OXFORD (GIGA)n SCIENCE

GigaScience, 2025, **14**, 1–14
DOI: 10.1093/gigascience/giaf097
Data Note

First chromosome-level genome assembly of the colonial chordate model *Botryllus schlosseri* (Tunicata)

Olivier De Thier [1,2], Marie Lebel [3], Mohammed M.Tawfeeq [1,2], Roland Faure [1,2], Philippe Dru [3], Simon Blanchoud [4], Alexandre Alié [3], Federico D. Brown [5], Jean-François Flot [1,2,*], and Stefano Tiozzo [3,*]

[1]Evolutionary Biology & Ecology, C.P. 160/12, Université libre de Bruxelles (ULB), Avenue F.D. Roosevelt 50, B-1050 Brussels, Belgium
[2]Interuniversity Institute of Bioinformatics in Brussels – (IB)[2], B-1050 Brussels, Belgium
[3]CNRS, Sorbonne Université, Laboratoire de Biologie du Développement de Villefranche Sur-mer (LBDV - UMR7009) , IMEV - 181 Chemin du Lazaret, F-06230 Villefranche-sur-Mer , France
[4]Department of Biology, University of Fribourg, CH-1700 Fribourg, Switzerland
[5]Departmento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo - SP 05508-090, Brazil
*Correspondence address. Stefano Tiozzo, Institut de la Mer de Villefranche (IMEV), Laboratoire de Biologie du Développement de Villefranche-sur-Mer (LBDV), Sorbonne Université, 06230 Villefranche-sur-Mer, France. E-mail: stefano.tiozzo@imev-mer.fr; Jean-François Flot, Université libre de Bruxelles (ULB), C.P. 160/12, Avenue F.D. Roosevelt 50, 1050 Brussels, Belgium. E-mail: jean-francois.flot@ulb.be

Abstract

Background: *Botryllus schlosseri* (Tunicata) is a colonial, laboratory model tunicate recognized for its remarkable developmental diversity, its regenerative abilities, and its peculiar genetically determined allorecognition system governed by a polymorphic locus controlling chimerism and cell parasitism.

Results: We report the first chromosome-level genome assembly of *B. schlosseri* subclade A1. By integrating long and short reads with Hi-C scaffolding, we produced both a phased diploid genome assembly and a conventional collapsed consensus sequence of 533 Mb. Of this total length, 96% belonged to 16 chromosome-scale scaffolds, with a BUSCO completeness score of 91.4%. We then compared our assembly with other high-quality tunicate genomes, revealing some synteny conservation but also extensive genomic rearrangements and a general loss of colinearity.

Conclusions: The chromosome-level resolution of this assembly enhances our understanding of genome organization in colonial modular organisms. Comparative analyses highlight the dynamic nature of tunicate genomes, with conserved macrosynteny yet extensive microsyntenic rearrangements and scrambling, underscoring their rapid evolutionary trajectory. This high-quality genome assembly provides a valuable resource for exploring the unique biological features of colonial chordates, including their exceptional regenerative abilities and complex allorecognition system.

Keywords: budding, regeneration, chimerism, ascidian, coloniality, model organism

Introduction

> Each member of the colony is an individual animal, but the colony is another individual animal, not like the sum of its individuals [...]. So a man of individualistic reason, if he must ask, "Which is the animal?" must abandon his particular kind of reason and say, "Why, it's two animals and they aren't alike any more than the cells of my body are like me. I am much more than the sum of my cells, and, for all I know, they are much more than the division of me."
> —John Steinbeck, *The Log from the Sea of Cortez*

In the subphylum Tunicata, the sister group of vertebrates [1], colonial species reproduce both sexually and asexually through various forms of budding. Through budding, new functional bodies emerge from adult somatic cells and tissues. Regardless of variations in budding modes among tunicate species [2] and of whether development occurs through asexual budding or sexually via embryogenesis, the basic body plan of adult tunicates is broadly conserved across the entire subphylum [3]. In colonial tunicates, asexually generated individuals generally remain physically connected, forming colonies. Colony formation, clonal reproduction, and modular organization have important physiological, ecological, and evolutionary implications. For example, modular organization supports rapid growth on hard, space-limited substrates, outperforming solitary forms. Morphological plasticity enables colony-level adaptation to predation, damage, or environmental changes. Furthermore, uniparental reproduction, including budding, likely provides a selective advantage for rapid colonization on invasion fronts or in disturbed habitats (reviewed in [4]). Like many other colonial tunicates, *Botryllus schlosseri* (Pallas, 1766) (NCBI:txid30301) can generate a functional adult body via 3 distinct developmental pathways. The first one involves sexual reproduction, where the fertilized egg passes through a larval stage and develops into an initial colony founder. The second pathway is asexual propagation, where the founder zooid continuously reproduces through palleal (aka peribranchial) budding, forming a colony of hundreds of zooids connected by the vascular system (a network of extracorporeal vessels within a cellulose-based extracellular matrix, the so-called tunic [5]; Fig. 1). Lastly, if all zooids and buds are removed from a *B. schlosseri* colony, new buds can regenerate from the vascular system in a process known as vascular budding, allowing asexual propagation and eventual colony

Received: February 19, 2025. **Revised:** May 13, 2025. **Accepted:** July 27, 2025

Figure 1: Colony of *Botryllus schlosseri* (photograph by Stefano Tiozzo). Scale bar: 1 mm.

reformation [6–8]. Zooids within a single colony are genetically identical clones. However, wild colonies often come into contact and fuse, resulting in chimeras where circulating cells carry different genotypes. These mixed pools of circulating cells contribute to sexual and, according to some authors, asexual and regenerative development [9–11]. During chimerism, donor cells may entirely replace the host's germline or somatic cells, a phenomenon termed germ cell or somatic cell parasitism, respectively [10, 12, 13]. As a result, zooids within a chimeric colony are not always clonemates.

Botryllus schlosseri was introduced to laboratories over half a century ago [14] as a model to study asexual development, regeneration [15], allorecognition, and chimerism [16, 17]. Over recent decades, a dedicated scientific community has emerged, advancing breeding techniques and developing imaging and molecular biology tools to better study this species [8, 9, 18–21]. Several anatomical descriptions and staging methods have been proposed [5, 22], and extensive transcriptomic databases for various developmental stages and tissues have been generated [8, 23–27]. In 2013, a draft genome of *B. schlosseri* was published [28], but it lacked the completeness and continuity required by today's assembly standards [29]. In this study, we present a high-quality, chromosome-level collapsed assembly as well as a chromosome-scale haplotype-resolved assembly for *B. schlosseri*. This new resource offers a robust platform for investigating the developmental and regenerative processes, complex allorecognition, chimerism, and cell parasitism of this colonial chordate.

Results and discussion

Sequencing and genome size estimation

Genomic DNA was extracted from a laboratory-reared colony, referred to as clone E*, derived from a single zygote and therefore nonchimeric. Sequencing libraries from clone E* yielded 489 million Illumina (short) paired-end 150-bp reads, 2.4 million PacBio HiFi (long) reads with an N50 length of ∼9.5 kb (max length of ∼50 kb), and 10.9 million ONT (long) reads with an N50 length of ∼10.3 kb (max length of ∼205 kb) (Table 1).

Based on *k*-mer analyses, the genome size was estimated to be around 500 Mbp with a heterozygosity of 3.63% (Supplementary Fig. S1), whereas Feulgen densitometry (a histochemical approach) yielded an estimate of ∼492 Mbp (using 1 pg = 978 Mbp; Supplementary Fig. S5). Both genome size estimates were concordant but notably smaller than a previous cytofluorimetry-based estimation of 725 Mb [30] and than the first genome assembly obtained by Voskoboynik et al. [28], which had a size of 580 Mbp.

Table 1: Sequencing technologies used to sequence *B. schlosseri*'s genome (clone E*), and related read statistics

Technology	Total size (Gbp)	Number of reads	N50 (bp)	Coverage
Illumina	73.2	488,906,094	150	146
Illumina Hi-C	15.9	106,488,252	150	32
PacBio HiFi (round 1)	7.9	1,170,137	8,711	16
PacBio HiFi (round 2)	10.8	1,218,052	10,151	22
ONT (R9.4.1)	58.9	10,888,103	10,320	118

Table 2: Assembly statistics for all the scaffolds and for the 16 longest ones

Measure	All scaffolds	16 longest scaffolds
Length (Mbp)	533	513
No. of sequences	254	16
N50 (Mbp)	30	31
GC (%)	40.52	40.46
No. of annotated genes	22,275	21,677
BUSCO Complete	91.6%	91.4%
(Single, Duplicated)	(90.7%, 0.9%)	(90.7%, 0.7%)
BUSCO Fragmented	3.1%	3.1%
BUSCO Missing	5.3%	5.5%

An initial collapsed genome assembly was obtained using hifiasm [31] (RRID:SCR_021069); it had a size of 570 Mbp and comprised 930 contigs with an N50 length of 4.9 Mbp. In this assembly, BlobToolKit (RRID:SCR_023351) identified 452 contigs (totaling 37 Mbp) as putative contamination and mitochondrial sequences (see next section), which were subsequently removed. Of these 37 Mbp, approximately half were attributed to members of the bacterial phylum Pseudomonadota (Supplementary Fig. S6). We identified 28 contigs that belonged to spore-forming unicellular parasites of the microsporidia group [32]. To our knowledge, this represents the first report of this fungal group in a tunicate species. However, we cannot rule out the possibility that these sequences may have been assigned incorrectly or originate from contaminants present in the water rather than from parasitized *Botryllus* tissues. The remaining contigs were corrected using CRAQ [33], which detects and breaks misassembled contigs; this raised the total number of contigs in the assembly from 478 to 516. We then performed Hi-C scaffolding using YaHS [34] (RRID:SCR_022965), which reduced the number of sequences to 256, before running CRAQ again on the scaffolded assembly: this time, 4 misassembled contigs were detected and broken. Finally, a manual curation was performed, resulting in an assembly made up of 16 major scaffolds, labeled Bs1 to Bs16, containing around 96% (513 Mbp) of the total sequence length (533 Mbp) (Table 2, Supplementary Table S3, Figs. 2 and 3). The number and relative lengths of these 16 major scaffolds were consistent with the published karyogram of *B. schlosseri* [35], with the exception of Bs16, which was notably longer in our assembly (Supplementary Fig. S14). The full assembly pipeline is summarized in Fig. 4 and detailed in the Methods section.

