

Acceleration of genome rearrangement in clitellate annelids

Darrin T. Schultz^{1,*}, Elizabeth A.C. Heath-Heckman^{2,3,*}, Christopher J. Winchell^{4,*}, Dian-Han Kuo⁵, Yun-sang Yu⁶, Fabian Oberauer¹, Kevin M. Kocot^{7,8}, Sung-Jin Cho^{6,+}, Oleg Simakov^{1,+}, David A. Weisblat^{4,+}

Affiliations

¹Department of Neuroscience and Developmental Biology, University of Vienna, Vienna 1010, Austria.

²Department of Integrative Biology, Michigan State University, East Lansing, MI, USA

³Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, USA

⁴Department of Molecular and Cell Biology, University of California, 385 Weill Hall, Berkeley, CA 94720-3200, USA

⁵Department of Life Science & Museum of Zoology, National Taiwan University, No. 1 Section 4 Roosevelt Rd., Taipei 10617, Taiwan.

⁶Department of Biological Sciences and Biotechnology, Chungbuk National University, Cheongju, 28644, Republic of Korea

⁷Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA

⁸Alabama Museum of Natural History, University of Alabama, Tuscaloosa, AL 35487, USA

*Equal contribution

⁺Correspondence to: Sung-Jin Cho (sjchobio@chungbuk.ac.kr), Oleg Simakov (oleg.simakov@univie.ac.at), David A. Weisblat (weisblat@berkeley.edu)

Keywords: Annelida, Clitellata, Synteny, Genome, Genome Evolution, Chromosome, Hox Genes

ORCID (alphabetical):

Cho,	Sung-Jin	https://orcid.org/0000-0001-6126-6310
Heath-Heckman,	Elizabeth A.C.	https://orcid.org/0000-0001-5265-9032
Kocot,	Kevin M.	https://orcid.org/0000-0002-8673-2688
Kuo,	Dian-Han	https://orcid.org/0000-0002-6461-2562
Oberauer,	Fabian	https://orcid.org/0009-0006-3076-2016
Schultz,	Darrin T.	https://orcid.org/0000-0003-1190-1122
Simakov,	Oleg	https://orcid.org/0000-0002-3585-4511
Weisblat,	David A.	https://orcid.org/0000-0002-9617-4130
Winchell,	Christopher J.	https://orcid.org/0009-0008-9815-7538
Yu,	Yun-Sang	https://orcid.org/0009-0005-8252-4602

ABSTRACT

Comparisons of multiple metazoan genomes have revealed the existence of ancestral linkage groups (ALGs), genomic scaffolds sharing sets of orthologous genes that have been inherited from ancestral animals for hundreds of millions of years (Simakov et al. 2022; Schultz et al. 2023). These ALGs have persisted across major animal taxa including Cnidaria, Deuterostomia, Ecdysozoa and Spiralia. Notwithstanding this general trend of chromosome-scale conservation, ALGs have been obliterated by extensive genome rearrangements in certain groups, most notably including Clitellata (oligochaetes and leeches), a group of easily overlooked invertebrates that is of tremendous ecological, agricultural and economic importance (Charles 2019; Barrett 2016). To further investigate these rearrangements, we have undertaken a comparison of 12 clitellate genomes (including four newly sequenced species) and 11 outgroup representatives. We show that these rearrangements began at the base of the Clitellata (rather than progressing gradually throughout polychaete annelids), that the inter-chromosomal rearrangements continue in several clitellate lineages and that these events have substantially shaped the evolution of the otherwise highly conserved Hox cluster.

INTRODUCTION

While changes in developmental processes are the immediate/proximal cause of changes in body plan during evolution, the underlying/ultimate cause is changes in the genomic information that is largely responsible for programming the developmental processes. A broad spectrum of genomic changes is possible, ranging from relatively frequent (point mutations and inversions) to much rarer (gene duplications, chromosomal translocations and whole genome duplications). Large differences in the frequency of successful (inherited) inversions versus translocations result in the observation that, over large evolutionary timescales, gene co-linearity among species is lost while chromosomal synteny is preserved, i.e. orthologous genes among species move back and forth along their respective chromosomes while largely maintaining their chromosomal identity, even among species as divergent as cnidarians and bilaterians, that have been evolving independently for over 600 MY (Simakov et al. 2022). Thus, conservation of synteny has allowed the inference of 29 ancestral chromosomes/chromosome arms (often referred to as ancestral linkage groups, ALGs) in the last common metazoan ancestor (Simakov et al. 2022; Schultz et al. 2023).

Exceptions to syntenic conservation exist, however. In particular, when the first three spiralian genomes were sequenced (Simakov et al. 2013), the mollusc *Lottia* and the polychaete annelid *Capitella* [separated by 534-636 MY (dos Reis et al. 2015)] showed strong synteny, but the genome of the leech *Helobdella* [a clitellate annelid separated by 476-636 MY from *Capitella* (dos Reis et al. 2015)] showed no synteny with either species; the dramatic acceleration of genome rearrangements in *Helobdella* extended even to the highly conserved Hox cluster, which exhibits multiple duplications and deletions along with atomization of the ancestral spiralian 11 gene cluster itself relative to *Lottia* and *Capitella*.

This dramatic loss of chromosomal synteny is of interest for several reasons. A priori, the relaxation of evolutionary constraints on genome architecture enhances evolutionary rates and may contribute to the evolution of novelty at genomic and morphological levels. The mechanisms by which synteny is lost are also of considerable interest. Inter-chromosomal translocations are frequently lethal because they can lead to an aberrant pairing of chromosomes during meiosis and/or gene dosage defects in the resulting gametes and embryos (Wright 1941). How might this problem have been circumvented in the lineage leading to *Helobdella*? Do the accelerated genome rearrangements result from an acceleration in all aspects of genome evolution for this taxon, or a specific relaxation of constraints on genome organization? Do the rearrangements seen in *Helobdella* result from a transient relaxation of constraints on genome organization followed by fixation somewhere in the lineage leading to *Helobdella*, perhaps within the Clitellata, or do they reflect an ongoing process?

To address these questions, we have compared the organization of 23 spiralian taxa for which well-assembled genomes are currently available (Figure 1A): 12 clitellate annelids, including 4 newly sequenced species (Figure 1B); 9 polychaete annelids; and 2 molluscan outgroups. We confirm a dramatic loss of chromosomal synteny between all the clitellate species with respect to polychaete and molluscan species. Even the most closely related polychaete species (*Capitella teleta*) sampled here shows no increased loss of ALGs (Simakov

et al. 2022). Moreover, this loss of synteny reflects an acceleration of genome rearrangements and is accompanied by elevated rates of protein sequence evolution among several of these taxa. Finally, comparisons among the available clitellate genomes suggest that the acceleration of genome rearrangements reflects an ongoing relaxation of constraints on genome organization, rather than a transient event at the base of the Clitellata. We speculate that this change results from a combination of developmental, physiological and/or ecological factors associated with the invasion of freshwater and terrestrial habitats by clitellate annelids.

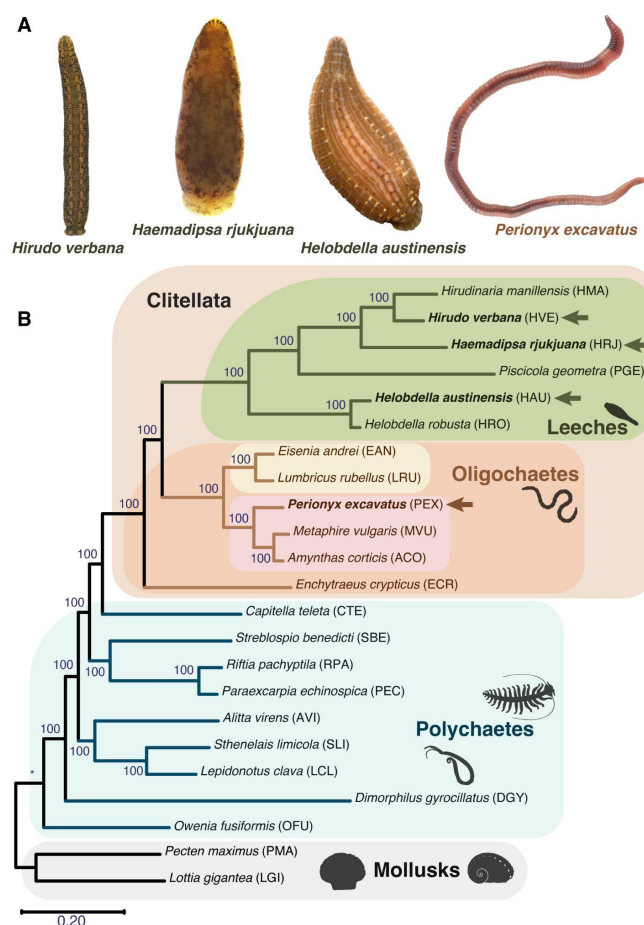


Figure 1: New clitellate genomes and clitellate-annelid relationships. A. Four newly sequenced clitellate annelids (dorsal views, anterior is up). *Hirudo verbana* (HVE, small adult 4.5, cm long). *Haemadipsa rjukjuana* (HRJ, adult, 2 cm long). *Helobdella austinensis* (HAU, adult, 1.5 cm long). *Perionyx excavatus* (PEX, adult, 11 cm long). **B.** Phylogenetic relationships (Maximum Likelihood tree based on 250 orthologous genes) among the 23 species (2 mollusks; 9 polychaete annelids; and 12 clitellate annelids- 6 oligochaetes and 6 leeches) used in this study (see Materials and Methods for details). Each species is assigned a three letter identifier. Photography credits: *H. verbana* and *H. austinensis* by Christopher J. Winchell © 2024. *H. rjukjuana* and *P. excavatus* by Sung-Jin Cho © 2024.

RESULTS AND DISCUSSION

Four new annelid genomes

Helobdella austinensis (HAU) is a small glossiphoniid leech, first identified in Austin Texas, USA (Kutschera et al. 2013), that feeds on freshwater snails and other aquatic invertebrates. It is a sibling species to *H. robusta* (HRO), which was among the first three spiralian to be sequenced (Simakov et al. 2013). HAU was sequenced because it largely replaced HRO as an experimental model due to its resistance to die-offs in lab culture. 90% of the 187 Mb Mb genome was assembled onto 41 scaffolds ≥ 1 Mb in length, using Illumina sequencing and HiRise assembly pipeline (Dovetail).

Haemadipsa rjukjuana (HRJ) is a sanguivorous land leech found in Taiwan, Japan and Korea (Lai, Nakano, and Chen 2011; Won et al. 2014). To our knowledge, this is the first haemadipsid leech species to be sequenced. Chromosome-level genome assembly analysis was achieved using PacBio sequencing and proximity ligation (Hi-C, Dovetail). The 139.8 Mb genome was assembled into 33 scaffolds, with 99% of the assembly on 11 chromosomes.

Hirudo verbana (HVE) is a hirudinid leech, one of two species of European medicinal leech (Siddall et al. 2007); it is widely used for the analysis of neural circuits underlying behavior, and by plastic surgeons for relief of vasocongestion resulting from reconstructive surgeries (Siddall et al. 2007; Kraemer et al. 1988; Houschyar et al. 2015). Previous analyses of its genome assemblies (Babenko et al. 2020; Kvist et al. 2020) had yielded short scaffolds. Here, a combination of PacBio sequencing and proximity ligation (Dovetail) yielded near chromosomal scale assembly. Roughly half of the 194 Mb genome was assembled onto 7 scaffolds, and 90% was assembled onto 35 scaffolds ≥ 1 Kb in length.

