Sea anemone genomes reveal ancestral metazoan chromosomal macrosynteny

Bob Zimmermann[1,2,#], Juan D. Montenegro[1,2,#], Sofia M.C. Robb[3], Whitney J. Fropf[3], Lukas Weilguny[1], Shuonan He[3], Shiyuan Chen[3], Jessica Lovegrove-Walsh[1], Eric M. Hill[3], Cheng-Yi Chen[3], Katerina Ragkousi[3,4], Daniela Praher[1], David Fredman[1], Darrin Schultz[1], Yehu Moran[1,5], Oleg Simakov[1,2], Grigory Genikhovich[1], Matthew C. Gibson[3]*, Ulrich Technau[1,2,6]*

[1]Department of Neurosciences and Developmental Biology, Faculty of Life Sciences, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria
[2]Research platform SinCeReSt, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria
[3]Stowers Institute for Medical Research, Kansas City, MO 64110 USA
[4]Department of Biology, Amherst College, Amherst, MA 01002 USA
[5]*Current address:* The Alexander Silberman Institute of Life Sciences, Faculty of Science, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel
[6]Max Perutz laboratories, University of Vienna, Dr. Bohrgasse 5, 1030 Vienna

[#] these authors contributed equally
*To whom correspondence should be addressed: ulrich.technau@univie.ac.at, mg2@stowers.org

Abstract

Draft genome sequences of non-bilaterian species have provided important insights into the evolution of the metazoan gene repertoire. However, there is little information about the evolution of gene clusters, genome architectures and karyotypes during animal evolution. In this regard, slowly evolving anthozoan Cnidaria, the sister group of Bilateria, are particularly informative. Here we report chromosome-level genome assemblies of two related cnidarians, the sea anemones *Nematostella vectensis* and *Scolanthus callimorphus*. We find a robust set of 15 chromosomes with a clear one-to-one correspondence between the two species. Both sea anemone genomes show remarkable chromosomal conservation with other cnidarians, several bilaterians and the sponge *Ephydatia muelleri*, allowing us to reconstruct ancestral cnidarian and metazoan chromosomal blocks, consisting of at least 19 and 16 ancestral linkage groups, respectively. We show that, in contrast to Bilateria, the Hox and NK clusters of investigated cnidarians are largely disintegrated, despite the presence of staggered *hox/gbx* expression in *Nematostella*. This loss of microsynteny conservation may be facilitated by shorter distances between cis-regulatory sequences and their cognate transcriptional start sites. In line with that, we find no clear evidence for topologically associated domains, suggesting fundamental differences in long-range gene regulation compared to vertebrates. These data suggest that large sets of ancestral metazoan genes have been retained in ancestral linkage groups of some extant lineages, yet, higher order gene regulation with associated 3D architecture may have evolved only after the cnidarian-bilaterian split.

Introduction

Comparative genomics and epigenomics have provided fundamental insights into the evolution of gene regulation in multicellular organisms. Among basally branching animals, this led to the surprising discovery of conserved gene repetoires[1–3], of microsyntenic gene clusters and patterns of DNA methylation and histone modification codes[4–6]. Recent applications of long-read sequencing technologies and proximity ligation methods such as Hi-C facilitated the genome assembly on chromosomal level of many organisms. These chromosome-level genome assemblies have presented the opportunity to compare the content and localization of homologous genes between distantly related species and the reconstruction of ancestral linkage groups. In vertebrates and several other bilaterians, topological analysis has revealed three-dimensional chromosomal architecture[7], organized as chromosomal compartments and topologically associated domains (TADs)[8]. TADs have been shown to act as boundaries of gene regulation[9–12]. Analyses on chromosome evolution and three-dimensional genome structure have mainly focused on bilaterians, and it remains unclear whether non-bilaterian animals share the topological signatures of bilaterian chromosomes. In this regard, studying Cnidaria, the sister clade to Bilateria, is a crucial step to understanding the evolution of animal genomes.

Cnidaria constitute a large clade of basally branching Metazoa, dating back between 590 and 690 Mya[13–15]. Their robust phylogenetic position as the sister group to Bilateria makes them the key group to study the evolution of bilaterian features, such as axis organization, mesoderm formation and central nervous system development[16]. The starlet sea anemone, *Nematostella vectensis* (Figure 1a - Hexacorallia; Actiniaria; Edwardsiidae), has been developed into an important model organism[17–19] and in 2007 became the first non-bilaterian animal to have a draft scaffold-level genome assembled[1]. The *Nematostella* genome revealed uncanny conservation of gene content to vertebrates as well as the first observations of large-scale macrosyntenic conservation by way of comparisons with human chromosomes to *Nematostella* scaffolds[1].

Since the release of the *Nematostella* genome, genomes of the representatives of all five cnidarian classes have become available[2,20–27] providing valuable insight into various aspects of the cnidarian gene complement and genome organization. However, these genomes originated from distantly-related species, and few genomic studies of Cnidaria have sought to search for genomic conservation signals of recently-diverged species, a cornerstone of comparative genomics[28]. For example, no genome sequence of an edwardsiid sea anemone has been reported, aside from that of *Nematostella*. Another sea anemone in the family Edwardsiidae is the "worm sea anemone" *Scolanthus callimorphus* (Figure 1b), dwelling in European intertidal zones[29,30], which according to our molecular clock calculations has separated from *Nematostella vectensis* approximately 174 Mio years ago, to date its closest sequenced relative (EDF1, see Materials and Methods for details).

Here, we report chromosome-level genome assemblies and new gene models for the model organism *Nematostella vectensis* and of the closely-related species *Scolanthus callimorphus*. We show that the macrosyntenic localization of orthologous genes across the chromosomes are highly conserved across many cnidarians and bilaterians, allowing us to reconstruct the minimal set of ancestral eumetazoan chromosomal linkage blocks. Notably, the macrosyntenic conservation contrasts with the loss of most of the microsyntenic blocks, including the integrity of the Hox cluster. Moreover, in *Nematostella*, which like all non-bilaterians lacks the TAD boundary protein CTCF[4,31], HiC analyses did not reveal any obvious TAD-like structures, as known in Bilateria. We propose that the proximity of most cis-regulatory regions to their target genes is an ancestral metazoan feature, whereas large-scale three-dimensional structures, such as TADs, only became necessary in bilaterians with larger genomes and more distantly located enhancers.

Results

High Quality Chromosome-Level Assemblies of Two Edwardsiid Genomes

Using short-read sequencing and a *k*-mer coverage model, we estimated the genome length of *Nematostella* at 244 Mb (EDF2), which is substantially shorter than previously suggested at 450 Mb[1]. This discrepancy could be in part attributed to the previous use of four haplotypes in sequencing, lower coverage and read length. The genome of the sea anemone *Exaiptasia pallida* is similar in length to *Nematostella*[32], while the estimated 414 Mb of the *Scolanthus* genome is at present the largest sequenced sea anemone genome (Figure 1c). Using PacBio long-read

sequencing and high-throughput conformation capture (Hi-C), we then assembled chromosome-level *Nematostella* and *Scolanthus* genomes, which surpass the quality of the original *Nematostella* genome in terms of contiguity, accuracy and mappability (see Supplementary Text, EDF2,3 for details).

Gene completeness as measured by alignment of single-copy pan-metazoan genes to the assemblies using the BUSCO method[33] was comparable to the previous assembly (EDF2i), however in order to completely assess this, we generated new gene models. Using a combination of IsoSeq and RNASeq data, we identified 24,525 gene models and 36,280 transcripts (see Supplementary Text for details). BUSCO analysis showed that the gene set was more complete than a previously-generated gene set[34]. In addition, the new gene models showed a better mappability to a publicly available single-cell RNA-Seq data set[35,36] (EDF4).

To facilitate the usage of the newly assembled genomes, we established a publicly accessible genome browser. Both new genome assemblies and associated data are available for browsing, downloading, and BLAST at SIMRbase (https://simrbase.stowers.org). The *Nematostella vectensis* genome assembly, referred to as Nvec200, has an abundance of aligned track data, including the newly generated gene models, a large collection of published RNAseq and ChIP-seq analyses, as well as 145 ultra-conserved non-coding elements (UCNEs) shared between *Nematostella* and *Scolanthus* (Supplementary Text; EDT8).

Comparison of *Nematostella* Chromosomes to Metazoan Genomes

We identified 15 chromosomes in the new genome assemblies of *Nematostella* and *Scolanthus*, and we numbered these chromosomes from 1 to 15 according to decreasing size (Figure 1d,e). This is in line with the previous estimates based on the number of *Nematostella* metaphase plates[1] and the analysis of *Nematostella* chromosome spreads[37]. Most chromosomes, according to their homologous pair, corresponded in length but are much larger in *Scolanthus* (Figure 4d). This was accounted for by a large fraction of unclassified and potentially lineage-specific repeat sequences (EDT1, Supplementary text). No obvious heteromorphic sex chromosomes could be identified based on read mapping depth, or from the level of heterochromatin usually associated with non-recombinant sex chromosomes. Ascertaining the sex-determining regions of the genome will require more data and detailed analyses.

Next we wished to determine the extent to which the *Nematostella* and *Scolanthus* chromosomes exhibit conservation of gene content and order (micro- and macrosynteny). Indeed, each of the 15 chromosomes of both species shared a majority of orthologous genes with a single corresponding chromosome in the other species (Figure 2a). We found that 8117 of 8692 mutual best BLAST hits between *Nematostella* and *Scolanthus* were retained on their respective chromosomes, implying a one-to-one homology between all 15 chromosomes. However, gene order was largely lost from the most recent common ancestor (MRCA), which we estimate to have diverged approximately 174 Mya (EDF1).

To assess the macrosyntenic conservation between anthozoans we compared the *Nematostella* chromosomes to those of the sea anemone *Exaiptasia pallida,* the stony coral *Acropora millepora* and the soft coral *Xenia sp*. (Figure 2a). While both *Exaiptasia* and *Acropora* are only assembled on the scaffold level, we observed that the gene content of the scaffolds suggest a similar karyotype to *Nematostella*. The chromosome-level genome assembly of the octocoral soft coral *Xenia sp.* also appears to have 15 chromosome-scale scaffolds[20]. However, despite the apparent identical number of chromosomes, the chromosome-scale scaffolds of *Xenia* showed surprisingly little homology to those of the hexacoral *Nematostella*. We discovered many translocations and fusions, with only 3 chromosomes demonstrating a one-to-one relationship (Figure 2a).

Notably, the hydrozoan *Hydra vulgaris,* which belongs to the clade of Medusozoa, also has 15 chromosomes[38]. While our analyses revealed many ancestral linkage groups in the comparison with *Hydra*, we also detected numerous chromosomal rearrangements, including splits and fusions (Figure 2a, EDF5). Specifically, we observe that *Nematostella* chromosomes 2 and 3 are partially or completely split in other cnidarian genomes. By comparison, the scyphozoan jellyfish *Rhopilema* has 21 chromosomes, supported with karyotype images[21,39]. We observed that all *Rhopilema* and *Nematostella* chromosomes show a clear 1-to-1, 1-to-2 or, in a single case, a 1-to-3 macrosyntenic correspondence, which can be traced to the ancestral linkage groups (EDF5,7). These comparisons among cnidarians suggest that the history of cnidarian chromosomes may be more complex than previously envisioned. This is in line with recent interpretations of cnidarian chromosome comparisons[40].

We then compared the chromosome-scale scaffolds of *Nematostella* with that of a bilaterian, the cephalochordate *Branchiostoma floridae*. Cephalochordates are early branching chordates lacking the two rounds of whole-genome duplication and allotetraploidization suggested for vertebrates[41,42], as represented by their single Hox cluster[43] (Figure 2a). Strikingly, *Branchiostoma* and *Nematostella* chromosomes retain extensive macrosynteny since their divergence from the bilaterian-cnidarian MRCA. The large-scale macrosynteny retention detected in *Branchiostoma* becomes less obvious once we focus on more recently-branching bilaterian clades. Vertebrates, such as the early branching teleost fish *Lepisosteus oculatus*[44] and humans, underwent two or more rounds of genome duplications and showed additional translocation events from the MRCA (Figure 2a). Among protostomes, we observed that the lophotrochozoan Ram's horn worm *Streplosbio benedicti* and the sea scallop *Patinopecten yessoensis*[45], as well as the ecdysozoan millipede *Trigoniulus corallinus*[46], retained many macrosyntenic links (Figure 2a).

The high degree of macrosynteny conservation with *Branchiostoma* motivated us to compare *Nematostella* chromosomes to those of the representative of an even older lineage, the Porifera (sponges), which branched off prior to the cnidarian-bilaterian split (EDF1) and is considered by many to be the sister group to all other animals[47,48]. Strikingly, the comparison with the recently-assembled chromosomes of the sponge *Ephydatia muelleri* revealed many shared ancestral linkage groups (Figure 2a).

Reconstruction of Ancestral Linkage Groups

The observed patterns of conserved macrosynteny between chromosomes of multiple distantly-related species and the varying degrees of gene shuffling and chromosomal rearrangements in others (e.g. *Drosophila, Caenorhabditis,* humans) prompted us to reconstruct the set of predicted ancestral linkage groups for the MRCA of cnidarians, bilaterians and metazoans (Figure 2b). First, we inferred gene orthology across all genomes investigated. For each of Metazoa, Bilateria and Cnidaria, we determined ancestral linkage groups by determining genes with high chromosomal linkage throughout each clade. Our method finds ALGs which maximize their "modularity", i.e. groups with high chromosomal linkage within ALGs and low linkage between ALGs, by evaluating multiple scenarios (see Methods and Supplemental Methods for details) to infer "ancestral linkage groups" (ALGs). We can then color the oxford plots with the genes in these blocks suggesting a pattern of splits and fusions in extant chromosomes from ancestral chromosomes.

We identified 17 ALGs present in the ancestor of cnidarians, although we could find similar modularity in scenarios of 16-18 ALGs (EDF14). These correspond well to the extant cnidarian chromosomes (Figure 2c, EDF5), however, while the *Rhopilema* and edwardsiid chromosomes appear highly representative of the ancestral cnidarian karyotype, the soft coral *Xenia* chromosomes appear to have undergone more chromosomal translocation events. Our results also indicate that although both Hydrozoa and Edwardsiidae have a clear 15 chromosome karyotype, their chromosomes originate from distinct fusions of cnidarian ALGs. For the ancestor of Bilateria, we identified 16 ALGs (Figure 2c). In a previous study 17 ALGs were proposed[14], which resulted in similarly high modularity in our analysis (EDF14). By comparison with the sponge *Ephydatia,* we reconstructed a minimum of 16 metazoan ALGs for the metazoan ancestor, which were maintained in the MRCA of Cnidaria, Bilateria and the sponge (Figure 2b,c).

