Comparative genomic analysis of *Cristatella mucedo* provides insights into *Bryozoan* evolution and nervous system function

Viktor V Starunov[1,2†], Alexander V Predeus[3†*], Yury A Barbitoff[3], Vladimir A Kutiumov[1], Arina L Maltseva[1], Ekatherina A Vodiasova[4], Andrea B Kohn[5], Leonid L Moroz[5*], Andrew N Ostrovsky[1,6*]

[1] Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg State University, Universitetskaya nab. 7/9, 199034, St. Petersburg, Russia

[2] Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, 199034, St. Petersburg, Russia

[3] Bioinformatics Institute, Kantemirovskaya 2A, 197342, St. Petersburg, Russia

[4] A.O. Kovalevsky Institute of Biology of the Southern Seas, Russian Academy of Sciences, Leninsky pr. 38/3, 119991, Moscow, Russia

[5] The Whitney Laboratory for Marine Bioscience, University of Florida, 9505 Ocean Shore Blvd, St Augustine, FL 32080, USA

[6] Department of Paleontology, Faculty of Earth Sciences, Geography and Astronomy, University of Vienna, Althanstrasse 14, 1090, Vienna, Austria

[†] These authors contributed equally to the study.

[*] To whom correspondence should be addressed: predeus@bioinf.me, moroz@whitney.ufl.edu, a.ostrovsky@spbu.ru

Abstract

The modular body organization is an enigmatic feature of different animal phyla scattered throughout the phylogenetic tree. Here we present a high-quality genome assembly of *Cristatella mucedo,* a unique freshwater bryozoan with mobile colonies, making it a first sequenced genome of the phylum Bryozoa. Using PacBio, Oxford Nanopore, and Illumina sequencing, we were able to obtain assembly with N50 of 4.1 Mb. Comparative genome analysis suggests that, despite larger genome size and higher number of genes, *C. mucedo* possesses a less diverse set of proteins compared to its immediate relatives. Gene family and pathway overrepresentation analysis were used to find candidate targets involved in bryozoan nervous system and locomotion. We used RNA sequencing to identify genes upregulated in various parts of the colony, as well as during the differentiation from frozen statoblasts, and validated several of these targets using *in situ* hybridization. Overall, analysis of the first Bryozoan genome allows important insights into the evolution of nervous system and modular body organization.

Introduction

Recent development of molecular methods allowed to study animal genomes of most known phyla and stimulated impressive progress in our understanding of their phylogenies and evolution. One of the biggest gaps in genome-based phylogeny is Bryozoa – medium-sized phylum of microscopic aquatic invertebrates comprising about 6000 recent and more than 15000 fossil species and having a long fossil history beginning in the early Ordovician[1]. The phylum consists of three classes: exclusively freshwater, non-skeletal Phylactolaemata, exclusively marine, calcified Stenolaemata and predominantly marine, but occasionally brackish- and freshwater, Gymnolaemata. Recent studies indicate the class Phylactolaemata is a sister group to the clade consisting of Stenolaemata and Gymnolaemata[2–4] (Figure 1A). Current molecular data are limited to individual genes or mitochondrial genomes, however, and generated controversial results[5–10].

From an evolutionary point of view, bryozoans are of extreme interest because of their modular organization. Modularity (coloniality) is scattered throughout the animal phylogenetic tree being independently evolved multiple times among cnidarians, hemichordates, tunicates and kamptozoans[11]. What makes Bryozoa especially interesting is that it is the only almost exclusively colonial animal group (with the exception of one genus that secondarily became solitary), having higher diversity of the colonial growth forms and constructions than in any other modular group.

Bryozoan colony consists of modules called zooids that are comprised of a cystid (body wall) and a polypide (retractile ciliated tentacular crown associated with U-shaped gut and retractor muscles) each. The tentacle crown is conventionally termed lophophore (Figure 1b). The subject species of our study was *Cristatella mucedo* Cuvier, 1798 with holarctic distribution (Figure 1c). *C. mucedo* forms caterpillar-like free-living colonies with thick gelatinous ectocyst and parallel rows of zooids could reach up to 8 cm in length and 4-5 mm in width. A unique ability of *C. mucedo* colonies is crawling, or gliding, that is performed via the activity of a colony basal part (foot) provided with two perpendicular muscular layers with a plexus of multipolar neurons sandwiched in between[11]. Intriguingly, *C. mucedo* motion is also responsive to light, which is especially puzzling given the absence of morphologically defined photoreceptors (summarized in Shunkina et al. 2015[11]). The only cerebral ganglion is located near the pharynx, with the nerve cords extending into the horse-shoe lophophore with two arms, the gut, and the cystid wall. Finally, a unique and characteristic feature of

Phylactolaemata, and *C. mucedo* in particular, is production of the dormant 'buds' (statoblasts), able to survive freezing and other harsh conditions (Figure 1 d-f[12,13]).

