BMC Genomics

RESEARCH

Open Access

The genome of a vestimentiferan tubeworm (*Ridgeia piscesae*) provides insights into its adaptation to a deep-sea environment

Muhua Wang[1,2†], Lingwei Ruan[3†], Meng Liu[4†], Zixuan Liu[1], Jian He[1,2], Long Zhang[1], Yuanyuan Wang[1], Hong Shi[3], Mingliang Chen[3], Feng Yang[3], Runying Zeng[3], Jianguo He[1,2*], Changjun Guo[1,2*] and Jianming Chen[3,5*]

Abstract

Background Vestimentifera (Polychaeta, Siboglinidae) is a taxon of deep-sea worm-like animals living in deep-sea hydrothermal vents, cold seeps, and organic falls. The morphology and lifespan of *Ridgeia piscesae*, which is the only vestimentiferan tubeworm species found in the hydrothermal vents on the Juan de Fuca Ridge, vary greatly according to endemic environment. Recent analyses have revealed the genomic basis of adaptation in three vent- and seep-dwelling vestimentiferan tubeworms. However, the evolutionary history and mechanism of adaptation in *R. piscesae*, a unique species in the family Siboglinidae, remain to be investigated.

Result We assembled a draft genome of *R. piscesae* collected at the Cathedral vent of the Juan de Fuca Ridge. Comparative genomic analysis showed that vent-dwelling tubeworms with a higher growth rate had smaller genome sizes than seep-dwelling tubeworms that grew much slower. A strong positive correlation between repeat content and genome size but not intron size and the number of protein-coding genes was identified in these deep-sea tubeworm species. Evolutionary analysis revealed that *Ridgeia pachyptila* and *R. piscesae*, the two tubeworm species that are endemic to hydrothermal vents of the eastern Pacific, started to diverge between 28.5 and 35 million years ago. Four genes involved in cell proliferation were found to be subject to positive selection in the genome of *R. piscesae*.

Conclusion *Ridgeia pachyptila* and *R. piscesae* started to diverge after the formation of the Gorda/Juan de Fuca/Explorer ridge systems and the East Pacific Rise. The high growth rates of vent-dwelling tubeworms might be derived from their small genome sizes. Cell proliferation is important for regulating the growth rate in *R. piscesae*.

Keywords Vestimentiferan tubeworm, *Ridgeia piscesae*, Hydrothermal vent, Genome evolution, Deep-sea adaptation

†Muhua Wang, Lingwei Ruan and Meng Liu contributed equally to this work.

*Correspondence:
Jianguo He
lsshjg@mail.sysu.edu.cn
Changjun Guo
gchangj@mail.sysu.edu.cn
Jianming Chen
chenjm@mju.edu.cn

[1] State Key Laboratory for Biocontrol, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Marine Sciences, Sun Yat-Sen University, Zhuhai 519082, China

[2] China-ASEAN Belt and Road Joint Laboratory On Mariculture Technology, Guangdong Province Key Laboratory for Aquatic Economic Animals, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China

[3] State Key Laboratory Breeding Base of Marine Genetic Resources, Key Laboratory of Marine Genetic Resources of Ministry of Natural Resources, Fujian Key Laboratory of Marine Genetic Resources, Ministry of Natural Resources, Third Institute of Oceanography, Xiamen 361005, China

[4] Novogene Bioinformatics Institute, Beijing 100083, China

[5] Fujian Key Laboratory On Conservation and Sustainable Utilization of Marine Biodiversity, Fuzhou Institute of Oceanography, Minjiang University, Fuzhou 350108, China

Introduction

The discovery of deep-sea hydrothermal vents and cold seeps, as well as their associated ecosystems, has revolutionized our view of biology and understanding of the energy sources that fuel primary productivity on Earth [1–3]. Hydrothermal vents are areas on the ocean floor where hot, anoxic, chemical-rich water is expelled into the cold, oxygen-rich deep ocean [4]. Cold seeps are areas where methane, hydrogen sulfide, and other hydrocarbons seep or emanate as gas from deep geologic sources [5]. Both hydrothermal vents and cold seeps are characterized by high hydrostatic pressure, darkness, a lack of oxygen and photosynthesis-derived nutrients, and high concentrations of toxic chemicals [6]. Organisms inhabiting around hydrothermal vents and cold seeps develop unique characters to adapt to these deep-sea reducing environments [7, 8]. Due to the complete absence of light, hydrothermal vent and cold seep ecosystems are driven by chemosynthesis instead of photosynthesis [9, 10]. The process is completed by chemosynthetic microorganisms, which cooperate with a variety of macrobenthos to form chemosynthetic symbioses and contribute to primary production supporting the ecosystem [11].

Vestimentifera (Polychaeta, Siboglinidae) is a taxon of deep-sea worm-like animals living in deep-sea hydrothermal vents, cold seeps, and organic falls [12]. The body of the adult vestimentiferan tubeworm is enclosed in a chitinous tube that is closed at the posterior end. Vestimentiferan tubeworms lack a digestive tract and rely on symbiosis with chemoautotrophic microorganisms, which inhabit a specialized internal organ, to meet their metabolic needs [13]. The first discovery of chemoautotrophic symbionts in *Riftia pachyptila*, a vestimentiferan tubeworm inhabiting hydrothermal vents on the East Pacific Rise (EPR), initiated the intensive study of these deep-sea tubeworms [1]. The body of the adult tubeworm comprises four main parts. The anteriorly located branchial plume is the primary site of gas exchange with the environment. Below the plume is the vestimentum, where the heart, gonopores and a simplified brain are located. The trophosome, which is primarily composed of symbiont-containing bacteriocytes and blood vessels, is located below the vestimentum. The segmented opisthosoma is located below the vestimentum [14–16].

Lifespan varies greatly between vent- and seep-dwelling tubeworms [17]. The lifespan of *R. pachyptila*, which thrives in relatively strong and continuous diffuse hydrothermal flow, was estimated to be less than 10 years [18]. In contrast, *Lamellibrachia luymesi*, which lives around cold seeps in the Gulf of Mexico, can live for up to 250 years [19]. In general, vent-dwelling tubeworms grow faster than seep-dwelling tubeworms. The growth rate of *R. pachyptila* can reach approximately 160 cm yr^{-1}, while the growth rate of *L. luymesi* is only approximately 3 cm yr^{-1} [20–22]. A previous analysis revealed that both vent- and seep-dwelling tubeworms have high rates of cell proliferation, and the variation in growth rates is attributed to the variation in apoptosis between vent- and seep-dwelling tubeworms, where apoptosis is substantially downregulated in vent-dwelling species [23].

The Juan de Fuca Ridge in the northeast Pacific Ocean is characterized by broad heterogeneity in chemical environments, ranging from vigorous, high-temperature vents to diffuse flow [24]. The hydrothermal vents on the Juan de Fuca Ridge provide numerous biotic habitats, which support the growth of a large quantity of endemic organisms [25]. Although the biomass of the endemic hydrothermal vent fauna is high, there are only a few macrofaunal species dominating a particular vent community [26, 27]. *Ridgeia piscesae* Jones (1985) is the only vestimentiferan tubeworm species in the hydrothermal vents on the Juan de Fuca Ridge. The tubeworm occurs at high density in most vents and acts as an ecosystem-structuring species by providing habitats for several other organisms and serving as a primary producer through chemosynthetic endosymbiosis [8, 28].

Ridgeia piscesae adopts different strategies to thrive in diverse environments in the vents on Juan de Fuca Ridge. First, two extreme growth forms (morphotypes) of *R. piscesae*, "short-fat" and "long-skinny", were discovered in geologically and chemically diverse vent fields [29]. The tube of the "short-fat" morphotype has a generally constant diameter of 2–3 cm, while the tube diameter of the "long-skinny" morphotype decreases from the anterior to posterior [30]. "Short-fat" *R. piscesae* prefers a relatively high-flow vent fluid of high temperature (up to 30 ℃) and high concentrations of sulfide. The "long-skinny" morphotype adapts to ambient temperature (2 ℃) and low concentrations of sulfide in areas of diffuse hydrothermal fluids [31]. This morphotype of *R. piscesae* can thrive in areas of diffuse vent fluids by acquiring sulfide using buried posterior tube sections [28]. Second, the lifespan of *R. piscesae* varies greatly according to the endemic environment [18]. The species can grow with high growth rates ranging from 6 to 95 cm yr^{-1} under favorable conditions and can grow very slowly when exposed to low levels of vent flow and sulfide [32, 33]. Strong phenotypic plasticity and a flexible lifespan allow *R. piscesae* to survive in diverse habitats and make it a unique species in the family Siboglinidae. Recent genomic analyses have revealed the genetic basis of adaptation in three vent- and seep-dwelling vestimentiferan tubeworms [34–36]. Although it plays a critical role in supporting the vent ecosystem, the evolutionary history and mechanism of adaptation in *R. piscesae*, a unique deep-sea tubeworm, remain to be investigated.

Results

Genome assembly and annotation

The samples of the "long-skinny" morphotype of *R. piscesae* were collected from the Cathedral deep-sea hydrothermal vent, Main Endeavor Field of the Juan de Fuca Ridge (47° 56' N, 129° 05' W, 2,181 m depth). Short-insert paired-end (180 bp, 300 bp and 500 bp) and long-insert mate-pair (2 kb, 5 kb, 10 kb and 15 kb) sequencing libraries were constructed and sequenced on the Illumina HiSeq 2000 platform. A total of 247.74 Gb of sequencing data was generated (Supplementary Table 1). Based on the *k*-mer distribution of 180-bp paired-end Illumina reads, the genome size was estimated to be 694.79 Mb with a heterozygosity of 1.2% (Supplementary Fig. 1). The final assembly of the *R. piscesae* genome was 574.96 Mb with a contig N50 size of 10.42 kb and a scaffold N50 size of 230.23 kb (Supplementary Table 2).