Perionyx excavatus (PEX) is a megascolecid earthworm. It probably originated from the Indian subcontinent, but it is commonly used in vermicomposting and has been collected at locations around the world (Hendrix et al. 2008). A chromosome-level genome assembly was achieved using PacBio sequencing and proximity ligation (Hi-C, Dovetail). The genome was assembled into 645 scaffolds with a size of 837 Mb, and 22 scaffolds corresponding to chromosomes were larger than 10 Mb, with a coverage of 92.16%.

Widespread loss of chromosomal synteny among clitellate annelids

Our analysis confirms the previous notion that chromosomes are generally conserved in the animal kingdom and can be represented by "algebraic" combinations of 29 ancestral animal linkage groups (Hendrix et al. 2008; Simakov et al. 2022). While the overall fusion rates can be different among animal clades, the overall background rate is relatively low among most of the metazoans (few fusions per hundreds of millions of years). However, some lineages show 'peaks' in such rearrangements that seem to occur in a time-restricted manner. For example, at the base of coleoid cephalopods, there has been a large-scale ALG reshuffling that resulted in a new set of linkage groups that have remained comparatively stable so far and only underwent species-specific fusions (Albertin et al. 2022).

The well-conserved synteny that is evident in mollusks and polychaete annelids has also been lost for all the clitellate genomes surveyed here (Figure 2). Our genomic sampling

suggests that the pace of genome rearrangements was dramatically accelerated at some point in the lineage leading to earthworms and leeches after it diverged from the capitellid lineage. In contrast to most other taxa, however, our findings for clitellates are striking in that not only the ancestral metazoan linkage group complement has been scrambled, but the inter-chromosomal scrambling may still be ongoing in at least some of the clitellate lineages (Figure 3).

Additional translocation-rich events happened at the base of each of the major clitellate lineages, including Megascolecidae, Lumbricidae, and Hirudinida. Rearrangements are also evident within the recent branches of these taxa. For example, in the leech lineages, multiple species-specific translocations can be observed. This enhanced rate of rearrangements is accompanied by an increased rate of protein sequence evolution among leeches (Figure 1A). Our findings are thus more consistent with a maintained acceleration of rearrangements in this group rather than a transient burst of accelerated rearrangements at the base followed by a return to relative stasis.

On top of the enhanced inter-chromosomal translocation rate, our analysis also detected whole genome duplications, confirming the already reported duplication in *Metaphire* (Jin et al. 2020). We observed the same duplicated karyotype in *Amyntas*, but not in *Perionyx*, which narrows the likely timing of this duplication event. Together, these data put clitellates forward as a model system to study the effect of chromosomal rearrangements in evolution.

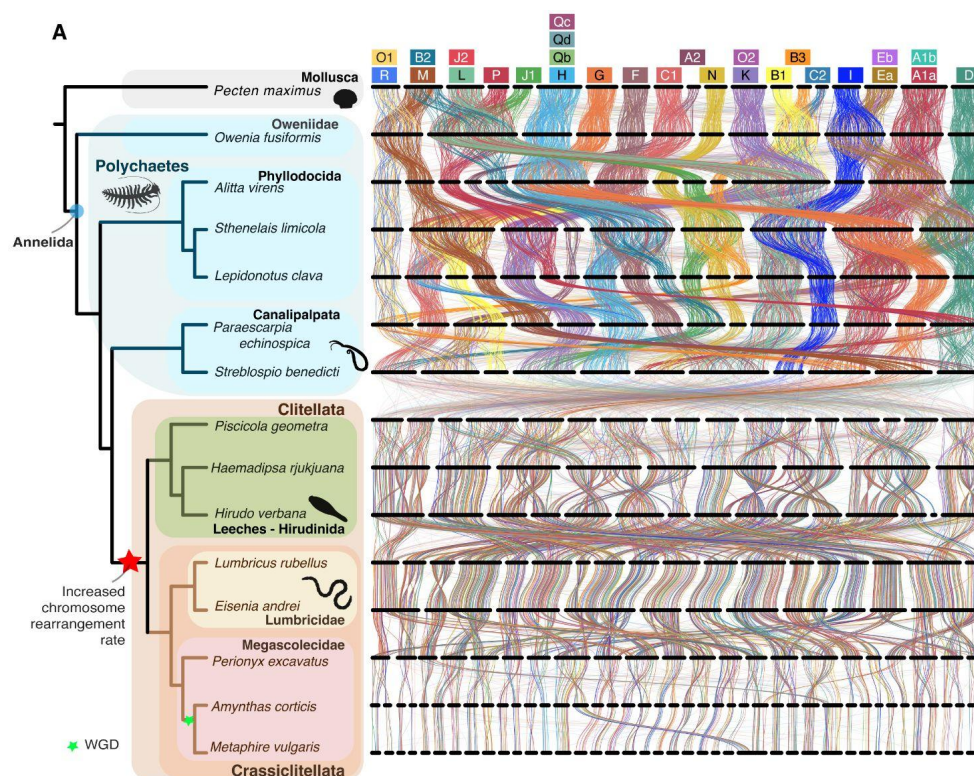
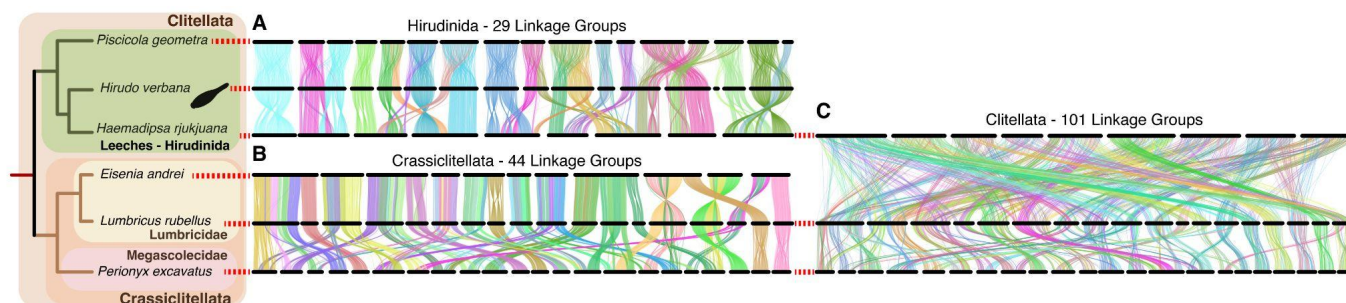


Figure 2: Rapid genome rearrangement in clitellate annelids. A. Synteny plot showing orthologs (curved vertical lines) on chromosomal scaffolds (horizontal black bars) between pairs of species. Orthologs are colored based on bilaterian-cnidarian-sponge ancestral linkage groups (BCnS ALGs). Orthologs on significantly-related chromosomes are opaque (Fisher's exact test ≤ 0.05), and orthologs on non-significant chromosome pairs are translucent. All major annelid clades, except the Clitellata, retain the BCnS ALGs that are conserved in other metazoans, including the scallop *Pecten maximus*. The genomes of extant clitellates show that there have been extensive structural rearrangements in the branch leading to clitellates, and within the clitellates.



Atomization of the Hox cluster among clitellates suggests that clitellate genome rearrangements are ongoing

Previous work has shown that Hox genes in *Helobdella robusta* are not organized in a clustered and linear array on a chromosome, as observed in many bilaterian genomes. Instead, they are broken into a few subclusters and singlets, with a number of duplicates and gene losses (Simakov et al. 2013). This “atomized” Hox configuration was observed in all clitellate species examined here; furthermore, they differ in various degrees among species (Supplementary Table 1).

To determine the extent of Hox gene translocations in the extensively rearranged clitellate genomes, we compared Hox inventories on syntenic scaffolds in three separate clitellate subgroups: lumbricids, megascolecids, and leeches. For each subgroup, we chose one anchor species to which other subgroup members were compared. The anchor species was chosen as the one with the fewest Hox-containing genome scaffolds (*Eisenia* for lumbricids and *Haemadipsa* for leeches), or as the taxon presumed to branch basally among sampled members of a subgroup (*Perionyx* for megascolecids). We grouped each Hox-containing scaffold in the anchor species with syntenic scaffolds of other subgroup members into “synteny sets” (Supplementary Table 1).

Comparing the Hox inventories within synteny sets allowed us to differentiate whether syntenic clitellate chromosomes contain conserved sub-clusters or random assortments of Hox genes (Figure 4). For the lumbricid oligochaetes (represented by *Lumbricus* and *Eisenia*), three out of six synteny sets have identical Hox inventories (Supplementary Table 1): set 4 (Hox3, two Scr, Post2), set 5 (Lab, Hox3, Scr, Lox5), and set 6 (Post1). In synteny set 1, each genus possesses Pb and Post2, but they differ with respect to a third gene, which in *Eisenia* is Scr and in *Lumbricus* is an additional copy of Post2. Both genera possess 12 genes in synteny set 3, but relative to *Eisenia*, *Lumbricus* is missing three copies of Lox4 and two copies of Scr, has gained an additional copy of Lab and of Post2, and possesses two copies of Dfd and one of Lox2 (Dfd and Lox2 were not found on any *Eisenia* scaffold).

Within megascolecid oligochaetes, *Metaphire* and *Amyntas* are within the Pheretima complex, a morphologically derived group within Megascolecidae (“Origin and Diversification of Pheretimoid Megascolecoid Earthworms in the Japanese Archipelago as Revealed by Mitogenomic Phylogenetics” 2023; Sims and Easton 1972) Sato et al. 2023 (“Origin and Diversification of Pheretimoid Megascolecoid Earthworms in the Japanese Archipelago as Revealed by Mitogenomic Phylogenetics” 2023; Sims and Easton 1972), and exhibit a shared whole genome duplication relative to *Perionyx* (Figure 2). Thus: 1), most individual chromosomes in *Perionyx* correspond to two paralogous chromosomes in *Amyntas* and in *Metaphire*, and 2) each of these paralogous chromosomes in *Amyntas* has an orthologous chromosome in *Metaphire*. Among these three megascolecid genomes, the Hox cluster remnants on the paralogous and orthologous chromosomes exhibit similarities consistent with their respective ancestries, but also differences, suggesting recent or ongoing inter-chromosomal translocations (Supplementary Table 1).

Thus, consistent with the whole genome duplication, we found that each of the eleven Hox-containing scaffolds in *Perionyx* is syntenic with two scaffolds in both *Metaphire* and *Amyntas* (Supplementary Table 1). The Hox inventories across 12 megascolecid synteny sets (Supplementary Table 1) were somewhat labile, however. In multiple cases, the presence of a particular Hox gene or set of Hox genes within a given synteny set was taxon-specific: *Perionyx* (Scr in set 1; Hox3 and Scr in set 3; Post2 in set 9; Post2, Lox2, and Lox4 in set 11), *Metaphire* (Pb in set 3; Pb, Lox2, Hox3 in set 6; Hox3 in set 12), *Amyntas* (Lox5 in set 6; Lox4 in set 7), and *Perionyx* plus *Metaphire* (Post2 in set 2; Scr in set 4; Lox2 in set 7; Scr and Hox3 in set 10). Despite these differences, we found that all three megascolecid species had in common the presence of: Pb and Post2 in set 1; Lab and Lox5 in set 3; Antp in set 4; Post1 in set 5; Post2, Lox4, Dfd, and Lab in set 6; Hox3 and Scr in set 7.