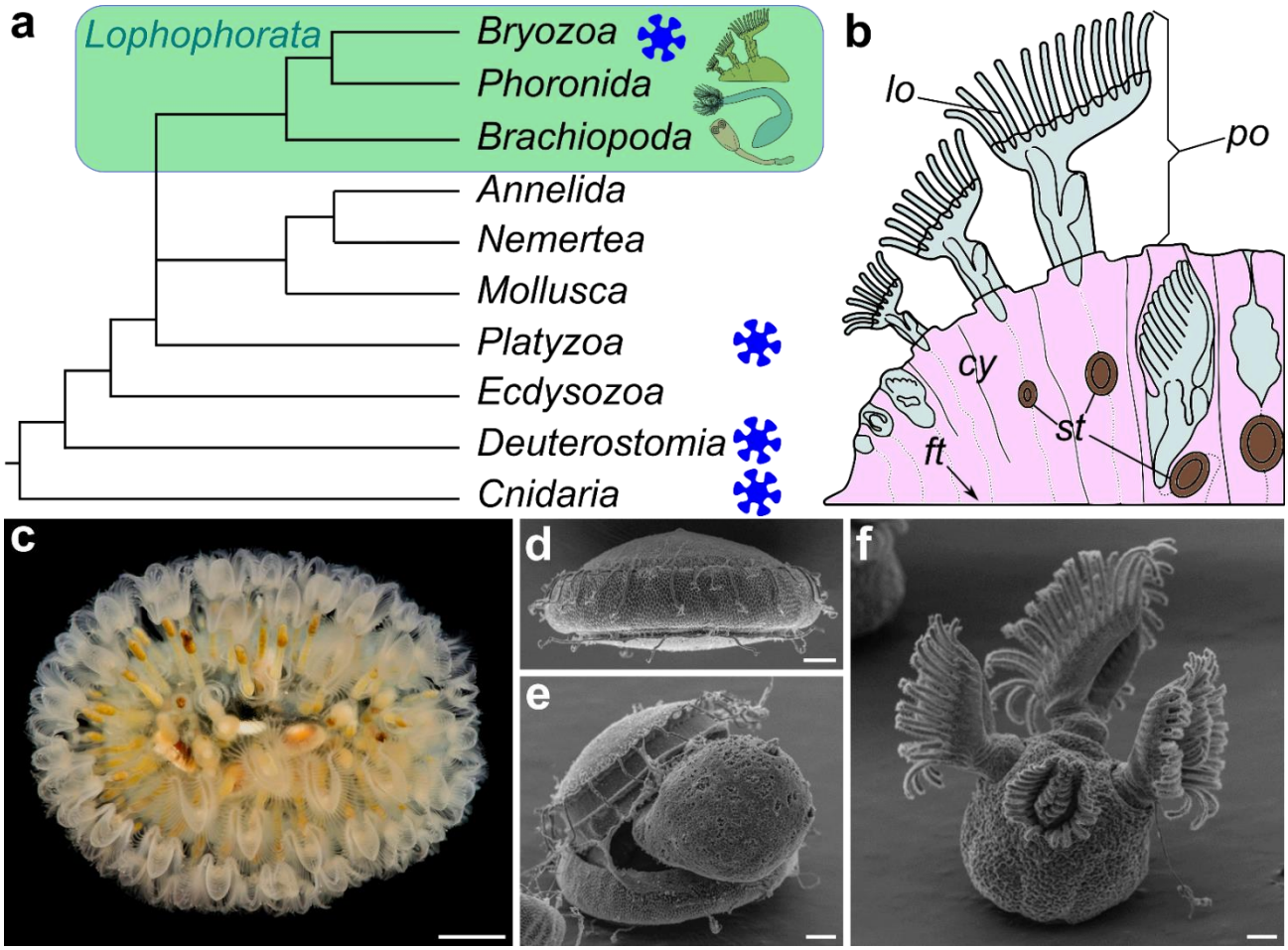

Figure 1. General morphology and current view of phylogenetic placement of *Cristatella mucedo*. **a**, Simplified scheme of animal phylogeny, Lophophorata position (according to Nesnidal, 2013), and distribution of modular body organization. Asterisks label the clades where modular organization is known. **b**, Generalized scheme of *Cristatella mucedo* colony; **c**, Photomicrograph of a juvenile colony; **d-f**, Development of a young colony from a statoblast. Scale bar: **c** - 1 mm, **d-f** - 100μm. cy - cystid, ft - foot, lo - lophophore, po - polypide, st - statoblasts.

Here we report the first high-quality genome assembly of the phylactolaemate bryozoan, *Cristatella mucedo*. Taken together with bulk transcriptomic data from developmental stages and morphologically defined parts of colonies, these data allow us to obtain a high-quality genome annotation. Using the three published annotated genomes of Brachiopoda, Nemertea, and Phoronida[12,13], we performed comparative genomic analysis of the four species, and identified under- and over-represented gene families.

Results

Genome assembly and annotation

Genome sequencing was done using Oxford Nanopore, PacBio, and Illumina sequencers, with a combined coverage of 88x for long reads and 80x for short reads. Several assembly strategies were employed and evaluated using contiguity (N50) and BUSCO scores. We note that regardless of the assembly strategy used, N50 with combination of PacBio and OxfordNanopore was at least 8 times higher compared to assemblies using PacBio reads only. The best selected strategy resulted in a genome with contig N50 of 4.116 Mb, and Table 1 describes comparative statistics of published genomes of closest lophotrochozoans[12,13].

Species	***C. mucedo***	***N.geniculatus***	***P. australis***	***L. anatina***
Phylum	Bryozoa	Nemertea	Phoronida	Brachiopoda
Common name	Moss animals	Ribbon worms	Horseshoe worms	Lamp shells
Genome size (Mb)	574	859	498	406
Sequencing coverage	170-fold	265-fold	227-fold	226-fold
Number of scaffolds	986	11,108	3,984	2,677
Scaffold N50 (kb)	4116	239	655	460
Contig N50 (kb)	4116	23.6	71.4	58.2
GC content, %	46.9	42.9	39.3	36.4
Repeats (%)	47.0	37.5	39.4	23.3
Number of genes	35,950	43,294	20,473	29,907
Gene density (per Mb)	62.6	50.4	41.1	73.7
Mean gene size (bp)	9,655	8,223	14,590	7,725
Mean transcript size (bp)	1,445	1,448	1,587	1,551
Mean intron per gene	5.7	5.2	7.4	7.3
Mean intron size (bp)	1,415	1,308	1,744	840

Table 1. Comparison of genome statistics of *C. mucedo* genome with three published genomes of the lophotrochozoans most closely related to bryozoa (*N. geniculatus, P. australis,* and *L. anatina*).

Overall, the genome of *C. mucedo* contains the highest percentage of repeats, although these differences could be partially attributed to repeat identification method. In other basic statistics, *C.*

mucedo appears to be closer to *N. geniculatus* in genome size, mean gene size, gene number, GC content, and mean number of introns per gene.

Gene repertoire analysis

In order to compare results with the most relevant published data, we fully reproduced gene family analysis of Luo et al[13], with the exception of adding *C. mucedo* proteins, removal of 4 species which proteomes were not available from NCBI, and using OrthoFinder software for improved sensitivity of gene family detection[14]. Overall, our analyzed dataset contained predicted proteins from 28 species. While overall picture was strikingly similar, our results show interesting differences from previously published gene family numbers, especially in the overall number of unique gene families, and gene family overlap (Figure 2a). *C. mucedo* was shown to possess a strikingly low number of unique gene families - 7,785, which is 1700 fewer than the next species in comparison, *P. australis*). This might suggest that the miniature size and simple structure of bryozoan zooids has resulted in genome simplification, leading to the loss of many ancestral gene families.

Similarly to Luo et al[13], we have used principal component analysis to investigate the relationship between the studied 28 species based on the number of genes belonging to each identified orthologous group (as identified by OrthoFinder). As in previous analysis, vertebrate genomes were visibly separated from lophotrochozoans and invertebrate deuterostomes; as could be expected, C. mucedo has also clustered with lophotrochozoans (Figure 2b,c).