A total of 87.4% sequencing reads could be aligned unambiguously to the assembled *R. piscesae* genome sequence, covering 99.74% of the assembly (Supplementary Table 3). In addition, 99.63% of Trinity assembled sequences (unigenes) could be aligned to the assembly (Supplementary Table 4). A benchmarking universal single-copy orthologs (BUSCO) assessment of the integrity of the genome assembly against the metazoan core gene set showed that the completeness of the genome was 91.7% (90.4% complete and 1.3% fragmented) (Supplementary Table 5). These results demonstrated that the completeness of the *R. piscesae* genome is comparable to that of the previously published tubeworm genomes (Table 1) [34–36].

Transposable elements (TEs) accounted for 30.17% of the *R. piscesae* genome assembly, with long interspersed elements (LINEs, 8.08%) as the most abundant class of TEs (Supplementary Table 6). The *R. piscesae* genome encodes 24,096 protein-coding genes, of which 95.54% are annotated based on known proteins in diverse public protein databases (Supplementary Table 7).

Phylogenomic analyses

To infer the evolutionary history of *R. piscesae*, a maximum-likelihood (ML) phylogenetic tree was constructed using single-copy orthologs of *R. piscesae* and 14 metazoans with *Adineta vaga* as an outgroup (Fig. 1, Supplementary Fig. 2, Supplementary Table 8). Two vent-dwelling tubeworms (*R. piscesae* and *R. pachyptila*) formed a clade. *Paraescarpia echinospica* and *L. luymesi* from cold seeps are basal to the vent clade. These results corroborate the view that vent-dwelling tubeworms might be derived from their seep-dwelling relatives [37, 38].

Ridgeia piscesae is endemic to the Gorda/Juan de Fuca/Explorer (GFE) ridge systems, and *R. pachyptila* is discovered on the EPR. The subduction of the Farallon-Pacific Ridge separated the GFE and EPR between 28.5 and 35 Ma ago. Molecular clock analysis revealed that the divergence time of *R. pachyptila* and *R. piscesae* was approximately 33.7 million years (Ma), suggesting that the divergence time between these two species might have been close to the separation time of the two ridge systems. The divergence time of *L. luymesi* and the other three tubeworms was estimated to be approximately 65.1 Ma, corroborating the view that modern vestimentiferan tubeworms started to diverge during the early Cenozoic Era [34, 35, 39].

Genome evolution of vestimentiferan tubeworms

It has been demonstrated that several factors, including repeat content and number of genes, contribute to the variation in genome sizes among different organisms [40, 41]. In addition, previous reports proposed that the differences in genome sizes among deep-sea tubeworms might be attributable to the numbers of repetitive elements and genes [34, 36]. The assembled genome size of *R. piscesae* is similar to that of *R. pachyptila* but smaller than those of two seep-dwelling tubeworms (*L. luymesi* and *P. echinospica*) (Table 1). The genomes of cold seep-dwelling tubeworms have more TEs, especially DNA transposons, LINEs and LTR retrotransposons, than those of hydrothermal vent-dwelling

Table 1 Genome assembly statistics of four deep-sea vestimentiferan tubeworms

	Ridgeia piscesae	*Riftia pachyptila* [36]	*Paraescarpia echinospica* [34]	*Lamellibrachia luymesi* [35]
Assembled genome size (Mb)	574.9	560.7	1090.9	687.7
Contig N50 (kb)	10.4	2870.3	253.6	24
Scaffold N50 (kb)	230.2	-	67,235.3	372.9
BUSCO (%)	92.6	99.4	95.1	95.8
Repeat content (%)	30.2	29.9	55.1	38.2

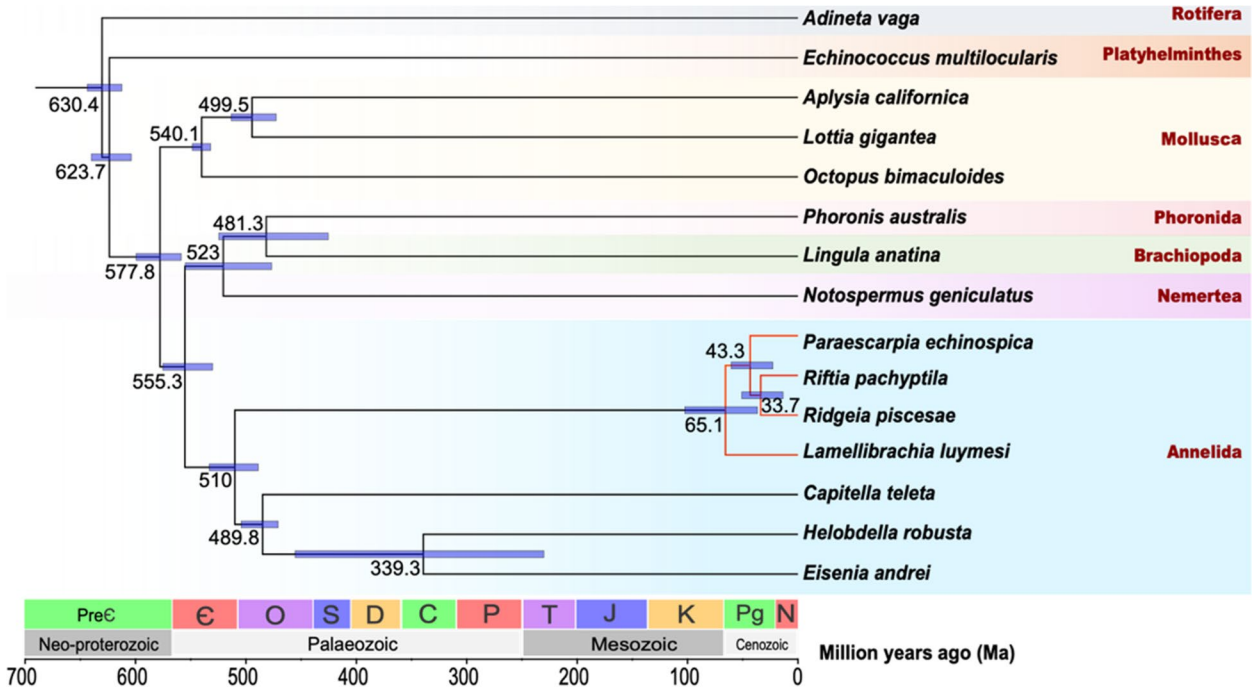

Fig. 1 A species tree of R. piscesae and 14 metazoans. Single-copy orthologs were used to reconstruct the phylogenetic tree. The divergence time between species pairs was listed above each node, and 95% confidence interval of the estimated divergence time was denoted as blue bar. *R. piscesae* diverged from *R. pachyptila* approximately 33.7 Ma ago

tubeworms (Fig. 2a). TEs accounted for 38.2% and 55.1% of the *L. luymesi* and *P. echinospica* genomes, and they constituted 30.2% and 29.9% of the genomes of *R. piscesae* and *R. pachyptila*, respectively. A strong positive correlation ($R^2 = 0.98$, $P = 0.0052$) was identified between genome size and repeat content in these four species (Fig. 2b), suggesting that TEs are a major contributor to genome size evolution in vestimentiferans.

Repeat landscape plots indicated that TE activity is different between *R. pachyptila* and three other tubeworm species (Fig. 3). There are recent expansions of TEs in the genomes of *L. luymesi*, *P. echinospica*, and *R. piscesae* but not in *R. pachyptila*. The main contributors to recent TE expansions in *L. luymesi*, *P. echinospica*, and *R. piscesae* appear to have been LINEs and DNA

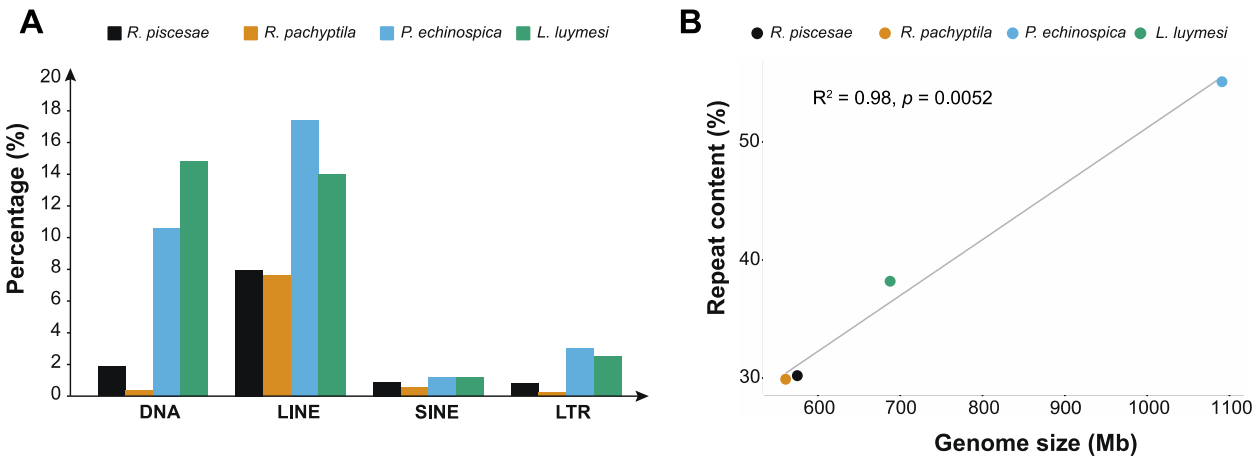

Fig. 2 Genome size evolution in four vestimentiferan tubeworms. **A** Comparison of the occurrence and composition of repetitive elements in the genomes of 4 vestimentiferan tubeworms. **B** The relationship between repeat contents and genome sizes in 4 vestimentiferan tubeworms. A strong positive correlation (R2 = 0.98, *P* = 0.0052) was identified between genome sizes and repeat contents in these four species