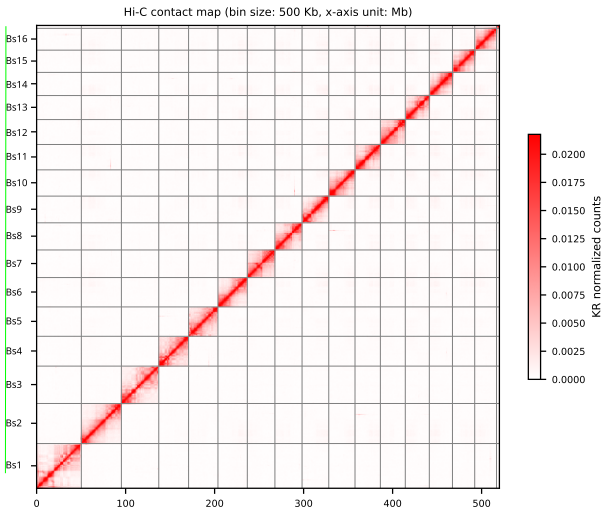

Figure 2: Hi-C heatmap of the collapsed assembly of the *Botryllus schlosseri* genome showing 16 chromosome-scale scaffolds. The figure was generated using the visualization module of HapHiC [36].

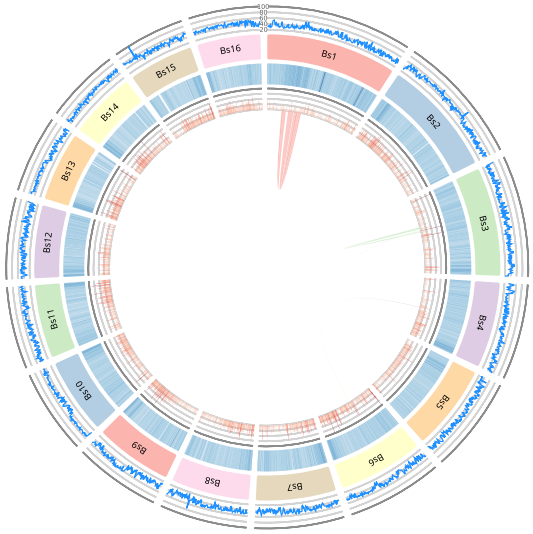

Figure 3: Circos plot of the distribution of several genomic characteristics along the 16 longest scaffolds (labeled Bs1 to Bs16) of the collapsed assembly (made using AccuSyn [37]). Each layer of the circle represents, from the inside to the outside, the synteny blocks detected by MCScanX [38], histograms of gene density, heatmaps of the presence of repetitive elements, the scaffold names in clockwise order, and the sequencing depth of HiFi reads.

The completeness of our assembly was assessed using the BUSCO tool [39] (RRID:SCR_015008, v5.4.4) with the metazoa_odb10 dataset, which returned a genome completeness of 91.6% (including 0.9% of duplicated marker genes), compared to 74.4% (including 23.7% of duplicated marker genes) for the assembly by Voskoboynik et al. [28] (Fig. 5). The high duplication score of the previously available assembly indicates that its larger size (580 Mbp vs. 533 Mbp) was caused by incompletely collapsed haplotypes [40]. Synteny analysis performed using MCScanX [38] (RRID:SCR_022067) highlighted the presence of 2 large-scale genomic palindromes located within Bs1 and a smaller one in Bs3 (displayed in red and green in the innermost layer of Fig. 3). To find out whether these palindromes may have resulted from assembly artifacts caused by uncollapsed haplotypes [41],

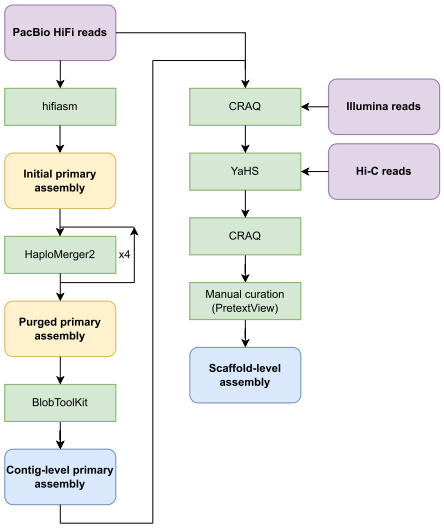

Figure 4: Assembly pipeline for the collapsed genome assembly (see Methods).

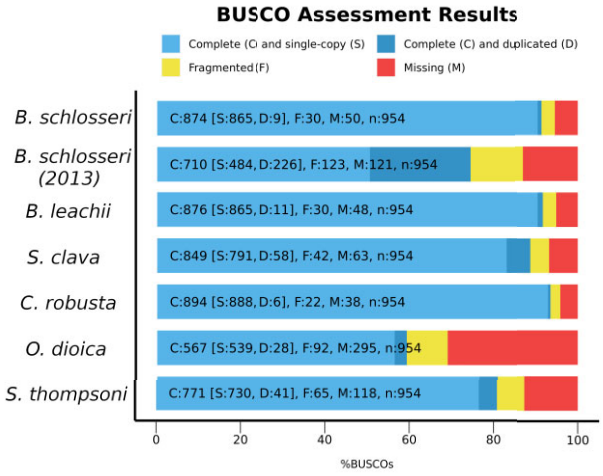

Figure 5: Orthology assignment in previous tunicate genome projects. Proportion of BUSCO genes detected or missed in the new genome assembly of *B. schlosseri* compared to the previous assembly (*B. schlosseri* [2013] [28]) and other reference genomes.

we checked the sequencing depth profiles across these regions (Supplementary Figs. S11–S13), as well as the localization of the duplicated BUSCO genes along the chromosomes, and did another run of CRAQ, this time using ONT as long reads (with higher coverage compared with the HiFi reads used in the previous rounds). There was no significant difference in the number of duplicated BUSCO genes within Bs1 and Bs3 compared to other genomic regions, and CRAQ did not detect structural errors in these scaffolds either. This suggests that the palindromes observed are real, with potential biological significance that will require further investigation.

Molecular identification as subclade A1

B. schlosseri is considered a species complex comprising 5 genetically distinct clades (A to E), each representing a cryptic species with its own characteristic geographic distribution [42, 43]. Detailed analysis of cytochrome *c* oxidase subunit I (COI) mitochondrial sequences divides clade A into 3 distinct subclades: A1,

Table 3: Classes of repeats in the *Botryllus schlosseri* genome. RepeatMasker summary table for the collapsed genome assembly of *Botryllus schlosseri* showing the percentages of identified repeat classes.

Repeat class	Percentage of genome
Long Interspersed Nuclear Elements (LINEs)	**4.52%**
LINE1	0.15%
LINE2	2.06%
Long Terminal Repeats (LTRs)	**1.34%**
DNA elements	**7.24%**
hAT-Charlie	2.96%
TcMar-Tigger	0.01%
Unclassified	**46.03%**
Total interspersed repeats	**59.12%**
Simple repeats	**3.94%**
Low complexity	**0.02%**
Total	**63.09%**

A2, and A3 [44]. The complete mitochondrial DNA of clone E∗ was recovered and assembled as a single circular contig. Our mitogenome assembly shares 99.95% identity with the published mitochondrial sequence assigned to the *B. schlosseri* subclade A1 [44]. Notably, this subclade includes the sc6ab specimen used by Voskoboynik et al. [28] to generate the previous reference assembly of *B. schlosseri*. Our mitogenome assembly further shares 99.7% nucleotide identity with that reference sequence. Phylogenetic analyses based on a COI fragment used as DNA barcode for ascidians ([44]) confirmed that sample E∗ belongs to subclade A1 (Supplementary Fig. S7), a group that is both widely distributed and employed as a laboratory model worldwide.

Structural and functional annotation

Using a *de novo* repeat library created by RepeatModeler (RRID:SCR_015027), RepeatMasker (RRID:SCR_012954) detected that around 63% of the novel *B. schlosseri* collapsed genome assembly consists of repetitive elements, which is close to the 65% of repeats found in the previously published assembly [28]. Most of these were interspersed repeats (see Table 3). A relatively high abundance of repetitive sequence was also reported in other colonial tunicates. For instance, *Salpa thompsoni* and *Salpa aspera*, both colonial species, possess a larger genome (742 Mb and 901 Mb, respectively) and an higher repeat content (ca. 80%) compared to solitary tunicates such as *Ciona robusta* (ca. 160 Mb, about 20–25% repeats) or *Oikopleura dioica*, which has a compact genome of 70 Mb with only ca. 15% repetitive content. This pattern suggests that colonial tunicates exhibit a greater genomic expansion and a larger repeat content than their solitary counterparts. Yet, the colonial *Botrylloides diegensis*, which carries a relatively small genome [45], and the solitary *S. clava*, with 46.6% repetitive elements, represent notable exceptions. Additional high-quality genome assemblies across a broader range of tunicate species will be essential to confidently assess the possible association between coloniality and repeat content [46–48].