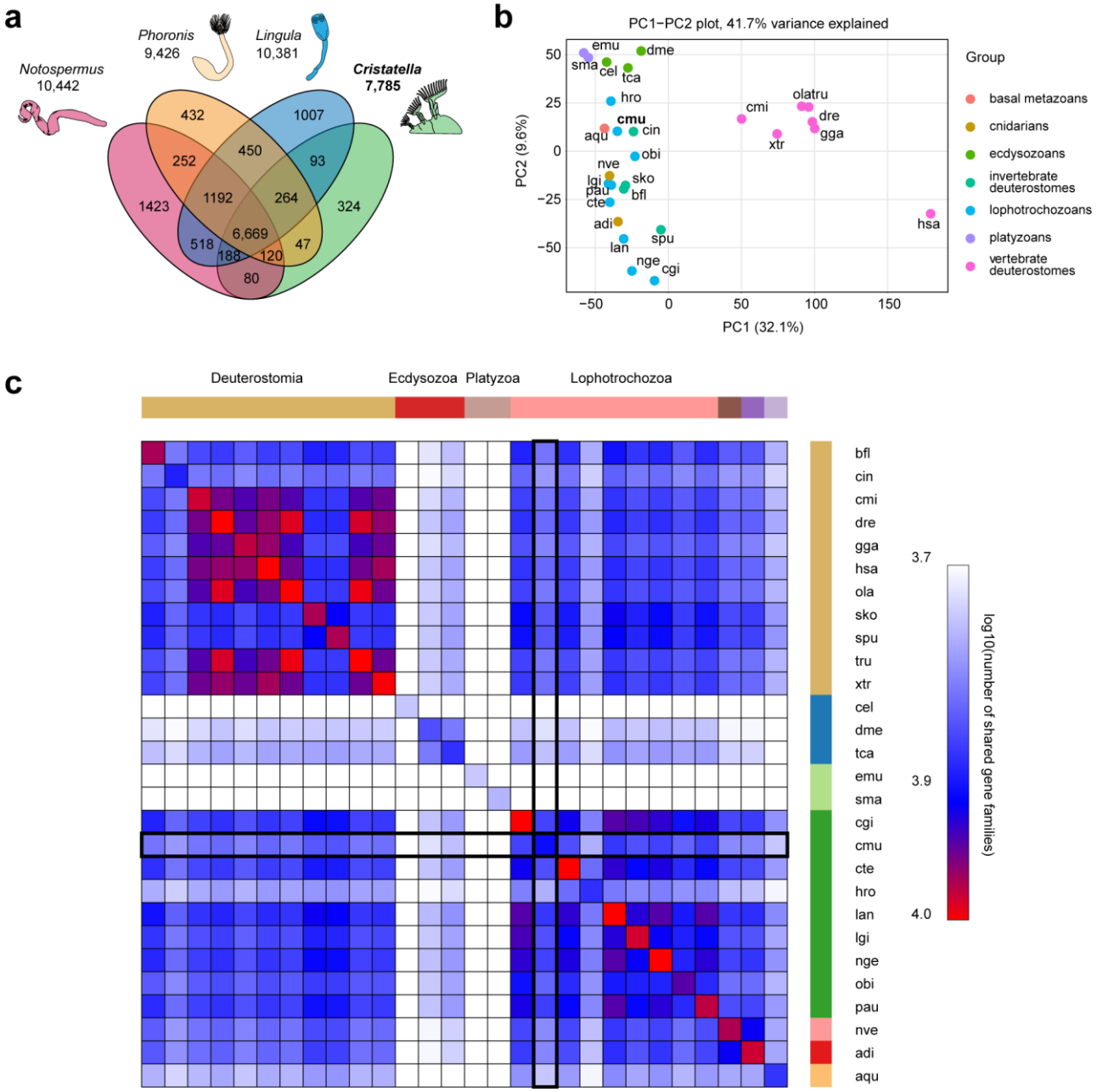

Figure 2. Comparative analysis of C. mucedo and 27 other metazoan species using OrthoFinder orthologous groups. **a**, Venn diagram of shared and unique gene families of C. mucedo and three closely related lophotrochozoans. **b**, Principal component analysis of OrthoFinder gene family sizes. **c**, Matrix of shared gene families among selected metazoans. Abbrebiations: adi, *Acropora digitifera;* aqu, *Amphimedon queenslandica;* bfl, *Branchiostoma floridae;* cel, *Caenorhabditis elegans;* cgi, *Crassotrea gigas;* cin, *Ciona intestinalis;* cmi, *Callorhynchus milii;* cte, *Capitella teleta;* dme, *Drosophila melanogaster;* dre, *Danio rerio;* emu, *Echinococcus multilocularis;* gga, *Gallus gallus;* hro, *Helobdella robusta;* hsa, *Homo sapiens;* lan, *ingula anatina;* lgi, *Lottia gigantea;* nge, *Notospermus geniculatus;* nve, *Nematostella vectensis;* obi, *Octopus bimaculoides;* ola, *Oryzias latipes;* pau, *Phoronis australis;* sko, *Saccoglossus kowalevskii;* sma, *Schistosoma mansoni;* spu, *Strongylocentrotus purpuratus;* tca, *Tribolium castaneum;* tru, *Takifugu rubripes;* xtr, *Xenopus tropicalis*.

Gene families over- and underrepresented in Bryozoa

We then compared the annotated proteome of *C. mucedo* with the proteomes of three Lophotrochozoan species that are closely related to Bryozoa (*Notospermus geniculatus*, *Phoronis australis*, and *Lingula anatina*) by analyzing overrepresentation and underrepresentation of Gene Ontology (GO) terms and PFAM domain annotations (see Methods). The most strongly enriched GO terms comprised ice binding (antifreeze) proteins (165 genes in *C. mucedo* compared to only 2 in *P. australis*), ATP-gated ion channels (14 genes compared to 3 in *N. geniculatus*), and histone proteins (almost two-fold increase in nucleosomal protein count compared to *N. geniculatus*) (Figure 3a, top). At the same time, we observed a striking underrepresentation (and sometimes complete absence) of genes that confer transposon activity and DNA integration, ADP-binding proteins, as well as a wide variety of protein families related to signal transduction and ion transport (including calcium channels, GABA receptors, and cAMP biosynthesis genes) (Figure 3a, bottom). In the PFAM domain analysis, we observed a strong overrepresentation of Runt domain (a conserved element of transcriptional regulators), as well as high enrichment of histone protein domains and domains involved in carbohydrate metabolism (Figure 3b).