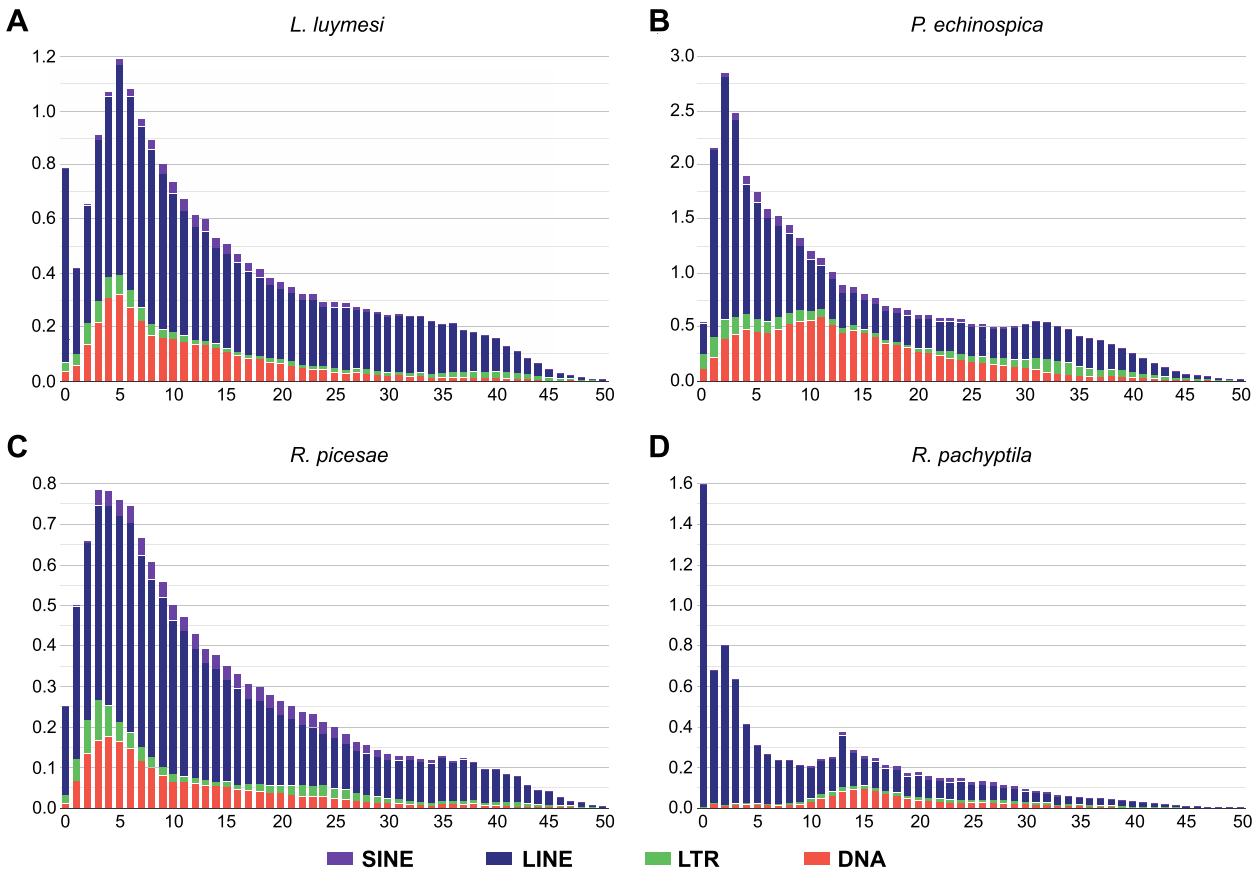

Fig. 3 Transposable element-accumulation profile in the genomes of four vestimentiferan tubeworms. There are recent expansions of TEs in the genomes of *L. luymesi*, *P. echinospica*, and *R. piscesae*, but not in *R. pachyptila*

transposons. Nonetheless, only LINEs were expanded recently in the genome of *R. pachyptila* (Fig. 3d).

The number of annotated gene models in the *R. piscesae* genome (24,096) is similar to those in the genomes of *R. pachyptila* (25,984) and *P. echinospica* (22,642) but smaller than that in the *L. luymesi* genome (38,998). Introns account for 220.1 Mb and 204.7 Mb of the genomes of two seep-dwelling tubeworms (*L. luymesi* and *P. echinospica*, respectively), as well as 264.8 Mb and 234.5 Mb of the genomes of two vent-dwelling tubeworms (*R. pachtypila* and *R. piscesae*, respectively) (Supplementary Table 9). The average length of introns in the genome of *R. piscesae* is longer than that in the genomes of the other three species. Additionally, two vent-dwelling tubeworms with smaller genome sizes had higher ratios of intron/exon length than the seep-dwelling tubeworms. Thus, gene number and intron size do not contribute to the differences in genome sizes between seep- and vent-dwelling tubeworms.

A previous study revealed that *R. pachyptila* experienced reductive evolution with more contracted than expanded gene families in the genome [36]. Gene-family analysis of four tubeworm species identified a core set of 10,225 gene families (Fig. 4a). In total, 601 and 279 lineage-specific gene families were identified in *R. piscesae* and *R. pachyptila*, respectively, which are much smaller than the numbers in *L. luymesi* (1181) and *P. echinospica* (1045). Additionally, gene-family analysis of 12 lophotrochozoans revealed that the numbers of expanded gene families were substantially smaller than those of contracted gene families in the two vent-dwelling tubeworms, while more gene families were expanded than contracted in their seep-dwelling counterparts (Fig. 4b). These results indicate that the genomes of vent-dwelling tubeworms are characterized by gene loss.

Hox genes are a set of conserved regulators that specify regions of the body plan of an embryo along the anterior–posterior axis in metazoans [42]. One of the Hox genes (*Antp*) plays a role in the development of the posterior segment of several marine annelids [43]. Loss of *Antp* was apparent across all four tubeworm genomes (Supplementary Fig. 3), corroborating the view that the loss of *Antp* contributes to the reduced segmentation of the posterior region of juvenile worms in vestimentiferans [34].

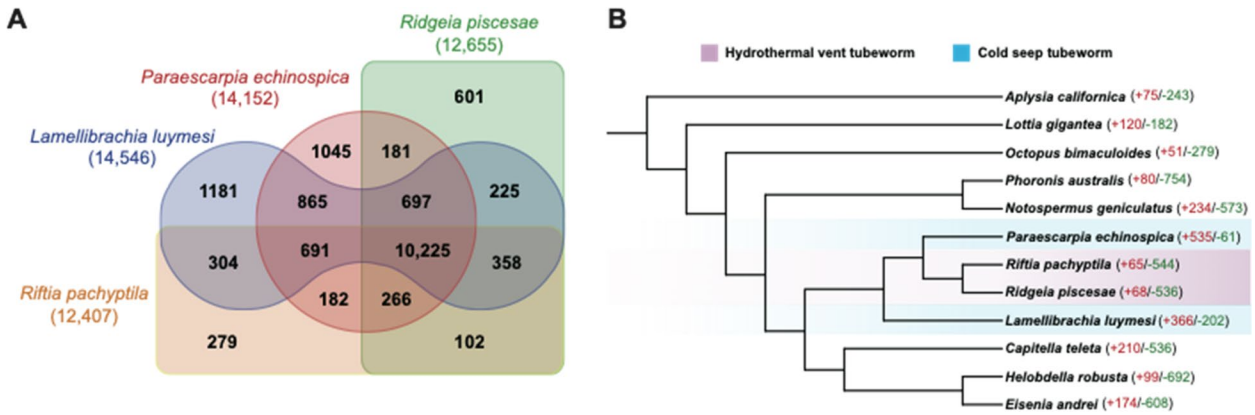

Fig. 4 Protein family evolution in four vestimentiferan tubeworms. **A** Venn diagram of shared and unique gene families in four vestimentiferan tubeworm species. Lineage-specific gene families of *R. piscesae* and *R. pachyptila* are much less than those of *L. luymesi* and *P. echinospica*. **B** Gene family expansion/contraction analysis of 4 vestimentiferan tubeworms and 8 other lophotrochozoans. The numbers of protein families that were significantly expanded (red) and contracted (green) ($P < 0.05$) in each species are denoted beside the species names

The *Lox2* gene is missing from the genome of *L. luymesi* but present in the genomes of three other tubeworms, suggesting that the loss of this gene was a lineage-specific event.

Genomic basis of deep-sea adaptation

Hemoglobins (Hbs) in vestimentiferan tubeworms, which bind oxygen and sulfide simultaneously and provide substrate for chemosynthesis by the symbionts, facilitate the adaptation of these species to deep-sea reducing environments. Four heme-containing chains were identified (A1, A2, B1, and B2) in hemoglobins of vestimentiferans [44]. To elucidate the evolution of Hbs in vestimentiferans, we identified Hb genes in the genomes of four tubeworm species. A single copy each of A2 and B2 Hb genes, as well as two copies of A1 genes, was identified in each of the tubeworm genomes (Fig. 5). The free cysteine residues in the A2 and B2 chains contribute to the sulfide-binding ability of the vestimentiferan Hbs [45]. A2 and B2 Hb genes are highly expressed in the muscle of the vestimentum of *R. piscesae*, suggesting their role in binding H_2S in this species (Supplementary Table 10). Previous studies revealed that the group of B1 Hbs was significantly expanded in *L. luymesi*, *P. echinospica*, and *R. pachyptila* [34–36]. With 17 identified genes, the group of B1 Hbs was also expanded in the genome of *R. piscesae*. Additionally, free cysteine was identified at the same position in 6 B1 Hb genes as in the A2 Hb genes of *R. piscesae* (Supplementary Fig. 4). The B1 genes with free cysteine are expressed in the muscle of the vestimentum of *R. piscesae*, corroborating the view that the free cysteines might also contribute to sulfide binding in the B1 hemoglobin chain of deep-sea tubeworms [35] (Supplementary Table 10). The expression levels of these B1 genes are highly variable, with only one gene highly expressed in the muscle of the vestimentum. This indicates that the vestimentum of *R. piscesae* might not be the major organ where B1 globins bind H_2S.