In order to visualize how the chromosomes of extant species have undergone major splits and fusions, we projected the reconstructed ALGs of the respective cnidarian, bilaterian and metazoan ancestor to the chromosomes of the extant species. We found that many ALGs correspond to the same chromosomes across multiple species and lineages (Figure 2c; EDF5-7). To determine the extent of this, we further explored the relationships between metazoan ALGs and those of Cnidaria and Bilateria (Figure 2c). Remarkably, many ancestral chromosomes exhibited a 1-to-1 correspondence across all predicted ancestral lineages (Figure 2c) and carried through to the extant lineages (Figure 2d).

Chromosomal Organization of the NK and extended Hox gene clusters

The chromosome-level assembly of the *Nematostella* genome allowed us to address the evolution of specific gene clusters. Prominent examples of clusters of homeodomain transcription factor coding genes ancestral for Bilateria include the SuperHox cluster, the ParaHox cluster, the NK/NK-like cluster as well as *NK2* group genes located separately[43,49,50]. It has been hypothesized that all of them originated from a single gene cluster, which then disintegrated during evolution[50]. Our analysis revealed that *Nematostella* possesses a separate ParaHox cluster of two genes, (*Gsx* and *Xlox*/*Cdx)* on chromosome 10, and a SuperHox cluster on chromosome 2 containing *Hox*, *Evx*, *Mnx*, and *Rough*, as well as more distant *Mox* and *Gbx*[51] (Figure 3, EDF8, EDT6). We

identified an *NK* cluster on chromosome 5 containing *NK1*, *NK5*, *Msx*, *NK4*, *NK3*, *NK7*, *NK6*, a more distant *Lbx*, a possible highly derived *Tlx-like* gene and, intriguingly, *Hex,* which is also linked to the NK cluster in the hemichordate *Saccoglossus kowalevskii*[52] and in the cephalochordate *Branchiostoma floridae*. Similar to Bilateria, the *NK2* genes were clustered separately and found on the chromosome 2 (Figure 3, EDF8-9, EDT6). In contrast, in the earlier-branching sponges, neither ParaHox nor extended Hox cluster genes exist, and only the NK cluster is present with a single *NK2/3/4* gene, two *NK5/6/7* genes, an *Msx* ortholog, as well as possible *Hex* and *Tlx* orthologs[53], (Figure 3, EDF10). Taken together, this allows us to propose that the last common ancestor of Cnidaria and Bilateria possessed an NK-cluster on a chromosome different from the one carrying the SuperHox cluster, and a separate NK2 cluster, possibly on the same chromosome as the SuperHox cluster (Figure 3). The hypothesized SuperHox-NK Megacluster[43], if it ever existed, must have both formed and broken apart during the time after the separation of the sponge lineage, but before the origin of the cnidarian-bilaterian ancestor (Figure 3a, Supplementary Text). The lack of selection pressure in favor of microsynteny conservation is clearly illustrated by the comparison of the divergent Hox clusters of *Scolanthus* and *Nematostella*. Although located on homologous chromosomes, the gene order, orientation, and the number of intervening genes differs drastically between these two species (Figure 3b). In contrast to the atomized Hox clusters of *Nematostella* and *Scolanthus*, the Hox cluster of their very distant anthozoan relative *Xenia*, is compact and contains “anterior” Hox genes *HoxA* and *HoxB* immediately next to the non-anterior Hox gene *HoxE* (Figure 3b) Aside from the tandem duplications, this compact state likely represents the ancestral organization. Similarly, *HoxE* and *HoxB* are immediate neighbors in the genome of the jellyfish *Rhopilema*, although the *Rhopilema* Hox cluster shows some evidence of disintegration (Figure 3b).

Topologically associating domains are not detected in either sea anemone genome

In the past decade, high-resolution chromosome conformation capture has increased interest in topologically associating domains (TADs), recurring chromosomal-folding motifs evidenced by signals in Hi-C contact maps[54]. Flanking regions of TADs are positively correlated with CCCTC-binding factor (CTCF) binding sites. Interestingly, no CTCF ortholog has been detected in non-bilaterian animals[31,54], but previous studies of non-bilaterian animals have proposed the existence of TAD-like structures[47].

Similar to previous studies, we sought to identify putative TADs within the edwardsiid genomes. Initial inspections of the Hi-C maps of both *Scolanthus* and *Nematostella*, however, revealed that both species appeared to lack TAD-like structures (Figure 4a; EDF11) and instead the Hi-C contact frequency decayed smoothly as a function of read pair distance along the chromosomes (EDF11). We quantified these findings by measuring the strength of topological boundaries, measured by insulation score, and found the topological boundaries to be weaker in the *Nematostella* genome compared to the fly (EDF12, $p = 8.38\times10^{-18}$, Mann-Whitney *U* test) and mouse genomes ($p = 2.93\times10^{-55}$, Mann-Whitney *U* test). We also verified that there was a lack of bilaterian-like TADs and weak topological domain boundaries in an independent *Nematostella* chromosome-scale genome assembly and Hi-C dataset produced in parallel to ours (EDF11[55]).

TAD structures are often well-conserved around gene clusters that control spatio-temporal gene expression, especially of developmental regulators, such as the TADs that separate the anterior and posterior hox genes in the mouse HoxD cluster[56]. We therefore explored whether TAD-like structures existed around Hox clusters in *Nematostella*. *Nematostella* has 6 Hox genes, one is on chromosome 5, while the other 5 Hox genes are located on chromosome 2, but in two clusters, separated by hundreds of genes between (Figure 4). We analyzed the genomic regions surrounding both partial hox clusters but again we could not detect any significant TAD-like structure.

Correlation of gene regulation with three-dimensional genome organization

We note that the *Nematostella* genome is fairly gene-dense (10±4 genes per 100kb), leaving relatively little intergenic sequence. It is conceivable that most relevant cis-regulatory elements are located in close proximity to the regulated gene. This is supported by the observation that many *Nematostella* transgenic reporter constructs faithfully mimic endogenous expression patterns with only 1-3 kilobases of upstream promoter sequence. Cis-regulatory regions are reflected by ATAC-seq peaks, which detects open chromatin. We therefore wished to assess whether there is a correlation between the distance of ATAC-seq peaks and the closest gene (i.e. transcriptional start or end site) and the genome size. We compared several species, where chromosome assemblies and ATACseq data are available and - as expected - we found that there is indeed a correlation between distance and of ATAC-seq peaks to the next gene and the genome size of the organism. However, there is also a correlation between small genomes and the absence of clear TAD structures, for instance in *C. elegans*, sponges and cnidarians. Notably, at least one cnidarian genome, *Hydra*, is relatively large (1.27 Gb), yet does not show typical TAD structures, although smaller-scale 3D structures could be observed[57]. Thus, within bilaterians, species with small genomes may show a tendency to lack TAD structures, while in non-bilaterians, which lack CTCF, they may not organize into stable TAD-like structures. At present, *Hydra* is the only available large non-bilaterian genome with relatively short distances between ATAC-seq peaks to the next gene.

Since TADs not only assure maintenance and directionality of enhancers to the regulated gene, they frequently also encompass more than one gene. In vertebrates, many TADs are conserved between species, suggesting that their maintenance is under selection pressure. This predicts that in organisms with TADs the number of microsyntenic genes is higher than in organisms lacking TADs. To test this, we compared the number of microsyntenies in different cnidarians, chordates and protostomes, roughly representing similar times of divergence. We find that, while the amount of microsyntenies consisting mostly of two consecutive orthologous genes did not show any difference between clades, microsyntenies consisting of at least three or four orthologous genes were reduced in cnidarians. No microsyntenies of four and more genes were found in any pair of species spanning the anthozoan-medusozoan split (e.g., *Nematostella*-*Rhopilema*), whereas around a dozen such regions are shared among distantly branching spiralian and chordate-vertebrate ("deuterostome") genomes (e.g., 11 blocks between the scallop *Pecten maximus* bilaterian and the annelid *Streblospio benedicti*, and 14 blocks for amphioxus *Branchiostoma floridae* to chicken *Gallus gallus* split) (Figure 4f). A similar trend was observed

for microsyntenies where intervening genes were allowed to be part of the blocks. For a commonly used threshold of up to 5 intervening genes[58], *Nematostella-Rhopilema* species pair showed 7 blocks, whereas *Pecten-Streblospio* had 82, and amphioxus to chicken had 45 microsyntenic blocks retained (Figure 4f, EDT12). Together, this data indicates higher retention and evolutionary exploration of longer syntenic regions in bilaterian genomes, potentially allowed for by the maintained distal topological architecture.

Discussion

The assembly of two high quality, chromosome-level edwardsiid sea anemone genomes has illuminated several intriguing aspects about chromosomal evolution, the NK and extended Hox clusters, the conservation of non-coding elements and the status of topologically associated domains in the common ancestor to cnidarians and bilaterians. In addition, the highly improved *Nematostella* genome and manually curated gene annotations will prove to be an invaluable resource for future studies of both coding and non-coding regions, structural variants among populations and continued development of functional tools for this model organism.

Nearly all members of the extended Hox cluster were distributed among distant, isolated microsyntenic blocks on chromosome 2 of *Nematostella* (pseudo-chromosome 4 in *Scolanthus*), with the single exception of *HoxF/Anthox1*, located on chromosome 5 (Figure 3; Supplementary Text). This indicates a lack of proximity constraint on the Hox genes in Cnidaria, contrasting with the situation in Bilateria. In addition, while a staggered spatiotemporal pattern of Hox expression along the secondary, directive axis of the *Nematostella* larva and polyp can be observed[59], unlike Bilateria, there is no correlation between expression and cluster position[60]. Notably, *HoxF/Anthox1* is not only located on a different chromosome but it is also the only Hox gene expressed in the (aboral) ectoderm, while all other Hox genes are expressed in the inner endomesodermal cell layer[61,62].

The dispersed NK and extended Hox clusters may be due to the diminishment or, possibe absence of higher-order chromosome organization at the level of microsynteny. In line with this, it was recently observed that the HoxD cluster boundaries in the mouse genome are marked by two TAD boundaries[56], and the cluster's intra-TAD gene order is deemed to be under selective pressure[63]. The lack of a CTCF gene in the *Nematostella* genome led us to hypothesize that the cnidarians might lack TADs, as TAD presence has been attributed to the appearance of CTCF[63]. Moreover, CTCF is absent not only in cnidarians but also in earlier branching ctenophores and sponges, which provides a possibility that the existence of TADs might represent a bilaterian-specific feature. While we were unable to detect any noticeable structure resembling the current definitions of TADs, it remains an open question as to whether larger or smaller structures, beyond the resolution of our data, could yet be detected. One study has suggested evidence for the higher-order chromosomal organization in a non-bilaterian, the sponge *Ephydatia*[47]. However, the contact maps resemble patterns we observe in our assemblies at the boundary of scaffolds or contigs, which can be the result of differential mappability from repetitive content or assembly issues. We therefore deliberately do not report any results from a TAD finder, since, after multiple rigorous rounds of manual assembly update, we can assert that the data we have generated do not qualitatively represent TAD boundaries per se, and most results would be likely false positives. While the precise definition of a TAD is still evolving[64,65], both data sets lack many characters of TADs identified in CTCF-containing genomes: hierarchical compartments, mammalian-specific "corner peaks" indicating strong interactions, and in our case, loop peaks and inter-contig compartments. This suggests that the presence of CTCF is necessary for the formation of TADs. The apparent absence of TADs in our analysis could also be explained by a higher degree of cell type variability. Therefore, we cannot exclude the possibility that performing the experiment with

a more homogenous cell population, or sequencing at a higher resolution, would reveal a signal on a smaller scale.

While microsynteny analyses reveal little conservation of the local gene order in the genomes of *Nematostella* and *Scolanthus*, macrosyntenic analysis of the edwardsiid chromosomes compared to available cnidarian genomes revealed a high level of conservation. We identified a stable set of 19 ALGs across all clades of sequenced cnidarian genomes. When compared to extant genomes, we can trace a small number of recombination events from the ALGs since the common cnidarian ancestor split an estimated 580 Mya. This stands in stark contrast to the history of, for example, the 326 Mya old ancestral genome of Amniota, which is estimated to have 49 distinct units, while the karyotypes of the extant amniote taxa consist of multiple translocated segments and variable chromosomes[66]. However, far more remarkable is the macrosynteny maintained between the edwardsiids, the early branching chordate *Branchiostoma*, and the sponge *Ephydatia*. Our analyses revealed clear one-to-one, one-to-few or few-to-one conservation of the chromosome-level linkages between cnidarians, sponges and early chordates, which suggests a striking retention of macrosyntenies throughout evolution of these animal lineages. This in turn allowed us to deduce a set of 16 ALGs of the last common metazoan ancestor, which was maintained in the cnidarian-bilaterian ancestor and gave rise to the 19 ancestral cnidarian chromosomes and the 17 ancestral bilaterian chromosomes[58]. The observed conservation of macrosyntenic linkages can result from a strong selective pressure to maintain intact chromosomes during meiosis, so that only local translocations within the chromosomes, scrambling local gene order, are favored. Changes in population sizes or asexual reproduction or selfing may facilitate chromosomal unit breakages. While such global chromosomal rearrangements have been observed for some clades (most notably dipterans but also nematodes and cephalopods), it still remains unclear whether additional gene regulatory constraints may exist at the whole-chromosomal level[67]. It is tempting to speculate that the emergence of the TADs in Bilateria may have restricted local rearrangements and at the same time released the constraints on maintaining the ancestral macrosyntenies conserved all the way back to the origin of multicellular animals.

The high level of macrosynteny maintaining ancestral chromosomal blocks over hundreds of millions of years contrasts with the low level of microsynteny conservation, even between more closely related cnidarians. Despite this reshuffling of genes within chromosomes, many developmental regulators involved in axial patterning show a conserved expression pattern, suggesting that their cis-regulatory elements have been maintained. For instance, in *Nematostella* and *Hydra*, which are separated over 500 Million years, all 12 Wnt genes, *brachyury,* and *foxA* are expressed at the oral pole, while *foxQ2* and *six3* are expressed aborally[68–71]. To ensure conservation of spatio-temporal gene expression patterns cis-regulatory elements must remain located in close proximity to the respective gene even when translocated to a new genomic position. This is supported by our distance analyses of ATACseq peaks. By contrast, bilaterians with large genomes, such as vertebrates, enhancers can be located up to several megabases away from the gene they regulate, often with several genes in between. To maintain the gene regulation by long distant enhancers, we predicted that the microsynteny conservation should be under higher selective pressure than in non-bilaterians. Indeed, microsynteny conservation is

higher in vertebrates than in cnidarians, in line with recent comparisons between skates, mouse and garfish[72]. We therefore propose a hypothetical evolutionary scenario, where close cis-regulation is ancestral to metazoans and maintained in non-bilaterians, which lack CTCF. It follows that CTCF evolved in the bilaterian common ancestor and was recruited to bind to the genome to act as an insulator and, together with cohesin, as a structural component defining recombination boundaries. This kept complex gene regulation by distant enhancers as a physical boundary for recombination events, but also acted as an insulator in TADs. Bilaterians that evolve a more compact genome would have a tendency to lose TADs and the dependence on CTCF. This seems to be the case in the *C. elegans* autosomes[73,74] and is also predicted for other small genomes, such as in urochordates.