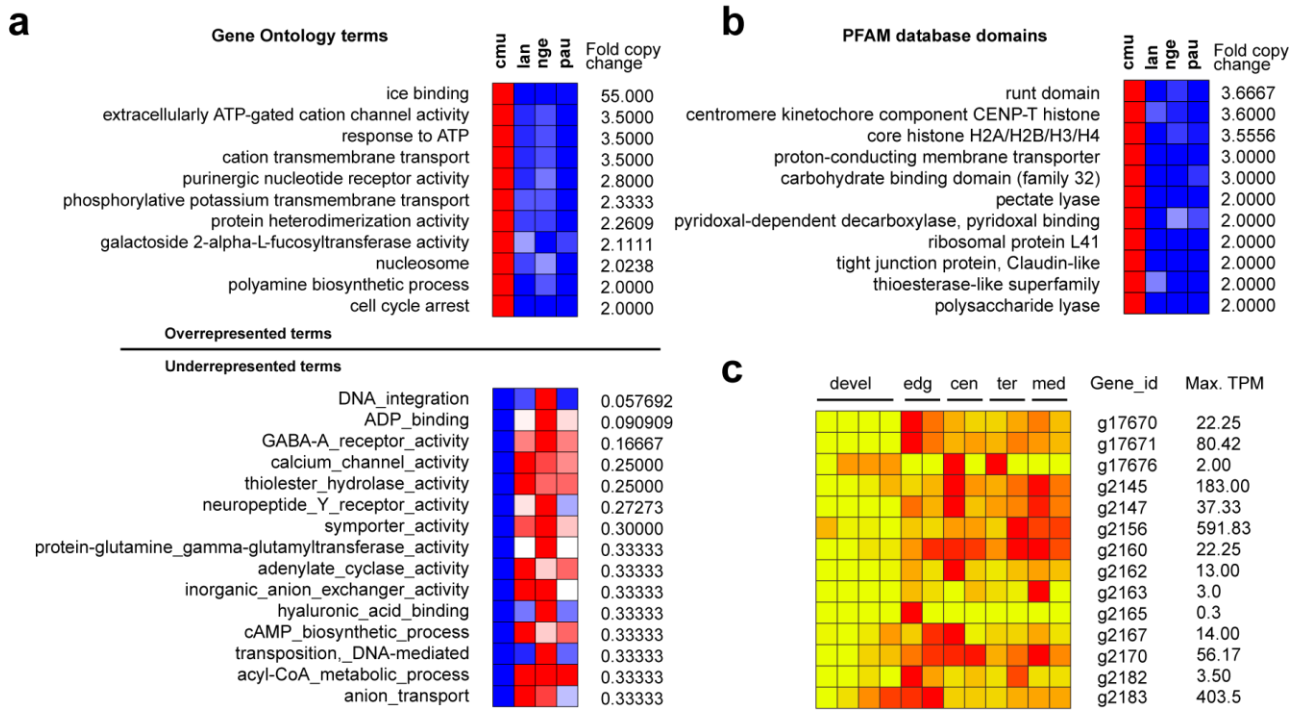

Figure 3. Targeted genome comparison of *C. mucedo* with other lophotrochozoan species. **a**, A heatmap showing InterProScan annotation counts for over- and underrepresented Gene Ontology terms in the

Cristatella proteome compared to three closely related lophotrochozoan species. Only groups showing 2-fold positive of 3-fold negative enrichment are shown. **b**, similarly to **a**, but showing overrepresented PFAM domains. **c**, Gene expression profiles of ATP-gated potassium channel genes identified as highly overrepresented in **a**. Abbreviations: devel, developmental stages, edg - colony edge, cen - central colony area, ter – terminal part of the colony, med - medial part of the colony (see Figure 4a). Maximum expression levels (TPM) across all RNA-Seq replicates are shown for all genes.

Comparison of specific gene clusters

Hox genes encode transcriptional factors, which play an important role in various morphogenetic processes[15]. While the exact mechanism of their function is not known, they control the regionalization of the anterior-posterior axis during embryonic development in both protostome and deuterostome clades. Same genes are also known to participate in various processes in adult organisms[16,17]. Expression patterns suggest a possible role of Hox genes in establishing morphological identity along the anterior posterior (AP) axis. In *C. mucedo* genome we found 7 putative Hox genes, located in two separate contigs. First cluster contains *Hox4*, *Lox5* and *Post2* homologs, while the second is composed of homologs of *Hox2*, *Hox4*, *Lox5* and *Post1*. The spatial collinearity was also maintained in both clusters. Such unusual situation may be a result of ancestral cluster duplication. It was previously reported that 6 Hox genes were found in another bryozoan, *Crisularia turrita*[18]. Similarly to *C. mucedo* in Crisularia *Hox4* is duplicated, which may suggest that the last common ancestor of phylactolaemate and gymnolaemate bryozoans already had a duplicated Hox cluster. On the other hand, *C. mucedo* does not have *Hox3*, which was present in Crisularia. The Hox gene content in both bryozoan species may suggest the parallel independent gene loss during evolution, which may be a result of modular body organization achievement. The colony usually does not have antero-posterior axis in terms of unitary bilaterian organisms. Here each zooid possesses its own A-P axis and the whole colony orientation should have maintained by other molecular mechanisms.

Transcriptome analysis

RNA-sequencing was employed to characterize transcriptional variation across several morphological locations, as well as stages of frozen statoblast differentiation. Two different dissection types were made. In the first case, the colony was divided into central and peripheral (edge)

zones. In the second case, oval colonies were divided into middle and terminal parts (Figure 4a). After transcriptome alignment, quantification, and normalization, k-means clustering identified groups of genes upregulated during differentiation and in various colony parts (Figure 4b). Resulting alignments and annotation were visualized in JBrowse (Figure 4c). For example, we assessed the expression profiles for genes with strongest proteome-wide enrichment shown on Figure 3. Of these, many genes had significant expression levels with differential expression in different colony parts. For example, 13 out of 14 predicted ATP-gated potassium channel genes were expressed at the level of 1 transcript per million (TPM), while three of them showed expression levels of > 100 TPM (Figure 3c). Notably, expression of one member of this family (g2183) was notably rising during the development of *C. mucedo*, with the highest expression of this gene in the edge part of the mature colony (Figure 4a).