Recent reports revealed that most enzymes related to amino acid biosynthesis were lost in *L. luymesi* and *R. pachyptila* [35, 36]. To gain better insight into the nutrient dependence of endosymbionts in vestimentiferans, we identified key enzymes involved in amino acid biosynthesis in the genomes of 4 tubeworms and 3 other annelid species (Fig. 6). All four tubeworms (*L. luymesi* and *R. pachyptila*, *R. piscesae* and *P. echinospica*) lack most key enzymes related to amino acid biosynthesis, corroborating the view that vestimentiferan tubeworms mainly rely on endosymbionts for synthesizing amino acids [35]. In addition to tubeworms, two other annelids (*Eisenia andrei* and *Helobdella robusta*) also lack most enzymes for amino acid biosynthesis, supporting the hypothesis that these two species acquire amino acids from food [46, 47].

The expansion of gene families is considered a major driver of adaptation and speciation [48]. Thus, we performed gene family expansion and contraction analysis with 4 vestimentiferan tubeworms and 8 other lophotrochozoans (Fig. 4b). In total, 10 gene families were significantly expanded in the genomes of all four tubeworms compared to the other 8 lophotrochozoans ($P < 0.05$) (Supplementary Table 11). Gene ontology analysis revealed that the expanded gene families were involved in chitin binding and innate immunity. Furthermore, 18 gene families were significantly expanded in the genomes of two cold-seep tubeworms. The expanded gene families were involved in DNA repair, innate immunity, and protein stability (Supplementary Table 12).

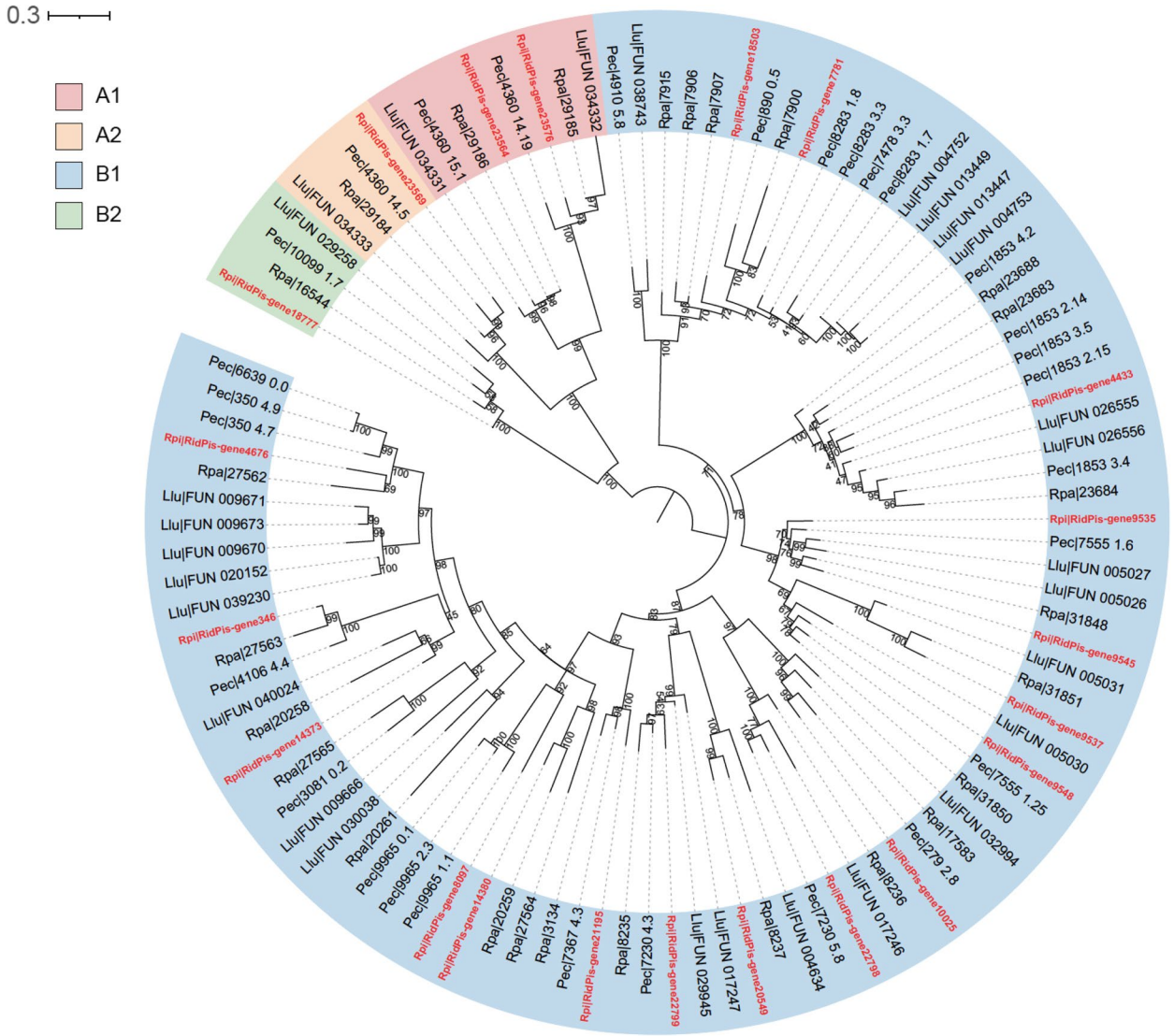

Fig. 5 Gene tree of Hb subunits A1, A2, B1, and B2 from four vestimentiferan tubeworms. The values near the nodes are ultrafast bootstrap (UFBoot) values. Genes from *R. piscesae* are labeled in red

In addition to expanded and contracted gene families, we also identified positively selected genes (PSGs) in the genome of *R. piscesae*. Compared with 11 other lophotrochozoan species, *R. piscesae* was found to harbor 9 PSGs. In total, 7 out of 9 PSGs were expressed in the vestimentum of *R. piscesae* (Supplementary Table 13). Interestingly, four genes (alkB homolog 2, alpha-ketoglutarate-dependent dioxygenase, *ALKBH2*; Derlin-1, *DERL1*; Ras-related and estrogen-regulated growth inhibitor, *RERG*; and AN1-type zinc finger protein 2B, *ZFAND2B*) involved in cell proliferation were subjected to positive selection in *R. piscesae*. Two of these genes (*ALKBH2* and *DERL1*) promote cell proliferation. ALKBH2 is responsible for protecting the genome from 1-meA damage by repairing the damage in double-stranded DNA [49]. *ALKBH2* promotes cell proliferation and is overexpressed in several types of tumor cells [50]. DERL1 participates in the endoplasmic reticulum (ER)-associated degradation response and unfolded protein response (UPR) [51]. *DERL1* is responsible for cell proliferation and promotes the progression of several types of cancers [52]. Interestingly, two other genes (*RERG* and *ZFAND2B*) inhibit cell proliferation. Overexpression of *RERG*, a member of the RAS superfamily of GTPases, inhibits cell proliferation and tumor formation [53, 54]. ZFAND2B reduces the abundance of IGF1R, a kinase that activates cell proliferation, in a proteasome-dependent manner [55]. These results suggest that, in addition

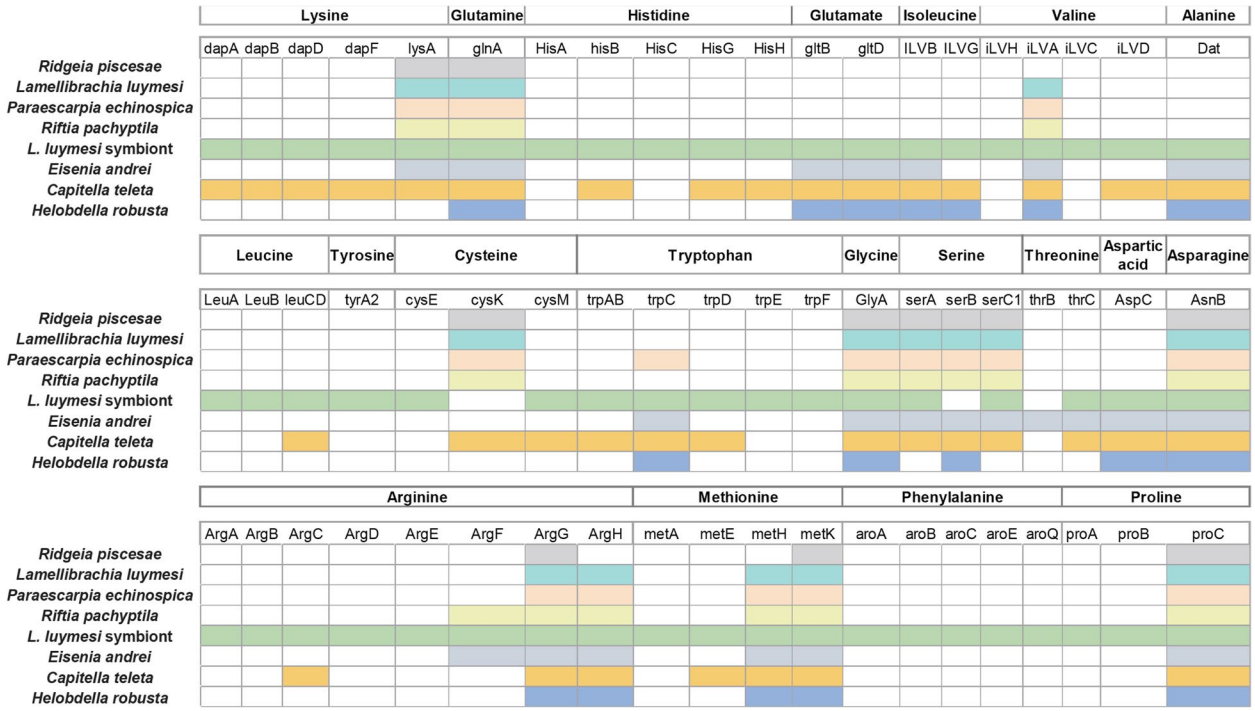

Fig. 6 The presence and absence of key amino acid biosynthesis genes in annelids and *L. luymesi* symbiont. Most key genes associated with amino acid biosynthesis are missing in the genomes of four tubeworms (*L. luymesi*, *P. echinospica*, *R. pachyptila*, and *R. piscesae*). These genes are presented in the genome of L. luymesi symbionts. Two other annelids (*E. andrei* and *H. robusta*) that acquire amino acids from food also lack some key genes associated with amino acid biosynthesis

to apoptosis, the regulation of cell proliferation also contributes to the variation in growth rates in *R. piscesae*.