Finally, we used *Haemadipsa rjukjuana* as the anchor species for comparing Hox organization in leeches (represented by *Helobdella*, *Piscicola*, *Haemadipsa*, *Hirudo*) and found conservation of Hox inventory within all three synteny sets (Supplementary Table 1). Set 1, contains seven Hox genes in two apparent sub-clusters (Lab-Scr-Lox5 and Lab-Dfd-Lox4-Post2) in all four genera. *Piscicola* Hox3 is the only other Hox gene in leech synteny set one. Set 2, containing Hox3 and an apparent Lox2-Scr subcluster, is conserved in *Haemadipsa*, *Helobdella*, and *Hirudo* but only partially conserved in *Piscicola*, for which no Lox2 gene was found and, as mentioned previously, seems to have moved its Hox3 gene from set 2 to set 1. Set 3, containing up to 8 Hox genes, including an apparent Dfd-Lox4-Lox2-Post2 subcluster, Antp, and three Scr genes, is perfectly conserved in *Haemadipsa* and *Helobdella*, but *Piscicola* and *Hirudo* are each missing two Scr genes. Moreover, the set 3 inventory of *Hirudo* is unique in that it appears to have lost Lox2 and Antp. Consistent with previous reports, we found no evidence of Pb or Post1 in any of these leech genomes (Simakov et al. 2013).

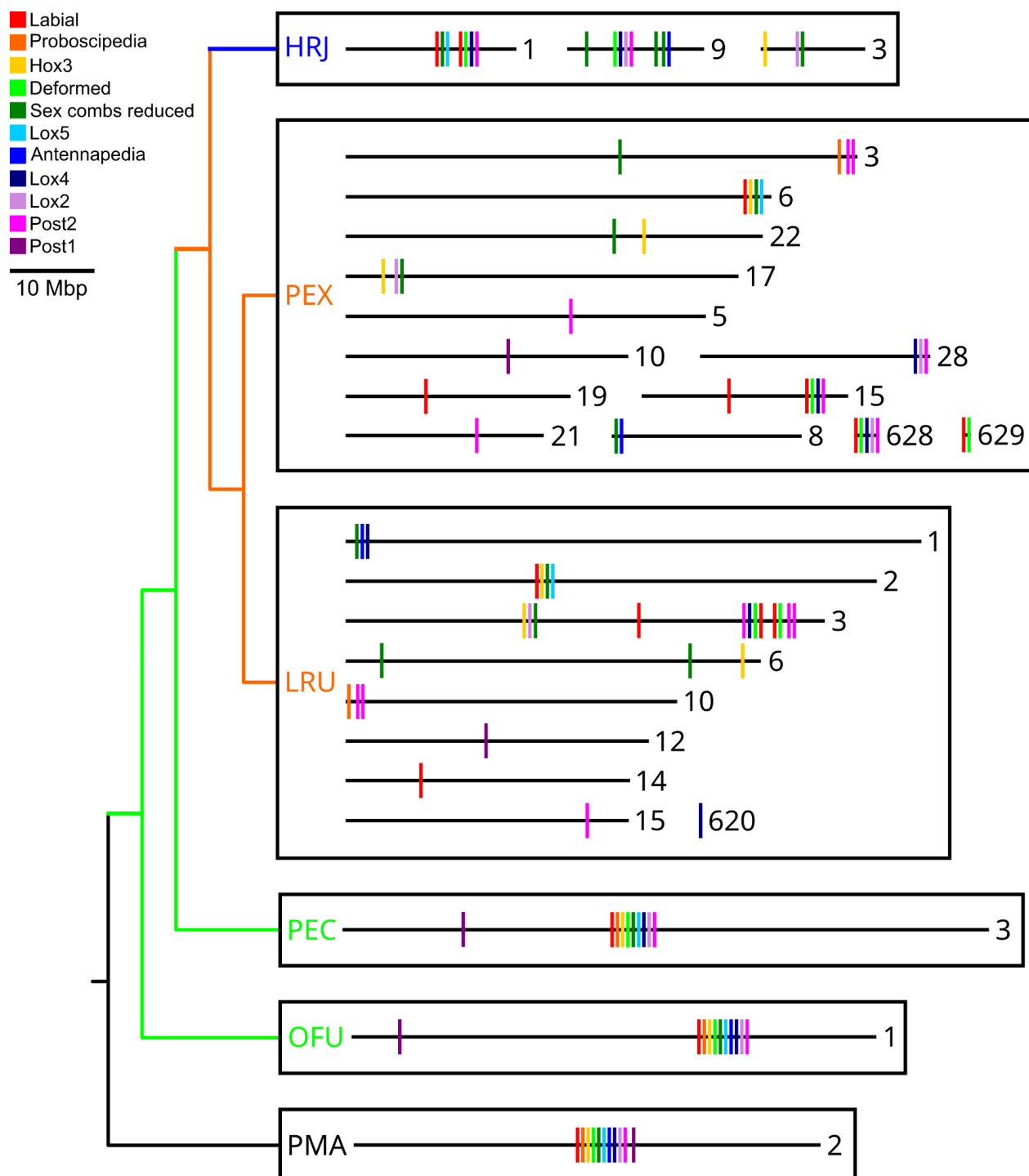


Figure 4: Dispersal of the Hox Cluster, exemplified in selected oligochaetes (orange) and leeches (blue), relative to polychaete (green) and molluscan (black) outgroups. Numbers refer to Hox-containing genome scaffolds, which are drawn to scale. Hox genes (colored vertical lines on scaffolds) are not to scale. Taxon names: PMA, *Pecten maximus*; OFU, *Owenia fusiformis*; PEC, *Paraescarpia echinospica*; LRU, *Lumbricus rubellus*; PEX, *Perionyx excavatus*; HRJ, *Haemadipsa rjukjuana*.

Potentially conserved Hox subclusters between oligochaetes and leeches

We also observed distinct associations of several sets of Hox genes across clitellate subgroups defined by their consistent colocalization on the same chromosomes.

1. Lab, Dfd, Lox4, and Post2 are contiguous within **lumbricid** synteny set 3 (*Lumbricus* only, as *Eisenia* is missing Dfd), **megascolecoid** synteny set 6, and **leech** synteny set 1. The chromosomes in these sets share some, but not much, synteny between subgroups.
2. Hox3, Scr, and Lox2 are contiguous within **megascolecoid** synteny set 7, **leech** synteny set 2, and possibly **lumbricid** synteny set 3.
3. Lab, Scr, and Lox5 (plus Hox3 in oligochaetes) are contiguous within **lumbricid** synteny set 5, **megascolecoid** synteny set 3 (*Perionyx* only), and **leech** synteny set 1.
4. Scr and Antp are contiguous within **megascolecoid** synteny set 4, **leech** synteny set 3, and possibly **lumbricid** synteny set 1 (Scr and Antp are unplaced in *Eisenia* but occur with Lox4 on *Lumbricus* scaffold 1).

Hox gene duplications and losses

As detailed above, the atomization of the Hox complement in clitellate annelids has been accompanied by numerous instances of gene duplication. In leeches, we also note two examples of Hox gene loss, namely *pb*, an anterior group gene (second in the ancestral cluster), and *post1* (a posterior group gene that comes last in the ancestral cluster).

A hallmark in the divergence of oligochaetes (earthworms) and true leeches from polychaetes is the reduction and then total loss of segmentally iterated bristles (chaetae) that serve to reduce slippage during peristaltic locomotion that is so familiar from earthworms. Within the canonical spiralian Hox cluster, *post1* occupies the terminal position of the posterior group and is expressed in the chaetal sacs of several annelid and brachiopod species (Fröbisch, Matus, and Seaver 2008; Kulakova et al. 2007; Schiemann et al. 2017) and is thought to play an integral role in chaetae function and development. Interestingly, we note that *Post1* is somewhat disproportionately separated from the rest of the Hox cluster in the bivalve mollusc (*Pecten maximus* - PMA, Figure 4)--specifically, the ratio of the length of the full 11 gene cluster to the length of the 10 gene cluster when *post1* is excluded is 1.4. In *Owenia fusiformis* - OFU, representing the deepest branch of polychaete annelids in this study, this ratio is increased to 78.2 and *post1* is now located upstream of the anterior group genes, indicative of either an inversion or transposition. The isolation of *post1* from other Hox genes is further increased in the clitellates--*post1* is commonly the only Hox gene on its scaffold in oligochaetes and is missing altogether from leech chromosomes. The absence of *post1* in the leeches is coincidental with their loss of chaetae. Further investigation into the role and consequence of *post1* and associated genes in chaetae-bearing annelids may help clarify whether clitellates lost chaetae due to a genomic deletion of *post1*, or if *post1* was lost by another function, such as relaxation of selective constraint after the loss of function at other loci. A second evolutionary

loss from the leech species examined here is the anterior group Hox gene *pb*, for which there is no obvious morphological correlate.

Factors that may contribute to relaxation of constraints on genome rearrangements in Clitellata

It is clear from our synteny analyses that the rate of genome rearrangements (more precisely, the inheritance and persistence of rearrangements that we measure in extant species) was elevated in the evolutionary branches leading up to, and within, the Clitellata. The protein phylogenomic results show that, despite the increased rate of genomic structural changes, the rate of protein evolution in the clitellate crown group is not particularly elevated relative to other annelids. The inference that the rate of genome rearrangements *remained* elevated within this group after initial divergence from all other annelids suggests that it does not reflect a transient process such as a wave of transposon activity that has since been silenced. The analysis of repetitive elements in the *Helobdella* genome also argues against that explanation (Simakov et al. 2013) and no effect on syntenic preservation in extremely repeat-rich genomes in vertebrates have been observed, e.g., in the lungfish (Meyer et al. 2021). The population-level or chromosome-scale sequencing of more genomes within the clitellates, and more broadly the annelids, will clarify the contemporary genome rearrangement rates in these clades.

We speculate that environmental factors associated with the emergence of clitellate annelids may have enabled the acceleration of heritable genome rearrangements in this clade. Freshwater and terrestrial habitats occupied by clitellates are apt to be less stable than the marine environment in which their polychaete ancestors evolved (Kuo, 2017). More ephemeral habitats may be expected to result in smaller and less stable populations, and smaller effective population size in turn facilitates the fixation of genomic changes through genetic drift (Viard, Justy, and Jarne 1997; Charbonnel et al. 2002). Another possible environmental factor contributing to increased rates of genome arrangement would be the presence of locally concentrating clastogenic agents (e.g., irradiation or chemicals) in the clitellate habitats, which are more heterogeneous and fragmented than the ancestral marine habitats.

We also speculate that evolutionary innovations associated with the emergence of clitellate annelids, namely simultaneous hermaphroditism, and self-fertilization in some taxa, might have concomitantly increased the rate at which genome rearrangements became fixed. Simultaneous hermaphroditism occurs in clitellate annelids, but is rare in polychaetes. The advantage of being able to mate with any con-specific adult in a small and/or low density population presumably offsets the energetic/physiological cost of developing two sets of reproductive organs (Heath, 1977). A more extreme adaptation to infrequent mating encounters is the capacity for self-fertilization. Self-mating, and to a lesser extent sibling matings, provide a possibility of rescuing chromosomal translocations that would otherwise prove lethal through defects in homolog pairing and/or gene dosage effects.