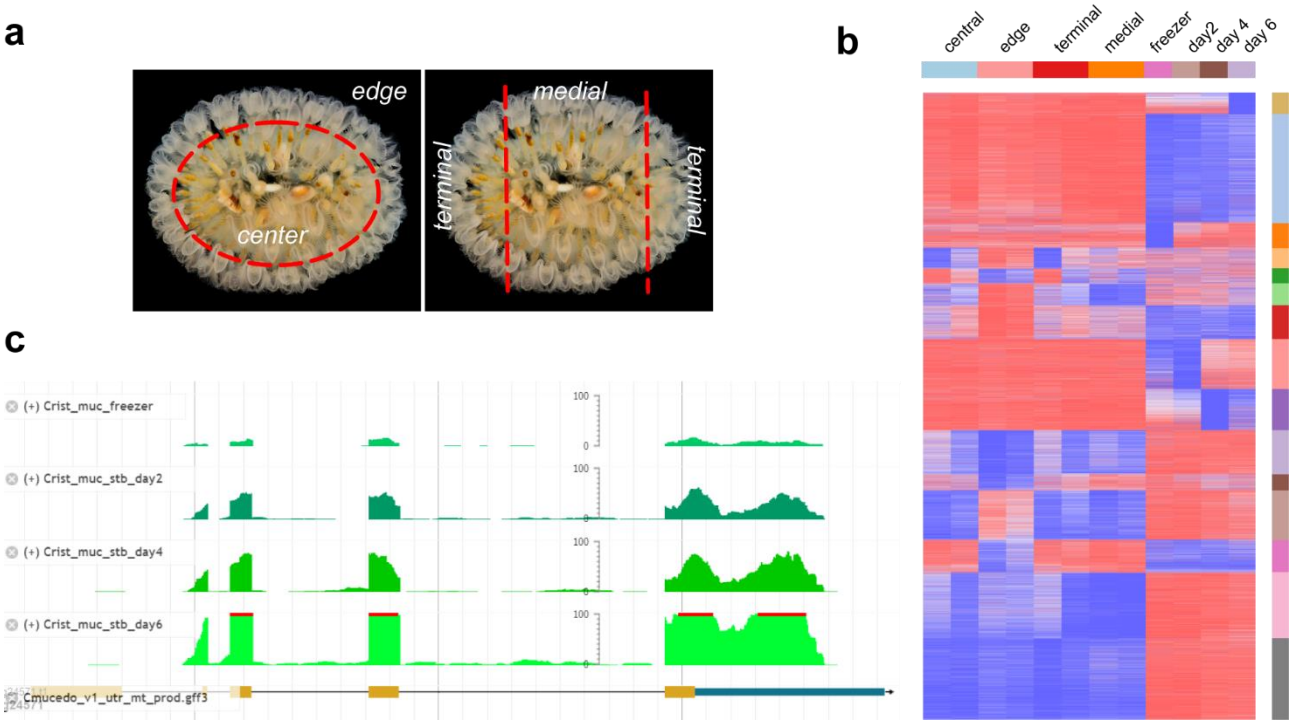

Figure 4. Overview of RNA-seq experiments preparation and analysis. **a**, Scheme of mature colonies dissection used to select central (cen), edge (edge), terminal (ends), or medial (med) fraction of the colony. **b**, Heatmap of top 15,000 most expressed genes, clustered into 15 co-expressed clusters using k-means clustering. **c**, JBrowse view of gene g24571, which is strongly induced after statoblast defrosting and during colony differentiation.

In situ hybridization

To assess the spatio-temporal pattern of gene expression we adopted the RNA in situ hybridization technique to *C. mucedo* colonies. Unfortunately, the statoblast walls are opaque and do not allow to follow the development. However, the method works well for young newly hatched colonies. We used *in situ* hybridization for several genes that were reported to tissue specific in transcriptomic analysis. The absence of the exoskeleton or any other dense structures and the overall animal transparency allow to analyze gene expression at the cellular level without any special clearing techniques that provides the great opportunity to study zooid development and the colonial growth. These results are shown in Figure 5.

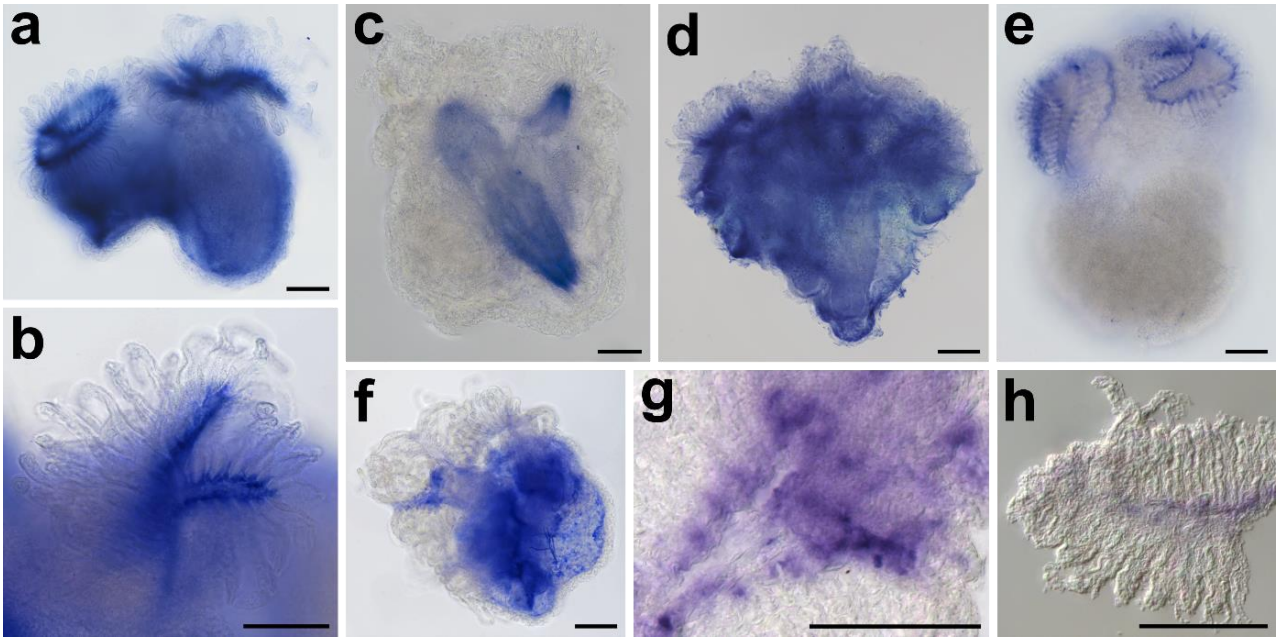

Figure 5. In situ RNA hybridizations of several characteristic genes. **a-b**, Gene g14272, showing expression in Lophophore arms; **c**, Gene g13715, expressed in digestive tract; **d**, Gene g18432, revealing expression in body wall; **e**, Gene g11575 showing specific expression in tentacle bases; **f-h**, Gene g30275, expressed in different tissues including lophophore. Scale bars: 100μm.