Discussion

In many hydrothermal vent and cold seep ecosystems, vestimentiferan tubeworms are among the dominant megafauna in habitats where hydrogen sulfide is present [56, 57]. *Ridgeia piscesae* is an ecosystem-structuring species and primary producer in the hydrothermal vents on the Juan de Fuca Ridge, where the biomass of the endemic fauna is high [8, 28]. In addition, *R. piscesae* has strong phenotypic plasticity, with two extreme morphotypes ("short-fat" and "long-skinny") diverging in several morphological characters found in this species. The "short-fat" morphotype prefers relatively high-flow vent fluid with high concentrations of sulfide, while the "long-skinny" morphotype survives in diffuse hydrothermal fluids with low concentrations of sulfide [31]. Strong phenotypic plasticity makes *R. piscesae* a unique species in the family Siboglinidae. In this study, we assembled and annotated a draft genome sequence of *R. piscesae* collected at the Cathedral vent of the Juan de Fuca Ridge.

Riftia pachyptila and *R. piscesae* are two tubeworm species endemic to hydrothermal vents in the eastern Pacific. *Ridgeia piscesae* and *R. pachyptila* are endemic to the GFE ridge systems and the EPR, respectively, which are separated due to the subduction of the Farallon-Pacific Ridge between 28.5 and 35 Ma. Phylogenomic analysis showed that the divergence time of *R. pachyptila* and *R. piscesae* was approximately 33.7 Ma, suggesting that the phenotypic divergence between these two species might be derived from adaptation to the local environments of the two ridge systems.

Genome sizes vary greatly among vent- and seep-dwelling tubeworms. It was proposed that natural selection and adaptive processes shape genome size evolution [58]. Previous analyses revealed that genome sizes are correlated with several phenotypic traits, including cell size and rates of metabolism and growth [59–62]. Thus, we studied the evolution of genome sizes in four vestimentiferan tubeworms. A strong positive correlation ($R^2 = 0.98$, $P = 0.0052$) between repeat content and genome size was identified in these tubeworm species. Among the four vestimentiferan tubeworm genomes, the *L. luymesi* genome has the most annotated gene models, while another seep-dwelling tubeworm (*P. echinospica*) genome has the fewest annotated gene models. In addition, the average length of introns in the genome of *R. piscesae* is also longer than that in the introns of the other three species' genomes. This suggests that repeat content

contributes to the variation in genome sizes in tubeworm species. However, the variation in genome sizes in tubeworms is not attributable to gene number and intron length, as proposed in a previous study [36].

Vent-dwelling tubeworms grow much faster than seep-dwelling tubeworms. The growth rates of *R. pachyptila* and *R. piscesae* can reach approximately 160 cm yr^{-1} and 95 cm yr^{-1}, respectively, while the growth rate of *L. luymesi* is only approximately 3 cm yr^{-1} [20–22, 32, 33]. We studied the factors that might contribute to the variation in growth rate in four vestimentiferan tubeworms. The lineage-specific gene families of *R. piscesae* (601) and *R. pachyptila* (279) are much smaller than those of *L. luymesi* (1181) and *P. echinospica* (1045). In addition, gene-family expansion and contraction analysis revealed that the numbers of expanded gene families were substantially smaller than those of contracted gene families in the two vent-dwelling tubeworms. A negative correlation between genome size and growth rate was identified in several species, as organisms with smaller genomes might undergo more rapid replication of their genome [62, 63]. A recent report showed that *R. pachyptila* underwent reductive evolution [36]. Our results indicate that both *R. pachyptila* and *R. piscesae* experienced reductive evolution. The small genome size of vent-dwelling tubeworms might contribute to their fast growth rates.

Previous immunohistochemical and ultrastructural cell cycle analyses revealed that both *L. luymesi* and *R. pachyptila* had extremely high cell proliferation activities. The divergence of growth rates between seep- and vent-dwelling tubeworms is attributable only to apoptosis. *Lamellibrachia luymesi* has balanced activities of proliferation and apoptosis in the epidermis, while apoptosis is substantially downregulated in this tissue in *R. pachyptila* [23]. Unlike those of other vestimentiferan tubeworms, the growth rates of *R. piscesae* vary greatly among individuals living in environments with different levels of vent flow and sulfide [32, 33]. Four genes involved in the regulation of cell proliferation were identified to be positively selected in *R. piscesae*. Interestingly, two of these genes promote cell proliferation, whereas the two other genes inhibit cell proliferation. This result indicates that both cell proliferation and apoptosis are involved in the regulation of growth in *R. piscesae*.

There is still room for improvement of *R. piscesae* genome (contig N50: 10.42 kb, scaffold N50: 230.23 kb), which may affect some genomic analyses, including underestimating the gene and repeat contents. The BUSCO estimate of the completeness of our assembly (91.7%) is comparable to that for previously published lophotrochozoan genomes. Additionally, 99.63% of Trinity assembled sequences could be aligned to our assembly. Thus, our assembly should be useful for exploring the genetic basis of deep-sea adaptation in this species. With two extreme morphotypes ("short-fat" and "long-skinny") adapted to different environments, *R. piscesae* is a unique species in the family Siboglinidae. It will be interesting to reveal the molecular basis of these two morphotypes adapt to their endemic environments through comparative transcriptomic analysis. Unfortunately, it is limited by the sample collection from deep-sea environments. We hope to improve the quality of genome assembly and perform comparative transcriptomic analysis between the two morphotypes in future studies.

Conclusions

Here, we assembled and annotated a draft genome of *R. piscesae* collected at the Cathedral vent of the Juan de Fuca Ridge. Evolutionary analysis suggested that the divergence between two vent-dwelling species (*R. piscesae* and *R. pachyptila*) might have been close in time to the separation of the GFE ridge systems and the EPR. Comparative genomic analysis showed that vent-dwelling tubeworms with a higher growth rate had smaller genome sizes than seep-dwelling tubeworms that grow much slower, suggesting that the high growth rates of vent-dwelling tubeworms are derived from their small genome sizes. The variation in the genome sizes of these deep-sea tubeworms is attributed to the repeat content but not the intron sizes and numbers of protein-coding genes. Finally, four genes involved in cell proliferation were found to be subject to positive selection in the genome of *R. piscesae*, indicating that cell proliferation is important for regulating the growth rate in this species.

Methods

Sampling and sequencing

The samples of "long-skinny" morphotype of *R. piscesae* were obtained during *Alvin* dive 4243 from the deep-sea hydrothermal vent at Cathedral vent, Main Endeavor Field of the Juan de Fuca Ridge (47° 56' N, 129° 05' W, 2,181 m depth) on August 9, 2006. Genomic DNA (gDNA) was extracted from vestimentum muscle of the specimen using a standard phenol/chloroform extraction protocol and broken into random fragments for whole-genome shotgun (WGS) sequencing. Agarose gel electrophoresis was used to check the quality of the gDNA, and Qubit system was used to quantify the gDNA. Short-insert paired-end libraries (180 bp, 300 bp and 500 bp) were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, USA) according to the standard protocol, respectively. Large-insert mate-pair libraries (2 kb, 5 kb, 10 kb and 15 kb) were prepared following the Cre-lox recombination-based protocol [64]. All DNA

libraries were subjected to paired-end sequencing on the Illumina Hiseq 2000 platform (Illumina). Muscle samples from vestimentum were also collected for constructing RNA sequencing (RNA-seq) library. Total RNA was extracted with TRIzol reagent (Molecular Research Center, USA). Paired-end library for RNA-seq was constructed using the Paired-End Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) and sequenced on the Illumina Hiseq 2000 platform (Illumina).

Genome assembly

NGS QC toolkit (v2.1) [65] was used to evaluate the quality of raw sequencing reads and filter high-quality reads. High-quality reads were obtained by filtering out the following types of reads: (1) reads with > = 10% unidentified nucleotides (N); (2) reads with adaptor contamination; (3) reads with > = 20% bases having Phred quality score < = 5; (4) duplicated reads generated by PCR amplification during the library construction process.

The size and heterozygosity of *R. piscesae* genome were estimated using the high-quality short-insert paired-end reads (180 bp) by *k*-mer frequency-distribution method. The number of *k*-mers and the peak depth of *k*-mer sizes at 19 was obtained using GenomeScope2 (v1.0.0) [66]. Due to the high heterozygosity of *R. piscesae* genome, a modified version of SOAPdenovo [67] was implemented for genome assembly. In brief, all short-insert paired-end reads were applied for contig assembly. Depth of coverage was obtained for each contig using SOAPdenovo with the parameters '-e 1 -M 0 -R', and the contigs with depth less than 60 were identified as heterozygous contigs. All WGS reads were aligned to the heterozygous contigs using SOAPdenovo. And links were generated between heterozygous contigs when supported by a minimum of three read pairs. Heterozygous contigs were clustered into bubble clusters based on the orientation and distance between heterozygous contigs. If two contigs represented two potential haplotypes in a bubble structure, the longer one was retained to ensure the integrity of contig assembly.