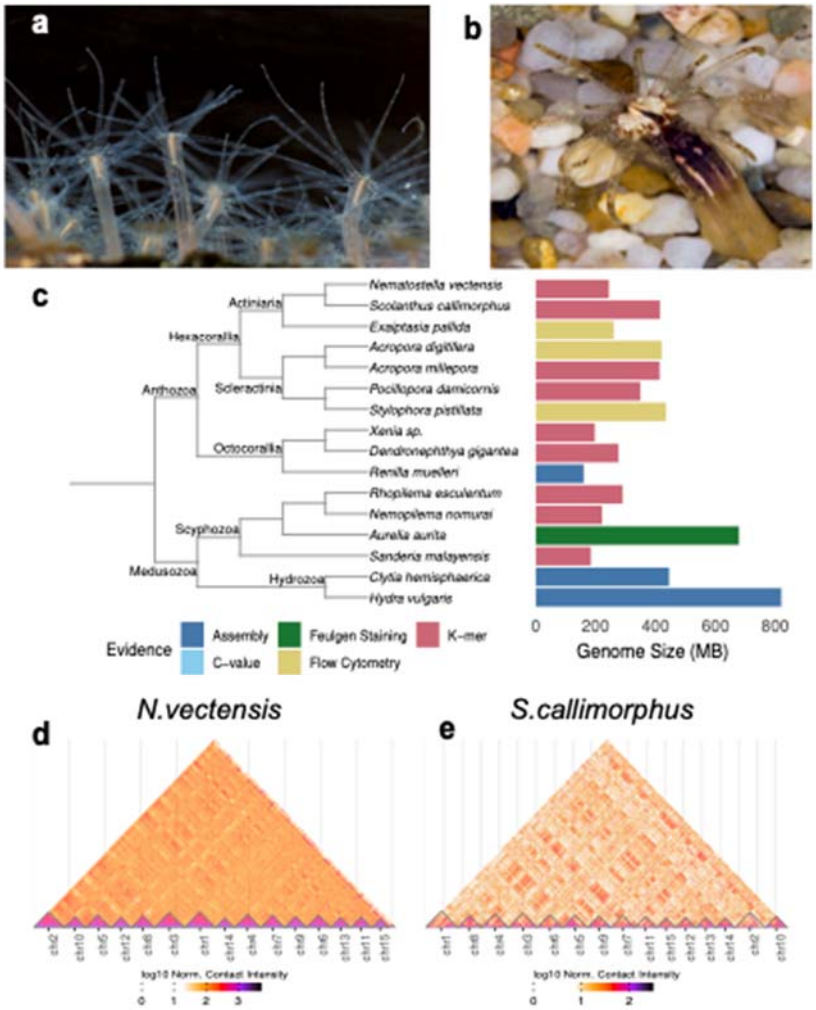


Figure 1. a) *Nematostella vectensis*. photo credit: Patrick RH Steinmetz b) *Scolanthus callimorphus*. photo credit: Robert Reischl c) Genome size estimates of cnidarian genomes taken from[2,21,22,24–26,32,39,75–79].

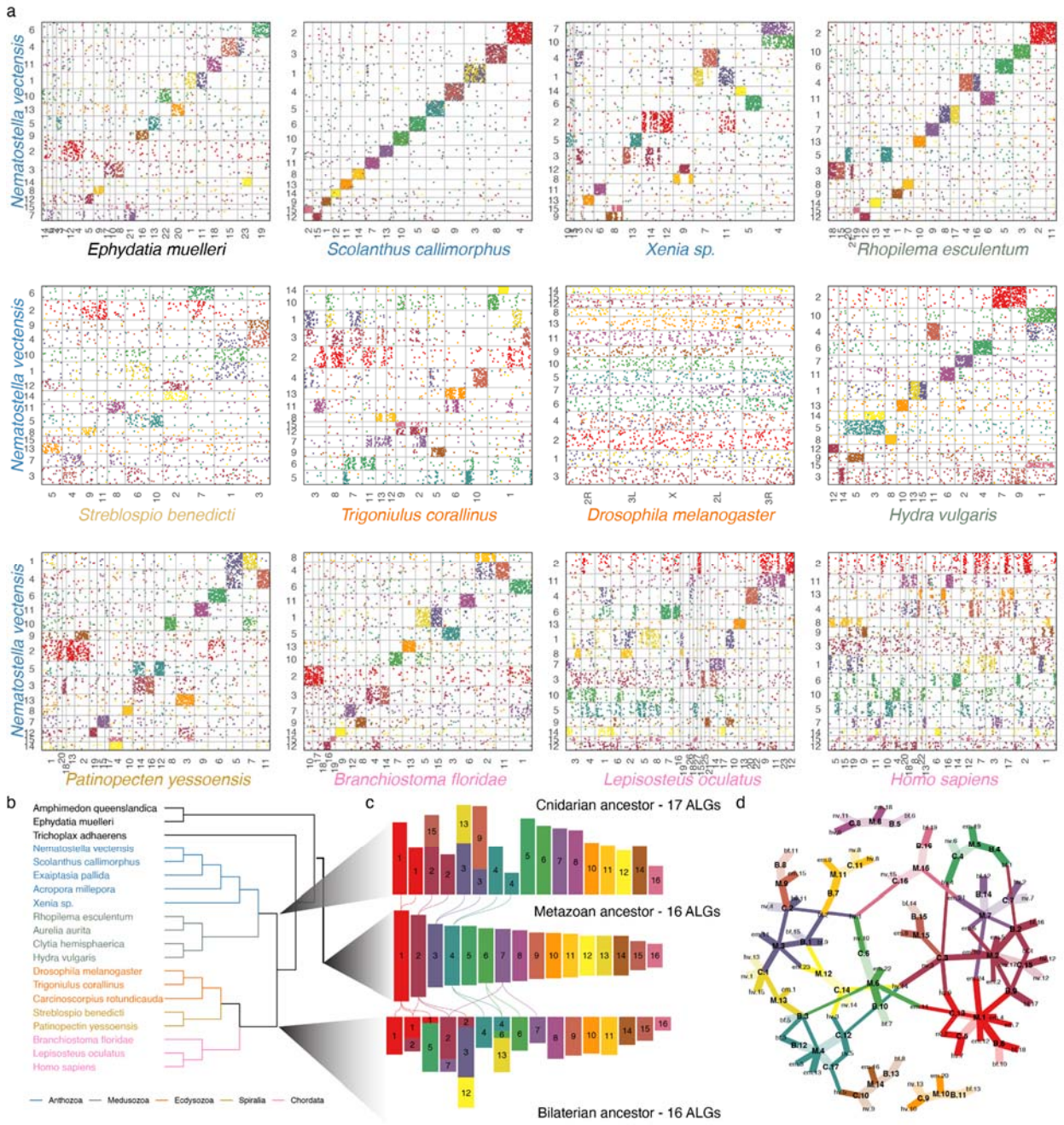


Figure 2. a) Oxford plots of the macrosyntenic relationships of *Nematostella vectensis* (*y*-axis) to genomes of various metazoans (*x*-axis). Pairwise orthologous are ordered by their position on the chromosome labeled on the axis. Dots (genes) are colored according to their membership to the metazoan ancestral linkage groups (ALGs) shown in (c). Species names are colored according to clade shown in (b). b) Genomes represented in this figure and EDF5-7. c) Relationships between the ALGs colored according to the metazoan ALG. Lines are drawn between ALGs to represent fissions. d) Graph representing the relationships of metazoan (M), bilaterian (B) and cnidarian (C) ALGs to each other and extant genomes of *Ephydatia* (em), *Nematostella* (nv), *Rhopilema* (re) and *Branchiostoma* (bf). Edges are colored by the metazoan ALG in (b) from which its path has a source. Line width represents the fraction of the extant genome or derived ALG (bilaterian or cnidarian) content of the ancestral genome. Faded edges represent near one-to-one relationships (> .8).

477

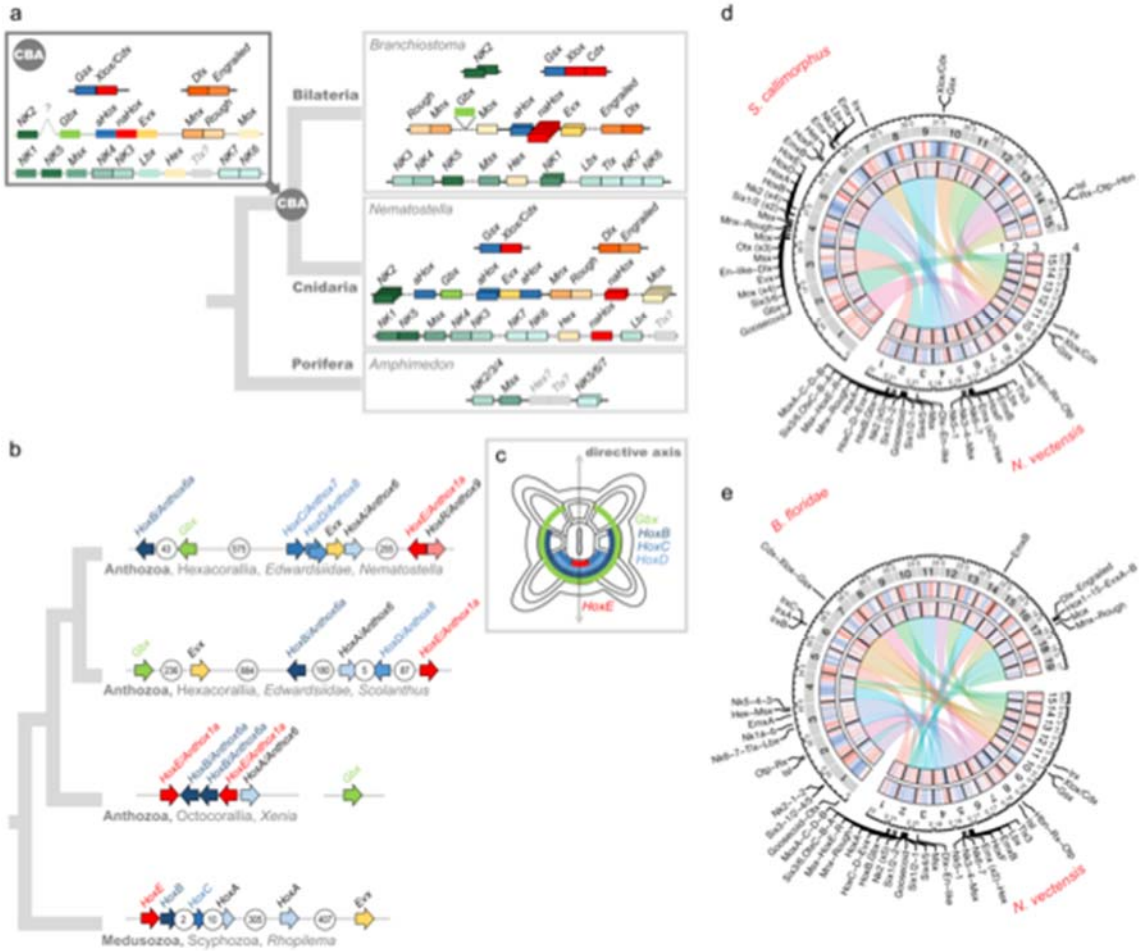

a
CBA
Bilateria
Cnidaria
Porifera
Branchiostoma
Nematostella
Amphimedon
b
Anthozoa, Hexacorallia, Edwardsiidae, Nematostella
Anthozoa, Hexacorallia, Edwardsiidae, Scolanthus
Anthozoa, Octocorallia, Xenia
Medusozoa, Scyphozoa, Rhopilema
c
directive axis
HoxB
HoxC
HoxD
HoxE
d
S. callimorphus
N. vectensis
e
B. floridae
N. vectensis

Figure 3. Evolution of a selection of Antennapedia class homeobox gene clusters. a) Composition of the extended Hox, ParaHox and NK clusters in the sponge *Amphimedon*, the sea anemone *Nematostella*, in a chordate *Branchiostoma*, and the deduced cluster composition of the cnidarian-bilaterian ancestor (CBA). Greyed-out genes with question marks have uncertain orthology. Genes shown as borderless boxes have an uncertain position relative to neighboring cluster members. Stacked boxes represent clusters of paralogs of the indicated ancestral gene. Genes in immediate proximity are indicated by abutting boxes. Linked genes of the same class separated by 1 to 50 intervening genes are connected with solid lines, over 50, with dashed lines. Grey intergenic connectors in the CBA indicate that the distances and the number of the intervening genes between the cluster members cannot be estimated. *NK2* of the CBA may be linked to the extended Hox cluster. Since *Branchiostoma Gbx* remained unplaced in the chromosome-level assembly, its position was taken from the scaffold-level assembly of *Branchiostoma lanceolatum*[80]. A two-gene ParaHox CBA scenario is shown although a three-gene ParaHox CBA scenario is possible based on evidence from Scyphozoa[21]. b) Organization of the *Nematostella* Hox cluster in comparison to the Hox clusters of *Scolanthus*, the octocoral *Xenia* and a scyphozoan jellyfish *Rhopilema* indicates loss of microsynteny. c) Staggered expression of *Gbx* and Hox genes along the directive axis of *Nematostella* (oral view) partially reflects the position of the genes on the chromosome. Arrows show the direction of transcription for each of the genes. The number of intervening genes is indicated in white circles. (d-e) Chromosomal relationships, genomic content and locations of NK and extended Hox cluster and other landmark homeobox genes of (d) *Nematostella* and *Scolanthus* and (e) *Nematostella* and *Branchiostoma*. 1) Chord diagram of macrosyntenic relationships of chromosomes based on the inferred ancestral linkage groups. Chords represent the proportion of the chromosome's gene content in common with the compared genome, not the positional information. 2) Scaled and centered gene density relative to the respective genome. Red is high and blue is low. 3) Scaled and centered density of interspersed repeat elements relative to the respective genome. Red is high and blue is low. 4) Locations of the landmark homeobox genes.