Discussion

We presented here a high-quality genome assembly of *Cristatella mucedo*, a freshwater bryozoan. To the best of our knowledge, this is the first published genome of any bryozoan, which gives an exciting opportunity for comparative genomics to gain insights into the evolution of this large and biologically interesting phylum. Using novel long-read approaches allowed us to obtain

near-chromosomal level of assembly contiguity, and combining the data with genomic and transcriptomic short read sequencing has ensured that the assembly has low error rate and is accurately annotated.

Comparative genomic analysis of *C. mucedo*, *L. anatina*, *P. australis*, and *N. geniculatus* have shed some light onto protein family content of these genomes. We discovered that C. mucedo has a much smaller protein diversity compared to the three other species, with 7,785 gene families (Figure 2), despite having more individual genes (35,950). This might be a result of genome degradation due to lifestyle miniaturization and simplification of bryozoan zooids, or because of increasing reliance on symbiotic bacteria to produce necessary metabolites.

We discovered a high proportion of genes related to nervous system among sets of over- and underrepresented GO terms and PFAM domains (Figure 3). This observation might reflect the peculiar properties of the bryozoan nervous system. For example, we found only 1 protein sequence with predicted gamma-aminobutyric acid (GABA) receptor function, which is unexpected given that GABA serves as a universal signaling molecule found in both plants and animals (reviewed in Fait et al[19]). Furthermore, we observed a general underrepresentation of calcium channel genes and cAMP biogenesis genes, suggesting that the role of Ca^{2+} signaling and cAMP-dependent signaling is also decreased in bryozoans. At the same time, we observed a substantial expansion of the ATP-gated potassium channel genes in the *C. mucedo* genome, with these genes having generally high levels of expression and a certain differential expression profile across colony sections (Figure 3c). This might suggest that ATP serves as one of the key signaling molecules in the colonial nervous plexus of the *Cristatella* 'colonial wall'. Such role of ATP in signaling between colony parts and zooids might coordinate the movement of a colony. Importantly, ATP-gated ion channels are important metabolic regulators in humans[20], and are considered to be specialized sensors of extracellular environment[21].

An overwhelming enrichment of antifreeze proteins with ice binding capacity in the *C. mucedo* genome is expected given the ability of statoblasts to survive harsh conditions, in particular freezing in both natural and laboratory environments. In our laboratory, statoblasts were successfully hatched after 5 years in the freezer at -20°C. This makes C. mucedo an exciting candidate to study for new genes that confer extreme resistance to cold environments.

The disruption of the Hox gene cluster might explain the absence of antero-posterior polarity in *C. mucedo colonies* and overall miniaturization and simplification of zooids compared with free-living lophotrochozoan animals such as phoronids or annelids[13,22,23].

Altogether this makes *C. mucedo* a new perspective model to study bilaterian relationships, evolution and development of modular organisms, as well as colony integration processes. The genome and transcriptomes presented here, together with the comparative genomic analysis, provide an insight into lophophorate evolution and modular body organization.

Materials and methods

Animal collection and storage

The colonies of *Cristatella mucedo* were collected in the pools of the Petrodvortsovyi District of St. Petersburg. The animals were kept for several days in tanks filled with filtered freshwater to empty the digestive tracts. The water was changed one or two times a day. The statoblasts were collected from mature colonies in autumn and were frozen in distilled water at -20°C.

Nucleic acid extraction and library preparation

The DNA was extracted from living colonies by different alternative methods. Typically, 1-2 colonies were taken for every extraction procedure. For PacBio and Illumina sequencing the DNA was extracted using Qiagen genomic tip 20/G (10223, Qiagen, Hilden, Germany). For Oxford Nanopore 1D sequencing we used Evrogen extractDNA blood kit (BM011, Evrogen, Moscow, Russia) with preliminary digestion with proteinase K. For Rapid Run Nanopore sequencing we used a modified phenol-chloroform DNA extraction protocol without any subsequent cleaning.

The library preparation for the Illumina sequencing was carried out using Illumina TruSeq DNA PCR-Free Library Preparation Kit (20015962, Illumina, San-Diego, California, USA). For Nanopore sequencing libraries were synthesized using SQK-LSK108 Ligation Sequencing Kit or SQK-RAD002 Rapid Sequencing Kit (Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's instructions. The DNA sequencing was performed with Illumina HiSeq 4000, Oxford Nanopore Technologies MinION, and PacBio Sequel instruments.

The RNA extractions were carried out from different zones of the colony. Two different dissection types were made. In the first case, the colony was divided into central and peripheral (edge) zones. In the second case, elongated colonies were divided into middle and terminal parts. The dissection schemes for every type of experiment are presented in Figure 4a. The RNA was extracted using Zymo Research Quick RNA mini extraction kit (R1050, Zymo Research, Irvine, California, USA). For every sample, two biological replicates were made. The enrichment of mRNA fraction were done using NEBNext® Poly(A) mRNA Magnetic Isolation Module (E7490S, New England BioLabs, Ipswich, Massachusetts, USA) The libraries were synthesized using NEBNext Ultra

Directional RNA Library Prep Kit for Illumina (E7760, New England BioLabs). Paired-end sequencing was carried out using Illumina HiSeq 2500.

To obtain developmental stages the statoblasts were defrosted, rinsed in distilled water and kept in Petri dishes with 50 ml of distilled water at 18°C. Usually, in 6-7 days, the young colonies were hatched from statoblasts. The following stages were taken to extract RNA: 0h - just after thawing (AT), 2 days AT, 4 dAT, and 6dAT. The RNA was extracted using Zymo Research Quick RNA mini extraction kit (R1050, Zymo Research). The libraries were synthesized with TruSeq Stranded Total RNA Library Prep (20020596, Illumina) and single-end sequencing was done on Illumina NextSeq 500 instrument.