To scaffolding the contigs, all short-insert paired-end and long-insert mate-pair reads were realigned onto the contig sequences using SOAPdenovo. Duplicated contigs that had high depth of coverage and conflicting connections to the unique contigs were masked during scaffolding. A hierarchical assembly strategy was used to construct contigs into primary scaffolds by adding the ascending insert size reads gradually. Finally, all short-insert reads were realigned onto the scaffold sequences to fill the gaps with the GapCloser program implemented in SOAPdenovo [68].

We also attempted to assemble the genome using two other assemblers (ABySS2, Platanus-Allee) [69, 70], but the genome quality and completeness were inferior to the SOAPdenovo assembly (Supplementary Table 14). Thus, we ignore the assemblies in the downstream analysis.

Genome quality assessment

To assess the completeness of the *R. piscesae* genome, high-quality short-insert paired-end reads were mapped to the genome assembly using Burrows-Wheeler Aligner (BWA) (v0.7.17) [71] with parameters of '-o 1 -e 5 -t 8 -n 15'. In addition, all RNA-seq reads were de novo assembled using Trinity (v2.9) [72]. The Trinity unigene assembled sequences with length > = 500 bp were mapped to the *R. piscesae* genome using BLAT (v35.1) [73] with default parameters and an identify cutoff of 90%. The completeness of the assembly was also evaluated using benchmarking universal single-copy orthologs (BUSCO) (v3.1.0) [74] with 978 metazoa single-copy orthologous genes (obd10).

Genome annotation

Tandem repeats in the genome were predicted using the program Tandem Repeats Finder (TRF) (v4.09) [75] with default parameters. Transposable elements (TEs) were identified using the homology-based and de novo prediction approaches. For homology-based prediction, RepeatMasker (v4.1.0) (http://www.repeatmasker.org/) were conducted to identify repeat sequences against the Repbase library. For de novo prediction, RepeatModeler (v2.0.1) (http://repeatmasker.org/RepeatModeler.html), LTR-Finder (v1.0.7) [76], RepeatScout (v1.0.5) [77] and Piler (v1.0) [78] were used to construct de novo repeat libraries. RepeatMasker (v4.1.0) was run against these libraries to search repeat elements.

Protein-coding genes in *R. piscesae* genome were predicted with three approaches: homology-based prediction, ab initio prediction and RNA-seq-based prediction. Protein-coding sequences of *Lottia gigantea*, *Helobdella robusta*, *Capitella teleta*, *Schistosoma mansoni*, *Caenorhabditis elegans*, *Anopheles gambiae*, *Drosophila melanogaster* and *Homo sapiens* were aligned to the *R. piscesae* genome using tblastn with a cut off E-value of 1e-5. GeneWise (v2.4) [79] was employed to predict gene models. For ab initio prediction, Augustus (v3.3.2) [80], Genscan [81], Geneid (v1.3) [82], GlimmerHMM (v3.0.4) [83] and SNAP [84] were used to predict genes on the repeat-masked genome. For RNA-seq-based prediction, unigenes generated using Trinity (v2.9) were aligned against the genome assembly with BLAT (v35.1) (identify > = 0.95 and align rate > = 0.95) [73]. In addition, the RNA-seq reads from were aligned to the *R. piscesae* genome using Tophat (v2.1.1) [85]. And gene structures were predicted using Cufflinks (v2.2.1) [86]. EvidenceModeler (EVM) (v1.1.1) [87] was used to integrate all

gene models derived from these three approaches into a non-redundant gene set.

Functional annotation was performed using BLASTP searches against SwissProt and TrEMBL databases [88] with a E-value cut-off of 1e-5. In addition, InterProScan (v5.4.0) [89] was used to screen proteins against five databases (Pfam, PRINTS, PROSITE, ProDom and SMART) to determine protein domains and motifs. Gene Ontology (GO) annotation of each gene was retrieved from the corresponding InterPro entry. In addition, KEGG annotation was performed using GhostKOALA [90].

Phylogenomic analysis

Protein sequences of 14 metazoan species (*Adineta vaga*, *Echinococcus multilocularis*, *Aplysia californica*, *Lottia gigantea*, *Octopus bimaculoide*, *Phoronis australis*, *Lingula anatina*, *Notospermus geniculatus*, *Capitella teleta*, *Helobdella robusta*, *Eisenia Andrei*, *Riftia pachyptila*, *Paraescarpia echinospica*, *Lamellibrachia lumysi*) were downloaded for gene family cluster analysis (Supplementary Table 8). The longest transcripts of each gene (more than 30 amino acids) were retained. All-to-all BLASTP was used to identify the similarities between retained protein sequences of these 14 metazoan species and *R. piscesae* (E-value threshold: 1e-7). OrthoFinder (v2.2.7) [91] was used to identify and cluster gene families among 15 species with default parameters. Gene clusters with > 100 gene copies in one or more species were removed. Protein sequences of all single-copy gene families were retrieved and aligned using MAFFT (v7.271) [92]. The alignments were trimmed using TrimAl (v1.2) [93]. The phylogenetic tree was reconstructed with the trimmed alignments using FastTree2 (v2.1.11) [94] with *Adineta vaga* as outgroup.

To estimate the divergent time, the trimmed alignments of single-copy orthologs among the 15 metazoan species were concatenated using PhyloSuite (v1.2.2) [95]. MCMCtree module of the PAML package (v4.9) [96] was used to estimate the divergent time with the concatenated alignment. The species tree of the 15 metazoan species was used as a guide tree, and the analysis was calibrated with the divergent time obtained from TimeTree database (minimum = 470.2 Ma and soft maximum = 531.5 Ma between *P. australis* and *L. anatina*) [97] and previous analyses (minimum = 470.2 Ma and soft maximum = 531.5 Ma between *A. californica* and *L. gigantea*; minimum = 532 Ma and soft maximum = 549 Ma for the first appearance of Mollusca; minimum = 476.3 Ma and soft maximum = 550.9 Ma for the appearance of capitellid-leech clade; minimum = 550.25 Ma and soft maximum = 636.1 Ma for the first appearance of Lophotrochozoa and Ecdysoa) [98–100].

Gene family expansion and contraction analysis

R8s (v1.7) was applied to obtain the ultrametric tree of 12 lophotrochozoan species (*C. teleta*, *H. robusta*, *E. andrei*, *L. gigantea*, *A. californica*, *N. geniculatus*, *A. californica*, *P. australis*, *R. pachyptila*, *P. echinospica*, *L. lumysi*, *R. piscesae*), which is calibrated with the divergent time between *C. teleta* and *L. gigantea* (688 Ma) obtained from TimeTree database. CAFÉ (v5) [101] was applied to determine the significance of gene-family expansion and contraction among 12 lophotrochozoan species based on the ultrametric tree and the gene clusters determined by OrthoFinder (v2.2.7). Gene families that were significantly expanded in each of four tubeworm species (*R. pachyptila*, *P. echinospica*, *L. lumysi*, *R. piscesae*) ($P < 0.05$) were annotated using PANTHER (v16.0) with the PANTHER HMM scoring tool (pantherScore2.pl) [102].

Homeobox gene analysis

Homeodomain sequences, which were retrieved from HomeoDB database [103], were aligned to *R. piscesae* genome assembly using tbalstn. Sequences of the candidate homeobox genes were extracted based on the alignment results. The extracted sequences were aligned against NCBI NR and HomeoDB database to classify the homeobox genes.

Hemoglobin gene family analysis

Protein sequences of hemoglobin A1, A2, B1, B2 chains of four tubeworm species were obtained with reference references using DIAMOND BLASTP [104] with a E-value cut-off of 1e-5. The sequences were annotated in NCBI NR database using BLASTP. And protein domains in these sequences were annotated by Pfamscan against Pfam-A.hmm database [105]. Sequences that have almost full length protein domains were aligned using MAFFT (v7.271) [106]. The alignments were trimmed using TrimAI (v1.2) [93]. The phylogenetic tree was reconstructed with the trimmed alignments using a maximum-likelihood method implemented in IQ-TREE2 (v2.1.2) [107]. The best-fit substitution model was selected by using ModelFinder algorithm [108]. Branch supports were assessed using the ultrafast bootstrap (UFBoot) approach with 1,000 replicates [109].

Identification of positively selected genes (PSGs)

We identified PSGs in the *R. piscesae* genome within single-copy orthologs among 12 lophotrochozoan species that were identified in gene-family expansion and contraction analysis. Protein sequences of all single-copy gene families were retrieved and aligned using MAFFT (v7.271) [92]. Phylogenetic tree of each family was reconstructed using IQ-TREE2 (v2.1.2) [107]. PSGs were

identified based on the phylogenetic trees using HyPhy (v2.5.30) with the adaptive Branch-Site Random Effects Likelihood (aBSREL) model [110].

Statistics and reproducibility

Alpha levels of 0.05 were regarded as statistically significant throughout the study, unless otherwise specified.

Abbreviations

EPR	East Pacific Rise
GFE	Gorda/Juan de Fuca/Explorer
BUSCO	Benchmarking universal single-copy orthologs
TEs	Transposable elements
LINE	Long interspersed element
ML	Maximum likelihood
Ma	Million years
Hb	Hemoglobin
PSG	Positively selected gene
ALKBH2	AlkB homolog 2, alpha-ketoglutarate-dependent dioxygenase
DERL1	Derlin-1
RERG	Ras-related and estrogen-regulated growth inhibitor
ZFAND2B	AN1-type zinc finger protein 2B
UPR	Unfolded protein response

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09166-y.