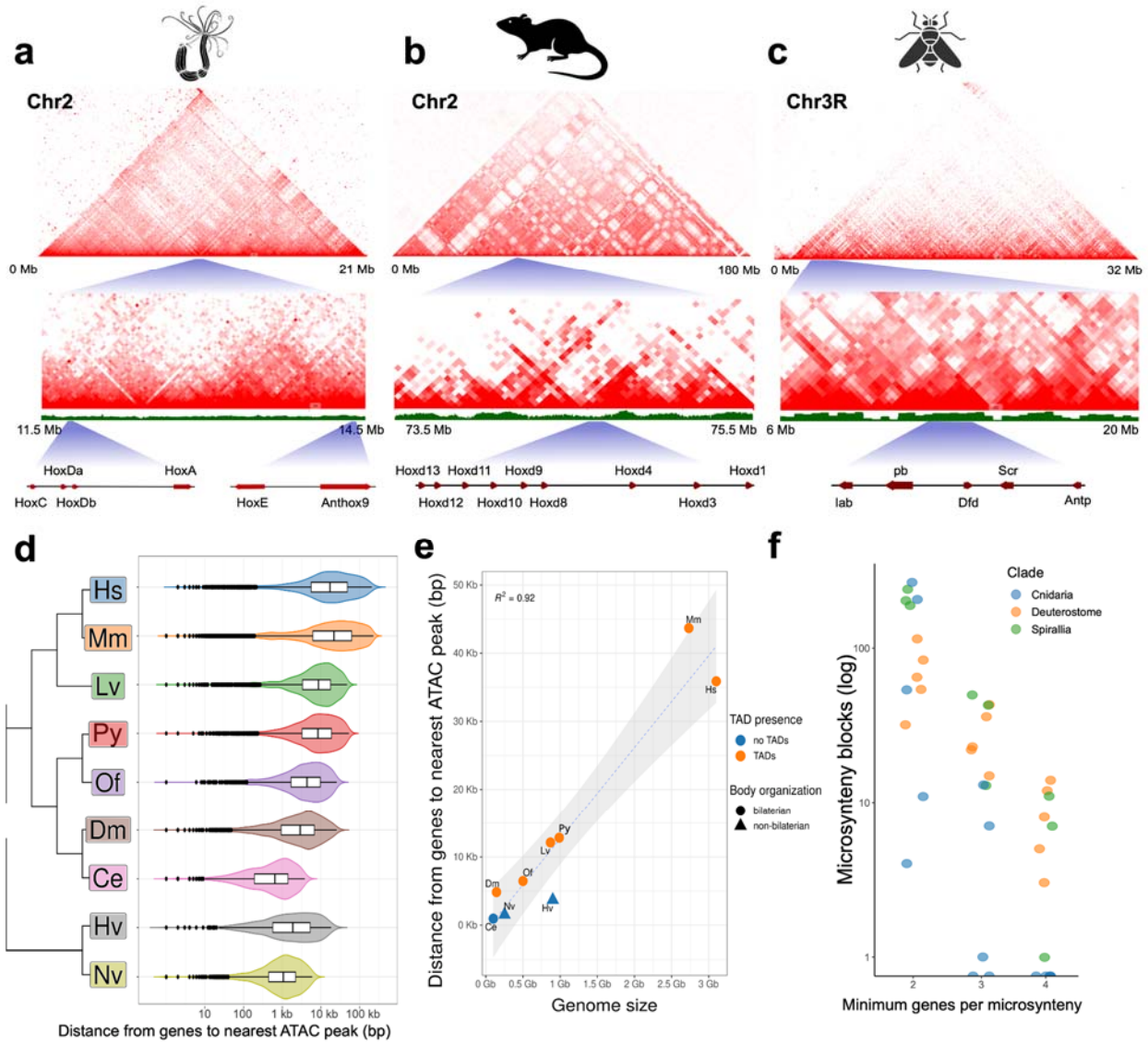


Figure 4. Hi-C contact maps. Contact maps (a) Nematostella Hox clusters, (b) Mouse HoxD cluster and (c) Fly Hox cluster show striking differences in the patterns of intrachromosomal interactions; d) distribution of the distance from ATACseq peaks to their closest peak for human (Hs), mouse (Mm), Sea urchin (*Lytechinus variegatus*, Lv), mussel (*Patinopecten yessoensis*, Py), polychete (*Owenia fisuformis*, Of), fly (Dm), *Caenorhabditis* (Ce), *Hydra vulgaris* (Hv) and *Nematostella vectensis* (Nv); e) correlation of mean gene-ATACseq peak distance and genome size for the same species; f) number of microsyntenic blocks of 2, 3 and 4 genes in Cnidarians, Deterostomes and Spiralians. The number of syntenic blocks falls rapidly in Cnidarians compared to Deuterostomes and Spiralians as the synteny length increases.

Acknowledgements

We thank Matthew Nicotra for providing us with the HMW DNA extraction protocol used for *Scolanthus.* We thank Robert Reischl for the photo of *Scolanthus* and Patrick R.H. Steinmetz and Hanna Kraus for the photo of *Nematostella vectensis* (Figure 1). Special thanks to Tatiana

Lebedeva for the cartoon drawings of animals used in this study. We are grateful to the Stowers Institute Molecular Biology Core facility, particularly Amanda Lawlor, Michael Peterson and Anoja Perera. This work was supported by grants of the Austrian Science Fund FWF (P24858; P21108) to U.T., support from the Stowers Institute for Medical Research to M.G. and an NIH Ruth L. Kirschstein NRSA (F32 GM131522) to E.M.H.. We are also grateful for the support of the CNRS Marine Station in Roscoff and the Assemble grant 227799 to U.T. for collecting *Scolanthus*.

Data Availability

All raw data is available via the National Center for Biotechnology Information under the accession PRJNA667495. The assembled genomes can be downloaded, browsed and searched on publicly available browsers at https://simrbase.stowers.org/starletseaanemone and https://simrbase.stowers.org/wormanemone. Code used to generate the analyses is available from the authors upon request. Publicly available data that were used in this manuscript include SRR1771322-SRR1771324 (mouse Hi-C), SRR10512944 (*D. melanogaster* Hi-C), and ERR8571699 (Wellcome Sanger *N. vectensis* Hi-C).

Materials and Methods

Animal Care and Source

Nematostella vectensis animals were cultured as previously described[81] at the University of Vienna and the Stowers Institute. Adult male and female individuals were verified by induction in isolation. *Scolanthus callimorphus* animals were collected at the Île Callot, Carantec, France. After transport, they were kept in seawater at 20°C and fed freshly hatched *Artemia salina* weekly or biweekly.

Sequencing

Short Read DNA-Seq

Genomic DNA samples were extracted from both adult male and female individual *Nematostella* adults using the DNeasy Blood and Tissue Kit (Qiagen). After purification, approximately 5 ug of genomic DNA was recovered from each sample. Following DNA extraction, samples were sheared and size selected for ~500 bp using a Blue Pippin Prep machine (Sage Science). Following size selection, sequencing libraries were created using a KAPA HTP Library Prep kit (Roche) and subjected to paired-end sequencing on an Illumina NextSeq 500. *Scolanthus* DNA samples for library preparation were aliquoted from high molecular weight extractions, described below.

High Molecular Weight DNA Extraction and Library Prep

Nematostella high molecular weight DNA was extracted at Dovetail Genomics. Samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20kb) for PacBio Sequel was constructed using SMRTbell Template Prep Kit 1.0 (PacBio, Menlo Park, CA, USA) using the manufacturer recommended protocol. The

pooled library was bound to polymerase using the Sequel Binding Kit 2.0 (PacBio) and loaded onto PacBio Sequel using the MagBead Kit V2 (PacBio). Sequencing was performed on the PacBio Sequel SMRT cell, using Instrument Control Software Version 5.0.0.6235, Primary analysis software Version 5.0.0.6236 and SMRT Link Version 5.0.0.6792, yielding 24.67 Gb over 3,050,403 subreads.

High molecular weight DNA from a single *Scolanthus callimorphus* adult animal was extracted using a modified Urea-based DNA extraction protocol[82,83]. A whole animal was flash frozen and ground with mortar and pestle. While frozen, drops of buffer UEB1 (7M Urea, 312.5 mM NaCl, 50 mM Tris-HCl pH 8, 20 mM EDTA pH 8.1% w:v N-Lauroylsarcosine sodium salt) were added and crushed with the tissue. Tissue was incubated in a final volume of 10 mL UEB1 at RT for 10 minutes. Three rounds of phenol-chloroform extraction were performed followed by DNA precipitation by addition of 0.7 volume isopropanol. The pellet was transferred to a fresh tube and washed twice in 70% EtOH and twice more in 100% EtOH, dried, and resuspended in TE buffer.

A library for PacBio sequencing was then prepared from the high molecular weight sample using the SMRTbell® Express Template Prep Kit v1. The libraries were then sequenced on a PacBio Sequel machine over 3 SMRT Cells, yielding a total of 22.85 Gb over 1,474,285 subreads. An aliquot of the same sample was used to prepare a library using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina. This was then subjected to 50 cycles of single-end sequencing in one flow cell lane using an Illumina HiSeq 2500 system.

Chicago libraries

2 Chicago libraries were prepared as described previously[84]. For each library, ~500 ng of HMW gDNA (mean fragment length = 100 kbp) was reconstituted into chromatin in vitro and fixed with formaldehyde. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq 2500 (rapid run mode). The number and length of read pairs produced for each library was: 116 million, 2x101 bp for library 1; 35 million, 2x101 bp for library 2. Together, these Chicago library reads provided 125 x sequence coverage of the genome (1-100 kb pairs).

Chromatin was extracted from a single *Nematostella vectensis* adult male and *Scolanthus callimorphus* adult (unknown sex) nuclei using the Phase Genomics Proximo Hi-C animal protocol. After proximity ligation and purification, 16 ng and 9 ng of DNA was recovered, respectively. For library preparation 1 µl of Library Reagent 1 was added 12 PCR cycles were performed. The final library was subjected to 150 total cycles of paired-end sequencing using an Illumina NextSeq 550 machine yielding a total of 13.5 gigabases.

Hi-C sequencing, *Scolanthus callimorphus* PacBio library preparation and sequencing, Scolanthus Illumina DNA library preparation and sequencing and adult *Nematostella vectensis* RNA library preparation and sequencing was performed at the VBCF NGS Unit (https://www.viennabiocenter.org/facilities). *Nematostella vectensis* DNA size selection, library preparation, and sequencing were performed by the Molecular Biology Core at the Stowers Institute for Medical Research.

Developmental and adult *Nematostella* RNA sequencing was performed as follows. *Nematostella* were spawned and eggs were de-jellied and fertilized as previously described[81]. Spawning and embryo development took place at 18°C. Eggs and embryos from different stages were collected (300 per sample) in duplicate as indicated: eggs (within 30min of spawn), blastula (7.5hpf), gastrula (23.5hpf) and planula (72hpf). Eggs and embryos were collected in eppendorf tubes and centrifuged to a pellet at 21,000 x g for 1 min. All seawater was quickly removed and pellets were resuspended in 150ml lysis buffer (RLT buffer supplied by the Qiagen RNeasy kit (#74104), supplemented with □-mercaptoethanol). The samples were homogenized with an electric pestle (1 min continuous drilling) and further supplemented with 200 ml of the above lysis buffer. Homogenized samples were then transferred into QIAshredder columns (Qiagen #79654) and centrifuged at 21,000 x g for 2 min. The flow throughs were supplemented with 1 ml 70 % ethanol and transferred to RNeasy columns and were processed according to the Qiagen RNeasy protocol. Quality and integrity of the RNA was evaluated using the Agilent RNA 600 pico kit (Agilent Technologies) and RNA samples were stored at -80°C until further processing. cDNA libraries were then constructed for polyA stranded sequencing. The resulting libraries were sequenced on Illumina HiSeq using paired end runs (RapidSeq- 2x150bp).

Genome Assembly

Size estimates for *Nematostella vectensis* and *Scolanthus callimorphus* were derived using Genomescope[85], taking the result of the highest *k* (56 and 18) which converged under the model.

Initial assemblies based on PacBio sequencing of *Nematostella* and *Scolanthus* were generated using canu version 1.8[86] with the parameters `rawErrorRate=0.3 correctedErrorRate=0.045`.

Nematostella haplotigs were removed using Purge Haplotigs[87]. First, the source PacBio reads were aligned onto the canu assembly using minimap2[88] using the parameters `-ax map-pb --secondary=no`. Following this a coverage histogram was generated using the Purge Haplotigs script `readhist`. Per the documented Purge Haplotigs protocol, lower, mid and high coverage limits were found by manual inspection of the plotted histogram to be 12, 57 and 130 respectively. All initial contigs marked as suspect or artifactual were removed from further analysis with the Purge Haplotigs script `purge`.

Due to lower sequencing coverage of *Scolanthus*, diploid per-scaffold coverage could not be deconvolved from haploid, and therefore Purge Haplotigs could not be used. Removal of redundant contigs was performed with Redundans version 0.14a[89] using the parameters `--`

noscaffolding --nogapclosing --overlap 0.66. Only contigs marked in the reduced version of the genome were used in further analysis.

The input de novo assembly, shotgun reads, and Chicago library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assembly[84]. Shotgun and Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper (http://snap.cs.berkeley.edu). The separations of Chicago read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold. After scaffolding, shotgun sequences were used to close gaps between contigs.

Repetitive DNA and Hi-C scaffolding

Repetitive DNA was found using two strategies. First, known repeats found in repbase[90] were searched in the assemblies using RepeatMasker[91] using the parameters -s -align -e ncbi in addition to -species nematostella for *Nematostella* and -species edwardsiidae for *Scolanthus*. Second, novel repeat sequences were found using RepeatModeler version 2.0[92]. After generating the repeat library, genomes' repeat regions were detected with the corresponding library using the same parameters in RepeatMasker.

Hi-C sequences were aligned to the reduced and repbase masked genomes of *Nematostella* and *Scolanthus* using bwa mem[93,94] using the parameters -5SP. For *Nematostella,* an additional candidate assembly was generated by mapping Hi-C sequences to the Chicago library scaffolded sequences using repbase masking (dovetail_standardmask) in addition to the contig-based scaffolding (contig_standardmask). Duplicate reads were marked with the samblaster utility[95], and duplicate, secondary and supplementary mappings were removed with samtools. These mappings were used to generate initial chromosomal assemblies using Lachesis[96], specifying the restriction site GATC. Assemblies were manually reviewed using Juicebox Assembly Tools version 1.11.08[97]. Candidate assemblies were compared using the nucmer aligner with default parameters and visualized using mummerplot[98]. Assemblies were converted over to Juicebox format using juicebox_scripts (https://github.com/phasegenomics/juicebox_scripts). In the case of *Scolanthus*, duplicate regions were clipped, and the resulting contigs were subjected to another round of alignment, assembly and review.

Nematostella scaffold correctness was assessed using REAPR[99]. *Nematostella* assembly nemVec1 was downloaded from the JGI website[1]. Sequences from the adult male and adult female (see Sequencing) were aligned to nemVec1 and the *Nematostella* genome after scaffolding with Chicago libraries using SMALT as well as the REAPR tool perfectmap using an expected insert size of 400, as determined from fragment analysis. Error-free bases and contiguity after breaking the genome were extracted from the results.

Genome and gene model set assembly and completeness was assessed using BUSCO version 3.0.2[33], using the gene set metazoa_odb9 as the standard.

Gene Models

Nematostella, Scolanthus and *M. senile* paired end sequences obtained from previous studies[34,100] and data available on BioProject PRJNA430035 were used to generate *de novo* assembled transcripts.