In Situ RNA hybridisation

The cDNA libraries were synthetized from the total RNA matrix using Evrogen MINT reverse transcription kit. The fragments from the transcriptome were amplified by PCR and cloned with pGEM®-T Easy (A1360, Promega, Madison, Wisconsin, USA) or P-drive (231124, Qiagen) vector systems. The probe synthesis was performed with DIG RNA Labeling Kit (SP6/T7) (11175025910, Roche, Basel, Switzerland) according to the manufacturer's instructions. For the RNA *in situ* hybridisation we used a modified protocol, used for an annelid *Alitta virens*[16].

Scanning electron microscopy

For scanning electron microscopy, the animals were relaxed by adding drop by drop 2% water chloralhydrate solution and fixed in 2.5% glutaraldehyde in 0.01 M phosphate buffer (PB, pH 7.2). After fixation, the specimens were rinsed 1–3 times in PB and post-fixed for 1h in 1% OsO4 in the same buffer. The specimens were dehydrated in acetone series of increasing concentration, critical point dried, coated with platinum, and examined under FEI Quanta 250 scanning electron microscope (FEI Company, The Netherlands).

Genome assembly

Numerous assembly strategies attempted to obtain best genome assembly were evaluated by contig N50 and BUSCO scores. In the best selected method, both Oxford Nanopore (22x) and PacBio

(66x) reads were combined and assembled using Flye v2.4.2[24] without any polishing. Resulting assembly contained 1153 contigs and had N50 of 3.97 Mb.

Assembly obtained from Flye was then iteratively polished in three steps. First, two rounds of mapping with minimap2 v2.17[25] and polishing with Racon v1.3.0[26] was done using PacBio reads only. Second, minimap2 and Racon v1.3.0 were used for two rounds of polishing with combined Illumina genome sequencing reads (since Racon only supports single-end reads). Finally, bowtie2 v2.3.5[27] and Pilon v1.23[28] was used to polish the assembly until convergence using paired-end reads. Convergence was achieved after 7 rounds of polishing.

To accurately assemble and annotate mitochondrion, the following strategy was used. We first identified two linear contigs containing mitochondrial genes, according to blast search. After this, we selected all PacBio, Oxford Nanopore, and Illumina reads that were successfully mapped to these contigs by minimap2 v2.17. Using the obtained subset of reads, we have assembled complete circular mitochondrial genome with Unicycler v0.4.4[29], a hybrid tool geared towards bacterial genome assembly. Using Oxford Nanopore + Illumina or PacBio + Illumina reads produced identical assemblies. Mitochondrial genome was annotated using MITOS2 web-server[30].

Assembly completeness was evaluated using BUSCO v3.0.2[31] in genome mode with "metazoa" OrthoDB database v9[32]. Final assembly had 918 (93.8%) complete BUSCO single-copy orthologs, with 15 (1.5%) genes duplicated, 16 (1.6%) fragmented, and 44 (4.6%) missing. Genome assembly is available on NCBI (ID XX), and all sequencing reads are deposited under BioProject PRJNA592440/BioSample SAMN13426717 and will be released upon publication.

Repeat modelling and genome annotation

Repeats were modelled using RepeatModeller v1.0.11. The resulting library of repeats was used for repeat discovery and softmasking with RepeatMasker v4.0.9 with rmblast 2.9.0+ search (http://www.repeatmasker.org/). This resulted in 280776395 bp (47.0 %) of the genome identified as repeats. All paired-end transcriptome reads were then aligned using hisat2 v2.1.0[33] to the soft-masked genome. After this, the soft-masked version of the genome was annotated using BRAKER2 v2.1.2 using aligned RNA-Seq reads as hints[34].

Additionally, we have generated a genome-guided transcriptome assembly using Trinity v2.8.5[35]. This assembly was then aligned to the genome using hisat2 v2.1.0. We then identified putative 3' and 5'UTR regions using the overlap of Trinity assembly and gene models predicted by BRAKER2.

Comparative genomic analysis

Proteomes for 32 analyzed species were downloaded from NCBI. Orthologous groups were then identified using OrthoFinder v2.3.3[14]. To compare the proteome of *Cristatella mucedo* with the proteomes of the three recently sequenced Lophotrochozoan species (*Notospermus geniculatus*, *Phoronis australis*, and *Lingula anatina*) we annotated the protein sequences of each organism using InterProScan 5.36-75.0[36]. We then calculated the number of unique protein hits for each InterPro category, PFAM domain, and Gene Ontology (GO) term annotated by InterProScan. To analyze the overrepresentation of GO terms and PFAM domains in the *C. mucedo* proteome we calculated the fold copy change metric as follows:

$$FCC = \frac{k}{\max(l, n, p) + 1}$$

, where *c* is the number of hits in the *C. mucedo* proteome, and *l*, *n*, and *p* are the numbers of hits in the proteomes of *L. anatina*, *N. geniculatus*, and *P. australis*, respectively.

For the analysis of underrepresented terms, we computed a similar value:

$$FCC = \frac{k + 1}{\min(l, n, p)}$$

For GO term analysis, we also aggregated the hit counts for each level of the ontology.

Transcriptome alignment and quantification

Both paired-end and single-end RNA-seq reads were aligned to the genome using splice-aware aligner hisat2 v2.1.0. Reads were assigned to annotated genomic features using featureCounts v1.5.1[37] using "-O -M --fraction -s 2" options. Read counts were then normalized and visualized using Phantasus v1.7.2 (http://genome.ifmo.ru/phantasus-dev/). Raw reads and processed read counts are available in GEO Omnibus database (series GSE141747). Genome annotation and aligned

transcriptomic reads can also be seen in JBrowse v1.16.6[38] genomic browser (http://cristatella.bioinf.me/).

Author contributions

VVS, AVP, ALM and ANO planned the research and experiments. VVS, EAV and VAK collected living colonies and ststoblasts, sprouted young colonies from statoblasts and performed nucleic acid extractions. VVS and AVP performed ONT sequencing, LLM and ABK made PacBio and Partially Illumina sequencing. AVP and YAB assembled the genome and transcriptomes and performed all bioinformatic data analysis. VVS made scanning electron microscopy and immunohistochemistry. VVS and VAK performed *in situ* RNA hybridizations. AVP, VVS and OAN drafted the manuscript and prepared figures. All the members reviewed the manuscript and approved the final version.