Additional file 1: Supplementary Figure 1. Distributionof 19-mer frequency in *Ridgeia piscesae*genome. The short-insert paired-end reads (180 bp) were used to generate the19-mer frequency curve. Thehet-erozygous rate and the genome size were determined based on the *k*-merdistribution. **Supplementary Figure 2.** The phylogenetic tree of *R.piscesae* and 14 other lophotrochozoans.The tree wasreconstructed with single-copy orthologs using a maximum likelihood approach. The ultrafast bootstrap (UFBoots) value is listed above each of the nodes.**Supplementary Figure 3.** Genomic organization of *Hox* gene clusters in 4vestimentiferan tubeworms and 11 other metazoans. *Hox* genes are indicated as rectangles.The orientations of genes are indicated by arrows below the genes. The genecomposition and orientation of *Hox* clusters are consistent between two vent-dwellingtubeworms (*R. pachyptila* and *R. piscesae*), but slightly differentbetween vent- and seep-dwelling tubeworms. **Supplementary Figure 4.**Alignment of hemoglobins in four tubeworms. Each tubeworm has two copies of A1 chain, one copy of A2 chain, and onecopy of B2 chain in hemoglobins of tubeworms. A group of B1 chain in hemoglobinwere found in each of four species. Free cysteine was found in A2, B2, and B1chains in hemoglobin. **Supplementary Table 1.** Statistics of the genome sequencing data of *Ridgeia piscesae*. **Supplementary Table 2.** Statistic of the *R. piscesae* genome assembly. **Supplementary Table 3.** Assessment of genome coverage ratebased on short-insert paired-end reads remapping analysis. **SupplementaryTable 4.** Assessment of gene coveragerate using Trinity assembled sequences (Unigenes). **SupplementaryTable 5.** BUSCO evaluationof *R. piscesae* genome assembly. **Supplementary Table 6.** Summary of annotated repeats in *R.piscesae* genome. **Supplementary Table 7.** Statistics of functional annotated gene models in the genome of *R.piscesae*. **Supplementary Table 8.** Information of genomes used to perform phylogenomic analysis. **Supplementary Table 9.** Exon and intronlengths of genes in four Vestimentiferan tubeworms. **Supplementary Table 10.** Expression levels of hemoglobin genes withfree cysteine in *R. piscesae*. **Supplementary Table 11.** Gene families were significantly expandedin the genomes of all four tubeworms. **SupplementaryTable 12.** Gene families were significantly expanded in the genomes of twoseep-dwelling tubeworms. **Supplementary Table 13.** Positively selected genes (PSGs) in *R.piscesae*. **Supplementary Table 14.** Summary of genome assemblies using two other assemblers

Acknowledgements

We gratefully acknowledge the crews of *Alvin* dive 4243. We appreciate Prof. Huaiyang Zhou, Dr. Brian Marquardt, Dr. Helen White, Mr. Mark Spear, Prof. Debbie Kelley, Prof. Marv Lilley, Prof. Peter R. Girguis and other friendly scientists for their help during the expedition and sampling collection. We gratefully acknowledge the National Supercomputing Center in Guangzhou for provision of computational resources.

Authors' contributions

J.M.C, C.J.G and J.G.H. conceived of the project and designed research; L.R., M.L., Z.L, Y.W. assembled and annotated the genome; M.W, J.H., L.Z., H.S., M.C., Y.J., F.Y., and R.Z., performed the evolutionary analyses; M.W., J.M.C., C.J.G, and J.G.H. wrote the paper with contribution from all authors. The author(s) read and approved the final manuscript.

Funding

This work was supported by National Key R&D Program of China (2018YFC0310702), the Projects under Major State Basic Research Development Program of China (973 Program) (2015CB755906), National Natural Science Foundation of China (31900309), GuangDong Basic and Applied Basic Research Foundation (2019A1515011644, 2020A1515010330), Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311021006), and Project 2018N2001 from Department of Fujian Science and Technology and Program for Innovative Research Team in Science and Technology in Fujian Province University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the paper.

Availability of data and materials

Raw reads and genome assembly are accessible in NCBI under BioProject number PRJNA826206. Raw reads and genome assembly are also available at the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0002911.

Declarations

Ethics approval and consent to participate

The tubeworm used in our study is an invertebrate, so the approval according to the regulations on the use of tubeworm is unnecessary.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 11 October 2022 Accepted: 3 February 2023
Published online: 11 February 2023

References

1. Corliss JB, Dymond J, Gordon LI, Edmond JM, von Herzen RP, Ballard RD, Green K, Williams D, Bainbridge A, Crane K, et al. Submerine thermal springs on the galapagos rift. Science. 1979;203(16):1073–83.
2. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V, et al. Hydrogen is an energy source for hydrothermal vent symbioses. Nature. 2011;476(7359):176–80.
3. Paull CK, Hecker B, Commeau R, Freeman-Lynde RP, Neumann C, Corso WP, Golubic S, Hook JE, Sikes E, Curray J. Biological communities at the Florida escarpment resemble hydrothermal vent taxa. Science. 1984;226(4677):965–7.

4. Von Damm KL: Controls on the chemistry and temporal variability of seafloor hydrothermal fluids. In: Seafloor hydrothermal system: Physical, chemical, biological, and geological interactions. Edited by Humphris RA, Zierenberg LS, Thomson RE; 1995: 222–247.
5. Suess E. Marine cold seeps and their manifestations: geological control, biogeochemical criteria and environmental conditions. Int J Earth Sci. 2014;103(7):1889–916.
6. Levin LA. Ecology of cold seep sediments: Interactions of fauna with flow, chemistry and microbes. Oceanogr Mar Biol. 2005;43:1–46.
7. Grassle JF. Hydrothermal vent animals: distribution and biology. Science. 1985;229(4715):713–7.
8. Childress JJ, Fisher CR. The biology of hydrothermal vent animals: physiology, biochemistry, and autotrophic symbioses. Oceanogr Mar Biol. 1992;30:337–441.
9. Vanreusel A, Andersen AC, Boetius A, Connelly D, Cunha MR, Decker C, Hilario A, K.A. K, Maignien L, Olu K et al: Biodiversity of cold seep ecosystems along the European margins. Oceanography 2009, 22(1):110–127.
10. Stewart FJ, Newton IL, Cavanaugh CM. Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. Trends Microbiol. 2005;13(9):439–48.
11. Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. Nat Rev Microbiol. 2019;17(5):271–83.
12. Bright M, Lallier FH. The biology of vestimentiferan tubeworms. Oceanogr Mar Biol. 2010;48:213–65.
13. Vrijenhoek RC. Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. Mol Ecol. 2010;19(20):4391–411.
14. Jones ML. Riftia pachyptila Jones: observations on the vestimentiferan worm from the galapagos rift. Science. 1981;213(4505):333–6.
15. Hand SC. Trophosome ultrastructure and the characterization of isolated bacteriocytes from invertebrate-sulfur bacteria symbioses. Biol Bull. 1987;173(1):260–76.
16. Schulze A. Comparative anatomy of excretory organs in vestimentiferan tube worms (Pogonophora, Obturata). J Morphol. 2001;250(1):1–11.
17. Lutz RA, Kennish MJ. Ecology of deep-sea hydrothermal vent communities: a review. Rev Geophys. 1993;31(3):211–42.
18. Urcuyo IA, Bergquist DC, MacDonald IR, VanHorn M, Fisher CR. Growth and longevity of the tubeworm Ridgeia piscesae in the variable diffuse flow habitats of the Juan de Fuca Ridge. Mar Ecol Prog Ser. 2007;344:143–57.
19. Bergquist DC, Williams FM, Fisher CR. Longevity record for deep-sea invertebrate. Nature. 2000;403(6769):499–500.
20. Fisher CR, Urcuyo IA, Simpkins MA, Nlx E. Life in the slow lane: growth and longevity of cold-seep vestimentiferans. Mar Ecol. 2008;18(1):83–94.
21. Thiebaut E, Huther X, Shillito B, Jollivet D, Gaill F. Spatial and temporal variations of recruitment in the tube worm Riftia pachyptila on the East Pacific Rise (9 degrees 50 ′N and 13 degrees N). Mar Ecol Prog Ser. 2002;234:147–57.
22. Shank TM, Fornari DJ, Von Damm KL, Lilley MD, Haymon RM, Lutz RA. Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50′N, East Pacific Rise). Deep Sea Res Part II Top Stud Oceanogr. 1998;45(1–3):465–515.
23. Pflugfelder B, Cary SC, Bright M. Dynamics of cell proliferation and apoptosis reflect different life strategies in hydrothermal vent and cold seep vestimentiferan tubeworms. Cell Tissue Res. 2009;337(1):149–65.
24. Tivey MK, Stakes DS, Cook TL, Hannington MD, Petersen S. A model for growth of steep-sided vent structures on the endeavour segment of the Juan de Fuca Ridge: results of a petrologic and geochemical study. J Geophys Res-Sol Ea. 1999;104(B10):22859–83.
25. Tsurumi M, Tunnicliffe V. Tubeworm-associated communities at hydrothermal vents on the Juan de Fuca Ridge, northeast Pacific. Deep Sea Res Part I Oceanogr Res Pap. 2003;50(5):611–29.
26. Lutz RA, Desbruyeres D, Shank TM, Vrijenhoek RC. A deep-sea hydrothermal vent community dominated by Stauromedusae. Deep Sea Res Part II Top Stud Oceanogr. 1998;45:329–34.
27. Tunnicliffe V, McArthur AG, McHugh D. A biogeographical perspective of the deep-sea hydrothermal vent fauna. Adv Mar Biol. 1998;34:353–442.
28. Urcuyo IA, Massoth GJ, Julian D, Fisher CR. Habitat, growth and physiological ecology of a basaltic community of Ridgeia piscesae from the Juan de Fuca Ridge. Deep Sea Res Part I Oceanogr Res Pap. 2003;50(6):763–80.
29. Southward EC, Tunnacliffe V, Black M. Revision of the species of Ridgeia from northeast Pacific hydrothermal vents, with a redescription of Ridgeia piscesae Jones (Pogonophora: Obturata=Vestimentifera). Can J Zool. 1995;73:282–95.
30. Jones M. On the status of the phylum-name, and other names, of the vestimentiferan tube worms. Proc Biol Soc Wash. 1987;100:1049–50.
31. Carney SL, Flores JF, Orobona KM, Butterfield DA, Fisher CR, Schaeffer SW. Environmental differences in hemoglobin gene expression in the hydrothermal vent tubeworm, Ridgeia piscesae. Comp Biochem Physiol B Biochem Mol Biol. 2007;146(3):326–37.
32. Tunnicliffe V, Embley RW, Holden JF, Butterfield DA, Massoth G, Juniper SK. Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca Ridge. Deep Sea Res Part I Oceanogr Res Pap. 1997;44(9–10):1627–44.
33. Sarrazin J, Robigou V, Juniper SK, Delaney JR. Biological and geological dynamics over four years on a high-temperature sulfide structure at the Juan de Fuca Ridge hydrothermal observatory. Mar Ecol Prog Ser. 1998;153(1):5–24.
34. Sun Y, Sun J, Yang Y, Lan Y, Ip JC, Wong WC, Kwan YH, Zhang Y, Han Z, Qiu JW, et al. Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm Paraescarpia echinospica. Mol Biol Evol. 2021;38(10):4116–34.
35. Li Y, Tassia MG, Waits DS, Bogantes VE, David KT, Halanych KM. Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm Lamellibrachia luymesi. BMC Biol. 2019;17(1):91.
36. de Oliveira AL, Mitchell J, Girguis P, Bright M. Novel insights on obligate symbiont lifestyle and adaptation to chemosynthetic environment as revealed by the giant tubeworm genome. Mol Biol Evol. 2022;39(1):msab347.
37. Black MB, Halanych KM, Maas PAY, Hoeh J, Hashimoto D, Desbruyeres D, Lutz RA, Vrijenhoek RC. Molecular systematics of vestimentiferan tubworms from hydrothermal vents and cold-water seeps. Mar Biol. 1997;130:141–9.
38. Halanych KM. Molecular phylogeny of siboglinid annelids (a.k.a. pogonophorans): a review. Hydrobiologia. 2005;535:297–307.
39. Little CTS, Vrijenhoek RC. Are hydrothermal vent animals living fossils? Trends Ecol Evol. 2003;18(11):582–8.
40. Lynch M. The Origins of Genome Architecture. Sunderland, MA: Sinauer Associates; 2007.
41. Niu S, Li J, Bo W, Yang W, Zuccolo A, Giacomello S, Chen X, Han F, Yang J, Song Y, et al. The Chinese pine genome and methylome unveil key features of conifer evolution. Cell. 2022;185(1):204-217 e214.
42. Pearson JC, Lemons D, McGinnis W. Modulating Hox gene functions during animal body patterning. Nat Rev Genet. 2005;6(12):893–904.
43. Bakalenko NI, Novikova EL, Nesterenko AY, Kulakova MA. Hox gene expression during postlarval development of the polychaete Alitta virens. EvoDevo. 2013;4(1):13.
44. Zal F, Suzuki T, Kawasaki Y, Childress JJ, Lallier FH, Toulmond A. Primary structure of the common polypeptide chain b from the multi-hemoglobin system of the hydrothermal vent tube worm Riftia pachyptila: an insight on the sulfide binding-site. Proteins. 1997;29(4):562–74.
45. Bailly X, Jollivet D, Vanin S, Deutsch J, Zal F, Lallier F, Toulmond A. Evolution of the sulfide-binding function within the globin multigenic family of the deep-sea hydrothermal vent tubeworm Riftia pachyptila. Mol Biol Evol. 2002;19(9):1421–33.
46. Pokarzhevskii AD, Zaboyev DP, Ganin GN, Gordienko SA. Amino acids in earthworms: are earthworms ecosystemivorous? Soil Bio & Biochem. 1997;29(3):559–67.
47. Bradley MDK, Reynolds JD. Diet of the leeches Erpobdella octoculata (L) and Helobdella stagnalis (L) in a lotic habitat subject to organic pollution. Freshw Biol. 1987;18(2):267–75.
48. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, Jordar VS, Maiti R, Kodira CD, Neafsey DE, Zeng QD, et al. Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. Genome Res. 2009;19(10):1722–31.
49. Aas PA, Otterlei M, Falnes PO, Vagbo CB, Skorpen F, Akbari M, Sundheim O, Bjoras M, Slupphaug G, Seeberg E, et al. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature. 2003;421(6925):859–63.