Trinity version 5.0.2[101] was run on each library using the flags `--min_contig_length 200 --min_kmer_cov 2`. For those which had a strand-specific library preparation, the flag `--SS_lib_type RF` was applied. To reduce redundancy, cd-hit version 4.6.8[102,103] was applied with the flags `-M 0 -c 1`. Transdecoder version 5.0.6[104] was used to detect open reading frames in the resulting reduced set of transcripts. Transcript abundance was quantified using salmon version 1.2.1[105] using the flags `--seqBias --useVBOpt --discardOrphansQuasi --softclip`.

For PacBio Iso-seq, 12 *Nematostella* RNA samples were collected over the course of multiple developmental stages, adult tissues and regeneration time points. For developmental stages, zygotes spawned by a single batch of wildtype colony were kept at 22°C, and collected at 0 hpf, 24 hpf, 48 hpf, 72 hpf and 7 dpf. Adult tissues were collected from sex-sorted, sexually mature wildtype individuals kept at 22°C. The male and female mesenteries were harvested separately by surgically opening the body column and carefully peeling off the attached body column tissues. Adult oral discs were collected by surgical removal of tentacles as well as the attached pharyngeal regions. Regeneration was induced by amputating the oral part of a sexually mature individual at the mid-pharyngeal level. Regenerating tissues close to the wound were collected at 4 hpa and 12 hpa, respectively. All the samples were deep frozen and lysed using TRIzolTM reagent (Invitrogen). Phenol-chloroform extraction was performed to remove undissolved mesoglea from adult tissues. Directzol™ RNA Miniprep Plus Kit (Zymo) was then used to purify total RNA from the aqueous phase. For each sample, 2 µg of total RNA with RIN > 7 was submitted to UC Berkeley for Iso-seq library construction.

RNA Libraries were sequenced at UC Berkeley using PacBio Sequel-II system. Raw subreads bams were processed and demultiplexed using PacBio's isoseq v3.2 conda pipeline. The steps include consensus generation, primer demultiplexing, polyA refinement and data clustering using default parameters. This resulted in the generation of 406,317 high quality HIFI reads and used to build Nvec200 transcriptome.

HIFI reads were mapped to the *Nematostella* genome using minimap2[88] using parameters (`-ax splice -uf --secondary=no`) to obtain the primary best alignments. Reads were then grouped and collapsed down to potential transcripts using PacBio's cDNA_Cupcake toolkit and TAMA[106]. Based on PacBio's guideline, transcripts with degraded 5' reads and have less than 10 FL counts were removed. Chimeric transcripts were then analyzed to find potential fusion genes. For reads that didn't map to the genome, de novo transcriptome assembly was performed using graph-based tool Cogent with kmer size equals 30. . Cupcake and TAMA results were merged into non-redundant gene models using stringtie v2[107]. Deep RNA-seq reads from 4 developmental

stages: egg, gastrula, pos-gastrula and planula were aligned to the genome using STAR v 2.7.3a[108]. Read alignments outside of the isoseq gene models were extracted with bedtools v2.29.2[109] and used for reference-guided transcriptome assembly using Stringtie2. Final gene models were obtained by merging Isoseq models and RNAseq models and manually corrected using previously cloned full length CDS from Nematostella vectensis in NCBI (EDT7). Finally, transdecoder v5.5.0 was used to produce CDS annotation using a minimum protein length of 50 amino acids and prioritizing ORFs with significant similarity to any family in the PFAM database[110]. Alignment of the protein candidates to the PFAM database was done using Hmmer v3.1b2[111]. RNAseq libraries from the *N. vectensis* developmental time series were downloaded and aligned to the new genome using STAR v2.7.3a[108] with standard parameters. Mapping and assignment efficiency was measured using featureCounts from the subread package[112] with the "-p" flag for paired-end libraries.

Evidence for *Scolanthus* gene models were taken from RNA-sequencing and repeats. *Scolanthus* RNA-seq reads (see Sequencing) were mapped to the *Scolanthus* contigs using STAR version 2.7.3a[108]. These mappings were used as evidence for intron junctions to generate putative gene models and estimating hidden Markov model parameters using BRAKER2[113,114]. Gene models were then refined using Augustus version 3.3.3[115] using extrinsic evidence from STAR splice junctions and the location of repeats from RepBase (see Genome Assembly) as counter-evidence for transcription. These models were filtered with the following criteria: 1) genes completely covered by RepeatModeler repeats (see Genome Assembly) were removed 2) predicted gene models were required to be either supported by external RNA-seq evidence as reported by Augustus or have a predicted ortholog as reported by Eggnog-mapper[116]. This resulted in a set of 24,625 gene models. Transcription factor identity was inferred by aligning the predicted protein sequences to Pfam A domains version 32.0[117] using hmmer version 3.3[111]. Transcription factor families were based on domains curated in a previous work[118].

Extended Hox cluster, NK cluster and ParaHox genes were found with BLAT[119] matches of published models[51,61,120–125] to the nv1 genome, taking the best hits. If an NVE gene model[34] corresponded to the matched genomic region, its location in the nv2 genome was then determined for macrosynteny analysis. In cases where no published gene was known, reciprocal BLAST hits between the bilaterian and cnidarian counterpart were taken as evidence for orthology.

Divergence Estimates

Single copy orthologs were detected by collecting common complete and duplicated BUSCO genes present in the *Scolanthus* and *Nematostella* genomes. Where duplicated BUSCOs were present, the transcript with the highest score was taken. This resulted in a total of 541 orthologs. BUSCOs found in genomes obtained from previous studies[2,21,22,26,27,32,41,44,47,126–129] were used to generate multiple alignments. Genes were aligned with mafft version 7.427 using the E-INS-i model and a maximum 1000 refinement iterations[130]. Alignments were trimmed using trimAl version 1.4.rev15 using the "gappyout" criteria[131]. A maximum likelihood tree was inferred using iqtree version 2.0.6, using the model finder partitioned on each gene, constrained to nuclear protein models[132]. Divergence estimates were determined using r8s version 1.8.1 using the

Langley-Fitch likelihood method[133]. Age ranges were estimated by fixing the split between Bilateria and Cnidaria at 595.7 and 688.3 Mya[134].

Single copy orthologs were detected by collecting common complete BUSCO genes present in the *A. millepora*, *A. digitifera*, *E. pallida*, *M. senile*, *Scolanthus* and *Nematostella* genomes. This resulted in a total of 229 orthologs.

Ultraconserved Elements

In order to determine noncoding elements conserved between *Scolanthus* and *Nematostella*, genomes repeat-masked from both *de novo* and repbase repeats were blasted using NCBI BLAST+ version 2.10.0[135], using the flags `-evalue 1E-10 -max_hsps 100000000 -max_target_seqs 100000000 -task megablast -perc_identity 0 -template_length 16 -penalty -2 -word_size 11 -template_type coding_and_optimal`. Additionally, the `-dbsize` parameter was set to the estimated genome size. Candidate hits were then filtered using criteria loosely based on previous work[136]: for each high-scoring pair, a sliding window method was used to determine subsections of the alignment with at least 95 % identity, and extending these windows as long as the identity remains at this level. *Nematostella* elements mapping to more than one locus in the *Scolanthus* genome were reduced to the longest locus pair in both genomes. Elements mostly mapping to coding sequence were removed, and the remaining elements were classified as intron or non-coding, depending on location. Recurring UCE sequences that were not identified by RepeatModeler or RepeatMasker were detected with blastclust version 2.2.26 requiring the length of hit to cover at least 90 % of either sequence for linkage.

Macrosynteny Analysis

Branchiostoma floridae gene models and sequences were retrieved from the published study[41]. All against all comparisons were performed with OMA standalone version 2.5[137].Genomes were downloaded from previous studies[2,3,20,22,26,32,44–47,76,126–129,138–141].

Ancestral genome reconstruction was carried out using a graph based approach. In brief, genes were summarized into multi-species orthologous groups, and these comprised the nodes, and orthology groups occurring on the same chromosome or scaffold of two different species were linked together. A consensus approach to community detection based on the Leiden algorithm[142] was used to determine ancestral linkage groups from this graph. For details, see the Supplementary Text.

TAD Sliding Window Analysis

Hi-C maps were generated for multiple species' genomes (g:) and reads (r:), including those of the ctenophore *Hormiphora californensis* (g: GCA_020137815.1, r: SRR13784181, SRR13784182)[143], the cnidarian *Hydra vulgaris* strain 105 (g: GCA_022113875.1, r: SRR14099165)[40], the cnidarian *Haliclystus octoradiatus* (g: GCA_916610825.1, r:

ERR6745733)[144], the *Nematostella* assembly and reads presented in this manuscript (also, Hi-C reads from the Darwin Tree of Life project, ERR8571699[145]), the cnidarian *Diadumene lineata* (g: GCA_918843875.1, r: ERR6688655)[146], the fire jellyfish *Rhopilema esculentum* (g: http://doi.org/10.5524/100720[39], r: SRR11649085[21]), the fly *Drosophila melanogaster* (g: assembly Release 6[128], r: SRR10512944[147]), the scallop *Pecten maximus* (g: http://dx.doi.org/10.6084/m9.figshare.10311068, r: SRR10119404)[148], the cephalochordate *Branchiostoma floridae* (g: GCA_000003815.2 Bfl_VNyyK, r: SRR12007919, SRR12059951)[41], and *Mus musculus* (g: GRCm39[149], r: SRR1771322-SRR1771324[150]).

The Hi-C maps were generated by mapping the raw paired-end Hi-C reads to the genome assemblies using chromap v0.2.3[151]. The pairs files were normalized using hicExplorer v3.6[152], and balanced using Cooler v0.8.10[153]. The insulation scores were then calculated using FanC v 0.9.23b[73] using 100kb bin sizes. The peaks and valleys in the insulation scores were also called with FanC v0.9.24[154]. We defined the delta between each peak and valley as the transition from a highly-interacting region to a poorly-interaction region as the strength of transitions of topological boundaries. We compared the distributions of these deltas with an uncorrected Mann-Whitney U two-sided test to test for significant differences in the median values.

Phylogenetic analysis of NK-like and SuperHox Genes

Sequences of the NK class proteins were MUSCLE-aligned with default settings in MEGA11[155], and trimmed using the Automated1 setting in TrimAl (v. 1.3)[131]. Trimmed sequences were used for calculating the NJ, and ML trees. The NJ tree was calculated in MEGA11[155], and the ML tree was calculated in IQ-TREE2[132] using the automatic model selection algorithm. Sequences of the SuperHox cluster genes were aligned as above but not trimmed. NJ and ML trees were calculated as above.

Nematostella gastrula ATAC-seq

Embryos were raised at 21°C until gastrula stage (24 hpf). Ten embryos were collected and washed with cold PBS, then cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1 % NP-40, 0.1% Tween-20, 0.01% Digitonin) was added. Embryos were disintegrated by pipetting and incubated in lysis buffer for 6.5 minutes on ice. Nuclei were then pelleted by centrifugation (0.5 rcf, 10 min, 4°C). The pellet was resuspended in 50 µl tagmentation mix (5x TD buffer, 0.1% Tween-20, 0.01% Digitonin, 2.5 µl Tn5 transposase). TD buffer and Tn5 transposase were kindly provided by the David Garfield lab. The transposition reaction was incubated at 37°C for 30 min with agitation at 1000 rpm and then cooled down on ice for 5 minutes. DNA purification was carried out with the QIAquick PCR kit (QIAGEN, #28104) according to the manufacturer's instructions, with an elution in 20 µl elution buffer. PCR amplification was performed for 12 cycles according to Buenrostro(2015). A final purification step was performed with the QIAquick PCR kit (QIAGEN, #28104), elution in 15 µl EB. The quality of the library was validated with an agarose gel, from which DNA was extracted with a peqlab gel extraction kit, eluted in 20 µl EB, and stored at -20°C. Library preparation and sequencing were done at Novogene and Vienna Biocenter Core Facility (VBCF).

Comparative ATACseq analysis

ATACseq peak calling was performed for the following species using publicly available datasets:

Species	SRA	Stage	Reference genome	Reference
Owenia fusiformis	SRR15898385	Blastula	GCA_903813345.2	[156]
Lytechinus variegatus	SRR18841698 SRR18841699 SRR18841700	Gastrula	GCF_018143015.1	[157]
Patinopecten yessoensis	SRR18361867 SRR18361868 SRR18361869 SRR18361870 SRR18361871 SRR18361873	Gastrula	GCF_002113885.1	[158]
Nematostella vectensis	SRR24055088	Gastrula	This study	This study

All peak callings were done following the Encode guideline for ATACseq. *Owenia* and *Nematostella* peak calls were done following the unreplicated data guidelines (https://www.encodeproject.org/pipelines/ENCPL344QWT/) while *Lytechinus* and *Petinopectin* following the replicated data guidelines (https://www.encodeproject.org/pipelines/ENCPL787FUN/). Briefly, raw reads were trimmed, aligned to their reference genome and filtered for mapq >=20. Psudoreplicates were produced for each library as well as for the merged alignment of multiple libraries and peak calling was done on the initial alignment, the merged alignment and all the pseudoreplicates. For unreplicated data, peaks were filtered based on p-signal <=0.05 and similar peak calls between the original alignment and at least one pseudoreplicate. For replicated data, final peaks were selected based on IDR scores between the merged library peaks and each replicate library.

ATACseq peaks for Human (hg19), mouse (mm10), Caenohabditis (Ce11) and Drosophila (dm6) were downloaded from the ChIP-atlas database (https://chip-atlas.org/, [159,160]). Only embryonic stage datasets were used in this study. Repetitive regions were identified using RepeatModeler and RepeatMasker for all genomes and peaks overlapping repeat regions were removed for downstream analysis. A custom R script was used to find the closest gene for each ATAC peak and measure the distances (available in https://github.com/jdmontenegro/NematostellaGenome).

All peaks overlapping genes were not considered in downstream analysis. Data points 1.5 times the interquartile range above the third quartile or below the first quartile were removed as outliers before plotting and regression. A simple linear regression was used to correlate genome size and mean intergenic distance for all species.

Microsynteny Analysis

Orthofinder version 2.5.4 was run to obtain orthogroups for the selected species. The microsynteny pipeline from Simakov et al[58] was run for each of the three clades (deuterostomes, spiralians, cnidarians) separately. For this a subset of orthogroups that had at least one gene per species in a given clade (e.g., every cnidarian species had an ortholog for the cnidarian clade micro-synteny analysis) was generated, to correct for any missing genes. The micro-synteny analysis was thus based on 6016, 4832, and 6539 orthogroups for cnidarians, spiralians, and deuterostomes, respectively. For the most strict profiling of gene block length, we run the micro-synteny pipeline without allowance for any intervening genes. The resulting blocks were then filtered to remove micro-syntenies composed of paralogous genes. The resulting total numbers of micro-syntenic blocks given different minimal required length (at least 2, 3, or 4 genes) are listed in EDT12.