Ab initio genome annotation using the BRAKER3 pipeline [49] (RRID:SCR_018964) initially predicted 16,966 coding genes, after which refinement using the PASA pipeline [50, 51] (RRID:SCR_014656) finally retrieved 22,275 genes coding for 30,813 proteins (see Table 4). This number is significantly lower than originally predicted for *B. schlosseri* (38,730 predicted genes [28]), probably

Table 4: Gene predictions and annotation statistics

Type	Number	Mean size (bp)	% genome
Gene	22,275	8,566.13	35.78
mRNA	30,813	10,576.62	N/A
CDS	237,200	199.16	8.86
Exon	241,815	289.83	13.14
5′ UTR	21,386	432.29	1.73
3′ UTR	20,985	648.00	2.55
Total	574,474	1,143.44	N/A

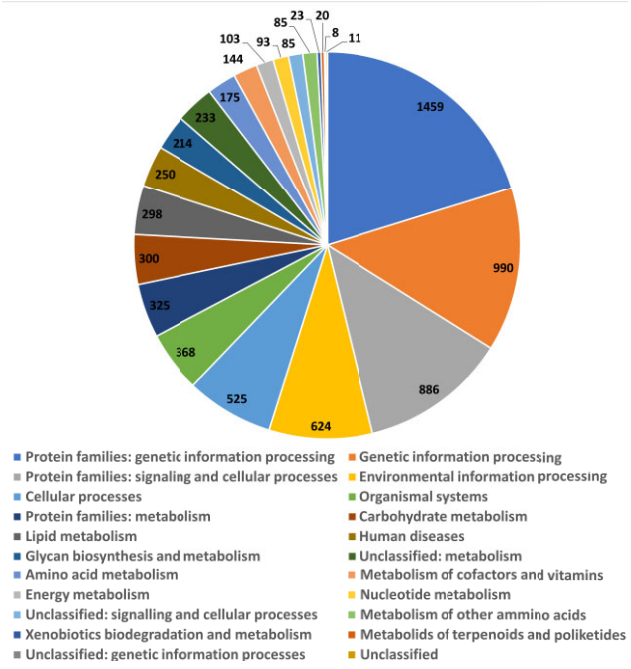

Figure 6: Pie chart of the assignation of the annotated genes of *Botryllus schlosseri* to KEGG functional categories using BlastKOALA [55].

due to the incomplete collapse of the previous assembly. In terms of completeness of the annotation, BUSCO retrieved 92.4% complete (79.7% single, 12.7% duplicated) and 1.8% fragmented metazoan genes when given all predicted isoforms, whereas it retrieved 92% complete (91% single, 0.9% duplicated) and 1.8% fragmented metazoan marker genes when filtered to only keep the longest isoform. Running BUSCO directly on the scaffold sequences yielded similar results (data not shown).

The functional annotation and orthology assignment [52], coupled with annotation of protein domains, motifs, and functional sites [53, 54], were written into gff3 and Genbank files. KEGG route-mapping assigned 7,221 genes over the annotated entries and distributed them across 21 KEGG categories (Fig. 6). Among them, the most prevalent ones include KEGG hierarchies dealing with genetic information processing (2,449/7,219, 22.92%), such as DNA replication, repair, recombination, transcription, translation, and regulation of gene expression; signaling and cellular processes (886/7,219, 12.27%); and environmental information processing (674/7,219, 8.64%), such as various cellular processes and signaling pathways involved in sensing, transducing (i.e., MAPK signaling, PI3K-Akt signaling, and cAMP signaling), responses to external signals (i.e., G-protein coupled receptors, receptor tyrosine kinases, and cytokine receptors), intracellular

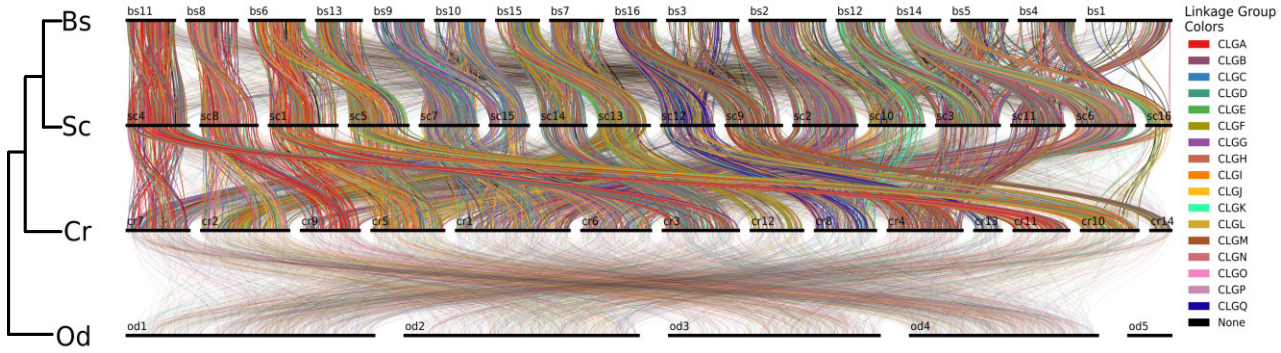

Figure 7: Synteny analyses using CLGs between *Botryllus schlosseri* (Bs), *Styela clava* (Sc), *Ciona robusta* (Cr), and *Oikopleura dioica* (Od). For each species, the horizontal black lines represent the chromosomes, while the colored vertical lines connect conserved orthologs between species pairs. Each color corresponds to one of the 17 ancestral CLGs identified in [58]. The opacity of the lines indicates the significance of the interaction between interspecies chromosomes, with solid colors representing significantly enriched conservation of synteny.

communication, and cell motility. The KEGG annotations provided for *B. schlosseri* are consistent and coherent with the functional annotation of the published complete genomes of other ascidian tunicates, such as *Styela clava*, *Ciona robusta*, and *Oikopleura dioica* (Supplementary Fig. S8).

Haplotype-resolved assembly

Given its heterozygosity level exceeding 3%, haplotype-resolved assemblies of *B. schlosseri* are crucial for studying differences between homologous chromosomes, such as structural variations. Using hifiasm with direct integration of Hi-C reads and subsequent scaffolding (Supplementary Fig. S9), we generated a pair of chromosome-scale, haplotype-resolved assemblies (haplotype 1 and haplotype 2), each organized into 16 major scaffolds (see Supplementary Fig. S10). With respective sizes of 496 Mbp and 494 Mbp, these assemblies are smaller than the collapsed assembly (533 Mbp). When considering only the 16 longest scaffolds, the sizes decrease to 480 Mbp for haplotype 1 and 464 Mbp for haplotype 2, compared to 513 Mbp for the collapsed assembly. Additionally, their BUSCO completeness scores are lower, with values of 90.9% and 91.2%, respectively, compared to 91.6% for the collapsed assembly. This is further reflected in their annotation results, where fewer genes were identified: 21,802 and 21,831 for haplotype 1 and haplotype 2, respectively, versus 22,275 for the collapsed assembly (see Supplementary Table S1). The observed differences in metrics, where the results for the haplotype-resolved assemblies are inferior to those for the collapsed assembly, may be attributed to misassemblies, particularly deletions. For example, when comparing the putative chromosome lengths (see Supplementary Table S2) for chromosomes 1 and 3, we observe a significant disparity in sizes between the 2 haplotypes, which may be attributed to incomplete sequence reconstructions during the assembly process. Such anomalies may additionally be observed when comparing the putative chromosome lengths of all assemblies with the karyogram of *B. schlosseri*, as described by Colombera [35] (see Supplementary Fig. S14). Notably, the sizes of the collapsed assembly appear to more closely match the expected distribution compared to the phased haplotypes. Furthermore, multiple structural variations between the 2 haplotypes, particularly small inversions (see Supplementary Figs. S15 and S16), seem to be present in the majority of the homologous chromosomes. However, as with the observed putative deletions, these may result from misassemblies and require further validation to enhance the quality of the haplotype-resolved assembly.

Synteny analyses

To assess macrosynteny conservation between *B. schlosseri* and other tunicates, we selected genomes that met 2 specific criteria: they were assembled at the chromosome level, ensuring comparable high-quality structural information, and they represented, as much as possible, the breadth of diversity within the tunicate subphylum. *S. clava* [56] belongs to the same order as *Botryllus* (Stolidobranchia), *C. robusta* [46] to a different order (Phlebobranchia), and *O. dioica* [47] to a different class of tunicates (Appendicularia) [57]. We used 17 groups of orthologous genes identified by Simakov et al. [58] as ancestral chordate linkage groups (CLGs). These groups of genes are thought to have remained physically linked since the divergence of the Olfactores lineage (which includes both vertebrates and tunicates) from cephalochordates. However, Oxford dot plots [59] revealed a general loss of syntenic equivalence [60] among tunicate genomes, even between *B. schlosseri* and *S. clava*, which share the same haploid chromosome number of 16. Despite this identical number of chromosomes, the comparison between the 2 stolidobranchs showed extensive chromosome rearrangements, including fissions and fusions with mixing [60, 61] (Fig. 7 and Supplementary Fig. S17). These rearrangements are even more pronounced in *C. robusta*, which has a haploid chromosome number of 14. The overall random distribution of ortholog pairs within blocks points to significant order scrambling, resulting in a loss of colinearity (i.e., the sequential order of genes along the same chromosome); the comparison with *O. dioica* shows a complete breakdown of both macrosynteny and colinearity, with CLGs fully scrambled and dispersed. The latter result is consistent with the very long and fast-evolving branch of Appendicularia compared to other tunicates [57], as well as with the extreme genome scrambling rate of Appendicularia compared to other tunicates and mammals [62]. The same analyses using a set of 29 linkage groups generally conserved among bilaterians, cnidarians, and sponges [60] yielded similar results (Supplementary Fig. S18). The extensive physical linkage of groups of orthologous genes has been shown to be conserved across highly divergent bilaterian phyla, including Chordata, Echinodermata, Mollusca, and Nemertea [60, 61]. Notably, our preliminary synteny analyses across 4 tunicate species reveal a highly dynamic genomic landscape, where syntenic equivalence, defined as one-to-one chromosomal correspondence regardless of gene order, is largely disrupted, even among species within the same family. Frequent chromosomal fission and fusion events further underscore the rapid evolutionary turnover of

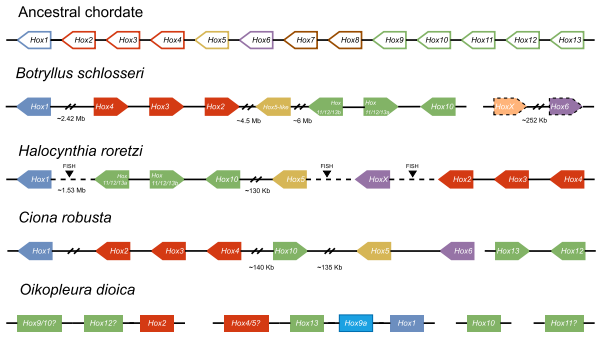

Figure 8: Representation of the Hox genes retrieved in the new assembly of *B. schlosseri* compared to the supposed original single Hox cluster of the chordate ancestor and other tunicates. Linked genes (present on the same scaffold) are connected by a solid line, while a dashed line is used when the linkage has been deduced using another method. When known, the transcription orientation is indicated by an arrow-shaped rectangle, which is surrounded by a dashed line when the Hox gene was retrieved with low confidence.

tunicate genomes. The increasing erosion of macrosynteny with phylogenetic distance suggests that patterns of conserved chromosomal linkage could serve as informative characters for phylogenetic inference. Interestingly, a similar pattern of genome rearrangements was recently reported in Bryozoa [61] and in clitellate annelids [63–65], pointing to a potential parallel and independent loss of the ancestral bilaterian genome architecture in these lineages and in tunicates. These observations raise compelling questions about the underlying mechanisms driving such rearrangements, which may reflect a relaxation of the selective constraints typically maintaining gene order in other metazoan groups [66].