50. Wilson DL, Beharry AA, Srivastava A, O'Connor TR, Kool ET. Fluorescence probes for ALKBH2 allow the measurement of DNA Alkylation repair and drug resistance responses. Angew Chem Int Ed Engl. 2018;57(39):12896–900.
51. Eshraghi A, Dixon SD, Tamilselvam B, Kim EJ, Gargi A, Kulik JC, Damoiseaux R, Blanke SR, Bradley KA. Cytolethal distending toxins require components of the ER-associated degradation pathway for host cell entry. PLoS Pathog. 2014;10(7): e1004295.
52. Dong QZ, Wang Y, Tang ZP, Fu L, Li QC, Wang ED, Wang EH. Derlin-1 is overexpressed in non-small cell lung cancer and promotes cancer cell invasion via EGFR-ERK-mediated up-regulation of MMP-2 and MMP-9. Am J Pathol. 2013;182(3):954–64.
53. Finlin BS, Gau CL, Murphy GA, Shao H, Kimel T, Seitz RS, Chiu YF, Botstein D, Brown PO, Der CJ, et al. RERG is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer. J Biol Chem. 2001;276(45):42259–67.
54. Ho JY, Hsu RJ, Liu JM, Chen SC, Liao GS, Gao HW, Yu CP. MicroRNA-382-5p aggravates breast cancer progression by regulating the RERG/Ras/ERK signaling axis. Oncotarget. 2017;8(14):22443–59.
55. Osorio FG, Soria-Valles C, Santiago-Fernandez O, Bernal T, Mittelbrunn M, Colado E, Rodriguez F, Bonzon-Kulichenko E, Vazquez J, Porta-de-la-Riva M, et al. Loss of the proteostasis factor AIRAPL causes myeloid transformation by deregulating IGF-1 signaling. Nat Med. 2016;22(1):91–6.
56. Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB. Prokaryotic cells in the hydrothermal vent tube worm riftia pachyptila jones: possible chemoautotrophic symbionts. Science. 1981;213(4505):340–2.
57. Tunnicliffe V. The nature and origin of the modern hydrothermal vent fauna. Palaios. 1992;7(4):338–50.
58. Gregory TR. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol Rev Camb Philos Soc. 2001;76(1):65–101.
59. Vinogradov AE. Nucleotypic effect in homeotherms: body-mass-corrected basal metabolic rate of mammals is related to genome size. Evolution. 1995;49(6):1249–59.
60. Cavalier-Smith T. Skeletal DNA and the evolution of genome size. Annu Rev Biophys Bioeng. 1982;11:273–302.
61. Wright NA, Gregory TR, Witt CC. Metabolic "engines" of flight drive genome size reduction in birds. Proc Biol Sci. 2014;281(1779):20132780.
62. Wyngaard GA, Rasch EM, Manning NM, Gasser K, Domangue R. The relationship between genome size, development rate, and body size in copepods. Hydrobiologia. 2005;532:123–37.
63. Tenaillon MI, Manicacci D, Nicolas SD, Tardieu F, Welcker C. Testing the link between genome size and growth rate in maize. PeerJ. 2016;4: e2408.
64. Van Nieuwerburgh F, Thompson RC, Ledesma J, Deforce D, Gaasterland T, Ordoukhanian P, Head SR. Illumina mate-paired DNA sequencing-library preparation using Cre-Lox recombination. Nucleic Acids Res. 2012;40(3):e24.
65. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE. 2012;7(2): e30619.
66. 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.
67. Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat Ecol Evol. 2017;1(5):120.
68. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1(1):18.
69. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, et al. ABySS 2.0: resource-efficient assembly of large genomes using a bloom filter. Genome Res. 2017;27(5):768–77.
70. Kajitani R, Yoshimura D, Okuno M, Minakuchi Y, Kagoshima H, Fujiyama A, Kubokawa K, Kohara Y, Toyoda A, Itoh T. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. Nat Commun. 2019;10(1):1702.
71. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
72. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-U130.
73. Kent WJ. BLAT–the BLAST-like alignment tool. Genome Res. 2002;12(4):656–64.
74. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.
75. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
76. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265–8.
77. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21(Suppl 1):i351-358.
78. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. Bioinformatics. 2005;21(Suppl 1):i152-158.
79. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14(5):988–95.
80. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33:W465–7.
81. Aggarwal G, Ramaswamy R. Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER. J Biosci. 2002;27(1 Suppl 1):7–14.
82. Parra G, Blanco E, Guigo R. GeneID in Drosophila. Genome Res. 2000;10(4):511–5.
83. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20(16):2878–9.
84. Korf I. Gene finding in novel genomes. BMC Bioinform. 2004;5:59.
85. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.
86. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
87. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR: Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol 2008, 9(1).
88. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000;28(1):45–8.
89. Mulder N, Apweiler R. InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol. 2007;396:59–70.
90. 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.
91. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.
92. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.
93. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
94. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3): e9490.
95. Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
96. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
97. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017;34(7):1812–9.
98. Donoghue P, Benton M, Yang ZH, Inoue J. Calibrating and constraining the molecular clock. J Vertebr Paleontol. 2009;29:89a–89a.

99. Benton MJ, Donoghue PCJ, Asher RJ, Friedman M, Near TJ, Vinther J: Constraints on the timescale of animal evolutionary history. Palaeontol Electron 2015, 18(1).
100. dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PC, Yang Z. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr Biol. 2015;25(22):2939–50.
101. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 2006;22(10):1269–71.
102. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 2017;45(D1):D183–9.
103. Zhong YF, Holland PW. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. Evol Dev. 2011;13(6):567–8.
104. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18(4):366–8.
105. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(D1):D222–30.
106. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
107. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
108. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
109. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22.
110. Pond SLK, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, et al. HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol. 2020;37(1):295–9.

Publisher's Note