Finally, we previously identified developmental considerations that might also contribute to the observed increase in genome rearrangements. In many animal species, dedicated germline precursor cells (PGCs) arise early in the embryonic cell divisions, which reduces the number of mitoses required to get from the zygote to the gametes that will found the next

generation. For example, PGCs in at least some polychaetes arise as the first two cells produced by the bilateral pairs of mesodermal stem cells, corresponding to eight and nine rounds of zygotic mitoses from the fertilized egg in the conserved spiralian cleavage patterns [reviewed by (Rebscher 2014)]. In contrast, the PGCs in clitellate embryos are not segregated from somatic lineages until after at least 24 rounds of zygotic mitoses and male and female PGC precursors share the same lineages through at least 23 rounds (Rebscher 2014; Kang, Pilon, and Weisblat 2002); (Oyama and Shimizu 2007; Cho, Vallès, and Weisblat 2014)--this long cell lineage delay to the segregation of PGCs increases both the opportunities for chromosomal aberrations to occur and the likelihood that any chromosomal alterations will be shared by male and female PGCs. Thus, we suggest that the increased rate of genome rearrangements that we have documented here for clitellate annelids could result from an increased probability for generating chromosomal abnormalities that are shared between the male and female gametes, combined with the capacity for rescuing such abnormalities through self-fertilization or sibling mating.

We note that other simultaneously hermaphroditic species are known, including mangrove killifish, and pulmonate molluscs, but sequenced representatives of these taxa do not appear to exhibit accelerated genome rearrangements (Supplementary Figure 1). Comparisons among these diverse taxa with respect to such variables population size and developmental details of PGC formation may provide further insights into the mechanisms of genome rearrangement acceleration.

CONCLUSIONS

In this manuscript, we report a comparison of key spiralian and in particular clitellate taxa in terms of their phylogeny and chromosomal evolution. Our analyses confirm the faster rates of protein sequence evolution in the leech lineages. Chromosomal-level synteny comparisons reveal that the clitellate lineage underwent substantial chromosomal reorganizations, fueled by almost complete rearrangement of the ancestral metazoan linkage groups. While such genome-wide events were likely time-restricted, some lineages (e.g., leeches) kept up a higher translocation rate between chromosomes after their initial divergence from other annelids. This is also reflected in the highly partitioned and partially duplicated Hox clusters. While the mechanisms behind the breakage of ancestral chromosomal synteny in animal genomes remain elusive, this paper identifies several key biological factors that may have driven such events in clitellates, including self-mating and late germline separation.

ACKNOWLEDGEMENTS

SJC acknowledges support by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A06046235, 2023R1A2C1007868), and by grants from the National Institute of Biological Resources (NIBR), funded by the Ministry of Environment (MOE) of the Republic of

Korea (NIBR202305101, NIBR202333201). DAW and DHK acknowledge support from Human Frontiers Science Program Grant RGP0060/2019. DTS and OS were supported by the European Research Council's Horizon 2020: European Union Research and Innovation Programme, grant no. 945026. EACH was supported by the National Institute of General Medical Sciences of the NIH under Award Number R35GM150478. The computational results of this work have been supported by using the Life Science Compute Cluster (LiSC) of the University of Vienna.

REFERENCES

- Albertin, Caroline B., Sofia Medina-Ruiz, Therese Mitros, Hannah Schmidbaur, Gustavo Sanchez, Z. Yan Wang, Jane Grimwood, et al. 2022. "Genome and Transcriptome Mechanisms Driving Cephalopod Evolution." *Nature Communications* 13 (1): 2427.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.
- Amorim, Mónica J. B., Yannick Gansemans, Susana I. L. Gomes, Filip Van Nieuwerburgh, and Janeck J. Scott-Fordsmand. 2021. "Annelid Genomes: Enchytraeus Crypticus, a Soil Model for the Innate (and Primed) Immune System." *Lab Animal* 50 (10): 285–94.
- Babenko, Vladislav V., Oleg V. Podgorny, Valentin A. Manuvera, Artem S. Kasianov, Alexander I. Manolov, Ekaterina N. Gafskaja, Dmitriy A. Shirokov, et al. 2020. "Draft Genome Sequences of Hirudo Medicinalis and Salivary Transcriptome of Three Closely Related Medicinal Leeches." *BMC Genomics* 21 (1): 331.
- Bao, Zhirong, and Sean R. Eddy. 2002. "Automated de Novo Identification of Repeat Sequence Families in Sequenced Genomes." *Genome Research* 12 (8): 1269–76.
- Barrett, Paul H. 2016. *The Works of Charles Darwin: V. 28: Formation of Vegetable Mould, Through the Action of Worms, with Observations on Their Habits (1881)*. Routledge.
- Benson, G. 1999. "Tandem Repeats Finder: A Program to Analyze DNA Sequences." *Nucleic Acids Research* 27 (2): 573–80.
- Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.
- Brůna, Tomáš, Katharina J. Hoff, Alexandre Lomsadze, Mario Stanke, and Mark Borodovsky. 2021. "BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database." *NAR Genomics and Bioinformatics* 3 (1): lqaa108.
- Brůna, Tomáš, Alexandre Lomsadze, and Mark Borodovsky. 2020. "GeneMark-EP+: Eukaryotic Gene Prediction with Self-Training in the Space of Genes and Proteins." *NAR Genomics and Bioinformatics* 2 (2): lqaa026.
- Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10 (December): 421.
- Chapman, Jarrod A., Isaac Ho, Sirisha Sunkara, Shujun Luo, Gary P. Schroth, and Daniel S. Rokhsar. 2011. "Meraculous: De Novo Genome Assembly with Short Paired-End Reads." Edited by Steven L. Salzberg. *PLoS One* 6 (8): e23501.
- Charbonnel, N., B. Angers, R. Rasatavonjizay, P. Bremond, C. Debain, and P. Jarne. 2002. "The Influence of Mating System, Demography, Parasites and Colonization on the Population Structure of Biomphalaria Pfeifferi in Madagascar." *Molecular Ecology* 11 (11): 2213–28.
- Charles, Darwin. 2019. *The Formation of Vegetable Mould, Through the Action of Worms, with*

Observations on Their Habits. Wentworth Press.

- Chin, Chen-Shan, Paul Peluso, Fritz J. Sedlazeck, Maria Nattestad, Gregory T. Concepcion, Alicia Clum, Christopher Dunn, et al. 2016. "Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing." *Nature Methods* 13 (12): 1050–54.
- Cho, Sung-Jin, Yvonne Vallès, and David A. Weisblat. 2014. "Differential Expression of Conserved Germ Line Markers and Delayed Segregation of Male and Female Primordial Germ Cells in a Hermaphrodite, the Leech *Helobdella*." *Molecular Biology and Evolution* 31 (2): 341–54.
- Criscuolo, Alexis, and Simonetta Gribaldo. 2010. "BMGE (Block Mapping and Gathering with Entropy): A New Software for Selection of Phylogenetic Informative Regions from Multiple Sequence Alignments." *BMC Evolutionary Biology* 10 (1): 1–21.
- Darbyshire, Teresa, John Bishop, Nova Mieszkowska, Patrick Adkins, Anna Holmes, Marine Biological Association Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al. 2022. "The Genome Sequence of the Scale Worm, *Lepidonotus Clava* (Montagu, 1808)." *Wellcome Open Research* 7 (December): 307.
- Darbyshire, Teresa, Mitchell Brennan, Sean McTierney, Marine Biological Association Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, and Darwin Tree of Life Consortium. 2023. "The Genome Sequence of the Segmented Worm, *Sthenelais Limicola* (Ehlers, 1864)." *Wellcome Open Research* 8 (January): 31.
- Di Franco, Arnaud, Raphaël Poujol, Denis Baurain, and Hervé Philippe. 2019. "Evaluating the Usefulness of Alignment Filtering Methods to Reduce the Impact of Errors on Evolutionary Inferences." *BMC Evolutionary Biology* 19 (1): 1–17.
- Doe, Jason. 2023. "The Genome Sequence of the Fish Leech, *Piscicola Geometra* (Linnaeus, 1761)." *Wellcome Open Research*, May. <https://europepmc.org/article/ppr/ppr666762>.
- Drábková, Marie, Kevin M. Kocot, Kenneth M. Halanych, Todd H. Oakley, Leonid L. Moroz, Johanna T. Cannon, Armand Kuris, et al. 2022. "Different Phylogenomic Methods Support Monophyly of Enigmatic 'Mesozoa' (Dicyemida + Orthonectida, Lophotrochozoa)." *Proceedings. Biological Sciences / The Royal Society* 289 (1978): 20220683.
- Dudchenko, Olga, Sanjit S. Batra, Arina D. Omer, Sarah K. Nyquist, Marie Hoeger, Neva C. Durand, Muhammad S. Shamim, et al. 2017. "De Novo Assembly of the Genome Using Hi-C Yields Chromosome-Length Scaffolds." *Science* 356 (6333): 92–95.
- Dudchenko, Olga, Muhammad S. Shamim, Sanjit S. Batra, Neva C. Durand, Nathaniel T. Musial, Ragib Mostofa, Melanie Pham, et al. 2018. "The Juicebox Assembly Tools Module Facilitates de Novo Assembly of Mammalian Genomes with Chromosome-Length Scaffolds for under \$1000." *bioRxiv*. <https://doi.org/10.1101/254797>.
- Durand, Neva C., Muhammad S. Shamim, Ido Machol, Suhas S. P. Rao, Miriam H. Huntley, Eric S. Lander, and Erez Lieberman Aiden. 2016. "Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments." *Cell Systems* 3 (1): 95–98.
- Emms, David M., and Steven Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20 (1): 238.
- Fletcher, Chris, Lyndall Pereira da Conceicao, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, and Darwin Tree of Life Consortium. 2023. "The Genome Sequence of the King Ragworm, *Alitta Virens* (Sars, 1835)." *Wellcome Open Research* 8 (July): 297.

- Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. 2020. "RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families." *Proceedings of the National Academy of Sciences* 117 (17): 9451–57.
- Fröbuis, Andreas C., David Q. Matus, and Elaine C. Seaver. 2008. "Genomic Organization and Expression Demonstrate Spatial and Temporal Hox Gene Colinearity in the Lophotrochozoan *Capitella* Sp. I." *PloS One* 3 (12): e4004.
- Geib, Scott M., Brian Hall, Theodore Derego, Forest T. Bremer, Kyle Cannoles, and Sheina B. Sim. 2018. "Genome Annotation Generator: A Simple Tool for Generating and Correcting WGS Annotation Tables for NCBI Submission." *GigaScience* 7 (4): 1–5.
- Guan, De-Long, Jie Yang, Ying-Kui Liu, Yuan Li, Da Mi, Li-Bin Ma, Zhe-Zhi Wang, Sheng-Quan Xu, and Qiang Qiu. 2019. "Draft Genome of the Asian Buffalo Leech *Hirudinaria Manillensis*." *Frontiers in Genetics* 10: 1321.
- Haas, Brian J., Arthur L. Delcher, Stephen M. Mount, Jennifer R. Wortman, Roger K. Smith Jr, Linda I. Hannick, Rama Maiti, et al. 2003. "Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies." *Nucleic Acids Research* 31 (19): 5654–66.
- Haas, Brian J., Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. "De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis." *Nature Protocols* 8 (8): 1494–1512.
- Haas, Brian J., Steven L. Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E. Allen, Joshua Orvis, Owen White, C. Robin Buell, and Jennifer R. Wortman. 2008. "Automated Eukaryotic Gene Structure Annotation Using EVIDENCEModeler and the Program to Assemble Spliced Alignments." *Genome Biology* 9 (1): R7.
- Heath-Heckman, Elizabeth, Shinja Yoo, Christopher Winchell, Maurizio Pellegrino, James Angstadt, Veronica B. Lammardo, Diana Bautista, Francisco F. De-Miguel, and David Weisblat. 2021. "Transcriptional Profiling of Identified Neurons in Leech." *BMC Genomics* 22 (1): 215.
- Hendrix, Paul F., Mac A. Callahan Jr, John M. Drake, Ching-Yu Huang, Sam W. James, Bruce A. Snyder, and Weixin Zhang. 2008. "Pandora's Box Contained Bait: The Global Problem of Introduced Earthworms." *Annual Review of Ecology, Evolution, and Systematics* 39 (1): 593–613.
- Houschyar, Khosrow S., Arash Momeni, Zeshaan N. Maan, Malcolm N. Pyles, Olivia S. Jew, Marion Strathe, and Andreas Michalsen. 2015. "Medical Leech Therapy in Plastic Reconstructive Surgery." *Wiener Medizinische Wochenschrift* 165 (19-20): 419–25.
- Jin, Feng, Zhaoli Zhou, Qi Guo, Zhenwen Liang, Ruoyu Yang, Jibao Jiang, Yanlin He, Qi Zhao, and Qiang Zhao. 2020. "High-Quality Genome Assembly of." *PeerJ* 8 (November): e10313.
- Kang, Dongmin, Marc Pilon, and David A. Weisblat. 2002. "Maternal and Zygotic Expression of a Nanos-Class Gene in the Leech *Helobdella robusta*: Primordial Germ Cells Arise from Segmental Mesoderm." *Developmental Biology* 245 (1): 28–41.
- Katoh, Kazutaka, and Daron M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30 (4): 772–80.
- Kenny, Nathan J., Shane A. McCarthy, Olga Dudchenko, Katherine James, Emma Betteridge, Craig Corton, Jale Dolucan, et al. 2020. "The Gene-Rich Genome of the Scallop *Pecten Maximus*." *GigaScience* 9 (5). <https://doi.org/10.1093/gigascience/giaa037>.
- Kraemer, B. A., K. E. Korber, T. I. Aquino, and A. Engleman. 1988. "Use of Leeches in Plastic and Reconstructive Surgery: A Review." *Journal of Reconstructive Microsurgery* 4 (5): 381–86.
- Krug, Patrick J., Serena A. Caplins, Krisha Algosio, Kanique Thomas, Ángel A. Valdés, Rachael Wade, Nur Leena W. Wong, Douglas J. Eernisse, and Kevin M. Kocot. 2022. "Phylogenomic Resolution of

- the Root of Panpulmonata, a Hyperdiverse Radiation of Gastropods: New Insight into the Evolution of Air Breathing." *Proceedings of the Royal Society B*, April. <https://doi.org/10.1098/rspb.2021.1855>.
- Kulakova, Milana, Nadezhda Bakalenko, Elena Novikova, Charles E. Cook, Elena Eliseeva, Patrick R. H. Steinmetz, Roman P. Kostyuchenko, et al. 2007. "Hox Gene Expression in Larval Development of the Polychaetes Nereis Virens and Platynereis Dumerilii (Annelida, Lophotrochozoa)." *Development Genes and Evolution* 217 (1): 39–54.
- Kutschera, U., H. Langguth, D-H Kuo, D. A. Weisblat, and M. Shankland. 2013. "Description of a New Leech Species from North America, *Helobdella Austinensis* sp. (Hirudinea: Glossiphoniidae), with Observations on Its Feeding Behaviour." *Mitteilungen Aus Dem Museum Fur Naturkunde in Berlin. Zoologische Reihe* 89 (2): 239–46.
- Kvist, Sebastian, Alejandro Manzano-Marín, Danielle de Carle, Peter Trontelj, and Mark E. Siddall. 2020. "Draft Genome of the European Medicinal Leech *Hirudo Medicinalis* (Annelida, Clitellata, Hirudiniformes) with Emphasis on Anticoagulants." *Scientific Reports* 10 (1): 9885.
- Lai, Yi-Te, Takafumi Nakano, and Jiun-Hong Chen. 2011. "Three Species of Land Leeches from Taiwan, *Haemadipsa Rjukjuana* Comb. N., a New Record for *Haemadipsa Picta* Moore, and an Updated Description of *Tritetrabdella Taiwana* (Oka)." *ZooKeys*, no. 139 (October): 1–22.
- Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science* 326 (5950): 289–93.
- Lomsadze, Alexandre, Vardges Ter-Hovhannisyan, Yury O. Chernoff, and Mark Borodovsky. 2005. "Gene Identification in Novel Eukaryotic Genomes by Self-Training Algorithm." *Nucleic Acids Research* 33 (20): 6494–6506.
- Marçais, Guillaume, and Carl Kingsford. 2011. "A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of K-Mers." *Bioinformatics* 27 (6): 764–70.
- Marchler-Bauer, Aron, Shennan Lu, John B. Anderson, Farideh Chitsaz, Myra K. Derbyshire, Carol DeWeese-Scott, Jessica H. Fong, et al. 2011. "CDD: A Conserved Domain Database for the Functional Annotation of Proteins." *Nucleic Acids Research* 39 (Database issue): D225–29.
- Martín-Durán, José M., Bruno C. Vellutini, Ferdinand Marlétaz, Viviana Cetrangolo, Nevena Cvetesic, Daniel Thiel, Simon Henriët, et al. 2021. "Conservative Route to Genome Compaction in a Miniature Annelid." *Nature Ecology & Evolution* 5 (2): 231–42.
- Martín-Zamora, Francisco M., Yan Liang, Kero Guynes, Allan M. Carrillo-Baltodano, Billie E. Davies, Rory D. Donnellan, Yongkai Tan, et al. 2023. "Annelid Functional Genomics Reveal the Origins of Bilaterian Life Cycles." *Nature* 615 (7950): 105–10.
- Meyer, Axel, Siegfried Schloissnig, Paolo Franchini, Kang Du, Joost M. Woltering, Iker Irisarri, Wai Yee Wong, et al. 2021. "Giant Lungfish Genome Elucidates the Conquest of Land by Vertebrates." *Nature* 590 (7845): 284–89.
- Minh, Bui Quang, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear. 2020. "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era." *Molecular Biology and Evolution* 37 (5): 1530–34.
- Mongiardino Koch, Nicolás. 2021. "Phylogenomic Subsampling and the Search for Phylogenetically Reliable Loci." *Molecular Biology and Evolution* 38 (9): 4025–38.
- Northcutt, Adam J., Eva K. Fischer, Joshua G. Puhl, Karen A. Mesce, and David J. Schulz. 2018. "An Annotated CNS Transcriptome of the Medicinal Leech, *Hirudo Verbana*: De Novo Sequencing to Characterize Genes Associated with Nervous System Activity." *PloS One* 13 (7): e0201206.
- Oliveira, André Luiz de, Jessica Mitchell, Peter Girguis, and Monika Bright. 2022. "Novel Insights on Obligate Symbiont Lifestyle and Adaptation to Chemosynthetic Environment as Revealed by the

- Giant Tubeworm Genome.” *Molecular Biology and Evolution* 39 (1).
<https://doi.org/10.1093/molbev/msab347>.
- “Origin and Diversification of Pheretimid Megascolecoid Earthworms in the Japanese Archipelago as Revealed by Mitogenomic Phylogenetics.” 2023. *Molecular Phylogenetics and Evolution* 182 (May): 107735.
- Oyama, Atsuko, and Takashi Shimizu. 2007. “Transient Occurrence of Vasa-Expressing Cells in Nongenital Segments during Embryonic Development in the Oligochaete Annelid Tubifex Tubifex.” *Development Genes and Evolution* 217 (10): 675–90.
- Price, Alkes L., Neil C. Jones, and Pavel A. Pevzner. 2005. “De Novo Identification of Repeat Families in Large Genomes.” *Bioinformatics* 21 Suppl 1 (June): i351–58.
- Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. “FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments.” *PloS One* 5 (3): e9490.
- Rebscher, Nicole. 2014. “Establishing the Germline in Spiralian Embryos.” *The International Journal of Developmental Biology* 58 (6-8): 403–11.
- Reis, Mario dos, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C. J. Donoghue, and Ziheng Yang. 2015. “Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales.” *Current Biology: CB* 25 (22): 2939–50.
- Roach, Michael J., Simon A. Schmidt, and Anthony R. Borneman. 2018. “Purge Haplotigs: Allelic Contig Reassignment for Third-Gen Diploid Genome Assemblies.” *BMC Bioinformatics* 19 (1): 1–10.
- Schiemann, Sabrina M., José M. Martín-Durán, Aina Børve, Bruno C. Vellutini, Yale J. Passamaneck, and Andreas Hejnol. 2017. “Clustered Brachiopod Hox Genes Are Not Expressed Collinearly and Are Associated with Lophotrochozoan Novelties.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (10): E1913–22.
- Schultz, Darrin T., Steven H. D. Haddock, Jessen V. Bredeson, Richard E. Green, Oleg Simakov, and Daniel S. Rokhsar. 2023. “Ancient Gene Linkages Support Ctenophores as Sister to Other Animals.” *Nature* 618 (7963): 110–17.
- Shao, Yong, Xiao-Bo Wang, Jin-Jin Zhang, Ming-Li Li, Shou-Song Wu, Xi-Yao Ma, Xue Wang, et al. 2020. “Genome and Single-Cell RNA-Sequencing of the Earthworm Eisenia Andrei Identifies Cellular Mechanisms Underlying Regeneration.” *Nature Communications* 11 (1): 2656.
- Short, Stephen, Amaia Green Etxabe, Alex Robinson, David Spurgeon, Peter Kille, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, and Darwin Tree of Life Consortium. 2023. “The Genome Sequence of the Red Compost Earthworm, Lumbricus Rubellus (Hoffmeister, 1843).” *Wellcome Open Research* 8 (August): 354.
- Siddall, Mark E., Peter Trontelj, Serge Y. Utevsy, Mary Nkamany, and Kenneth S. Macdonald. 2007. “Diverse Molecular Data Demonstrate That Commercially Available Medicinal Leeches Are Not Hirudo Medicinalis.” *Proceedings. Biological Sciences / The Royal Society* 274 (1617): 1481–87.
- Simakov, Oleg, Jessen Bredeson, Kodiak Berkoff, Ferdinand Marletaz, Therese Mitros, Darrin T. Schultz, Brendan L. O’Connell, et al. 2022. “Deeply Conserved Synteny and the Evolution of Metazoan Chromosomes.” *Science Advances* 8 (5): eabi5884.
- Simakov, Oleg, Ferdinand Marletaz, Sung-Jin Cho, Eric Edsinger-Gonzales, Paul Havlak, Uffe Hellsten, Dian-Han Kuo, et al. 2013. “Insights into Bilaterian Evolution from Three Spiralian Genomes.” *Nature* 493 (7433): 526–31.
- Simakov, Oleg, Ferdinand Marletaz, Jia-Xing Yue, Brendan O’Connell, Jerry Jenkins, Alexander Brandt, Robert Calef, et al. 2020. “Deeply Conserved Synteny Resolves Early Events in Vertebrate Evolution.” *Nature Ecology & Evolution* 4 (6): 820–30.

- Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12.
- Sims, R. W., and E. G. Easton. 1972. "A Numerical Revision of the Earthworm Genus *Pheretima* Auct. (Megascolecidae: Oligochaeta) with the Recognition of New Genera and an Appendix on the Earthworms Collected by the Royal Society North Borneo Expedition." *Biological Journal of the Linnean Society. Linnean Society of London* 4 (3): 169–268.
- Stanke, Mario, Oliver Keller, Irfan Gunduz, Alec Hayes, Stephan Waack, and Burkhard Morgenstern. 2006. "AUGUSTUS: Ab Initio Prediction of Alternative Transcripts." *Nucleic Acids Research* 34 (Web Server issue): W435–39.
- Sun, Yanan, Jin Sun, Yi Yang, Yi Lan, Jack Chi-Ho Ip, Wai Chuen Wong, Yick Hang Kwan, et al. 2021. "Genomic Signatures Supporting the Symbiosis and Formation of Chitinous Tube in the Deep-Sea Tubeworm *Paraescarpia Echinospica*." *Molecular Biology and Evolution* 38 (10): 4116–34.
- The Darwin Tree of Life Project, Mark Blaxter, Nova Mieszkowska, Federica Di Palma, Peter Holland, Richard Durbin, Thomas Richards, et al. 2022. "Sequence Locally, Think Globally: The Darwin Tree of Life Project." *Proceedings of the National Academy of Sciences* 119 (4): e2115642118.
- The UniProt Consortium, Alex Bateman, Maria-Jesus Martin, Sandra Orchard, Michele Magrane, Shadab Ahmad, Emanuele Alpi, et al. 2022. "UniProt: The Universal Protein Knowledgebase in 2023." *Nucleic Acids Research* 51 (D1): D523–31.
- Viard, F., F. Justy, and P. Jarne. 1997. "THE INFLUENCE OF SELF-FERTILIZATION AND POPULATION DYNAMICS ON THE GENETIC STRUCTURE OF SUBDIVIDED POPULATIONS: A CASE STUDY USING MICROSATELLITE MARKERS IN THE FRESHWATER SNAIL *BULINUS TRUNCATUS*." *Evolution; International Journal of Organic Evolution* 51 (5): 1518–28.
- Vurture, Gregory W., Fritz J. Sedlazeck, Maria Nattestad, Charles J. Underwood, Han Fang, James Gurtowski, and Michael C. Schatz. 2017. "GenomeScope: Fast Reference-Free Genome Profiling from Short Reads." *Bioinformatics* 33 (14): 2202–4.
- Walker, Bruce J., Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, et al. 2014. "Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement." *PloS One* 9 (11): e112963.
- Wang, Xing, Yi Zhang, Yufeng Zhang, Mingming Kang, Yuanbo Li, Samuel W. James, Yang Yang, et al. 2021. "Amynthes Corticis Genome Reveals Molecular Mechanisms behind Global Distribution." *Communications Biology* 4 (1): 135.
- Won, Sohyun, Bae Keun Park, Baek Jun Kim, Hye Won Kim, Jun Gu Kang, Tae Seo Park, Hong Yul Seo, Ye Eun, Ki Gyoung Kim, and Joon Seok Chae. 2014. "Molecular Identification of *Haemadipsa Rjukjuana* (Hirudiniformes: Haemadipsidae) in Gageo Island, Korea." *The Korean Journal of Parasitology* 52 (2): 169–75.
- Wright, Sewall. 1941. "On the Probability of Fixation of Reciprocal Translocations." *The American Naturalist*, November. <https://doi.org/10.1086/280996>.
- Zakas, Christina, Nathan D. Harry, Elizabeth H. Scholl, and Matthew V. Rockman. 2022. "The Genome of the Poecilogonous Annelid *Streblospio Benedicti*." *Genome Biology and Evolution* 14 (2). <https://doi.org/10.1093/gbe/evac008>.
- Sato C, Nendai N, Nagata N, Okuzaki Y, Ikeda H, Minamiya Y, Sota T. 2023. "Origin and diversification of pheretimoid megascolecoid earthworms in the Japanese Archipelago as revealed by mitogenomic phylogenetics". *Mol Phylogenet Evol.* 182:107735. doi: 10.1016/j.ympev.2023.107735. Epub 2023 Feb 17. PMID: 36805472.

- Darwin, Charles. 1881. The Formation of Vegetable Mould, Through the Action of Worms, with Observations on Their Habits. John Murray.
- Heath, D. J. 1977. "Simultaneous Hermaphroditism; Cost and Benefit." 1977. *Journal of Theoretical Biology* 64 (2): 363–73.
- Kuo, Dian-Han. 2017. "The Polychaete-to-Clitellate Transition: An EvoDevo Perspective." *Developmental Biology* 427 (2): 230–40.
- Sato, Chika, Noriko Nendai, Nobuaki Nagata, Yutaka Okuzaki, Hiroshi Ikeda, Yukio Minamiya, and Teiji Sota. 2023. "Origin and Diversification of Pheretimid Megascolecoid Earthworms in the Japanese Archipelago as Revealed by Mitogenomic Phylogenetics." *Molecular Phylogenetics and Evolution* 182 (May): 107735.
- Viard, F., F. Justy, and P. Jarne. 1997. "The Influence Of Self-Fertilization And Population Dynamics On The Genetic Structure Of Subdivided Populations: A Case Study Using Microsatellite Markers In The Freshwater Snail *Bulinus truncatus*." *Evolution; International Journal of Organic Evolution* 51 (5): 1518–28.

METHODS

Genome database preparation

The following chromosome-scale and scaffold-level annelid and mollusk genomes, as well as their genome annotations and protein sequences, were downloaded for this project: *Helobdella robusta* (GCF_000326865.2) (Simakov et al. 2013), *Hirudinaria manillensis* (GCA_034509925.1) (Guan et al. 2019), *Metaphire vulgaris* (GCA_018105865.1) (Jin et al. 2020), *Amyntas corticis* (GCA_900184025.1) (Wang et al. 2021), *Eisenia andrei* (GWHACBE000000000) (Shao et al. 2020), *Enchytraeus crypticus* (GCA_905160935.1) (Amorim et al. 2021), *Capitella teleta* (GCA_000328365.1) (Simakov et al. 2013), *Streblospio benedicti* (GCA_019095985.1) (Zakas et al. 2022), *Dimorphilus gyrotilatus* (GCA_904063045.1) (Martín-Durán et al. 2021), *Owenia fusiformis* (GCA_903813345.2) (Martín-Zamora et al. 2023), *Lottia gigantea* (GCF_000327385.1) (Simakov et al. 2013), *Pecten maximus* (GCA_902652985.1) (Kenny et al. 2020), *Riftia pachyptila* (<https://phaidra.univie.ac.at/detail/o:1220865>) (de Oliveira et al. 2022), *Lepidonotus clava* (GCA_936440205.1) (Darbyshire et al. 2022), *Sthenelais limicola* (GCA_942159475.1) (Darbyshire et al. 2023), *Alitta virens* (GCA_932294295.1) (Fletcher et al. 2023), *Paraescarpia echinospica* (GCA_020002185.1) (Sun et al. 2021), *Lumbricus rubellus* (GCA_945859605.1) (Short et al. 2023), and *Piscicola geometra* (GCA_943735955.1) (Doe 2023). Additionally, we downloaded the following genomes of the simultaneously hermaphroditic species from NCBI: *Biomphalaria glabrata* (GCA_947242115.1) (The Darwin Tree of Life Project et al. 2022), and *Kryptolebias marmoratus* (GCF_001649575.2).

The genomes were prepared for use with the odp software package (Schultz et al. 2023) by extracting the protein coordinate information to rbh files.

Sample collection

Adult earthworm *Perionyx excavatus* were obtained from the Nan-ji Water Reclamation Center (37° 35' 13" N, 126° 50' 45" E) in Kyunggi-do, South Korea. Adult leech *Haemadipsa rjukjuana* were collected on Mt. Dock-Sil (altitude 639m) on the island Gageo-do (E 125° 07', N 34° 04'), South Korea. The earthworms and leeches were collected in compliance with local regulations, and collected animals were maintained in an animal incubation room under controlled conditions. *Helobdella austinensis* (Kutschera et al. 2013) adults were originally collected from Shoal Creek in Austin, TX (within 1 kilometer of 30° 17' 41" N, 97° 44' 44" W) in 1998 and have been in continuous laboratory culture since. The collection of *H. austinensis* leeches was performed in compliance with Texas regulations (TAC Title 31, PT 2 TPWD, Ch 69, Subchapter J, Rule §69.302). Adult *Hirudo* were obtained from a commercial supplier (Leeches.com).

Haemadipsa rjukjuana and *Perionyx excavatus* methods

***H. rjukjuana* and *P. excavatus* DNA and RNA extraction** The MagAttract HMW DNA Kit (Qiagen, catalog no.67563) was used to extract DNA from the muscle tissue was sampled from an adult *H. rjukjuana* with a body length of 2 cm, and an adult *P. excavatus* with a body length of 11 cm. DNA was eluted with 100ul of Solution AE. Total RNA was extracted from six

individuals of each species using mirVana miRNA Isolation Kit (Ambion) according to the manufacturer's recommended procedures.

***H. rjukjuana* and *P. excavatus* DNA Short-read Sequencing** We first used a NanoDrop spectrometer to confirm that the *H. rjukjuana* and *P. excavatus* DNA was pure, with an OD260/280 ratio of 1.8–2.0. We also verified that the DNA was intact, using a 1% agarose gel electrophoresis. Lastly, we quantified the DNA with a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

Whole-genome shotgun sequencing libraries were prepared according to the Illumina Truseq Nano DNA Library prep protocol. Briefly, 0.2µg of high molecular weight (HMW) DNA was randomly sheared using the Covaris S2 system to a median insert size of 350 basepairs. The fragmented DNA was prepared according to the manufacturer's protocol, and the quality of the amplified libraries was verified by capillary electrophoresis (Bioanalyzer, Agilent). The libraries were quantified using qPCR with SYBR Green PCR Master Mix (Applied Biosystems), then were pooled in equimolar amounts. The pool was sequenced on an Illumina NovaSeq 6000 system following the standard Illumina protocols for 2x150 bp sequencing.

***H. rjukjuana* and *P. excavatus* RNA Short-read Sequencing** The *H. rjukjuana* and *P. excavatus* RNA purity was determined by assaying 1 µl of total RNA extract on a NanoDrop8000 spectrophotometer. The RNA Integrity Number (RIN) value was calculated using an Agilent Technologies 2100 Bioanalyzer. The mRNA sequencing libraries were prepared according to the manufacturer's instructions for random hexamer-primed RNA-seq libraries, and were amplified with indexing primers (Illumina Truseq stranded mRNA library prep kit). The quality of the amplified libraries was verified by capillary electrophoresis (Bioanalyzer, Agilent). The RNA-seq libraries were quantified using qPCR with SYBR Green PCR Master Mix (Applied Biosystems). After quantification the libraries were pooled in equimolar amounts and sequenced on an Illumina NovaSeq 6000 system following the provided protocols for 2x100 bp sequencing.

***H. rjukjuana* and *P. excavatus* Long-read Sequencing** Using the Covaris G-tube we generated 20Kb fragments by shearing genomic DNA according to the manufacturer's recommended protocol. Using the AMPureXP bead purification system to remove the small fragments. A total of 5µg for each sample was used as input into library preparation. The SMRTbell library was constructed by using SMRTbell™ Template Prep Kit 1.0 (PN 100-259-100). Using the BluePippin Size selection system we remove the small fragments to create a large-insert library. After sequencing primers were annealed to the SMRTbell template, DNA polymerase was bound to the complex (Sequel Binding Kit 2.0) and purified with SMRTbell Clean Up Columns v2 Kit-Mag: PN 01-303-600. The purification step was performed after polymerase binding to remove excess unbound polymerases and polymerase molecules bound to small DNA inserts. The MagBead Kit was used to bind the library complex with MagBeads before sequencing. The SMRTbell library was sequenced using one Pacific Biosciences 8M SMRT cell using Sequel Sequencing Kit 2.1 on 600 minute movie (continuous long read mode) using the Pacific Biosciences Sequel 2 sequencing platform.