Hox gene analyses

Hox genes are a subset of homeobox genes that play important developmental roles in the specification of body segments along the anterior-posterior axis. Their arrangement into a syntenic cluster colinear with gene expression is conserved across Bilateria, with some exceptions [67]. In the new collapsed assembly, we retrieved 10 *B. schlosseri* Hox genes, which is consistent with draft genomes of other ascidian tunicates [68]. Orthology of *B. schlosseri* Hox genes was assessed using phylogenetic analyses, as in Sekigami et al. [69], based on Hox tree topology among the tunicates *C. robusta* and *Halocynthia roretzi*, the cephalochordate *Branchiostoma lanceolatum*, and 3 vertebrate species. The names of the *B. schlosseri* Hox genes were assigned based on their proximity to the ones of *C. robusta* (Supplementary Figs. S19 and S20). However, most branches had low bootstrap support, and therefore including more tunicates as well as vertebrate species will be necessary to resolve the complex evolution of the Hox gene cluster across tunicates [68]. Although Hox genes are colinear between cephalochordates and vertebrates, it is not the case for tunicates [70]. In the tunicate species studied thus far, Hox clusters exhibit divergences in terms of colinearity and synteny relative to the ancestral chordate cluster [68]. In contrast to previous data [28, 45], our new assembly revealed that *B. schlosseri*'s Hox genes are less scattered than previously described, suggesting improved contiguity in the new genome assembly. Eight of them are grouped on the second largest scaffold (Bs2), yet for some of them at a relatively large distance, whereas 2 other ones are found on the 15th largest scaffold (Bs15) (Fig. 8). Comparison with 2 tunicate ascidians, belonging to the same (*H. roretzi* [69]) and a different (*C. robusta* [46]) order, revealed partially conserved synteny as well as inversions and transpositions across the 3 species (Fig. 8). These observations agree with the general trend of synteny conservation despite loss of colinearity observed for CLGs [58] and are also consistent with the phylogenetic relationships among the species sequenced [2, 57]. Yet, the limited availability of chromosome-level genome assemblies continues to hinder a clear picture of the evolutionary dynamics of the Hox clusters across tunicates. Altogether, these findings show that *B. schlosseri* follows the general tunicate trend of dispersed and rearranged Hox clusters, but with a more clustered configuration than previously thought. This could reflect lineage-specific retention of partial clustering and provides a more refined view of the dynamic genomic architecture in tunicates. While colinearity was clearly lost, partial synteny and clustering remain, offering a potential window into the mechanisms and consequences of Hox cluster disintegration during chordate evolution.

Conclusion

Tunicate genomes are known for their rapid evolution, featuring high rates of molecular divergence and extensive genomic rearrangements, and they are generally remarkably compact compared to vertebrates, though genome size varies among tunicate species [71]. Additionally, while some tunicates exhibit high levels of repetitive elements, others show moderate repeat content [45, 66]. Despite these variations, tunicate genomes share conserved noncoding elements, reflecting deep regulatory constraints within this diverse subphylum [72]. Although solitary tunicates such as *Ciona* and *Oikopleura*, along with other species, have been instrumental in shaping our understanding of tunicate genomes, colonial tunicates remain relatively understudied at the genomic level. Colonial species also introduce unique biological questions related to allorecognition, asexual reproduction, and whole-body regeneration. As a widely used model for colonial tunicates, *B. schlosseri* provides an essential reference for studying these processes, making a high-quality genome assembly particularly valuable. Comparative synteny analyses highlight both conserved and highly rearranged genomic features across tunicates, reinforcing the notion of their exceptional genomic plasticity. By making this resource available, we aim to facilitate future research into the evolutionary and functional genomics of chordates, also highlighting unique adaptations that define tunicate biology.

Methods

Sampling, DNA isolation, and sequencing

Isogenic colonies of *B. schlosseri* were raised on glass slides in the marine-culture system described in Langenbacher et al. [21]. Genomic DNA was extracted from the colony labeled E* using Qiagen's MagAttract HMW DNA Kit (67563). Libraries were prepared and sequencing was performed at Novogene for Illumina 2 × 150-bp paired-end (PE) reads, at the Next Generation Sequencing Platform of the University of Bern (Switzerland) and Leiden Genome Technology Center (Leiden, Netherlands) for HiFi PacBio long reads in round 1 and round 2, respectively (PacBio Sequel II, SMRT-bell library), and at UCAGenomix (Valbonne, France) for Oxford Nanopore (ONT) long reads (on a FLO-PRO002 flow cell with R9.4.1 pore proteins, using the SQK-LSK109 ligation sequencing kit). Nanopore base calling was performed using Guppy (RRID:SCR_023196, v3.2.10). A Hi-C library was prepared using the Arima High Coverage HiC Kit (A410110), followed by the Arima

HiC+ Kit (A510008, A303011), and sequenced using Illumina (2 × 150 bp).

Data preprocessing

PacBio HiFi reads were processed with HiFiAdapterFilt v2.0.1 [73] to remove adapter sequences, while Porechop (RRID:SCR_016967, v0.2.4) was used to trim basic adapters from ONT reads. For Illumina reads, quality trimming and adapter clipping were performed using Trimmomatic [74] (RRID:SCR_011848, v0.39), while quality check, prior to and after trimming, was done using FastQC (RRID:SCR_01458 v0.11.5).

Genome size estimation

The genome size of colony E* was measured using an improved Feulgen protocol [75] by comparison with 2 standards of known C-values: *Periplaneta americana* (3.41 pg) [76] and *Lasius niger* (0.30 pg) [77]. In brief, the protocol steps included chopping the tissues of each specimen into tiny pieces using a sterilized razor blade with a few drops of 40% acetic acid, then leaving them for 48 hours in the dark, and immersing the processed slides into fixation reagent (85:10:5 volumes of methanol/formaldehyde/acetic acid), then hydrolyzing them (using hydrochloric acid 5M) and staining them (using Schiff's reagent).

A digital camera (5 megapixels) mounted on a compound microscope with a 100× objective was used for imaging the slides. During the photography sessions, we maintained constant camera settings for exposure and gain, white balance calibration parameters, microscope light intensity, light condenser, and focal lens positions. In the image analysis protocol, we first outlined the nuclear boundary using the polygon tool in ImageJ [78], then extracted from ImageJ the area size of the nucleus (ASN) and the mean gray value of the nucleus (GVN). Next, we outlined in ImageJ a doughnut-shaped area surrounding the same nucleus and used it to extract the mean gray value of its background (GVB). This process was repeated for up to 30 nuclei per sample. The difference between GVB and GVN is an estimate of the average optical density (OD) of a nucleus; multiplying it by its ASN yields its integrated optical density (IOD), which is proportional to the amount of DNA in this nucleus.

Comparison of IOD values of the sample with those of the standards allows us to calculate the genome size of the sample, provided that 2 assumptions are verified: (i) all the nuclei of a given specimen contain about the same amount of DNA, and (ii) the IODs of nuclei of the standards are proportional to their known C-values. To check the first assumption of the method, we used a R script to plot for each specimen the 1/OD values of their nuclei versus their ASN values and verify that the resulting linear regression passed through the origin of the plot (Supplementary Fig. S3). To check the second assumption, we plotted the average IOD of each standard versus their known C-value and verified that the resulting line passed through the origin of the plot (Supplementary Fig. S4). As both assumptions of the method were met, we proceeded to estimating the C-value of the sample: for that, we divided the IOD of each nucleus of each standard by its known genome size, resulting in a set of 60 integrated optical densities divided by C-values (IOD/C). Finally, we used a R script to divide each of the 30 IODs of the sample by each of the 60 IOD/C values of the standards, then plotted the distribution of the resulting 1,800 estimated C-values of the sample and took the mode of its Gaussian kernel density as the most likely genome size.

A genome size estimation based on the *k*-mer spectrum of the Illumina reads was also performed using KMC v3.2.1 [79] and the GenomeScope2.0 [80] web server, with a *k*-mer size of 21 and a *k*-mer count cutoff of 100,000.

Collapsed genome assembly

First, the PacBio HiFi reads were assembled into contigs using hifiasm [31] with the haplotype purging option disabled (option -l0 with hifiasm in HiFi-only assembly mode). Second, uncollapsed haplotypes were purged using multiple rounds of HaploMerger2 (release 20180603) [81] until the BUSCO duplication score stabilized. Third, nonmetazoan contigs were identified and removed from the assemblies using BlobToolKit v4.1.5 [82]. To this aim, contigs were aligned to the NCBI nucleotide database (accessed 18 March 2023) using BLAST [83] (RRID:SCR_001653, v2.13.0+) with the blastn command, as well as to the UniProt reference proteome database (accessed 23 March 2023) using DIAMOND [84] (RRID:SCR_016071, v2.1.6); contig HiFi coverage depth was computed using minimap2 v2.24-r1122 [85]. Using the "bestsumorder" rule of BlobToolKit, only the contigs assigned to the taxon "Chordata" or without a match ("no-hit") were kept. Finally, a BLASTN search for fragments of the mitochondrial genome among the contigs was performed using the published complete mitochondrial genome of *B. schlosseri* (RefSeq NC_021463.1) [28] to remove contigs showing at least 80% coverage and identity with the query sequence.