References

1. Putnam, N. H. *et al.* Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. *Science* **317**, 86–94 (2007).
2. Chapman, J. A. *et al.* The dynamic genome of Hydra. *Nature* **464**, 592–596 (2010).
3. Srivastava, M. *et al.* The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010).
4. Schwaiger, M. *et al.* Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* **24**, 639–650 (2014).
5. Gaiti, F. *et al.* Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. *Elife* **6**, e22194 (2017).
6. Nanty, L. *et al.* Comparative methylomics reveals gene-body H3K36me3 in Drosophila predicts DNA methylation and CpG landscapes in other invertebrates. *Genome Res.* **21**,

1841–1850 (2011).

7. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

8. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

9. Acemel, R. D., Maeso, I. & Gómez□Skarmeta, J. L. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *WIREs Developmental Biology* **6**, e265 (2017).

10. Anderson, E., Devenney, P. S., Hill, R. E. & Lettice, L. A. Mapping the Shh long-range regulatory domain. *Development* **141**, 3934–3943 (2014).

11. 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).

12. Rodríguez-Carballo, E., Lopez-Delisle, L., Yakushiji-Kaminatsui, N., Ullate-Agote, A. & Duboule, D. Impact of genome architecture on the functional activation and repression of Hox regulatory landscapes. *BMC Biol.* **17**, 55 (2019).

13. Cartwright, P. *et al.* Exceptionally Preserved Jellyfishes from the Middle Cambrian. *PLoS One* **2**, e1121 (2007).

14. Technau, U., Genikhovich, G. & Kraus, J. E. M. Cnidaria. in *Evolutionary Developmental Biology of Invertebrates 1* (ed. Wanninger, A.) 115–163 (Springer Vienna, 2015).

15. Zapata, F. *et al.* Phylogenomic Analyses Support Traditional Relationships within Cnidaria. *PLoS One* **10**, e0139068 (2015).

16. Technau, U. & Steele, R. E. Evolutionary crossroads in developmental biology: Cnidaria. *Development* **138**, 1447–1458 (2011).

17. Röttinger, E. Nematostella vectensis, an Emerging Model for Deciphering the Molecular and Cellular Mechanisms Underlying Whole-Body Regeneration. *Cells* **10**, (2021).

18. Rentzsch, F. & Technau, U. Genomics and development of Nematostella vectensis and other anthozoans. *Curr. Opin. Genet. Dev.* **39**, 63–70 (2016).
19. Layden, M. J., Rentzsch, F. & Röttinger, E. The rise of the starlet sea anemone *Nematostella vectensis* as a model system to investigate development and regeneration: Overview of starlet sea anemone *Nematostella vectensis*. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 408–428 (2016).
20. Hu, M., Zheng, X., Fan, C.-M. & Zheng, Y. Lineage dynamics of the endosymbiotic cell type in the soft coral Xenia. *Nature* **582**, 534–538 (2020).
21. Nong, W. *et al.* Jellyfish genomes reveal distinct homeobox gene clusters and conservation of small RNA processing. *Nat. Commun.* **11**, 3051 (2020).
22. Leclère, L. *et al.* The genome of the jellyfish Clytia hemisphaerica and the evolution of the cnidarian life-cycle. *Nature Ecology & Evolution* **3**, 801–810 (2019).
23. Khalturin, K. *et al.* Medusozoan genomes inform the evolution of the jellyfish body plan. *Nature Ecology & Evolution* **3**, 811–822 (2019).
24. Ohdera, A. *et al.* Box, stalked, and upside-down? Draft genomes from diverse jellyfish (Cnidaria, Acraspeda) lineages: Alatina alata (Cubozoa), Calvadosia cruxmelitensis (Staurozoa), and Cassiopea xamachana (Scyphozoa). *Gigascience* **8**, (2019).
25. Kim, H.-M. *et al.* The genome of the giant Nomura's jellyfish sheds light on the early evolution of active predation. *BMC Biol.* **17**, 28 (2019).
26. Gold, D. A. *et al.* The genome of the jellyfish Aurelia and the evolution of animal complexity. *Nature Ecology & Evolution* **3**, 96 (2019).
27. Shinzato, C. *et al.* Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320–323 (2011).
28. Wilson, T. J., Lazner, F., Kola, I. & Hertzog, P. J. The Mouse and the Genomic Era. in *Comparative Genomics* (ed. Clark, M. S.) 97–121 (Springer US, 2000).
29. Gosse, P. Scolanthus callimorphus. in *World List of Actiniaria,*.

30. Wilson, E. Scolanthus callimorphus Worm anemone. in *Marine Life Information Network: Biology and Sensitivity Key Information Reviews, [on-line]* (2005).
31. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences* **109**, 17507–17512 (2012).
32. Baumgarten, S. *et al.* The genome of Aiptasia, a sea anemone model for coral symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11893–11898 (2015).
33. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 227–245 (Springer, 2019).
34. Fredman, D., Schwaiger, M., Rentzsch, F. & Technau, U. Nematostella vectensis transcriptome and gene models v2.0. Preprint at https://figshare.com/articles/Nematostella_vectensis_transcriptome_and_gene_models_v2_0/807696 (2013).
35. Cole, A. G. *et al.* Muscle cell-type diversification is driven by bHLH transcription factor expansion and extensive effector gene duplications. *Nat. Commun.* **14**, 1747 (2023).
36. Steger, J. *et al.* Single-cell transcriptomics identifies conserved regulators of neuroglandular lineages. *Cell Rep.* **40**, 111370 (2022).
37. Guo, L. *et al.* An adaptable chromosome preparation methodology for use in invertebrate research organisms. *BMC Biol.* **16**, 25 (2018).
38. Zacharias, H., Anokhin, B., Khalturin, K. & Bosch, T. C. G. Genome sizes and chromosomes in the basal metazoan Hydra. *Zoology* **107**, 219–227 (2004).
39. Li, Y. *et al.* Chromosome-level reference genome of the jellyfish Rhopilema esculentum. *Gigascience* **9**, (2020).
40. Simakov, O. *et al.* Deeply conserved synteny and the evolution of metazoan chromosomes. *Science Advances* **8**, eabi5884 (2022).

41. Simakov, O. *et al.* Deeply conserved synteny resolves early events in vertebrate evolution. *Nature Ecology & Evolution* **4**, 820–830 (2020).
42. Dehal, P. & Boore, J. L. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLoS Biol.* **3**, e314 (2005).
43. Pollard, S. L. & Holland, P. W. H. Evidence for 14 homeobox gene clusters in human genome ancestry. *Curr. Biol.* **10**, 1059–1062 (2000).
44. Braasch, I. *et al.* The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **48**, 427–437 (2016).
45. Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution* **1**, 1–12 (2017).
46. Qu, Z. *et al.* Millipede genomes reveal unique adaptations during myriapod evolution. *PLoS Biol.* **18**, e3000636 (2020).
47. Kenny, N. J. *et al.* Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge Ephydatia muelleri. *Nature Communications* **11**, 3676 (2020).
48. Kapli, P. & Telford, M. J. Topology-dependent asymmetry in systematic errors affects phylogenetic placement of Ctenophora and Xenacoelomorpha. *Science Advances* **6**, eabc5162 (2020).
49. Butts, T., Holland, P. W. H. & Ferrier, D. E. K. The Urbilaterian Super-Hox cluster. *Trends Genet.* **24**, 259–262 (2008).
50. Ferrier, D. E. K. Evolution of Homeobox Gene Clusters in Animals: The Giga-Cluster and Primary vs. Secondary Clustering. *Frontiers in Ecology and Evolution* **4**, (2016).
51. Chourrout, D. *et al.* Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* **442**, 684–687 (2006).
52. Simakov, O. *et al.* Hemichordate genomes and deuterostome origins. *Nature* **527**, 459–465 (2015).
53. Larroux, C. *et al.* The NK Homeobox Gene Cluster Predates the Origin of Hox Genes. *Curr.*

Biol. **17**, 706–710 (2007).

54. Friedman, N. & Rando, O. J. Epigenomics and the structure of the living genome. *Genome Res.* **25**, 1482–1490 (2015).
55. Fletcher, C. *et al.* The genome sequence of the starlet sea anemone, Nematostella vectensis (Stephenson, 1935). *Wellcome Open Res.* **8**, 79 (2023).
56. Rodríguez-Carballo, E. *et al.* The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).
57. Cazet, J. F. *et al.* A chromosome-scale epigenetic map of the *Hydra* genome reveals conserved regulators of cell state. *Genome Res.* gr.277040.122 (2023).
58. Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526–531 (2013).
59. He, S. *et al.* An axial Hox code controls tissue segmentation and body patterning in Nematostella vectensis. *Science* **361**, 1377–1380 (2018).
60. McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. *Cell* **68**, 283–302 (1992).
61. 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).
62. Finnerty, J. R., Pang, K., Burton, P., Paulson, D. & Martindale, M. Q. Origins of Bilateral Symmetry: Hox and Dpp Expression in a Sea Anemone. *Science* **304**, 1335–1337 (2004).
63. Lazar, N. H. *et al.* Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. *Genome Res.* **28**, 983–997 (2018).
64. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
65. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Science Advances* **5**, eaaw1668 (2019).

66. Sacerdot, C., Louis, A., Bon, C., Berthelot, C. & Roest Crollius, H. Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol.* **19**, 166 (2018).

67. Clarence, T. *et al.* Robust 3D modeling reveals spatiosyntenic properties of animal genomes. *iScience* **26**, 106136 (2023).

68. Sinigaglia, C., Busengdal, H., Leclère, L., Technau, U. & Rentzsch, F. The bilaterian head patterning gene six3/6 controls aboral domain development in a cnidarian. *PLoS Biol.* **11**, e1001488 (2013).

69. Schwaiger, M. *et al.* An ancestral Wnt-Brachyury feedback loop in axial patterning and recruitment of mesoderm-determining target genes. *Nat Ecol Evol* **6**, 1921–1939 (2022).

70. Kusserow, A. *et al.* Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433**, 156–160 (2005).

71. Lengfeld, T. *et al.* Multiple Wnts are involved in Hydra organizer formation and regeneration. *Dev. Biol.* **330**, 186–199 (2009).

72. Marlétaz, F. *et al.* The little skate genome and the evolutionary emergence of wing-like fin appendages. *bioRxiv* 2022.03.21.485123 (2022) doi:10.1101/2022.03.21.485123.

73. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).

74. Kim, J. *et al.* Condensin DC loads and spreads from recruitment sites to create loop-anchored TADs in C. elegans. *Elife* **11**, (2022).

75. Voolstra, C. R. *et al.* Comparative analysis of the genomes of Stylophora pistillata and Acropora digitifera provides evidence for extensive differences between species of corals. *Sci. Rep.* **7**, 1–14 (2017).

76. Moya, A. *et al.* Whole transcriptome analysis of the coral Acropora millepora reveals complex responses to CO□-driven acidification during the initiation of calcification. *Mol. Ecol.* **21**, 2440–2454 (2012).

77. Cunning, R., Bay, R. A., Gillette, P., Baker, A. C. & Traylor-Knowles, N. Comparative

analysis of the Pocillopora damicornis genome highlights role of immune system in coral evolution. *Sci. Rep.* **8**, 16134 (2018).

78. Jiang, J. B. *et al.* A hybrid de novo assembly of the sea pansy (Renilla muelleri) genome. *Gigascience* **8**, (2019).
79. Jeon, Y. *et al.* The Draft Genome of an Octocoral, Dendronephthya gigantea. *Genome Biol. Evol.* **11**, 949–953 (2019).
80. Marlétaz, F. *et al.* Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64–70 (2018).
81. Fritzenwanker, J. H. & Technau, U. Induction of gametogenesis in the basal cnidarian Nematostella vectensis (Anthozoa). *Dev. Genes Evol.* **212**, 99–103 (2002).
82. Chen, J. & Dellaporta, S. Urea-based Plant DNA Miniprep. in *The Maize Handbook* (eds. Freeling, M. & Walbot, V.) 526–527 (Springer, 1994).
83. Sanders, S. M. *et al.* CRISPR/Cas9-mediated gene knockin in the hydroid Hydractinia symbiolongicarpus. *BMC Genomics* **19**, 649 (2018).
84. Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
85. Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
86. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
87. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).
88. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
89. Pryszcz, L. P. & Gabaldón, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113–e113 (2016).

90. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).

91. Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-4.0*. (2013).

92. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020).

93. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]* (2013).

94. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

95. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).

96. Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).

97. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**, 99–101 (2016).

98. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* 9 (2004).

99. Hunt, M. *et al.* REAPR: a universal tool for genome assembly evaluation. *Genome Biol.* **14**, R47 (2013).

100. Praher, D. *et al.* Conservation and turnover of miRNAs and their highly complementary targets in early branching animals. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20203169 (2021).

101. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

102. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-

generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

103. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

104. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.* **8**, (2013).

105. Srivastava, A. *et al.* Alignment and mapping methodology influence transcript abundance estimation. *bioRxiv* 657874 (2019).

106. Kuo, R. I. *et al.* Illuminating the dark side of the human transcriptome with long read transcript sequencing. *BMC Genomics* **21**, 751 (2020).

107. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).

108. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

109. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

110. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).

111. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

112. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

113. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).

114. Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: Incorporating Protein Homology Information into Gene Prediction with GeneMark-EP and AUGUSTUS. 1 (2018).

115. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC*

Bioinformatics **7**, 62 (2006).

116. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

117. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

118. de Mendoza, A. *et al.* Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proceedings of the National Academy of Sciences* **110**, E4858–E4866 (2013).

119. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656–664 (2002).

120. Matus, D. Q., Pang, K., Daly, M. & Martindale, M. Q. Expression of Pax gene family members in the anthozoan cnidarian, Nematostella vectensis: Pax gene expression in Nematostella vectensis. *Evol. Dev.* **9**, 25–38 (2007).

121. Ryan, J. F. *et al.* The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, Nematostella vectensis. *Genome Biol.* **7**, R64 (2006).

122. Mazza, M. E., Pang, K., Martindale, M. Q. & Finnerty, J. R. Genomic organization, gene structure, and developmental expression of three Clustered otx genes in the sea anemone Nematostella vectensis. *J. Exp. Zool. B Mol. Dev. Evol.* **308B**, 494–506 (2007).

123. Matus, D. Q., Thomsen, G. H. & Martindale, M. Q. Dorso/Ventral Genes Are Asymmetrically Expressed and Involved in Germ-Layer Demarcation during Cnidarian Gastrulation. *Curr. Biol.* **16**, 499–505 (2006).

124. Hudry, B. *et al.* Molecular insights into the origin of the Hox-TALE patterning system. *Elife* **3**, e01939 (2014).

125. Mazza, M. E., Pang, K., Reitzel, A. M., Martindale, M. Q. & Finnerty, J. R. A conserved cluster of three PRD-class homeobox genes (homeobrain, rx and orthopedia) in the Cnidaria and Protostomia. *Evodevo* **1**, 3 (2010).