***H. rjukjuana* and *P. excavatus* Dovetail Hi-C sequencing** Dovetail HiC libraries were prepared as described previously (Lieberman-Aiden et al. 2009), using a protocol first fixing tissue, then isolating nuclei. The digestion enzyme used was DpnII. The re-ligated DNA was then sheared to ~350 bp mean fragment size with a Covaris shearing machine, and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters, then enriched then amplified on streptavidin beads. The libraries were sequenced on an Illumina NovaSeq 6000 system.

***Haemadipsa rjukjuana* and *Perionyx excavatus* Genome Size Estimation** To estimate the genome size, we used the Illumina whole genome sequencing data to count 17-, 19-, and 21-mers with Jellyfish version 2.1.3 (Marçais and Kingsford 2011). We estimated the genome sizes using GenomeScope (Marçais and Kingsford 2011; Vurture et al. 2017) to obtain estimates for genome sizes, heterozygosity and duplication levels.

***Haemadipsa rjukjuana* and *Perionyx excavatus* Genome Assembly** For both species, the average coverage of continuous long read (CLR) PacBio sequences was about 96-fold. The average subread length was 16.4 Kb and N50 subread length was 23.2 Kb. The genomes were *De novo* assembled using FALCON-Unzip (Chin et al. 2016) with read lengths greater than the subread N50 value, 23.2 kbp. Duplicate haplotigs were removed using Purge Haplotigs (Roach, Schmidt, and Borneman 2018) with default parameters. The resulting assembly was error-corrected using Pilon (Roach, Schmidt, and Borneman 2018; Walker et al. 2014) with primary contigs to improve the quality of genome assembly results. We checked the assessment of the genome assemblies using BUSCO (Benchmarking Universal Single-Copy Orthologs) (Simão et al. 2015).

***Haemadipsa rjukjuana* and *Perionyx excavatus* Hi-C Scaffolding Analysis** The Hi-C reads were aligned to primary contigs using juicer version 1.5 (Durand et al. 2016). and the alignments were scaffolded using 3D-DNA (Durand et al. 2016; Dudchenko et al. 2017). The initial scaffolding results were manually reviewed for correcting mis-join and unplaced contigs using Juicebox (Dudchenko et al. 2018). Finally, 3D-DNA was used to regenerate chromosome-level genome assemblies.

***Haemadipsa rjukjuana* and *Perionyx excavatus* Repeat Annotation** A *de novo* repeat library was constructed using RepeatModeler v.1.0.3 (Bao and Eddy 2002), and RECON/RepeatScout v.1.0.5 (A. L. Price, Jones, and Pevzner 2005), with default parameters. Tandem Repeats Finder (Benson 1999) was also used to predict consensus sequences and classification data for each repeat. All repeats collected by RepeatModeler were searched against the UniProt/SwissProt database (The UniProt Consortium et al. 2022); transposons were excluded. Repetitive elements in the genome were identified using RepeatMasker v.4.0.9 with the repeat library from RepeatModeler.

***Haemadipsa rjukjuana* and *Perionyx excavatus* Gene Prediction and Annotation** Genome prediction was performed using EvidenceModeler (EVM) v.1.1.1 (Haas et al. 2008), which

integrates the results of multiple gene predictions. Repeat-masked genomes were used for *ab initio* gene prediction using GeneMark-ES v.4.68 (Lomsadze et al. 2005) and Augustus v.3.4.0 (Lomsadze et al. 2005; Stanke et al. 2006). Then, the hints for protein and *ab initio* predictions were extracted using protein sequences from Actinopterygii, a clade of bony fishes, in the UniProt/SwissProt protein database (The UniProt Consortium et al. 2022) using ProtHint v.2.6.0 (Brůna, Lomsadze, and Borodovsky 2020). The hints were used to perform protein predictions using GeneMark-EP+ v.4.68 (Brůna, Lomsadze, and Borodovsky 2020) and *ab initio* predictions using Augustus. To obtain transcriptome-level evidence, the PASA pipeline v.2.3.3 (Brůna, Lomsadze, and Borodovsky 2020; Haas et al. 2003) with transcriptome assembly data using Trinity v.2.8.5 (Haas et al. 2013) was used. EVM was used to integrate the *ab initio*, transcriptome, and protein prediction results to obtain the final gene prediction with the weights (ABINITIO_PREDICTION=1, PROTEIN=10, TRANSCRIPT=10). Finally, to predict changes in exons by the addition of untranslated regions (UTRs), the PASA pipeline with Iso-Seq data was used again. Genome Annotation Generator v.2.0.1 (Geib et al. 2018) was used for adding start/stop codon data and generating a gff file. The predicted genes were annotated by aligning them to the NCBI non-redundant protein (nr) database (Geib et al. 2018; Marchler-Bauer et al. 2011) using NCBI BLAST v.2.9.0 (Altschul et al. 1990) with a maximum e-value of 1×10^{-5} .

***Helobdella austinensis* and *Hirudo verbana* methods**

***H. austinensis* and *H. verbana* DNA extraction and library preparation** Dovetail Genomics isolated HMW DNA from *Hirudo verbana* testisacs, and from a single clutch of intact, unfed *Helobdella austinensis* individuals. The *Helobdella austinensis* DNA was processed into a whole genome shotgun (WGS) Illumina sequencing library, and was sequenced on a 2x150 run on an Illumina machine to a depth of 678 million read pairs. The *Hirudo verbana* DNA was processed into a Pacific Biosciences SMRT library and was sequenced on CLR mode.

***H. austinensis* and *H. verbana* genome assembly.** The *Helobdella austinensis* WGS Illumina reads were quality trimmed with Trimmomatic (Bolger, Lohse, and Usadel 2014) using the ILLUMINACLIP parameters “2:30:10 SLIDINGWINDOW:13:20 LEADING:20 TRAILING:20 MINLEN:23”. The genome size was estimated as above for *Perionyx excavatus* and *Haemadipsa rjukjuana*, except k-mer sizes of 19, 37, 49, 79, and 109 were used to also infer the k-mer suitability for the Meraculous genome assembler. Meraculous v2.2.4 (Chapman et al. 2011) was used to assemble the genome into contigs and scaffolds using default parameters, a k-mer size of 37, and an estimated genome size of 228 Mbp. Haplotigs were removed from the assembly as part of the Meraculous genome assembly process.

The *Hirudo verbana* Pacific Biosciences CLR reads were assembled using the same procedure as described above for *Perionyx excavatus* and *Haemadipsa rjukjuana*.

***H. austinensis* and *H. verbana* Dovetail Omni-C Library Preparation and Sequencing**

Dovetail “Omni-C” DNase I Hi-C libraries were prepared by Dovetail Genomics, now Cantata Bio, for both *Helobdella austinensis* and *Hirudo verbana* using their provided protocol. The

protocol includes chromatin fixation, DNase I digestion and random re-ligation with biotinylated bridge oligos, and streptavidin enrichment of DNA with re-ligation junctions. The sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. The libraries were sequenced on an Illumina HiSeqX platform to produce an approximately 30x sequence coverage of the estimated genome size.

***Helobdella austinensis* and *Hirudo verbana* Assembly Scaffolding with HiRise** The input *de novo* Illumina assembly for *H. austinensis* and the *de novo* Pacific Biosciences *H. verbana* assemblies, and their Dovetail Omni-C library reads, were used as input data for the HiRise genome scaffolder with standard parameters (Putnam et al, 2016). Sequences were curated post-scaffolding by Dovetail genomics such that the assemblies were largely composed of chromosome-level scaffolds.

***H. verbana* and *H. austinensis* genome annotation.** The chromosome-scale *Hirudo verbana* genome assembly was annotated by Dovetail genomics. The input data for *Hirudo verbana* annotation hints were existing Illumina RNA-seq data from dissected testisacs, nephridia (Northcutt et al. 2018), ganglia, and neuronal cell types (Heath-Heckman et al. 2021).

The chromosome-scale *Helobdella austinensis* genome was annotated using BRAKER2 v2.1.2 (Brůna et al. 2021), using protein sequences from *Hirudo verbana* as splice site hints.

Annotation of the *Piscicola geometra* genome.

The *Piscicola geometra* genome was downloaded from the Darwin Tree of Life Data Portal (<https://portal.darwintreeoflife.org>; accession GCA_943735955; “All Seq FASTA”). Because an annotation of this genome was not publicly available and no transcriptome data were available for this species at the time this work was performed, we annotated it using publicly available leech transcriptome assemblies from NCBI TSA (GBRF01.1, GGIQ01.1, GIVY01.1, GIVZ01.1, GIWB01.1, GIWE01.1, GIWF01.1, GIWG01.1, GIWH01.1, GIWI01.1) and transcriptomes from representatives of Piscicolidae and Ozobranchidae. For the latter, raw data were downloaded from NCBI SRA (SRR6766626, SRR10997450, SRR6766627, SRR6766628, SRR6766630, SRR6766631, SRR6766632, SRR10997434, and SRR6766629) and assembled with Trinity v2.8.4 (Haas et al. 2013) with the following optional flags: --trimmomatic --quality_trimming_params "ILLUMINACLIP:/kmk/scripts/trimmomatic/adapters.fasta:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25" --normalize_reads. Assembled transcriptomes were translated using TransDecoder (<https://github.com/TransDecoder>) following the approach used by (Kutschera et al. 2013; Drábková et al. 2022). Repeats in the genome assembly were annotated and softmasked with RepeatMasker and RepeatModeler (Flynn et al. 2020) using rmbast for both programs (<https://www.repeatmasker.org>). For RepeatModeler, a maximum genome sample size of 1M and the --LTRStruct option were used. For RepeatMasker, the slow and gccalc options were used. Translated transcriptomes were then mapped to the genome assembly with ProHint v2.6 (Brůna, Lomsadze, and Borodovsky 2020) with an e-value cutoff of 1×10^{-25} . Using the output of ProHint as evidence, annotation of

protein-coding genes was performed with BRAKER v2.1.6 (Brůna et al. 2021) with the following settings: “--epmode --softmasking --crf.”

Orthology assessment, alignment, and matrix construction. Our approach to generate datasets for phylogenomic analysis followed the bioinformatic pipeline of (Krug et al. 2022). We used OrthoFinder v2.4.0 (Brůna et al. 2021; Emms and Kelly 2019) to identify putatively orthologous sequences among taxa, removing sequences of <100 amino acids from fasta files and keeping the longest non-redundant sequence. Fasta files sampled for ≥75% of taxa were aligned with MAFFT v7.310 (Kato and Standley 2013), putatively mistranslated regions removed with HmmerCleaner (Di Franco et al. 2019), and alignments trimmed to remove ambiguously aligned regions with BMGE v1.12.2 (Criscuolo and Gribaldo 2010). Approximate ML trees were constructed for each alignment with FastTree v2 (M. N. Price, Dehal, and Arkin 2010), and PhyloPyPruner v0.9.5 (<https://pypi.org/project/phylopypruner>) was used to identify strictly orthologous sequences. We also selected the best 250 genes retained by PhyloPyPruner based on seven properties calculated in genesortR (Mongiardino Koch 2021).