To scaffold the assemblies, PacBio HiFi and Illumina reads were first mapped to the assemblies using minimap2. Putative misjoined regions were then identified and automatically split using CRAQ v1.0.9 [33] with default parameters, except for the addition of –break. Hi-C reads were subsequently mapped to the output of CRAQ using the Arima Genomics mapping pipeline script arima_mapping_pipeline.sh [86], and YaHS v1.2 [34] was run with default parameters to scaffold the assemblies. CRAQ was then applied to the results, and finally the scaffolds were manually curated using PretextMap (RRID:SCR_022023,v0.1.9) and PretextView (RRID:SCR_022024, v0.2.5). Metrics for the assemblies were computed using SeqKit v2.3.0 [87] (parameter stats -a). The quality and completeness were checked using KAT v2.4.2 [88] on *k*-mers from both PacBio HiFi and Illumina reads, as well as BUSCO v5.4.4 [89] (using the -m genome mode) with the metazoa_odb10 dataset.

Haplotype-resolved assembly

Two haplotype-resolved assemblies (haplotype 1 and haplotype 2) were generated using hifiasm in Hi-C Integrated Assembly mode, which directly integrates Hi-C reads. To refine the assemblies, uncollapsed sequences were purged for haplotype 1 using purge_dups [90], and BlobToolKit was employed, as with the collapsed assembly, to filter out contamination, resulting in contig-level assemblies (see Supplementary Figs. S9 and S6). The scaffolding process for haplotype 1 and haplotype 2 followed the same method as for the collapsed assembly, with the final scaffolds ordered based on alignment to the collapsed assembly rather than by descending size (see Supplementary Fig. S15).

Genome annotation

For all the assemblies, repetitive elements were identified using RepeatModeler and RepeatMasker pipeline. A *de novo* repeat library was generated using RepeatModeler2 v2.0.3 [91] and used as input for RepeatMasker (SCR_012954, v4.0.6) to detect, classify, and soft-mask repeats in the genomic sequences. RNA sequencing (RNA-seq) reads were aligned to the soft-masked assemblies using STAR v2.7.10b (default options) [92]. Based on the aligned transcripts, on a list of proteins from OrthoDB v11 for

Metazoa [93] as extrinsic evidence and on the soft-masked assemblies, genes were predicted and annotated using the BRAKER3 v3.06 pipeline for RNA-seq and protein data without training or gene prediction with untranslated region (UTR) parameters [49, 94–106]. A refinement of the initial BRAKER3 structural annotation and the addition of UTRs were then performed with an implementation of the PASA pipeline v2.4.1 [50], together with EVidenceModeler (EVM) [51] (RRID:SCR_014659, v2.1.0). A third of the RNA-seq reads of the Rodriguez et al. [23] transcriptome was aligned again to the assemblies and their BRAKER3 annotation using STAR (MAX_INTRON_SIZE=20000) [92] (RRID:SCR_015899, v2.7.10b) and assembled with StringTie [107] (RRID:SCR_016323, v2.2.1) using the BRAKER3 annotation as a reference. The PASA alignment assembly step was then run as described on its GitHub Wiki with the transcripts assembled by StringTie and independently with Trinity assemblies of publicly available RNA-seq reads [8, 23, 25]. TransDecoder [108] was run within PASA to identify coding sequences within the assembled transcripts. A consensus annotation of coding sequences (CDSs) was found by EVM by leveraging both the transcripts and coding sequences identified for each RNA-seq by PASA (evidence weights: 1 for BRAKER3 input, 5 for PASA transcripts and TransDecoder CDSs). The gene models were refined, with addition of the UTRs and isoforms, by running the PASA genome annotation step sequentially with each previously generated PASA database (using EVM output as the first reference, then the output of the previous PASA genome annotation run). Functional annotation was performed starting from the structural annotation obtained with the BRAKER3-PASA pipeline. Eggnog-mapper [52, 109] and Interproscan [53, 54] were used for orthology-based annotation (nr, KEGG, Gene Ontology terms) and for protein domains prediction, respectively. Both approaches were used as input for the Funannotate pipeline (RRID:SCR_023039, v1.8.15), yielding a gff3 and a GenBank file with functional annotations.

Mitochondrial genome assembly

The mitochondrial genome was reconstructed using NOVOPlasty [110] (RRID:SCR_017335, v4.3.1). A COI fragment from *B. schlosseri* clade A1 (GenBank MT731471.1) was used as a seed in combination with our Illumina reads as input.

Comparative genomics analyses

The genome assemblies and annotations for the comparison of the collapsed assembly with other tunicate species were retrieved from ANISEED [111] for *Botrylloides leachii*, *C. robusta*, and for the first assembly of *B. schlosseri*, while *O. dioica* originates from [47], *S. thompsoni* from [48], and *S. clava* from [56]. Macrosynteny analyses were performed using the odp tool [59]. For each species, analyses were based on the longest protein isoforms generated from their annotation file using the scripts agat_sp_keep_longest_isoform.pl and agat_sp_extract_sequences.pl from AGAT (RRID:SCR_027223, v0.7.0).

Phylogenetic analyses

COI fragments were retrieved from [44] and aligned with MUSCLE [112]. A maximum likelihood tree was generated using MEGA5 [113] with the model HKY+I+G followed by 1,000 bootstrap replicates. Phylogenetic analyses of *B. schlosseri* Hox genes were performed using sequences retrieved from Sekigami et al. [69]. First, the sequences were aligned using MUSCLE [112], as implemented in AliView [114], and then IQ-TREE 2 [115] was used to build a maximum likelihood phylogeny with the best-fit model JTT+R6 [116, 117], selected by ModelFinder [118], following the Bayesian information criterion [119] and with 10,000 ultrafast bootstrap replicates [120]. The same alignment was used to build a Bayesian tree using MrBayes (RRID:SCR_012067, V.3.2.7) (gamma-distributed rate variation across sites; mixed AA substitution models).

Additional Files

Supplementary Fig. S1. Genomescope2.0 results obtained with the Illumina reads, a *k*-mer length of 21, and a maximum counts of 100,000.

Supplementary Fig. S2. Output of the KAT comp tool comparing the *k*-mers found in the Illumina and HiFi reads to those present in the collapsed (top), haplotype 1 (middle), and haplotype 2 (bottom) assemblies of *B. schlosseri*. The *k*-mer completeness, based on the highest peak (corresponding here to heterozygous *k*-mers), is respectively (from top to bottom) 53.03%, 47.94%, and 46.92%. A perfectly correct haploid representation should have a *k*-mer completeness of 50%.

Supplementary Fig. S3. Linear regressions confirming that the total amount of DNA coloration per nucleus is constant for each species, regardless of nuclear size.

Supplementary Fig. S4. Linear regression confirming that the integrated optical density of each standard is proportional to its known C-value.

Supplementary Fig. S5. Genome size histogram of *Botryllus schlosseri* obtained using Feulgen microphotodensitometry.

Supplementary Fig. S6. BlobPlots of the assemblies of *B. schlosseri*. *Initial* refers to results obtained before filtering out contamination. *Kept* represents the contigs retained in the assemblies before scaffolding, while *Removed* represents those discarded as contamination.

Supplementary Fig. S7. Maximum likelihood tree of *Botryllus schlosseri* clades and subclades reconstructed from COI sequences [44]. Branches shows boostrap values. Accession ID are indicated between parentheses.

Supplementary Fig. S8. Comparison of the percentage of genes of *Botryllus schlosseri*, *Ciona robusta*, *Oikopleura dioica*, and *Styela clava* assigned to different KEGG functional categories by BlastKOALA [55].

Supplementary Fig. S9. Assembly pipeline used to generate the contig-level assemblies of haplotype 1 and haplotype 2. The downstream steps (not shown) to produce scaffold-level assemblies are identical to those used for the collapsed assembly.

Supplementary Fig. S10. Hi-C heatmaps of the haplotype 1 (left) and haplotype 2 (right) assemblies, showing 16 chromosome-scale scaffolds for both.

Supplementary Fig. S11. Representation of the 2 largest palindromic regions on the sequence Bs1, based on the syntenic blocks identified by MCScanX [38] (shown in green and purple). Coverage was calculated using ONT reads, and the curve, which was smoothed using a rolling mean with a window size of 100,000 bp, does not show major deviations in the palindromic regions compared to the average coverage across the entire sequence (indicated by the dashed horizontal line). The gene names marking the start and end of each region are labeled. For example, the block extending from gene Boschl.Bs1.g184.t1 to Boschl.Bs1.g237.t1 (first green rightward arrow) is syntenic with the block from Boschl.Bs1.g370.t1 to Boschl.Bs1.g433.t1 (second green leftward arrow) in reverse order.

Supplementary Fig. S12. Representation of the large palindromic region on sequence Bs3. In (a), it is plotted in the same manner

as in Supplementary Fig. S11. In (b), the same data are shown without smoothing the coverage curve and without restricting the coverage scaling to 200×. The large peak around position 16.6 Mb corresponds to a region highly enriched in monomers likely to be centromeric repeats and is located between 2 putative topologically associating domains (see Supplementary Fig. S13).

Supplementary Fig. S13. (a) Tandem repeat region sizes along the sequence Bs3, based on monomers likely to be centromeric repeats and identified using quarTeT CentroMiner [123] on the collapsed assembly. A long repetitive region is observed between 16 and 17 Mb. (b) Zoom-in on the Hi-C heatmap of sequence Bs3, spanning from 12 to 22 Mb and displayed with PretextView (RRID:SCR_022024, v0.2.5), where 2 putative topologically associating domains (TADs) have been manually highlighted with red lines. The gap between the 2 putative TADs extends approximately from 16.514 to 16.595 Mb.