126. Ying, H. *et al.* The Whole-Genome Sequence of the Coral Acropora millepora. *Genome Biol. Evol.* **11**, 1374–1379 (2019).
127. Consortium*, The C. elegans Sequencing. Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science* **282**, 2012–2018 (1998).
128. Hoskins, R. A. *et al.* The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res.* **25**, 445–458 (2015).
129. Hydra 2.0 Web Portal. https://research.nhgri.nih.gov/hydra.
130. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
131. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
132. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
133. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
134. dos Reis, M. *et al.* Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Curr. Biol.* **25**, 2939–2950 (2015).
135. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
136. Dimitrieva, S. & Bucher, P. UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res.* **41**, D101–D109 (2013).
137. Roth, A. C. J., Gonnet, G. H. & Dessimoz, C. Algorithm of OMA for large-scale orthology inference. *BMC Bioinformatics* **9**, 518 (2008).
138. Zakas, C., Harry, N. D., Scholl, E. H. & Rockman, M. V. The genome of the poecilogonous annelid Streblospio benedicti. *bioRxiv* 2021.04.15.440069 (2021).

139. Shingate, P. *et al.* Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nat. Commun.* **11**, 2322 (2020).

140. Harris, T. W. *et al.* WormBase: a modern Model Organism Information Resource. *Nucleic Acids Res.* **48**, D762–D767 (2020).

141. Srivastava, M. *et al.* The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).

142. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

143. Schultz, D. T. *et al.* A chromosome-scale genome assembly and karyotype of the ctenophore *Hormiphora californensis*. *G3 (Bethesda)* **11**, jkab302 (2021).

144. Blaxter, M. *et al.* The genome sequence of the spotted kaleidoscope jellyfish, *Haliclystus octoradiatus* (James-Clark, 1863). *Wellcome Open Res.* **8**, 6 (2023).

145. Fletcher, C., da Conceicoa, L. P., of Life Consortium, D. T. & Others. The genome sequence of the starlet sea anemone, *Nematostella vectensis* (Stephenson, 1935). *Wellcome Open Research* **8**, 79 (2023).

146. Wood, C. *et al.* The genome sequence of the orange-striped anemone, *Diadumene lineata* (Verrill, 1869). *Wellcome Open Res.* **7**, 93 (2022).

147. Adams, M. *et al.* One fly-one genome: chromosome-scale genome assembly of a single outbred *Drosophila melanogaster*. *Nucleic Acids Res.* **48**, e75 (2020).

148. Kenny, N. J. *et al.* The gene-rich genome of the scallop *Pecten maximus*. *Gigascience* **9**, (2020).

149. Mouse Genome Sequencing Consortium *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).

150. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309 (2015).

151. Zhang, H. *et al.* Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat.*

Commun. **12**, 1–6 (2021).

152. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).

153. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).

154. 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).

155. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).

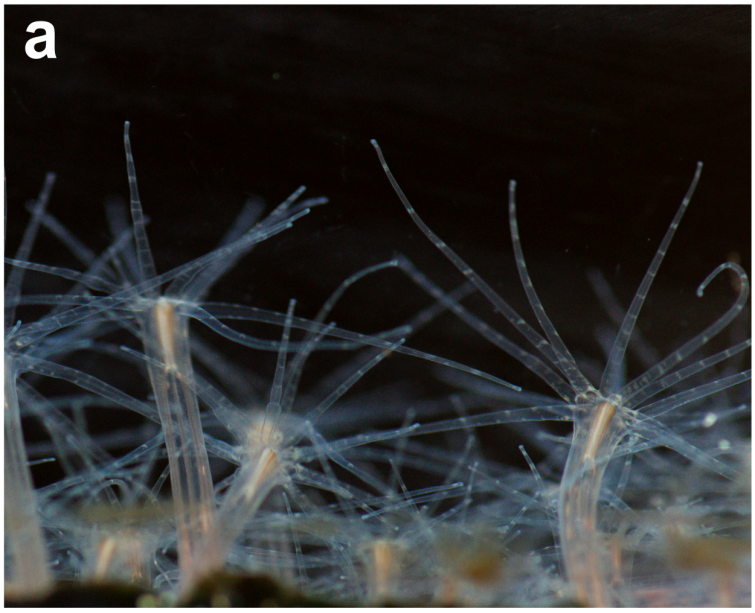
156. Seudre, O., Carrillo-Baltodano, A. M., Liang, Y. & Martín-Durán, J. M. ERK1/2 is an ancestral organising signal in spiral cleavage. *Nat. Commun.* **13**, 2286 (2022).

157. Davidson, P. L. *et al.* Recent reconfiguration of an ancient developmental gene regulatory network in Heliocidaris sea urchins. *Nat Ecol Evol* **6**, 1907–1920 (2022).

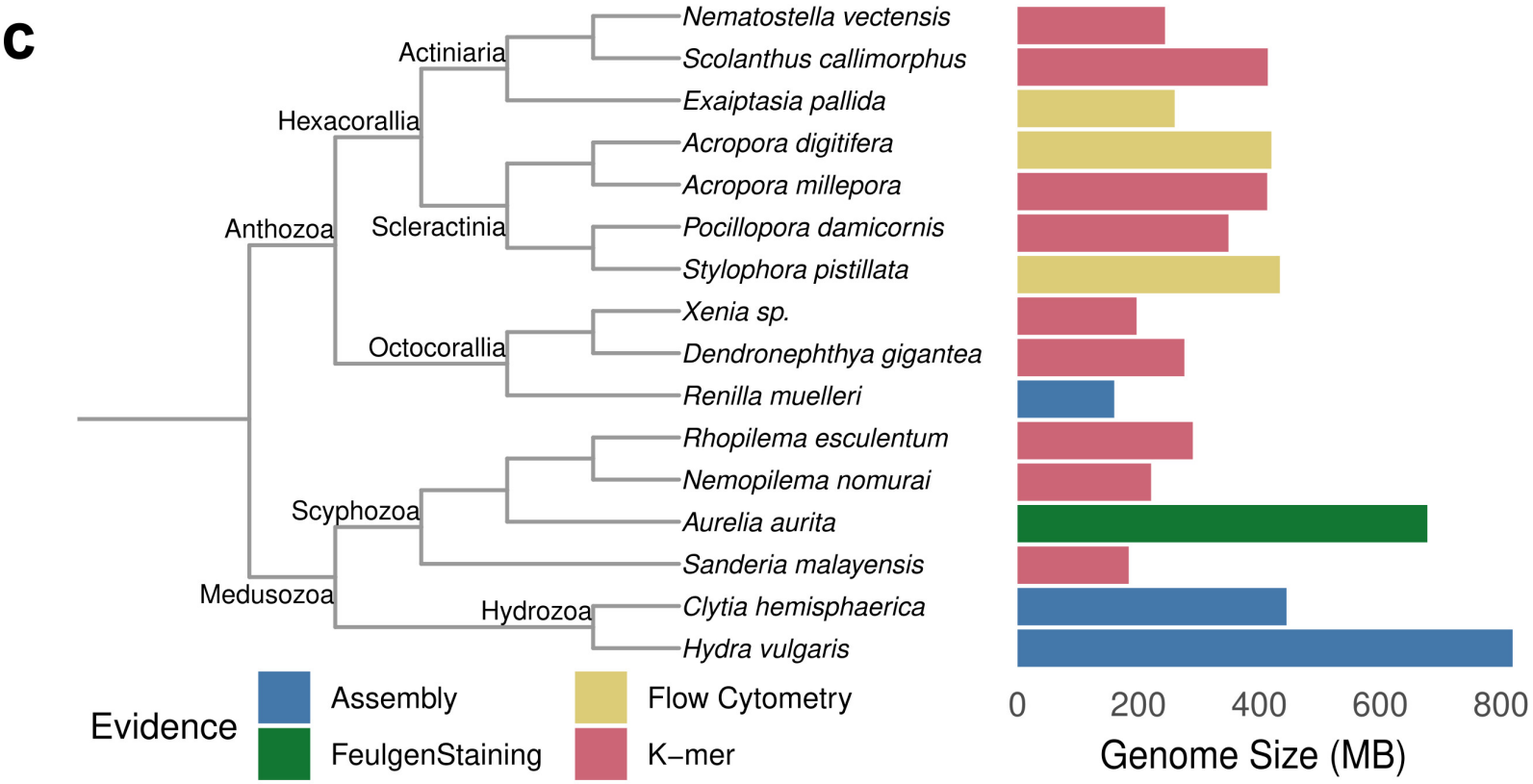
158. Han, W. *et al.* Ancient homomorphy of molluscan sex chromosomes sustained by reversible sex-biased genes and sex determiner translocation. *Nat Ecol Evol* **6**, 1891–1906 (2022).

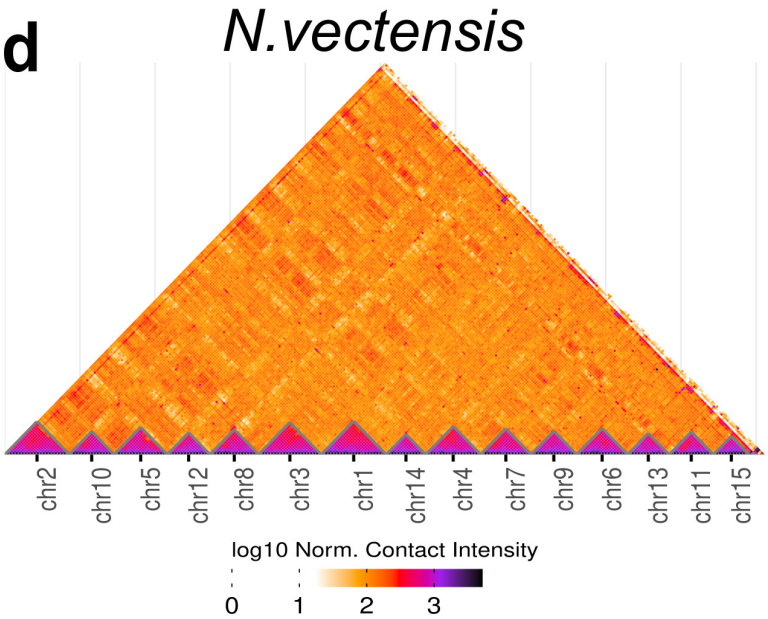
159. Zou, Z., Ohta, T., Miura, F. & Oki, S. ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Res.* **50**, W175–82 (2022).

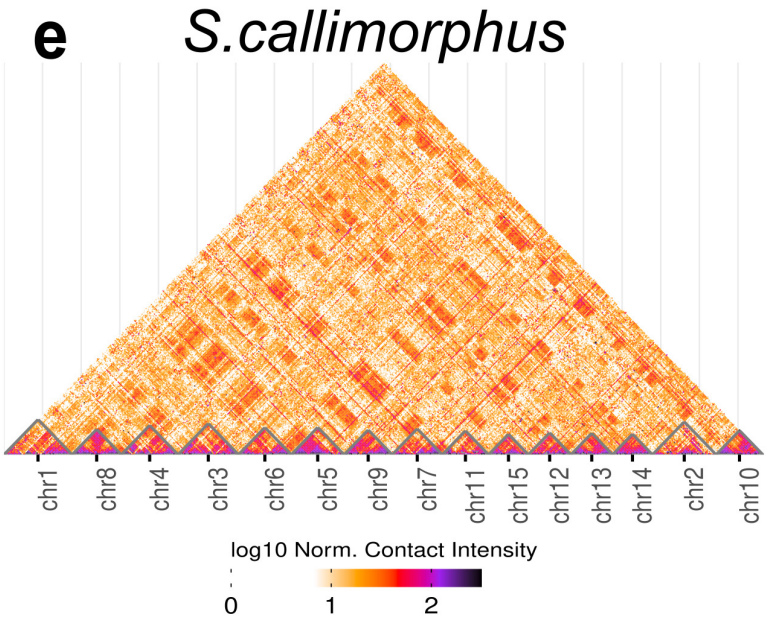
160. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19**, (2018).