Acknowledgements

The authors are grateful to Dmitrii Polev for Illumina library preparations, and Alexey Mirolubov for the help with scanning electron microscopy. The scientific research was performed at the Center for molecular and cell technologies, Center for Culturing Collection of Microorganisms, center "Biobank", and center "CHROMAS" of St. Petersburg State University and "Taxon" Research Resource Center of Zoological Institute RAS (http://www.ckp-rf.ru/ckp/3038/?sphra se_id=8879024).

The work was supported by the Russian Science Foundation (research grant 18-14-00086) to OAN, and by the Ministry of Education and Science of the Russian Federation grant No. 14.W03.31.0015 to LLM.

Ethical standards

All applicable international, national, and institutional guidelines for the care and use of animals were followed. I neither used endangered species nor were the investigated animals collected in protected areas.

References

1. Bock, P. E. & Gordon, D. P. Phylum Bryozoa Ehrenberg, 1831. In: Zhang, Z.-Q. (Ed.) Animal Biodiversity: An Outline of Higher-level Classification and Survey of Taxonomic Richness (Addenda 2013). *Zootaxa* **3703**, 67 (2013).
2. Fuchs, J., Obst, M. & Sundberg, P. The first comprehensive molecular phylogeny of Bryozoa (Ectoprocta) based on combined analyses of nuclear and mitochondrial genes. *Mol. Phylogenet. Evol.* **52**, 225–233 (2009).
3. Hausdorf, B., Helmkampf, M., Nesnidal, M. P. & Bruchhaus, I. Phylogenetic relationships within the lophophorate lineages (Ectoprocta, Brachiopoda and Phoronida). *Mol. Phylogenet. Evol.* **55**, 1121–1127 (2010).
4. Waeschenbach, A., Taylor, P. D. & Littlewood, D. T. J. A molecular phylogeny of bryozoans. *Mol. Phylogenet. Evol.* **62**, 718–735 (2012).
5. Waeschenbach, A., Telford, M. J., Porter, J. S. & Littlewood, D. T. J. The complete mitochondrial genome of Flustrellidra hispida and the phylogenetic position of Bryozoa among the Metazoa. *Mol. Phylogenet. Evol.* **40**, 195–207 (2006).
6. Jang, K. & Hwang, U. Complete mitochondrial genome of Bugula neritina (Bryozoa, Gymnolaemata, Cheilostomata): phylogenetic position of Bryozoa and phylogeny of lophophorates within the Lophotrochozoa. *BMC Genomics* **10**, 167 (2009).
7. Tsyganov-Bodounov, A., Hayward, P. J., Porter, J. S. & Skibinski, D. O. F. Bayesian phylogenetics of Bryozoa. *Mol. Phylogenet. Evol.* **52**, 904–910 (2009).
8. Sun, M. *et al.* The complete mitochondrial genome of Watersipora subtorquata (Bryozoa, Gymnolaemata, Ctenostomata) with phylogenetic consideration of Bryozoa. *Gene* **439**, 17–24 (2009).
9. Sun, M. *et al.* Complete mitochondrial genome of Tubulipora flabellaris (Bryozoa: Stenolaemata): The first representative from the class Stenolaemata with unique gene order. *Mar. Genomics* **4**, 159–165 (2011).
10. Shen, X. *et al.* Complete mitochondrial genome of Membranipora grandicella (Bryozoa: Cheilostomatida) determined with next-generation sequencing: The first representative of the suborder Malacostegina. *Comp. Biochem. Physiol. Part D Genomics Proteomics* **7**, 248–253 (2012).
11. Shunkina, K. V., Zaytseva, O. V., Starunov, V. V. & Ostrovsky, A. N. Comparative morphology of the nervous system in three phylactolaemate bryozoans. *Front. Zool.* **12**, 28 (2015).
12. Luo, Y.-J. *et al.* The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat. Commun.* **6**, 8301 (2015).
13. Luo, Y.-J. *et al.* Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. *Nat. Ecol. Evol.* **2**, 141–151 (2018).
14. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
15. Akam, M. Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. 7.
16. Novikova, E. L., Bakalenko, N. I., Nesterenko, A. Y. & Kulakova, M. A. Expression of Hox genes during regeneration of nereid polychaete Alitta (Nereis) virens (Annelida, Lophotrochozoa). *EvoDevo* **4**, 14 (2013).
17. Bakalenko, N. I., Novikova, E. L., Nesterenko, A. Y. & Kulakova, M. A. Hox gene expression during postlarval development of the polychaete Alitta virens. *EvoDevo* **4**, 13 (2013).

18. Passamaneck, Y. J. & Halanych, K. M. Evidence from Hox genes that bryozoans are lophotrochozoans. *Evol. Dev.* **6**, 275–281 (2004).

19. Fait, A., Yellin, A. & Fromm, H. 12 GABA and GHB Neurotransmitters in Plants and Animals. 15.

20. Schwiebert, E. M. & Zsembery, A. Extracellular ATP as a signaling molecule for epithelial cells. *Biochim. Biophys. Acta BBA - Biomembr.* **1615**, 7–32 (2003).

21. Fabbretti, E. ATP-Gated P2X3 Receptors Are Specialised Sensors of the Extracellular Environment. in *Protein Reviews* (ed. Atassi, M. Z.) vol. 1051 7–16 (Springer Singapore, 2017).

22. Kulakova, M. *et al.* Hox gene expression in larval development of the polychaetes Nereis virens and Platynereis dumerilii (Annelida, Lophotrochozoa). *Dev. Genes Evol.* **217**, 39–54 (2007).

23. Fröbius, A. C., Matus, D. Q. & Seaver, E. C. Genomic Organization and Expression Demonstrate Spatial and Temporal Hox Gene Colinearity in the Lophotrochozoan Capitella sp. I. *PLoS ONE* **3**, e4004 (2008).

24. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).

25. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

26. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).

27. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

28. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **9**, e112963 (2014).

29. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput. Biol.* **13**, e1005595 (2017).

30. Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).

31. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).

32. Zdobnov, E. M. *et al.* OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749 (2017).

33. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

34. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-Genome Annotation with BRAKER. in *Gene Prediction* (ed. Kollmar, M.) vol. 1962 65–95 (Springer New York, 2019).

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

36. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

37. 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).

38. Buels, R. *et al*. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol*. **17**, 66 (2016).