Supplementary Fig. S14. Comparisons between the 16 longest scaffolds from the collapsed, haplotype 1, and haplotype 2 assemblies and the karyogram of Colombera [35]. The lengths of the bars were calculated as the proportion (in percentage) of each chromosome's length relative to the total genome length. The order of scaffolds for haplotype 1 and haplotype 2 is based on the sizes of the scaffolds in descending order, rather than their alignment to the collapsed assembly.

Supplementary Fig. S15. D-GENIES [124] dot plots of the final alignments: haplotype 1 vs. the collapsed assembly (top), haplotype 2 vs. the collapsed assembly (middle), and haplotype 1 vs. haplotype 2 (bottom). These were used to assess synteny and guide scaffold ordering.

Supplementary Fig. S16. AccuSyn [37] representation of syntenic blocks identified using MCScanX [38] between the 16 largest scaffolds of haplotype 1 (left, with scaffold names ending in "A") and haplotype 2 (right, with scaffold names ending in "B") assemblies. Inverted blocks are highlighted in red.

Supplementary Fig. S17. Investigation of synteny conservation among tunicate genomes. In the first column, dot plots depict the chromosome-scale scaffolds of *Botryllus schlosseri* (x-axis) plotted against those of *Styela clava*, *Ciona robusta*, and *Oikopleura dioica* (y-axis). Each dot in the plot represents an ortholog, specifically a reciprocal best diamond blastp match between 2 species. The units of the x- and y-axes are the number of orthologous proteins: 9,813, 5,772, and 4,064 orthologs found between the 16 chromosome-scale scaffolds of *B. schlosseri* and the 16 of *S. clava*, the 14 of *C. robusta*, and the 5 of *O. dioica*, respectively. If there were chromosome breaks, Fisher's exact test (FET) was used to calculate the significance of the interactions between the subchromosomal pieces. Otherwise, FET was calculated on whole chromosomes. The opacity of the dots depicts the significance of FET. Dots that are a solid color are in cells with a FET *P* value less than or equal to 0.05. Dots that are translucent are in cells with a FET *P* value greater than 0.05. Dx and Dy values allow us to pinpoint places where there may be sudden breaks in synteny [58]. The second column of the figure depicts the same information as the first one, but plotted following chromosome base pair coordinates rather than gene index. This is better suited for visualizing gene-poor regions of the chromosomes.

Supplementary Fig. S18. Synteny conservation of bilaterian, cnidarian, and sponge linkage groups (BCnS LGs) between *Botryllus schlosseri* (Bs), *Styela clava* (Sc), *Ciona robusta* (Cr), and *Oikopleura dioica* (Od). For each species, the horizontal black lines represent the chromosomes, while the colored vertical lines connect conserved orthologs between species pairs. Each color corresponds to 1 of the 29 ancestral BCnS LGs identified in [60]. The opacity of the lines indicates the significance of the interaction between interspecies chromosomes, with solid colors representing significantly enriched conservation of synteny.

Supplementary Fig. S19. Phylogenetic analyses of Hox gene candidates of *Botryllus schlosseri*. The ML tree was generated using IQ-TREE 2 [115] by adding the *B. schlosseri* sequences to the alignment of Sekigami et al. [69] and keeping the homeodomains as well as the flanking 20 N-terminal and 7 C-terminal amino acids. Ultrafast bootstrap values are shown in red.

Supplementary Fig. S20. Phylogenetic analyses of Hox genes candidates of *Botryllus schlosseri*. The Bayesian tree was generated using MrBayes [125] by adding the *B. schlosseri* sequences to the alignment of Sekigami et al. [69] and keeping the homeodomains as well as the flanking 20 N-terminal and 7 C-terminal amino acids. Posterior probabilities are shown in red.

Supplementary Table S1. Metrics for the collapsed, haplotype 1, and haplotype 2 assemblies.

Supplementary Table S2. Comparison of the putative chromosome sizes (in kbp) across the 3 different assemblies. The putative chromosomes correspond to the 16 longest scaffolds, ordered in descending size for the collapsed assembly. For the haplotype 1 and haplotype 2 assemblies, the scaffold order is based on their alignment to the collapsed assembly, with percentages in parentheses indicating their size relative to the reference collapsed assembly.

Supplementary Table S3. Assembly statistics of the new collapsed assembly of *Botryllis schlosseri* compared to the existing chromosome-level reference assemblies of *Styela clava*, *Ciona robusta*, and *Oikopleura dioica*.

Abbreviations

ASN: area size of the nucleus; BLAST: Basic Local Alignment Search Tool; BUSCO: Benchmarking Universal Single-Copy Orthologs; CDS: coding sequences; CLG: chordate linkage group; EVM: EVidenceModeler; GVN: gray value of the nucleus; IOD: integrated optical density; KEGG: Kyoto Encyclopedia of Genes and Genomes; NCBI: National Center for Biotechnology Information; OD: optical density; ONT: Oxford Nanopore; PE: paired-end; RNA-seq: RNA sequencing; UTR: untranslated region.

Acknowledgments

We thank EMBRC-France and in particular Laurent Gilletta for isolating isogenic colonies and maintaining the aquaculture system. Thanks also to the Next Generation Sequencing Platform of the University of Bern (Switzerland) for providing part of the HiFi sequencing. We thank Aaron Reinke for pointing out the presence of macrosporidia sequences in the BloobToolKit analyses, Vitoria Tobias Santos for filtering part of the RNA-seq dataset used for the annotation, Carmela Gissi and Lino Ometto for useful scientific exchange, and the 3 reviewers for useful feedback.

Author Contributions

O.D.T. carried out the majority of the assembly and analyses. S.T., O.D.T., and J.F.F. conceived the project and drafted the manuscript with the contribution of M.L. M.M.T. conducted the Feulgen analyses. S.B. assisted with the initial stages of the assembly and provided part of the HiFi dataset. M.L. and P.D. handled the annotation and contributed to the analyses. A.A., F.D.B., and R.F. provided

valuable technical and scientific insights. S.T. and J.F.F. supervised the research. All authors reviewed and approved the final version of the manuscript.

Funding

This work was supported by ANR (ANR-14-CE02-0019-01 and ANR-24-CE02-2277), INSB-DBM and Sorbonne University AAP Emergence 2021 to S.T., FAPESP 15/50164-5 & 19/06927-5 to F.D.B., and the Fonds de la Recherche Scientifique (F.R.S.-FNRS) via PDR grant T.0078.23 to J.F.F.

Data Availability

The genomic and transcriptomic sequence data generated in this study are available under the BioProject accessions: PRJNA1225683. The gene expression data utilized in this study are available from The Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under the following accessions: GSE62112, GSE193805. All additional supporting data are available in the *GigaScience* repository, GigaDB [121], and in Octopus [122].

Competing Interests

The authors declare that they have no competing interests.

References

1. Delsuc F, Brinkmann H, Chourrout D, et al. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature. 2006;439(7079):965–68. https://doi.org/10.1038/nature04336.
2. Alié A, Hiebert LS, Scelzo M, et al. The eventful history of nonembryonic development in tunicates. J Exp Zool Part B Mol Dev Evol. 2021;336(3):250–66. https://doi.org/10.1002/jez.b.22940.
3. Stolfi A, Brown FD. Tunicata. In: Wanninger A, ed. Evolutionary developmental biology of invertebrates 6: Deuterostomia. Vienna, Austria: Springer; 2015:135–204. https://doi.org/10.1007/978-3-7091-1856-6_4.
4. Hiebert LS, Simpson C, Tiozzo S. Coloniality, clonality, and modularity in animals: the elephant in the room. J Exp Zool Part B Mol Dev Evol. 2021;336(3):198–211. https://doi.org/10.1002/jez.b.22944.
5. Manni L, Gasparini F, Hotta K, et al. Ontology for the asexual development and anatomy of the colonial chordate *Botryllus schlosseri*. PLoS One. 2014;9(5):e96434. https://doi.org/10.1371/journal.pone.0096434.
6. Sabbadin A, Zaniolo G, Majone F. Determination of polarity and bilateral asymmetry in palleal and vascular buds of the ascidian *Botryllus schlosseri*. Dev Biol. 1975;46(1):79–87. https://doi.org/10.1016/0012-1606(75)90088-3.
7. Nourizadeh S, Kassmer S, Rodriguez D, et al. Whole body regeneration and developmental competition in two botryllid ascidians. EvoDevo. 2021;12(1):15. https://doi.org/10.1186/s13227-021-00185-y.
8. Ricci L, Salmon B, Olivier C, et al. The onset of whole-body regeneration in *Botryllus schlosseri*: morphological and molecular characterization. Front Cell Dev Biol. 2022;10:173. https://doi.org/10.3389/FCELL.2022.843775.
9. Laird DJ, De Tomaso AW, Weissman IL. Stem cells are units of natural selection in a colonial ascidian. Cell. 2005;123(7):1351–60. https://doi.org/10.1016/j.cell.2005.10.026.
10. Laird DJ, De Tomaso AW. Predatory stem cells in the non-zebrafish chordate, *Botryllus schlosseri*. Zebrafish. 2005;1(4):357–61. https://doi.org/10.1089/zeb.2005.1.357.
11. Brown FD, Tiozzo S, Roux MM, et al. Early lineage specification of long-lived germline precursors in the colonial ascidian *Botryllus schlosseri*. Development. 2009;136(20):3485–94. https://doi.org/10.1242/dev.037754.
12. Pancer Z, Gershon H, Rinkevich B. Coexistence and possible parasitism of somatic and germ cell lines in chimeras of the colonial urochordate *Botryllus schlosseri*. Biol Bull. 1995;189(2):106–12. https://doi.org/10.2307/1542460.
13. Stoner DS, Weissman IL. Somatic and germ cell parasitism in a colonial ascidian: possible role for a highly polymorphic allorecognition system. Proc Nat Acad Sci USA. 1996;93(26):15254–59. https://doi.org/10.1073/pnas.93.26.15254.
14. Manni L, Anselmi C, Cima F, et al. Sixty years of experimental studies on the blastogenesis of the colonial tunicate *Botryllus schlosseri*. Dev Biol. 2019;448(2):293–308. https://doi.org/10.1016/j.ydbio.2018.09.009.
15. Kassmer SH, Rodriguez D, De Tomaso AW. Colonial ascidians as model organisms for the study of germ cells, fertility, whole body regeneration, vascular biology and aging. Curr Opin Genet Dev. 2016;39:101–6. https://doi.org/10.1016/j.gde.2016.06.001.
16. Taketa DA, De Tomaso AW. *Botryllus schlosseri* allorecognition: tackling the enigma. Dev Comp Immunol. 2015;48(1):254–65. https://doi.org/10.1016/j.dci.2014.03.014.
17. Nydam ML. Evolution of allorecognition in the Tunicata. Biology 2020;9(6):129. https://doi.org/10.3390/biology9060129.
18. Epelbaum A, Therriault TW, Paulson A, et al. Botryllid tunicates: culture techniques and experimental procedures. Aquat Invasions. 2009;4(1):111–20. https://doi.org/10.3391/ai.2009.4.1.12.
19. Gasparini F, Manni L, Cima F, et al. Sexual and asexual reproduction in the colonial ascidian *Botryllus schlosseri*. Genesis. 2015;53(1):105–20. https://doi.org/10.1002/dvg.22802.
20. Wawrzyniak MK, Matas Serrato LA, Blanchoud S. Long-term monitoring data logs of a recirculating artificial seawater based colonial ascidian aquaculture. Data Brief. 2021;38:107372. https://doi.org/10.1016/j.dib.2021.107372.
21. Langenbacher AD, Rodriguez D, Di Maio A, et al. Whole-mount fluorescent *in situ* hybridization staining of the colonial tunicate *Botryllus schlosseri*. Genesis. 2015;53(1):194–201. https://doi.org/10.1002/dvg.22820.
22. Manni L, Zaniolo G, Cima F, et al. *Botryllus schlosseri*: a model ascidian for the study of asexual reproduction. Dev Dyn. 2007;236(2):335–52. https://doi.org/10.1002/dvdy.21037.
23. Rodriguez D, Sanders EN, Farell K, et al. Analysis of the basal chordate *Botryllus schlosseri* reveals a set of genes associated with fertility. BMC Genom. 2014;15(1):1183. https://doi.org/10.1186/1471-2164-15-1183.
24. Campagna D, Gasparini F, Franchi N, et al. Transcriptome dynamics in the asexual cycle of the chordate *Botryllus schlosseri*. BMC Genomics. 2016;17(1):275. https://doi.org/10.1186/s12864-016-2598-1.
25. Ricci L, Chaurasia A, Lapébie P, et al. Identification of differentially expressed genes from multipotent epithelia at the onset of an asexual development. Sci Rep. 2016;6:27357. https://doi.org/10.1038/srep27357.