Phylogenomic analyses. We performed ML analyses on the complete dataset with IQ-Tree v2 (Mongiardino Koch 2021; Minh et al. 2020) using three strategies: 1) the best-fitting model of amino acid substitution for each partition (-m MFP), 2) Lanfear clustering (-m MFP+merge), and 3) the PMSF model (-m LG+C20+F+G) with the contree produced from the Lanfear clustering analysis specified with “-ft.” The 250-gene dataset was analyzed using the PMSF model with the contree produced by the Lanfear clustering analysis of the complete dataset as the guide tree. Topological support was assessed with 1,000 rapid bootstraps for all analyses. Pairwise patristic distances between taxa were calculated with Patristic v1.0. based on results of the analysis of the complete matrix using Lanfear clustering.

Pairwise species comparisons. The odp software v0.3.0 was used to perform reciprocal best blast hits between species pairs, to identify whether a given ortholog corresponded to a known bilaterian, cnidarian, and sponge ancestral linkage group (BCnS ALG), to identify significantly large linkage groups between species, and to create ribbon plots of multiple species. The protocols for these steps have been previously described (Schultz et al. 2023).

BLAST searches for Hox genes. We used the TBLASTN operation of BLAST+ 2.15.0 (Camacho et al. 2009) to find reciprocal best hits for Hox genes between an initial set of queries (Hox protein sequences of the polychaete *Capitella teleta*) and a subject genome. To identify potential Hox duplicates, we queried the *Capitella* genome with multiple top hits from the initial subject genome. In an effort to recover Hox genes whose homeoboxes contain introns, we used the TBLASTN commands: -max_intron_length 8000 and -evalue 1e-5.

Data Availability

Genomes and raw reads will be available under NCBI BioProject PRJNA1109803.

Code Availability

The odp software, and the scripts used to make dotplots, ribbon diagrams, and to infer ancestral linkage groups, is available on github <https://github.com/conchoecia/odp>.

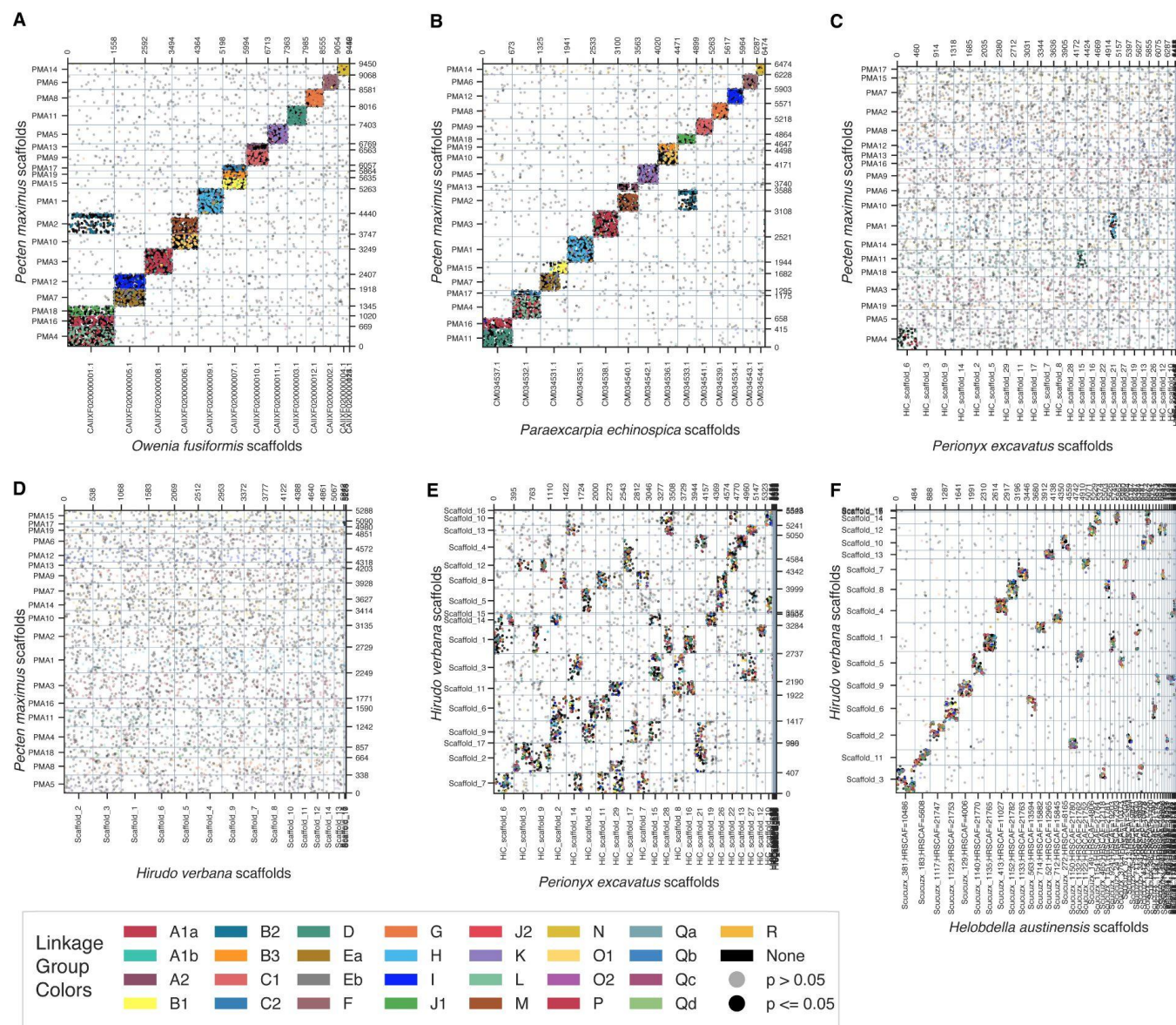
AUTHOR CONTRIBUTIONS

DAW, EACH, CJW, and DHK designed the scientific objectives of the study. KK and DTL conducted phylogenetic analyses. EACH and YSY performed *de novo* genome assembly and annotation. KK, DTS, FO, YSY, CJW, EACH and SJC prepared the genomic data. DTS wrote the code for synteny analyses. DTS and FO conducted the synteny analyses. CJW, EACH and DW performed the Hox evolutionary analysis. DAW, CJW and OS wrote the initial manuscript draft. All authors participated in interpreting the results, editing the manuscript, and creating the figures.

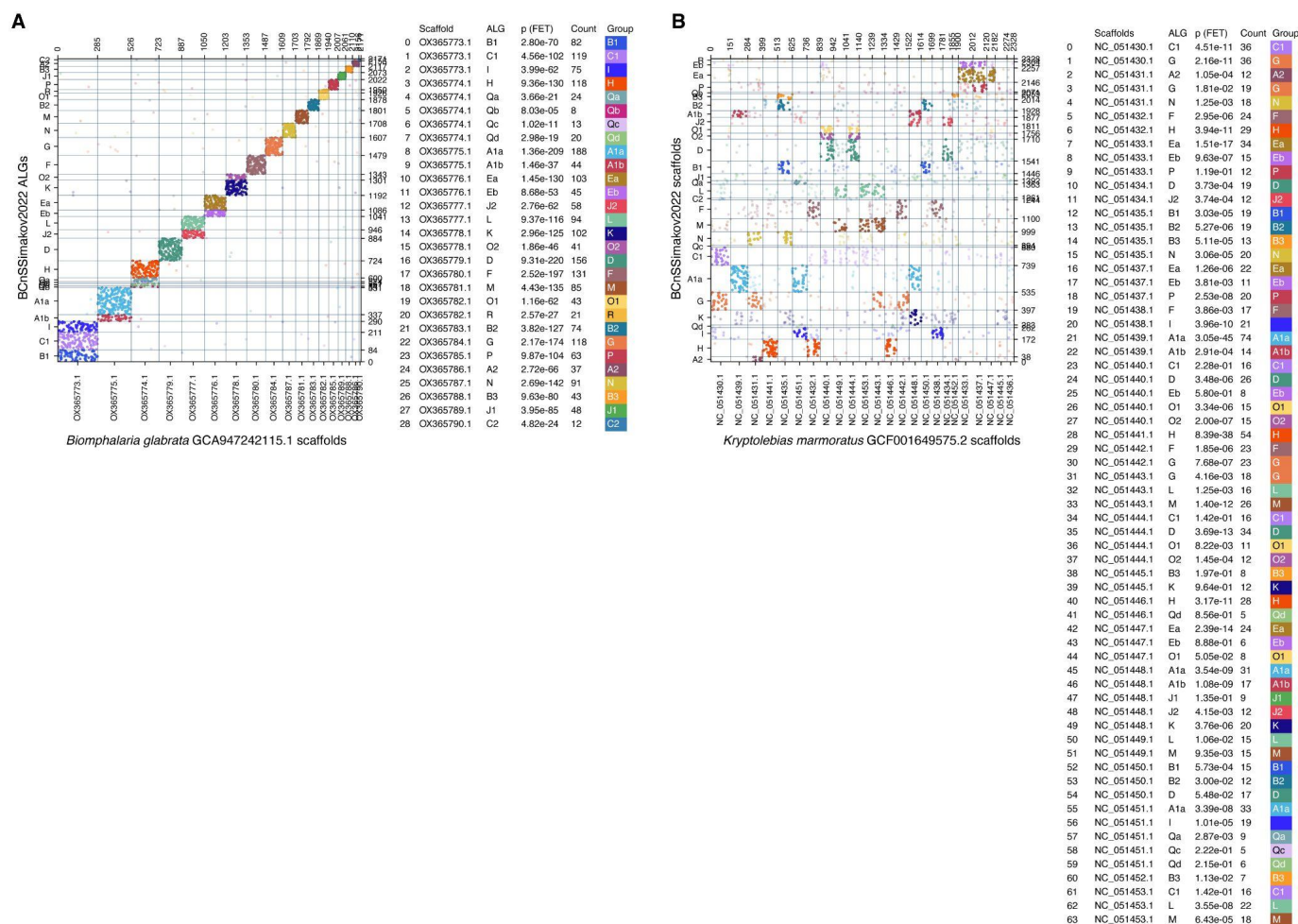
ETHICS DECLARATIONS

All authors declare that they have no competing interests.

Supplementary Information



Supplementary Figure 1: Two-species spiralian protein orthology dotplots. Oxford dotplots showing protein orthology shared between two species, identified with reciprocal best blastp hits by odp. Each axis shows the protein index along each chromosome of the depicted species. The dots are colored if they were identified by odp as being orthologous to a BCnS ALG ortholog. Black dots were orthologs found in these two species, but were not identified, or were a suboptimal hit for, BCnS ALG orthologs. **A.** The scallop *P. maximus* and the annelid *O. fusiformis* have a similar karyotype in which the BCnS ALGs are intact, predominantly on single chromosomes. **B.** The annelid *P. echinospica*, a more recently diverged relative of clitellate annelids, also has a karyotype in which the BCnS ALGs are intact. **C.** The BCnS ALGs have rearranged in the clitellates, shown here in the earthworm *P. excavatus*, and in **D.** the leech *H. verba*. **E.** Within clitellates the karyotype has changed significantly since the divergence of earthworms and leeches. **F.** Several chromosome fusion and fission events have occurred since the divergence of the leeches *Hirudo* and *Helobdella*.



Supplementary Table 1.

Hox inventories on syntenic genome scaffolds in three clitellate subgroups