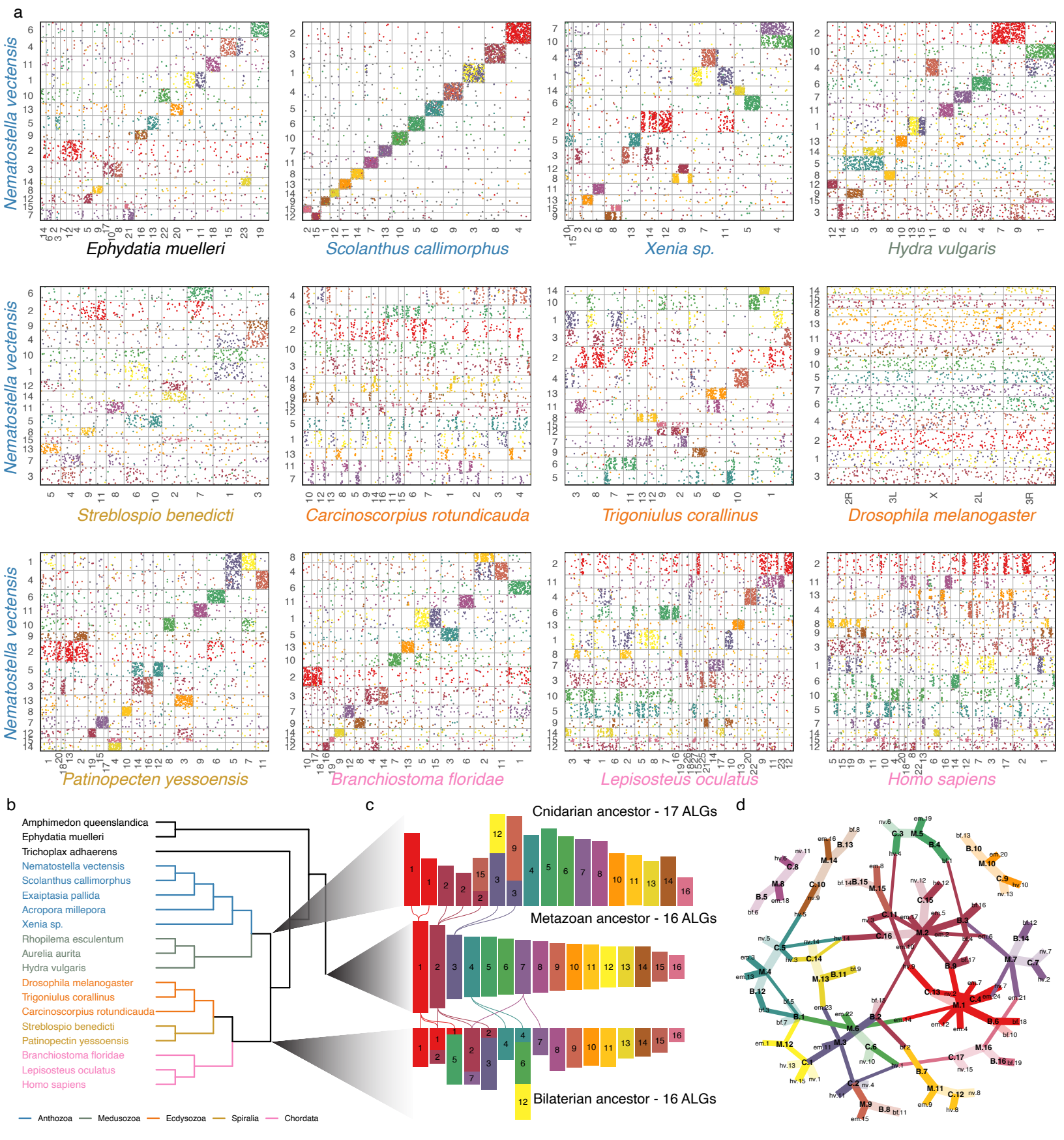

a

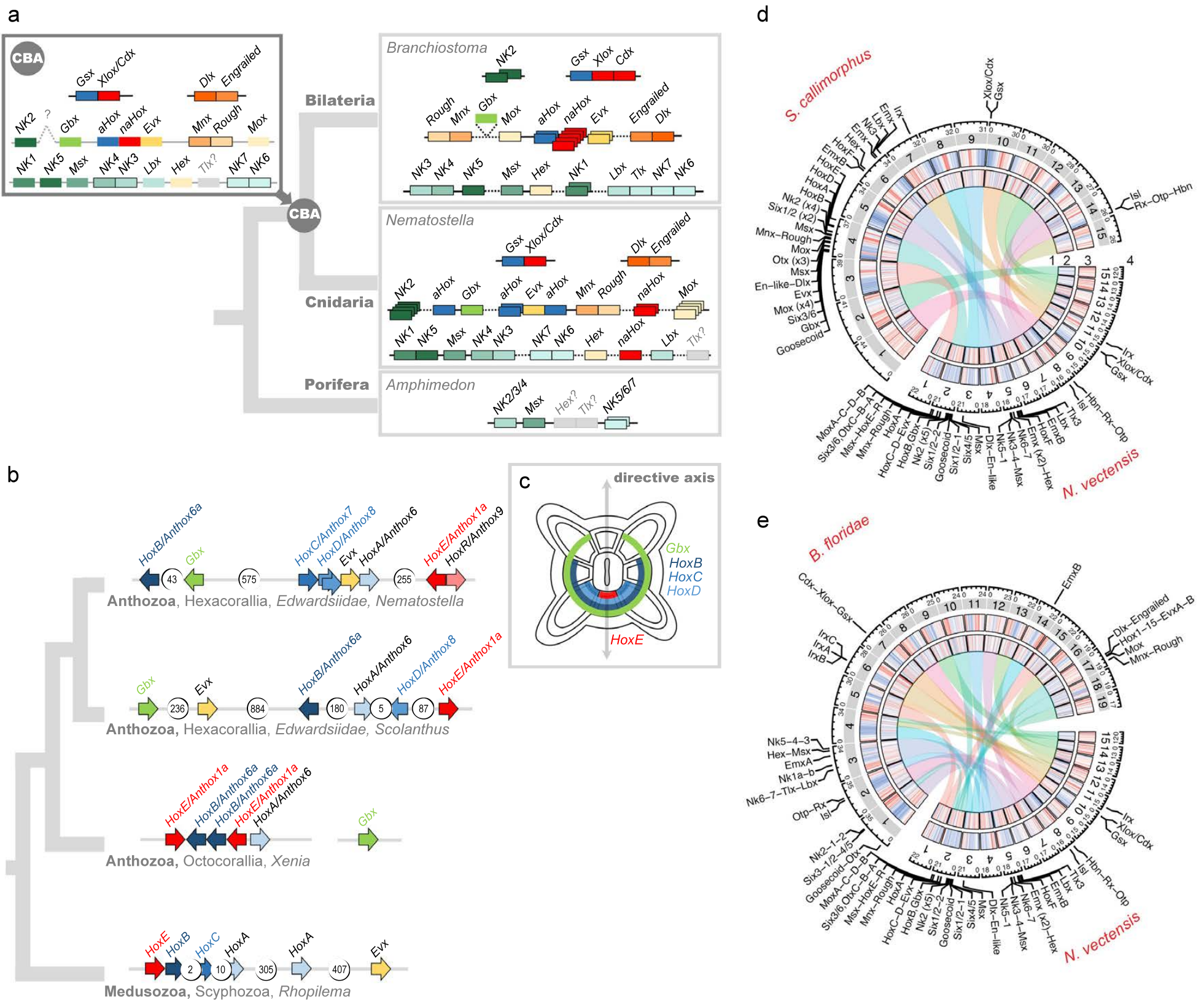
b

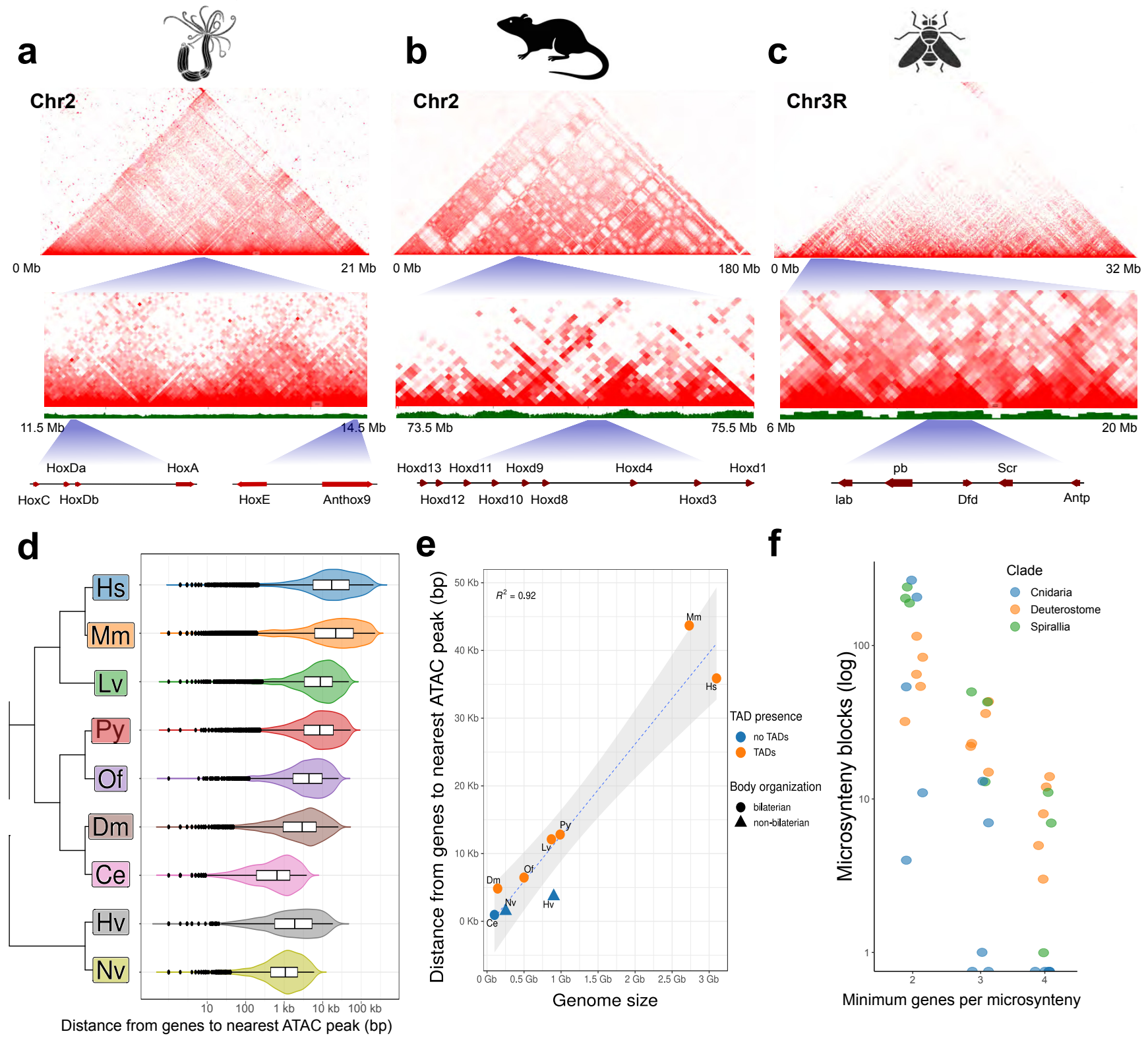

c
Nematostella vectensis
Scolanthus callimorphus
Exaiptasia pallida
Acropora digitifera
Acropora millepora
Pocillopora damicornis
Stylophora pistillata
Xenia sp.
Dendronephthya gigantea
Renilla muelleri
Rhopilema esculentum
Nemopilema nomurai
Aurelia aurita
Sanderia malayensis
Clytia hemisphaerica
Hydra vulgaris
Actiniaria
Hexacorallia
Scleractinia
Anthozoa
Octocorallia
Scyphozoa
Medusozoa
Hydrozoa
Evidence
Assembly
Flow Cytometry
FeulgenStaining
K−mer
0
200
400
600
800
Genome Size (MB)


d
N.vectensis
chr2
chr10
chr5
chr12
chr8
chr3
chr1
chr14
chr4
chr7
chr9
chr6
chr13
chr11
chr15
log10 Norm. Contact Intensity
0
1
2
3


e
S.callimorphus
chr1
chr8
chr4
chr3
chr6
chr5
chr9
chr7
chr11
chr15
chr12
chr13
chr14
chr2
chr10
log10 Norm. Contact Intensity
0
1
2


a
Nematostella vectensis
Ephydatia muelleri
Scolanthus callimorphus
Xenia sp.
Hydra vulgaris
Streblospio benedicti
Carcinoscorpius rotundicauda
Trigoniulus corallinus
Drosophila melanogaster
Patinopecten yessoensis
Branchiostoma floridae
Lepisosteus oculatus
Homo sapiens
b
Amphimedon queenslandica
Ephydatia muelleri
Trichoplax adhaerens
Nematostella vectensis
Scolanthus callimorphus
Exaiptasia pallida
Acropora millepora
Xenia sp.
Rhopilema esculentum
Aurelia aurita
Hydra vulgaris
Drosophila melanogaster
Trigoniulus corallinus
Carcinoscorpius rotundicauda
Streblospio benedicti
Patinopectin yessoensis
Branchiostoma floridae
Lepisosteus oculatus
Homo sapiens
Anthozoa
Medusozoa
Ecdysozoa
Spiralia
Chordata
c
Cnidarian ancestor - 17 ALGs
Metazoan ancestor - 16 ALGs
Bilaterian ancestor - 16 ALGs
d


a
CBA
Gsx
Xlox/Cdx
Dlx
Engrailed
NK2
?
Gbx
aHox
naHox
Evx
Mnx
Rough
Mox
NK1
NK5
Msx
NK4
NK3
Lbx
Hex
Tlx?
NK7
NK6
Bilateria
Branchiostoma
Cnidaria
Nematostella
Porifera
Amphimedon
NK2/3/4
Hex?
NK5/6/7
b
HoxB/Anthox6a
HoxC/Anthox7
HoxD/Anthox8
HoxA/Anthox6
HoxE/Anthox1a
HoxR/Anthox9
43
575
255
Anthozoa, Hexacorallia, Edwardsiidae, Nematostella
236
884
180
5
87
Anthozoa, Hexacorallia, Edwardsiidae, Scolanthus
Anthozoa, Octocorallia, Xenia
HoxE
HoxB
HoxC
HoxA
2
10
305
407
Medusozoa, Scyphozoa, Rhopilema
c
directive axis
Gbx
HoxB
HoxC
HoxD
HoxE
d
S. callimorphus
N. vectensis
Xlox/Cdx
Gsx
Emx
Lbx
Irx
Nk3
Hex
HoxF
EmxB
HoxE
HoxD
HoxA
HoxB
Nk2 (x4)
Six1/2 (x2)
Msx
Mnx–Rough
Mox
Otx (x3)
Msx
En–like–Dlx
Evx
Mox (x4)
Six3/6
Gbx
Goosecoid
Isl
Rx–Otp–Hbn
Irx
Xlox/Cdx
Gsx
Hbn–Rx–Otp
Isl
Tlx3
Lbx
EmxB
HoxF
Emx (x2)–Hex
Nk6–7
Nk3–4–Msx
Nk5–1
Dlx–En–like
Msx
Six4/5
Six1/2–1
Goosecoid
Six1/2–2
Nk2 (x5)
HoxB,Gbx
HoxA
HoxC–D–Evx
Mnx–Rough
Msx–HoxE–R
Six3/6,OtxC–B–A
MoxA–C–D–B
e
B. floridae
N. vectensis
Cdx–Xlox–Gsx
EmxB
IrxC
IrxA
IrxB
Dlx–Engrailed
Hox1–15–EvxA–B
Mox
Mnx–Rough
Nk5–4–3
Hex–Msx
EmxA
Nk1a–b
Nk6–7–Tlx–Lbx
Otp–Rx
Isl
Nk2–1–2
Six3–1/2–4/5
Goosecoid–Otx
MoxA–C–D–B
Six3/6,OtxC–B–A
Msx–HoxE–R
Mnx–Rough
HoxA
HoxC–D–Evx
HoxB,Gbx
Nk2 (x5)
Six1/2–2
Goosecoid
Six1/2–1
Six4/5
Msx
Dlx–En–like
Nk5–1
Nk3–4–Msx
Nk6–7
Emx (x2)–Hex
HoxF
EmxB
Lbx
Tlx3
Isl
Hbn–Rx–Otp
Irx
Xlox/Cdx
Gsx


a
Chr2
0 Mb
21 Mb
11.5 Mb
14.5 Mb
HoxDa
HoxA
HoxC
HoxDb
HoxE
Anthox9
b
Chr2
0 Mb
180 Mb
73.5 Mb
75.5 Mb
Hoxd13
Hoxd11
Hoxd9
Hoxd4
Hoxd1
Hoxd12
Hoxd10
Hoxd8
Hoxd3
c
Chr3R
0 Mb
32 Mb
6 Mb
20 Mb
pb
Scr
lab
Dfd
Antp
d
Hs
Mm
Lv
Py
Of
Dm
Ce
Hv
Nv
10
100
1 kb
10 kb
100 kb
Distance from genes to nearest ATAC peak (bp)
e
R^2 = 0.92
Distance from genes to nearest ATAC peak (bp)
50 Kb
40 Kb
30 Kb
20 Kb
10 Kb
0 Kb
0 Gb
0.5 Gb
1 Gb
1.5 Gb
2 Gb
2.5 Gb
3 Gb
Genome size
TAD presence
no TADs
TADs
Body organization
bilaterian
non-bilaterian
f
Clade
Cnidaria
Deuterostome
Spirallia
Microsynteny blocks (log)
100
10
1
2
3
4
Minimum genes per microsynteny