26. Rosental B, Kowarsky M, Seita J, et al. Complex mammalian-like haematopoietic system found in a colonial chordate. Nature. 2018;564(7736):425–29. https://doi.org/10.1038/s41586-018-0783-x.
27. Kowarsky M, Anselmi C, Hotta K, et al. Sexual and asexual development: two distinct programs producing the same tunicate. Cell Rep. 2021;34(4):108681. https://doi.org/10.1016/j.celrep.2020.108681.
28. Voskoboynik A, Neff NF, Sahoo D, et al. The genome sequence of the colonial chordate, *Botryllus schlosseri*. eLife. 2013;2:e00569. https://doi.org/10.7554/eLife.00569.
29. Lawniczak MKN, Durbin R, Flicek P, et al. Standards recommendations for the Earth BioGenome Project. Proc Natl Acad Sci U S A. 2022;119(4):e2115639118. https://doi.org/10.1073/pnas.2115639118.
30. De Tomaso AW, Saito Y, Ishizuka KJ, et al. Mapping the genome of a model protochordate. I. A low resolution genetic map encompassing the fusion/histocompatibility (Fu/HC) locus of *Botryllus schlosseri*. Genetics 1998;149(1):277–87. https://doi.org/10.1093/genetics/149.1.277.
31. Cheng H, Concepcion GT, Feng X, et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–75. https://doi.org/10.1038/s41592-020-01056-5.
32. Bojko J, Reinke AW, Stentiford GD, et al. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. Trends Parasitol. 2022;38(8):642–59. https://doi.org/10.1016/j.pt.2022.05.007.
33. Li K, Xu P, Wang J, et al. Identification of errors in draft genome assemblies at single-nucleotide resolution for quality assessment and improvement. Nat Commun. 2023;14(1):6556. https://doi.org/10.1038/s41467-023-42336-w.
34. Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023;39(1):btac808. https://doi.org/10.1093/bioinformatics/btac808.
35. Colombera D. The karyology of the colonial ascidian *Botryllus schlosseri* (Pallas). Caryologia. 1969;22(4):339–49. https://doi.org/10.1080/00087114.1969.10796353.
36. Zeng X, Yi Z, Zhang X, et al. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. Nat Plants. 2024;10(8):1184–200. https://doi.org/10.1038/s41477-024-01755-3.
37. Bandi V, Gutwin C, Siri JN, et al. Visualization tools for genomic conservation. In: Edwards D, ed. Plant bioinformatics: methods and protocols. New York, NY: Springer US; 2022:285–308.https://doi.org/10.1007/978-1-0716-2067-0_16
38. Wang Y, Tang H, DeBarry JD, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49. https://doi.org/10.1093/nar/gkr1293.
39. Simão FA, Waterhouse RM, Ioannidis P, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–212. https://doi.org/10.1093/bioinformatics/btv351.
40. Guiglielmoni N, Houtain A, Derzelle A, et al. Overcoming uncollapsed haplotypes in long-read assemblies of non-model organisms. BMC Bioinformatics. 2021;22(1):303. https://doi.org/10.1186/s12859-021-04118-3.
41. Simion P, Narayan J, Houtain A, et al. Chromosome-level genome assembly reveals homologous chromosomes and recombination in asexual rotifer *Adineta vaga*. Sci Adv. 2021;7(41):eabg4216. https://doi.org/10.1126/sciadv.abg4216.
42. López-Legentil S, Turon X, Planes S. Genetic structure of the star sea squirt, *Botryllus schlosseri*, introduced in southern European harbours. Mol Ecol. 2006;15(13):3957–67. https://doi.org/10.1111/j.1365-294X.2006.03087.x.
43. Bock DG, MacIsaac HJ, Cristescu ME. Multilocus genetic analyses differentiate between widespread and spatially restricted cryptic species in a model ascidian. Proc R Soc B Biol Sci. 2012;279(1737):2377–85. https://doi.org/10.1098/rspb.2011.2610.
44. Salonna M, Gasparini F, Huchon D, et al. An elongated COI fragment to discriminate botryllid species and as an improved ascidian DNA barcode. Sci Rep. 2021;11(1):4078. https://doi.org/10.1038/s41598-021-83127-x.
45. Blanchoud S, Rutherford K, Zondag L, et al. *De novo* draft assembly of the *Botrylloides leachii* genome provides further insight into tunicate evolution. Sci Rep. 2018;8(1):5518. https://doi.org/10.1038/s41598-018-23749-w.
46. Satou Y, Nakamura R, Yu D, et al. A nearly complete genome of *Ciona intestinalis* type A (*C . robusta*) reveals the contribution of inversion to chromosomal evolution in the genus *Ciona*. Genome Biol Evol. 2019;11(11):3144–57. https://doi.org/10.1093/gbe/evz228.
47. Bliznina A, Masunaga A, Mansfield MJ, et al. Telomere-to-telomere assembly of the genome of an individual *Oikopleura dioica* from Okinawa using Nanopore-based sequencing. BMC Genomics. 2021;22(1):222. https://doi.org/10.1186/s12864-021-07512-6.
48. Castellano KR, Batta-Lona P, Bucklin A, et al. *Salpa* genome and developmental transcriptome analyses reveal molecular flexibility enabling reproductive success in a rapidly changing environment. Sci Rep. 2023;13(1):21056. https://doi.org/10.1038/s41598-023-47429-6.
49. Gabriel L, Brůna T, Hoff KJ, et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 2024;34(5):769–77. https://doi.org/10.1101/gr.278090.123.
50. Haas BJ. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31(19):5654–66. https://doi.org/10.1093/nar/gkg770.
51. Haas BJ, Salzberg SL, Zhu W, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9(1):R7. https://doi.org/10.1186/gb-2008-9-1-r7.
52. Cantalapiedra CP, Hernández-Plaza A, Letunic I, et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38(12):5825–29. https://doi.org/10.1093/molbev/msab293.
53. Blum M, Chang HY, Chuguransky S, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021;49(D1):D344–54. https://doi.org/10.1093/nar/gkaa977.
54. Jones P, Binns D, Chang HY, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40. https://doi.org/10.1093/bioinformatics/btu031.
55. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726–31. https://doi.org/10.1016/j.jmb.2015.11.006.
56. Wei J, Zhang J, Lu Q, et al. Genomic basis of environmental adaptation in the leathery sea squirt (*Styela clava*). Mol Ecol Resour. 2020;20(5):1414–31. https://doi.org/10.1111/1755-0998.13209.

