academia Sinica Career Development Award (AS-CDA-112-L06) and a National Science and Technology Council Research Project Grant (113-2311-B-001-026) to Y.-J.L.

References

1. J. H. Renwick, The mapping of human chromosomes. *Annu. Rev. Genet.* **5**, 81–120 (1971).
2. B. Zimmermann, *et al.*, Topological structures and syntenic conservation in sea anemone genomes. *Nat. Commun.* **14**, 8270 (2023).
3. N. H. Putnam, *et al.*, The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008).
4. O. Simakov, *et al.*, Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).

5. S. Wang, *et al.*, Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* **1**, 120 (2017).

6. C.-Y. Lin, *et al.*, Chromosome-level genome assemblies of 2 hemichordates provide new insights into deuterostome origin and chromosome evolution. *PLoS Biol.* **22**, e3002661 (2024).

7. O. Simakov, *et al.*, Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* **4**, 820–830 (2020).

8. D. T. Schultz, *et al.*, Ancient gene linkages support ctenophores as sister to other animals. *Nature* **618**, 110–117 (2023).

9. T. D. Lewin, I. J.-Y. Liao, Y.-J. Luo, Annelid comparative genomics and the evolution of massive lineage-specific genome rearrangement in bilaterians. *bioRxiv* 2024.05.15.594353 (2024).

10. D. T. Schultz, *et al.*, Acceleration of genome rearrangement in clitellate annelids. *bioRxiv* 2024.05.12.593736 (2024).

11. C. Vargas-Chávez, *et al.*, A punctuated burst of massive genomic rearrangements by chromosome shattering and the origin of non-marine annelids. *bioRxiv* 2024.05.16.594344 (2024).

12. M. Ivankovic, *et al.*, A comparative analysis of planarian genomes reveals regulatory conservation in the face of rapid structural divergence. *bioRxiv* 2023.12.22.572568 (2023).

13. T. D. Lewin, *et al.*, Fusion, fission, and scrambling of the bilaterian genome in Bryozoa. *bioRxiv* 2024.02.15.580425 (2024).

14. E. Parey, *et al.*, The brittle star genome illuminates the genetic basis of animal appendage regeneration. *Nat. Ecol. Evol.* (2024).

15. D. Casane, P. Laurenti, Why coelacanths are not ‘living fossils’: a review of molecular and morphological data. *Bioessays* **35**, 332–338 (2013).

16. L. Gabriel, *et al.*, BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769–777 (2024).