57. Delsuc F, Philippe H, Tsagkogeorga G, et al. A phylogenomic framework and timescale for comparative studies of tunicates. BMC Biol. 2018;16(1):1–14. https://doi.org/10.1186/s12915-018-0499-2.
58. Simakov O, Marlétaz F, Yue JX, et al. Deeply conserved synteny resolves early events in vertebrate evolution. Nat Ecol Evol. 2020;4(6):820–30. https://doi.org/10.1038/s41559-020-1156-z.
59. Schultz DT, Haddock SHD, Bredeson JV, et al. Ancient gene linkages support ctenophores as sister to other animals. Nature. 2023;618(7963):110–17. https://doi.org/10.1038/s41586-023-05936-6.
60. Simakov O, Bredeson J, Berkoff K, et al. Deeply conserved synteny and the evolution of metazoan chromosomes. Sci Adv. 2022;8(5):eabi5884. https://doi.org/10.1126/sciadv.abi5884.
61. Lewin TD, Liao IJY, Chen ME, et al. Fusion, fission, and scrambling of the bilaterian genome in Bryozoa. Genome Res. 2025;35(1):78–92. https://doi.org/10.1101/gr.279636.124.
62. Plessy C, Mansfield MJ, Bliznina A, et al. Extreme genome scrambling in marine planktonic *Oikopleura dioica* cryptic species. Genome Res. 2024;34(3):426–40. https://doi.org/10.1101/gr.278295.123.
63. Vargas-Chávez C, Benítez-Álvarez L, Martínez-Redondo GI, et al. An episodic burst of massive genomic rearrangements and the origin of non-marine annelids. Nat Ecol Evol. 2025;9(7):1263–79. https://doi.org/10.1038/s41559-025-02728-1.
64. Lewin TD, Liao IJY, Luo YJ. Annelid comparative genomics and the evolution of massive lineage-specific genome rearrangement in bilaterians. Mol Biol Evol. 2024;41(9):msae172. https://doi.org/10.1093/molbev/msae172.
65. Schultz D, Heath-Heckman E, Winchell C, et al. Acceleration of genome rearrangement in clitellate annelids. bioRxiv. 2024. https://doi.org/10.1101/2024.05.12.593736.
66. Berna L, Alvarez-Valin F. Evolutionary genomics of fast evolving tunicates. Genome Biol Evol. 2014;6(7):1724–38. https://doi.org/10.1093/gbe/evu122.
67. Monteiro AS, Ferrier DEK. Hox genes are not always Colinear. Int J Biol Sci. 2006;2(3):95–103. https://doi.org/10.7150/ijbs.2.95.
68. DeBiasse MB, Colgan WN, Harris L, et al. Inferring tunicate relationships and the evolution of the tunicate Hox cluster with the genome of *Corella inflata*. Genome Biol Evol. 2020;12(6):948–64. https://doi.org/10.1093/gbe/evaa060.
69. Sekigami Y, Kobayashi T, Omi A, et al. Hox gene cluster of the ascidian, *Halocynthia roretzi*, reveals multiple ancient steps of cluster disintegration during ascidian evolution. Zool Lett. 2017;3:17. https://doi.org/10.1186/s40851-017-0078-3.
70. Gaunt SJ. Seeking sense in the Hox gene cluster. J Dev Biol. 2022;10(4):48. https://doi.org/10.3390/jdb10040048.
71. Caputi L. Evolutionary genomics of tunicates. Sci Rev Biol. 2024;3(2):22–32. https://doi.org/10.57098/SciRevs.Biology.3.2.3.
72. Sanges R, Hadzhiev Y, Gueroult-Bellone M, et al. Highly conserved elements discovered in vertebrates are present in non-syntenic loci of tunicates, act as enhancers and can be transcribed during development. Nucleic Acids Res. 2013;41(6):3600–18. https://doi.org/10.1093/nar/gkt030.
73. Sim SB, Corpuz RL, Simmonds TJ, et al. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. BMC Genomics. 2022;23(1):157. https://doi.org/10.1186/s12864-022-08375-1.
74. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.
75. M.Tawfeeq M, Swaelus U, Rodriguez Gaudray F, et al. Refining Feulgen: low-cost and accurate genome size measurements for everyone. bioRxiv. 2025. https://doi.org/10.1101/2025.08.29.673164.
76. Wang L, Xiong Q, Saelim N, et al. Genome assembly and annotation of *Periplaneta americana* reveal a comprehensive cockroach allergen profile. Allergy. 2023;78(4):1088–103. https://doi.org/10.1111/all.15531.
77. Vizueta J, Xiong Z, Ding G, et al. Adaptive radiation and social evolution of the ant. Cell. 2025. Online ahead of print. https://doi.org/10.1016/j.cell.2025.05.030.
78. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–75. https://doi.org/10.1038/nmeth.2089.
79. Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. Bioinformatics. 2017;33(17):2759–61. https://doi.org/10.1093/bioinformatics/btx304.
80. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11(1):1432. https://doi.org/10.1038/s41467-020-14998-3.
81. Huang S, Kang M, Xu A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics. 2017;33(16):2577–79. https://doi.org/10.1093/bioinformatics/btx220.
82. Challis R, Richards E, Rajan J, et al. BlobToolKit–interactive quality assessment of genome assemblies. G3 (Bethesda). 2020;10(4):1361–74. https://doi.org/10.1534/g3.119.400908.
83. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421. https://doi.org/10.1186/1471-2105-10-421.
84. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18(4):366–68. https://doi.org/10.1038/s41592-021-01101-x.
85. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37(23):4572–74. https://doi.org/10.1093/bioinformatics/btab705.
86. Ghurye J, Rhie A, Walenz BP, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 2019;15(8):e1007273. https://doi.org/10.1371/journal.pcbi.1007273.
87. Shen W, Le S, Li Y, et al. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 2016;11(10):e0163962. https://doi.org/10.1371/journal.pone.0163962.
88. Mapleson D, Garcia Accinelli G, Kettleborough G, et al. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics. 2017;33(4):574–76. https://doi.org/10.1093/bioinformatics/btw663.
89. Manni M, Berkeley MR, Seppey M, et al. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38(10):4647–54. https://doi.org/10.1093/molbev/msab199.
90. Guan D, McCarthy SA, Wood J, et al. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36(9):2896–98. https://doi.org/10.1093/bioinformatics/btaa025.

91. Flynn JM, Hubley R, Goubert C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117(17):9451–57. https://doi.org/10.1073/pnas.1921046117.
92. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.
93. Kuznetsov D, Tegenfeldt F, Manni M, et al. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 2023;51(D1):D445–51. https://doi.org/10.1093/nar/gkac998.
94. Lomsadze A. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005;33(20):6494–506. https://doi.org/10.1093/nar/gki937.
95. Stanke M, Schöffmann O, Morgenstern B, et al. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7(1):62. https://doi.org/10.1186/1471-2105-7-62.
96. Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 2014;42(15):e119. https://doi.org/10.1093/nar/gku557.
97. Gotoh O. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. Nucleic Acids Res. 2008;36(8):2630–38. https://doi.org/10.1093/nar/gkn105.
98. Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. 2012;40(20):e161. https://doi.org/10.1093/nar/gks708.
99. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60. https://doi.org/10.1038/nmeth.3176.
100. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2020;2(2):lqaa026. https://doi.org/10.1093/nargab/lqaa026.
101. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Research. 2020;9:ISCB Comm J–304. https://doi.org/10.12688/f1000research.23297.1.
102. Kovaka S, Zimin AV, Pertea GM, et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019;20(1):278. https://doi.org/10.1186/s13059-019-1910-1.
103. Stanke M, Diekhans M, Baertsch R, et al. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics. 2008;24(5):637–44. https://doi.org/10.1093/bioinformatics/btn013.
104. Hoff KJ, Lomsadze A, Borodovsky M, et al. Whole-genome annotation with BRAKER. In: Gene prediction: methods and protocols. No. 1962 in Methods in Molecular Biology. Springer; New York, NY, 2019:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.
105. Hoff KJ, Lange S, Lomsadze A, et al. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2016;32(5):767–69. https://doi.org/10.1093/bioinformatics/btv661.
106. Brůna T, Hoff KJ, Lomsadze A, et al. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021;3(1):lqaa108. https://doi.org/10.1093/nargab/lqaa108.
107. Pertea M, Pertea GM, Antonescu CM, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–95. https://doi.org/10.1038/nbt.3122.
108. Haas BJ, Papanicolaou A, Yassour M, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512. https://doi.org/10.1038/nprot.2013.084.
109. Huerta-Cepas J, Szklarczyk D, Heller D, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–14. https://doi.org/10.1093/nar/gky1085.
110. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45(4):e18. https://doi.org/10.1093/nar/gkw1060.
111. Dardaillon J, Dauga D, Simion P, et al. ANISEED 2019: 4D exploration of genetic data for an extended range of tunicates. Nucleic Acids Res. 2020;48(D1):D668–75. https://doi.org/10.1093/nar/gkz955.
112. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–97. https://doi.org/10.1093/nar/gkh340.
113. Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–39. https://doi.org/10.1093/molbev/msr121.
114. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30(22):3276–78. https://doi.org/10.1093/bioinformatics/btu531.
115. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–34. https://doi.org/10.1093/molbev/msaa015.
116. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Bioinformatics. 1992;8(3):275–82. https://doi.org/10.1093/bioinformatics/8.3.275.
117. Yang Z. A space-time process model for the evolution of DNA sequences. Genetics. 1995;139(2):993–1005. https://doi.org/10.1093/genetics/139.2.993.
118. Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–89. https://doi.org/10.1038/nmeth.4285.
119. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6(2):461–64. https://doi.org/10.1214/aos/1176344136.
120. Hoang DT, Chernomor O, von Haeseler A, et al. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22. https://doi.org/10.1093/molbev/msx281.
121. De Thier O, Lebel M, Tawfeeq M, et al. Supporting data for "First Chromosome-Level Genome Assembly of the Colonial Chordate Model *Botryllus schlosseri* (Tunicata)." GigaScience Database. 2025. http://dx.doi.org/10.5524/102718.
122. Dru Ph. Octopus: LBDV bioinformatics server. 2025. https://octopus.obs-vlfr.fr/public/Botryllus_genome/index.html (accessed August 22, 2025).
123. Lin Y, Ye C, Li X, et al. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identi-

fication. Hortic Res. 2023;10(8):uhad127. https://doi.org/10.1093/hr/uhad127.
124. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 2018;6:e4958. https://doi.org/10.7717/peerj.4958.
125. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572–74. https://doi.org/10.1093/bioinformatics/btg180.

Received: February 19, 2025. **Revised:** May 13, 2025. **Accepted:** July 27, 2025